Workplan for the Monitoring and Refinement of the Inpatient Rehabilitation Prospective Payment System

Melinda Beeuwkes Buntin, Grace Carter, Joan Buchanan, Orla Hayden, Susan Paddock, Daniel Relles, Gregory Ridgeway, Debra Saliba, Mark Totten, and Barbara Wynn

DRU-2769-CMS

March 2002

The RAND unrestricted draft series is intended to transmit preliminary results of RAND research. Unrestricted drafts have not been formally reviewed or edited. The views and conclusions expressed are tentative. A draft should not be cited or quoted without permission of the author, unless the preface grants such permission.

RAND is a nonprofit institution that helps improve policy and decisionmaking through research and analysis. RAND’s publications do not necessarily reflect the opinions or policies of its research sponsors.
CONTENTS

1. INTRODUCTION ..................................................... 1
   Payment Under the IRF PPS ..................................... 1
   Overview of Phase II Workplan ............................... 2

2. DATA ..................................................................... 4
   Linked IRF Files .................................................... 4
   Post Acute Claims .................................................. 5

3. REFINEMENT .......................................................... 8
   Modifications to the Fitting Algorithms ....................... 9
     Simultaneously Producing FRGs and Estimating Comorbidity
     Effects ............................................................. 9
     Transfer Cases ................................................. 11
     Validating the Models ....................................... 12
   Refining CMG Structure ......................................... 13
     Comorbidity Definitions, RIC Definitions, and Medical
     Complexity ....................................................... 13
     Varying Choice of Indices within RIC .................... 15
     Unobserved FIM Items ....................................... 15
     Unusual Cases .................................................. 16
     Weights .......................................................... 16
     Methodological Improvements in Facility-Level Adjustments
     Identifying And Accommodating Unduly Influential Facilities 18
     Additional Data on the Size of Teaching Programs ........ 21
     Additional Refinement Activities (Years 2 and 3) Once Post-IRF PPS
     Data Become Available ....................................... 22
     Further Assessment of Modeling Assumptions ............. 23

4. MONITORING .......................................................... 25
   Monitoring IRF-Specific Effects ................................ 25
     How Hospital-level Activities are Affected by Payment Elements 26
     Identifying Changes in Coding Procedure .................. 30
   Monitoring Effects Across Post Acute Care ................ 31
     Understanding Post Acute Wide Effects of New Payment Systems 32
     Post-Acute Care Access ...................................... 36
     PAC Payments and Outcomes ................................ 40
   Analytic Issues ................................................... 42
     The Monitoring System ...................................... 46

5. SPECIAL STUDIES ..................................................... 48
   Staff Time Measurement Study ................................ 48
   Compression ...................................................... 49
   Research Plan .................................................... 50
   Assessing the Predictive Value of MDS-PAC Items ........ 51
     Background ..................................................... 51
     Data and Methods .......................................... 53
   Functional Assessment ....................................... 55
   New Technology .................................................. 56
     Identifying New Technologies With Cost Implications .... 59
     Investigating Dispersion Of New Technologies And Cost Impacts 59
Evaluating Alternatives To Recognize New Technology In Payment System ........................................ 60

6. SCHEDULE AND DELIVERABLES ........................................... 63

REFERENCES ................................................................. 65
1. INTRODUCTION

This paper describes the analysis plan for phase II of RAND’s project on the Inpatient Rehabilitation Facility Prospective Payment System (IRF PPS). During this phase we will develop a system to monitor the operation of the IRF PPS and refine the payment parameters of the system.

In the Balanced Budget Act of 1997, Congress mandated that the Centers for Medicare and Medicaid Services (CMS) implement a PPS for inpatient rehabilitation. This new PPS was implemented in January 1, 2002. In the first phase of this project RAND performed research to support CMS’s efforts to design, develop, and implement this IRF PPS. That research, presented in Carter et al (2001) and Buchanan et al. (in press), included evaluations of alternatives for many of the parts of the payment system.

The new PPS will apply to rehabilitation hospitals and to distinct rehabilitation units of acute care hospitals that were excluded from the acute PPS. Medicare patients in such facilities receive intensive therapy (generally at least three hours per day). In addition, 75 percent of each facility’s patients must have one of 10 specified problems related to neurological or musculoskeletal disorders or burns.

PAYMENT UNDER THE IRF PPS

The unit of payment in the IRF PPS is a Medicare covered hospital stay, beginning with an admission to the rehabilitation hospital or unit and ending with discharge from that facility. Each case is classified into a Case Mix Group, or CMG. The CMGs are based on impairment, functional status as measured by items from the Functional Independence Measure (FIM), age, and comorbidities. The CMGs are assigned based on information in a new patient assessment instrument (the IRF PAI). Additional groups for deaths and atypically short stay cases are assigned based on claims data.

---

1 CMS is the successor to the Health Care Financing Administration, which was also known as HCFA.
The IRF PPS payment for a discharge in hospital $i$ in CMG $k$ is given by

$$F = R \cdot A_i \cdot W_k,$$

where $R$ is the national conversion factor, $A_i$ is the facility payment adjustment, and $W_k$ is the CMG relative weight. This payment is increased for outlier cases. Also, short stay transfer cases receive a payment for each day in the hospital plus a case level payment equal to one-half of one day’s payment.

Two-thirds of the IRF PPS payment amount plus one-third of the amount that would be paid under TEFRA will be paid for hospital fiscal years beginning from January 1 thru September 30, 2002. Facilities may choose to be paid entirely under the IRF PPS in these fiscal years. In the following years, all payments to all IRFs will be made under the IRF PPS. The initial value of $R$ was chosen to meet the statutory budget neutrality constraint that payment under the new RPPS should equal what payment would have been under TEFRA, as estimated by the CMS’s Office of the Actuary.

OVERVIEW OF PHASE II WORKPLAN

This project is funded under a phased contract. In Phase I we helped CMS design and develop the IRF PPS and its initial implementation. In Phase II we will help CMS develop a monitoring system and refine the payment system. Phase II of our contract is expected to last three years, although the funding decision will be made annually. In this workplan, we include the analyses that are planned or the entire 3-year period. However, since some of the research in the later years of Phase II will depend on both the findings of the initial analyses and the annual funding decisions, we provide somewhat more detail about the activities in the initial year.

We’ve divided the planned analysis during Phase II into three major areas: refinement, monitoring, and special studies. The refinement activities will use data from the operating IRF PPS to re-evaluate payment options and update payment parameters. We will also attempt to improve on the methodologies used to calculate payment parameters where possible. The initial development of these potential improvements in
methodology will be conducted during the first year of phase II, using pre-implementation data. Then, if CMS decides to use the new methodology, the actual updating of payment parameters can be done as rapidly as possible. This in turn will allow the update to use data that is as current as possible given the constraints of the rule production process. These refinement activities are discussed in section 3, below.

Monitoring activities will include building a baseline that describes post acute care prior to the implementation of the RPPS. We will also develop models that describe current patterns and trends in IRF PPS care and indicators of changes in patterns and trends. Monitoring activities are discussed in section 4.

The special studies discussed in section 5 are aimed at particular aspects of IRF care and the resources that it requires. For example, in the first year of phase II, we will examine how staff time varies by patient characteristics, the role of technology in changing case mix, costs, and outcomes in IRFs, and the ability of selected MDS-PAC items to predict costs. Although some of these activities may affect either refinement or monitoring activities, the connection depends strongly on the findings from the studies.

The data used for refinement and monitoring are discussed in section 2. The special studies will often require additional data that will be described along with each particular study in Section 5.

Phase II of our work is expected to last for 3 years -- until September 30, 2004. This work plan gives an overview of activities throughout the entire period, but presents greater detail for our activities during the first 15 months. The last section of our workplan, section 6, discusses the schedule of our research activities and reports during this first 15-month period.
2. DATA

Our Phase I final report was based on MEDPAR and FIM discharge data covering calendar years 1996 through 1999 and on cost report files covering PPS12 through 15. The FIM data came from UDSmr, from Caredata.com, and from HealthSouth Hospitals. For the monitoring and refinement phase of our research, we wish to add MEDPAR and FIM discharge data covering from January 2000 through as much of 2003 as is feasible. We will of course also need updated cost report information.

In addition to the MEDPAR-FIM and cost report data, we also need to acquire data that will tell us about the use of other forms of post-acute care--skilled nursing facilities, home health agencies, and outpatient rehabilitation services. The combined data will allow us to analyze trends in case mix, discharge setting, and cost, before and after implementation of the IRF PPS.

LINKED IRF FILES

To build our Phase I analytical data sets, we used probabilistic matching methods to link MEDPAR records to records from the FIM file that described the same discharge. The MEDPAR files covered calendar years 1996 through 1999 and grew over time from about 344,000 to 390,000 rehabilitation records per year. During this four-year period, the percent of IRFs providing FIM data grew from 42 to 62 percent. Our final merged files matched about 60 percent of all MEDPAR cases. We matched about 90 percent of MEDPAR cases to a FIM case for hospitals that participated in FIM data systems for an entire year. We were able to find a MEDPAR record for about 95 percent of the FIM cases where Medicare was listed as the primary payer.

The merged MEDPAR/FIM files enabled us to associate costs with functional status and other FIM data so we could build case classification models and analyze payment options.

We intend to again use probabilistic methods to create merged files for calendar years 2000 and 2001 in order to establish baseline trends describing IRF care before January 2002.
We will request 2000 and 2001 MEDPAR data—the latter to become complete and available around September 2002. (We have to wait until September so we can get the final mortality update for monitoring.) We are planning to ask HealthSouth and UDSmr for their data, which we will merge with MEDPAR using the methods described above.

POST ACUTE CLAIMS

Our requirements for the first year of Phase II of the study, however, go far beyond MEDPAR and PIM. To build the classification models we only had to look at experiences within rehabilitation hospitals. Now we must step back to the acute care setting and ask what happens to patients in certain clinical groups who may require post-acute care but did not necessarily enter a rehabilitation hospital. We will need MEDPAR to identify these patients (e.g., stroke patients, and hip replacement patients) and will use several CMS administrative files on post-acute care (PAC) to see what happens to them when they leave the hospital.

There are five types of post-acute care that we will attempting to measure—Inpatient Rehabilitation Facility (IRF), Skilled Nursing Facility (SNF), Home Health Agency (HHA), and rehabilitation provided in a hospital Outpatient Department (OPD) or by a physician supplier. SNF, HHA, and OPD each appears on its own set of standard analytic files. Each file will appear in at least two versions: Version H for the earlier years, Version I for the present.

- IRF (in 1996-2001: from linked MEDPAR/PIM data; in 2002-2003, from MEDPAR or CMS's bills and the IRF Patient Assessment Instrument). For 2002 and beyond, our file will describe 100 percent of the rehabilitation stays. If necessary, in order to obtain the data in a timely fashion, we will link the IRF PAI data to the discharge claim.

- SNF (1996-2003): Skilled Nursing Facility records also appear in MEDPAR, and we plan to use those.

- HHA (1996-2003): Under the HHA PPS, HHA records will describe 60-day episodes, usually with several visits per episode. They will also describe the services performed during that visit. Therapy services are identified by certain "revenue center" codes. They are not directly
linked to an acute care stay, but by defining timing rules (e.g., a series of visits beginning soon after discharge) one can connect them to such stays.

- OPD (1996-2003): Outpatient Department data are recorded within the Medicare Part B data system. OPD records describe visits—one record per visit—and the services performed during that visit. They are not directly linked to an acute care stay, but by defining timing rules (e.g., a series of visits beginning soon after discharge) one can connect them to such stays. We will restrict selection to cases with a revenue center code that indicates therapy. We will also select all records from Comprehensive Outpatient Rehabilitation Facilities (CORFs), which can be identified from their provider number.

- Physician/supplier files (1996-2003): We will use the CPT codes to restrict our analysis to claims related to therapy services and for selected DME items.

- Provider of Service files (1996-2003): We will use these files to identify salient characteristics of facilities.

Because of the volume of post acute claims, we will create a finder file of beneficiaries who had an acute inpatient stay. We want to avoid narrowing down the finder file in a way that will actually miss some heavy users of post-acute care. We think the safest and most flexible thing to do at this point is to begin with a given year—1999, the latest year now available and the year for which our current MEDPAR files are both complete and well-understood. Then we will:

- build a "finder" file consisting of all MEDPAR cases for that year; and
- acquire the entire SNF, HHA, OPD, and physician/supplier data listed above for that set of cases.
- acquire the full set of 1999 OPD and physician/supplier records for a sample of patients

Using MEDPAR as the finder file should produce Part B files for OPD, HHA, and physician/supplier that are a lot smaller than the original sizes. We will get to these data quickly, compute outcomes of interest, and see whether there are some data that clearly have no utility and could be eliminated. The time spent waiting for the rest of
the data will still be productive, because we will be using 1999 data to develop our analysis programs and to decide what is important to focus on. We may decide that we can build effective finder files defined by DRG (e.g., stroke, hip fracture, lower extremity joint replacement), predicted RIC (based on ICD9 procedure codes), presence at IRFs or SNFs, and a random sample of other cases. Or, we may decide to take all of HHA and SNF, but limit outpatient therapy bills to a sample.

The above files will be sufficient to establish a baseline of post acute care use. We will have six years of such information, and it should be sufficient to identify trends, such as the proportion of cases going to each form of post-acute care by type of impairment. As noted above, all of the files have had format changes during the 1996-2003 analysis period. But the change of greatest importance is the creation of a new measurement instrument for the IRF PPS--the Inpatient Rehabilitation Facility Patient Assessment Instrument. This will give us a full accounting of cases going to IRFs, so we will have data on the universe of cases and potential biases due to missing cases will be eliminated. It will also contain some additional types of measurements, which could help predict costs, including swallowing difficulties, stage of skin ulcers at admission, and pain level indicators. We discuss the types of analysis we plan to do with these data in the following section.
3. REFINEMENT

In this section, we discuss refinement of the CMGs and other payment parameters of the IRF PPS. Actual recommendations for refinement must await the availability of more than a year’s worth of data from the IRF PPS. The calculations require only a year of data. We want that year to begin later than Jan. 1, 2002 in order to allow for provider “learning effects” and the PPS implementation schedule which is based on hospital fiscal year. Thus in the first year of phase II we will concentrate on possible methodological improvements in the payment system. Subsequent years’ work will deal with the implications of the improved data and with calculations related to the actual refinement.

The existing IRF PPS payment system distinguishes between typical cases where the patient is discharged to the community after a stay of at least three days and all other atypical cases (CMS, 2001). The atypical cases account for about 20 percent of total costs. Typical cases were first classified into patient groups that are homogeneous with respect to resource use. Our measure of resource use is log wage-adjusted cost. In the first step used for the existing system, groups were derived which depend on type of impairment, functional status, and age. We call these function-related groups, or FRGs. Then, in a second step, a list of comorbidities separated into three tiers splits most FRGs with the results called FRGCs. Typical cases are thus classified. Adjustments for atypical cases were based on this initial typical case classification and on CMS policy decisions.

Improvements in the FRGCs formulation could propagate through the system. We focus below on ways to improve the overall assignment of costs based on two methods: the application of different fitting algorithms to a larger class of cases and modifications of the input data to the algorithm. The third section below covers updating and refining the weight calculation.

In the upcoming year, we will also explore improvements in the statistical methods we use for facility adjustments. In particular, we will build upon the analyses contained in Carter et al. (2001) which
identified outlier facilities that could be influential on facility payment adjustments. Also this year, we will try to obtain additional Residency Review Committee (RRC) data on indirect teaching so we can assess the reliability of the cost report teaching data, thereby addressing a concern raised at the TEP. The fifth and final subsection below discusses how the new data from the IRF PAI may affect the desirability of updating and refining the facility adjustment.

**Modifications to the Fitting Algorithms**

We plan to study further the methods that we use to form FRGs and estimate parameters. None of these modifications alters the structural form of the payment system but only how we arrive at the best FRG definitions. The improvements include the simultaneous estimation of comorbidity and FRG effects and the inclusion of data on transfer cases in the algorithm that creates FRGs.

**Simultaneously Producing FRGs and Estimating Comorbidity Effects**

We will attempt to improve upon the estimation process for the FRGs and comorbidity tier definitions. The CMGs now in use were defined in two separate stages. We used CART to construct the FRGs. We then tried to detect residual variation in log (cost) that we could attribute to the presence of patients’ comorbidities. However, it is possible that allowing for comorbidity effects would influence the FRGs that CART produces in such a way as to improve the ability of the CMGs to predict costs.

We intend to assess whether creating FRGs while simultaneously controlling for the effects of comorbidity tiers can improve the ability of CMGs to predict cost. If so, we would recommend that CMS use this new method in order to create refined CMGs.

We plan to develop and study a backfitting algorithm to simultaneously obtain FRGs and comorbidity adjustments. We wish to model log (cost) by developing a patient level model having the form

\[
\log(\text{cost}) = f(\text{age, motor, cognitive}) + \beta_1 \text{Tier 1} + \beta_2 \text{Tier 2} + \beta_3 \text{Tier 3} + \epsilon
\]
where \( f(\cdot) \) is a tree structured regression model adhering to the same restrictions detailed in our previous reports (monotonicity, size restrictions, etc.) and the tier covariates are 0/1 indicator variables.

Backfitting is a common optimization procedure used for fitting non-linear models to data. The generalized additive models (GAM), for example, use backfitting to obtain their model components. Backfitting algorithms have the common feature of breaking down the full regression problem into smaller regression problems that could be solved easily if we fix the values of other unknowns in the full model.

In our situation, if we have fixed FRGs then estimation of \( \beta \) is straightforward. We simply use linear least squares without an intercept to model the tree residuals. For a fixed value of \( \beta \) the tree algorithm is simple. We can use the standard CART algorithm with the stopping rule of our choice to model the residuals left by the linear model. Backfitting simply iterates between fitting the linear comorbidity effects and the tree structured FRGs. To summarize, the fitting will proceed as follows:

0. Fix definition of tiers.
1. Fit tree models within RICs to determine FRGs.
2. Compute residuals from the tree models, and regress these residuals on tier dummy variables within RIC.
3. Compute residuals from comorbidity models and refit tree models to these residuals.
4. Go to Step 2 and iterate the process to convergence.

For "smooth" regression problems this process is guaranteed to converge to the optimum since each iteration must reduce the error. Tree structured regression, however, is not smooth since the choice of the number of terminal nodes and the discrete splitting decisions can differ greatly from one iteration to the next. Nevertheless we are confident that some variant of the backfitting algorithm proposed here can satisfactorily converge to a useful and meaningful model.

We have experimented with this algorithm on some small data sets, and it appears to converge quite nicely. The comorbidity regression coefficients eventually do not change from iteration to iteration, and the trees stabilize quickly. This method could potentially affect the
structure of the trees. We intend to see by how much the individual-level predictions vary and in which RICs the simultaneous versus sequential methods make a difference.

Our assessment will also use validation sample distinct from the analysis sample. For example, we will create CMGs on the 1999 data and use them to predict costs found in the 2000 data. We will compare the accuracy of predictions of cost from CMGs created by the backfitting model with that from FRGs and tiers that are created sequentially. If the backfitting algorithm reliably and consistently improves cost prediction, then we will recommend that CMS use it in subsequent refinement of the CMGs.

Transfer Cases

We will also assess whether including data from transfer cases in CMG construction can improve the accuracy of cost predictions while maintaining current transfer case payment policy and the flexibility for CMS to refine that policy as it judges necessary. All transfer cases were omitted during the first FRG construction. It is possible that including data on these cases in CMG construction will improve the ability of CMGs to predict cost.

When including data on transfer cases, we need to reflect CMS’s payment policy. Transfer cases are paid differently depending on whether they are long stay or short stay transfers. The division into these two groups is based on the average LOS of the union of typical cases and long stay transfers. Long stay transfers are paid the same amount normally given for a patient in the same CMG who is discharged to the community. Short stay transfer cases are paid a per diem that is calculated from data on their CMG (CMS, 2001). Since the payment for transfer cases depends on the FRGC assignment, it may make sense to have these cases affect group definitions. After developing the backfitting for the situations previously described, we plan on assessing the value of incorporating the transfer patients that are currently treated separately into the CMG estimation stage.
We can build our FRGs and estimate comorbidity effects while including transfer cases using the cost model

\[
\log(\text{cost}) = f(\text{age, motor, cognitive}) + \\
\beta_1 \text{Tier 1} + \beta_2 \text{Tier 2} + \beta_3 \text{Tier 3} + \\
I(\text{short stay transfer}) \times \log\left(\frac{\text{LOS} + 0.5}{\text{average LOS}}\right) + \epsilon.
\]

Here \(I(\text{short stay transfer})\) is a 0/1 indicator variable that mathematically inserts the per diem adjustment for only the short stay transfer cases. We will consider giving short stay cases a fractional weight in the model as is now done in the relative weight calculation.

The assessment of including transfer cases in CMG construction will use similar methodology to that already discussed for the backfitting algorithm. We will examine the effect on individual CMG definitions and RICs. For example, we may find that including transfer cases may provide enough data to allow detection of additional groups of very expensive cases. We will also examine the accuracy of predictions of cost in validation samples that are separate from the fitting sample. We would assess the accuracy of payments for cases discharged to the community as well as for all cases and hospitals.

By simultaneously estimating the tier effects and groups and by including all of the cases whose payment is influenced by the definition of groups of these components we may enable CMS to choose to move to a payment formula that models hospital costs more closely. At any stage we can enforce policy constraints or fix certain parameters to ensure the system introduces no adverse incentives. For example the 0.5 in the numerator of the prorating formula reflects a policy decision. If CMS should wish to re-examine this policy, we could allow this to be a parameter and thus determine the extent to which a change in this formula could improve payment accuracy.

Validating the Models

As before, we will fit models for each year of data, testing on the other years. The most important tests will involve data for 2000 and 2001.
REFINING CMG STRUCTURE

In addition to the above methodological improvements in grouping we will reassess the input variables to the algorithm including definitions of RIC, comorbidity tier, and FIM indices. In the first and second year we will examine existing data to see if we can improve these items. In the second and third year, we will examine whether the improved coding and additional information available on the IRF PAI allows us to improve the definition of CMGs.

Comorbidity Definitions, RIC Definitions, and Medical Complexity

We will revisit the comorbidity analysis from at least two perspectives. First, we will examine whether the comorbidity effects differ enough under the simultaneous fitting algorithm to cause us to revisit our tier definitions. Second, are there additional variables that can simplify or improve on the way comorbidity payments are made. It may well be true that improved coding of comorbidities in the IRF PAI and the new questions on the IRF PAI can, along with the improved fitting algorithm, lead to a much more accurate way of paying for patients with comorbidities.

There are several ways in which we would like to pursue refinement of the comorbidity tiers and RICs in the first year of our project. First, we would like to systematically address all ICD-9-CM codes in the definition of RIC, rather than just the ones hypothesized by our clinical experts. The cancer analysis presented in our final report (Carter et al, 2001, Section 4) provides a model that could be applied to all the codes in each body system within the ICD-9-CM coding system.

Second, we would like to pursue the possibility of a complexity adjustment that better takes into account the cost of cases with multiple comorbidities. Our report showed that a substantial fraction of the cost of tier 1 conditions is due to multiple comorbidities. We hope to be able to develop a complexity model which accounts for these interactions without having the incentives for upcoding that are found in a simple count model.

Third, many medically complex cases are found in RIC 20, the miscellaneous RIC, as well as among patients with multiple comorbidities
and multiple impairments. By detailing the clinical characteristics of all patients in these groups, we may be able to discover sub-groups that should be treated as separate RICs.

Fourth, we will evaluate how the use of comorbidities in the determination of FRGs affects the definition of tiers and the decisions to use 3 tiers with effects that vary across only RIC.

When the IRF PAI data is available, we will repeat the analyses performed in the first year with the new data to verify how the improved data affects our conclusions. We will also examine new data available on the IRF PAI that could provide useful supplemental measures of comorbidity status and or function. These include pressure ulcers, pain levels, swallowing status and distance walked. All of these, except distance walked, are optional items. We will examine their performance on the subset of hospitals that provide responses to these items on all (or almost all) of their patients. We expect that we will not be able to actually incorporate these variables into CMG definitions, but they may provide important evidence that one or more of these items should be mandatory so that they could be used in case classification.

Finally, we would like to use the IRF PAI data to study the subset of comorbidities that were known at admission. In particular, we would like to assess the presence at admission of comorbidities that were removed from the payment list because they may be preventable. The only possible way to determine which comorbidities are present at admission from the IRF PAI form is to consider ICD-9-CM codes that are found in item 24 and not in item 47. We do fear, however, that the identification of diagnoses present at admission using this method may be inaccurate. It requires double entry of diagnoses found after admission in both items 24 and 47 and hospitals may object to the extra work. Despite the instructions, hospitals may neglect to fill in some diagnoses in item 47 because they do not believe the diagnosis is a complication or because they think that stating that a condition is a complication might be interpreted as poor care. Further, hospitals may not understand the instructions for item 24. Despite these potential problems, it may be that the data from the IRF PAI will be more complete and accurate than the data that we had on the initial database.
Varying Choice of Indices within RIC

In developing FRGs for the initial implementation of the IRF PPS, we noted that MART and GAM models demonstrated that there was some predictive information in the additional indices, but CART had been unable to use it. We want to study alternative indices further. If we preselect indices by choosing the appropriate set for each RIC, then we think that CART might be able to do better than with the existing motor and cognitive scores. We will be looking mostly at the breakdown proposed in Steinman and Jette (1997) of motor scores: transfers, locomotion, sphincter and bladder control, and self care. Within RIC, we will look for clinical indicators that one or more of these indices would be appropriate. We will also fit OLS models on the component scores as well as the indices to see which ones appear to be important. Having preselected the indices for each RIC, we will then attempt to tailor a CART model to each RIC with a limited number of indices given to CART.

Unobserved FIM Items

The new PAI allows staff to explicitly code a subset of items as unobserved. Those patients that have not been observed for most items on the FIM scale will be marked as unobserved rather than at the most dependent score. We will develop rigorous methods for handling this additional scoring information.

The first step to identifying and understanding coding changes will be to examine patterns of "zero response" -- e.g., whether IRF PAI scores of zero are clustered in certain RICs, types of facilities, or patients (e.g., more seriously ill patients). In order to understand the effect that the zero option should have on the definition of groups, we will examine several ways of treating these zeros. We will regard the zeros as missing data and examine alternative ways to impute non-zero values for the observed zero. Currently they are imputed at a value of 1. Alternatives to consider include the average of the item score in the RIC in non-missing cases and the average on other items for this patient. The best answer is likely to depend on which item has the missing value.
We will also look at more sophisticated methods such as multiple imputation (Rubin, 1987). It will be critical to understand why zero codes occur -- e.g., for some cases, zero codes may appear randomly, while for others the presence of a zero code may depend upon the patient's condition or on the "true" but unobserved non-zero value -- and to assess and model these assumptions. Given an understanding of why the zeros occur, we can then estimate what the score would have been if it had been observed and see if the improved scale increases the ability of the FRGs to predict cost.

Unusual Cases

Other CMS decisions about case classifications created: (1) an atypically short stay CMG for non-transfer cases with LOS of less than 3 days; (2) four CMGs for in-hospital deaths; (3) a single bundled payment whenever the same patient returned to the same IRF within 3 days; and (4) supplemental payment for outlier cases equal to an expected 3 percent of total IF PPS payments. We do not anticipate the need for CMS to modify any of these decisions. Nevertheless, when an entire year’s IRF PPS data are available we will examine it to see if there are any unintended consequences of these decisions. For example, we will look at the distribution of short LOS cases (say each of LOS 1 through 6) to see if there has been a large increase in cases with a LOS of 4 days under the IRF PPS. We will also examine the distribution of interruptions to see if their length has changed. The monitoring of LOS and outlier cases is discussed further in section 3.

WEIGHTS

The weights provide a measure of the relative resource use of each CMG so that payment for all cases is proportional to the cost of a typical case in the CMG. Given the expected improvement in the coding of cases and the availability of data from all cases rather than those at selected hospitals, we plan to analyze the effect of recalibrating the CMGs during the third year of our project using the most recent available full year’s data from the linked IRF PAI and claims data. At that time, if any of the activities discussed above lead CMS to redefine the CMGs, we would incorporate these changes too.
The weights used in the final rule (CMS, 2001) were 'hospital specific' relative weights -- one can think of this as if the calculation formula for weights contains an indicator variable for each hospital in the sample. We believed that this was necessary because the TEFRA payments allowed some hospitals to incur substantially greater costs than others. Further, we found much evidence that hospitals did vary in their treatment of similar patients and that hospital specific weights provided more accurate measures of the relative resource use of CMGs.

We will examine whether the IRF PPS has led to a greater homogeneity in the cost of cases in the same CMG at different hospitals. We expect that the accounting cost of the case will still be the best measure of resource use. We would expect to see greater homogeneity under the PPS because: (1) hospitals will improve their coding in response to the new training and the PPS incentives; (2) the PPS will put greater pressure to reduce costs on the most expensive hospitals and allow hospitals who were paid less than average under TEFRA to provide additional resources to patients; and (3) hospitals will modify their charge structure in order to improve their knowledge of the costs incurred by individual patients and, as a by-product, improve our measurement of accounting cost. If such greater homogeneity is found, we will compare using the payment factors at each hospital to control for its costliness with using hospital identity as the control.

The weights used in the final rule were 'decompressed' using a simple, ad-hoc formula based on the relationship at the hospital level between costs and Case Mix Index (CMI). We hope that improved data and methods will eliminate the need for such a fix after recalculation. Further, the staff time measurement study in Section 4 might lead to a redefinition of the accounting cost of each case that reflects the increased nursing time required by medically complex patients and those with very little independence.

**METHODOLOGICAL IMPROVEMENTS IN FACILITY-LEVEL ADJUSTMENTS**

In the upcoming year, we will focus on exploring improvements in the statistical methods we use for facility adjustments. In particular,
we will build upon the analyses contained in Carter et al. (2001) which identified outlier facilities that could be influential on facility payment adjustments. To do this, we will use state-of-the-art methods for identifying outlier facilities and assessing undue influence of facilities on regression parameters. Also this year, we will try to obtain additional Residency Review Committee (RRC) data on indirect teaching so we can assess the reliability of the cost report teaching data, thereby addressing a concern raised at the TEP.

Below, we discuss specific proposed improvements in the statistical modeling of facility adjustments, including examining alternative ways to identify and accommodate unduly influential facilities. We also detail some of the key issues with the variables used to predict facility-level cost that will be examined as part of refinement.

**Identifying And Accommodating Unduly Influential Facilities**

In our Final Report (Carter et al., 2002), we used fully-specified and payment multivariate linear regression models to develop facility-level adjustment payment factors for the IRP PPS. Additional analyses, many of which are described in the Final Report, supported the use of multivariate linear regression for developing facility-level adjustments. However, questions remain about the influence of those facilities which appeared to be outliers and thus were not perfectly described by the linear regression model. A major refinement activity this year will be to explore whether these outlying facilities are unduly influential on the regression payment parameters that are derived from the model, and, if so, how best to modify the multivariate regression model to accommodate these facilities.

It is typical to find a small number of outliers when fitting a linear regression model. Problems occur, however, when an outlier is influential. Thus, we will build upon the analyses presented in the Final Report that identified unusual facilities by assessing the influence of facilities on the payment adjustments. In particular, we will identify unduly influential facilities in the regression analysis - i.e., facilities that are both outliers and influential.
Once unduly influential facilities are identified, an issue will be whether to include them in the analysis or delete them. One would only want to delete them if it is clear that they do not come from the universe of IRFs that we are interested in -- for example, we may decide that including new facilities that have not been subject to the TEFRA limits in the regression model is undesirable, especially if they are unduly influential. However, omitting facilities from the payment model can be undesirable; these facilities are, after all, rehabilitation facilities, so their behavior is reflective at some level of rehabilitation care regardless of their idiosyncrasies. Also, it could be controversial to delete one subset of facilities while deciding to include others. We will analyze and assess the implications of including such IRFs in the facility adjustment model, and provide our findings to CMS.

One way to balance concerns about unduly influential facilities while retaining as much information as possible about the sample of rehabilitation facilities is to fit a linear regression model that identifies unduly influential points and downweights them accordingly; thus, the component of the behavior of these IRFs that is explained well by the overall linear regression model is retained with the rest regarded as error. One such model exists for Bayesian linear regression, which allows for relaxing the assumption that all facilities have errors that come from a common normal distribution; it allows for each facility to have an additional error term that reflects how well-represented it is by the common normal distribution. In this way, the contribution of a facility to a regression is determined in part by an additional variance weight (West, 1984).

We illustrate how this approach will be used in the refinement phase by replicating the fully-specified regression of Table 7.8 of Carter et al. (2002) and the final payment regression model (Model 10 of Carter et al., 2002, Chapter 7). In the first step of this analysis, the fully-specified Bayesian outlier regression model yielded a significant coefficient\(^2\) of 0.423 for indirect teaching, unlike the fully-specified

\(^2\) The 95% Bayesian probability interval for the indirect teaching coefficient did not contain zero.
regression of Table 7.8 of Carter et al. (2002) in which the teaching coefficient was 0.340 and was not statistically significant. The second step was to further examine IRFs that were identified as being extreme (i.e., those facilities that had large variance weights) by the Bayesian outlier regression model; these IRFs tended to be freestanding, larger in size, more urban/large urban, newer (more recent certification date), and proprietary than the sample as a whole. All of the 35 most extreme facilities were established providers and thus were subject to TEFRA limits. The mean of the indirect teaching variable (L1_RTCH, log(1+ir2adc)) among facilities with any teaching activity was 11.69, whereas among the most outlying 35 facilities, it was 7.26. Thus, the amount of teaching is lower among the extreme facilities that have any teaching than among the teaching facilities in the full sample; since these lowered amounts of teaching are downweighted in the outlier accommodation analysis, they are not driving the teaching coefficient toward zero as they are in the standard fully-specified regression.

The third step of the analysis was to revisit the final payment regression model and include indirect teaching this time. As it turned out, indirect teaching was insignificant in both the standard linear regression model and the Bayesian outlier model. While this example supports the use of the current payment regression model, it also demonstrates the potential usefulness of the Bayesian outlier regression model for future refinements of the payment system. The outlier model offered a way to identify facilities that were extreme, and thus provided a starting point for further examining characteristics that might explain why certain IRFs might be identified as outlying and/or influential in the analysis.

While the Bayesian outlier linear regression method is useful for identifying overall outlying and unduly influential facilities and downweighting them accordingly, it does not assess the influence of a particular facility (or group of facilities) on the regression coefficients. Case deletion methods are available for standard multivariate linear regression and can be explored, namely DFBETAS (Belsley, Welsch and Kuh, 1980). However, this method expresses the influence of each case on the regression coefficient in terms of its
standard error; while this can be informative about the mean change, it does not diagnose whether the facility is influential enough to render a variable statistically insignificant. A simulation-based method developed for Bayesian linear regression (Bradlow and Zaslavsky, 1997) that might be worth exploring and would allow one to assess not only the mean change in the payment adjustment variable but also whether the deletion of a facility would change the significance of a payment variable or whether deletion of one facility relative to deletion of another would yield significantly different payment variables. This might help to identify facilities with opposing effects on the regression analysis.

One major issue in the final report was how to distinguish between duly and unduly influential cases - e.g., a large facility might be flagged as influential simply because of its size and not because of its behavior, which is not what we desire to identify. In the fully-specified model of Table 7.8 of Carter et al. (2002), size was included as a predictor in the model, which should help account for this. However, the payment models do not contain a term for size; thus, one possible way to identify unduly influential hospitals in the payment models is to adjust for size by including it as an independent variable in diagnostic regressions.

Additional Data on the Size of Teaching Programs

We will continue to analyze the potential impact of teaching on costs. The TEP panel suggested that further examination of indirect teaching should be a refinement activity. Analysis will depend on having better data on interns and residents that serve in IRFs, and we plan to use data from the Residency Review Committee (RRC) to address this question; one of the TEP panel members offered to help us obtain data from the RRC that might help us assess the accuracy of the indirect teaching data that are available in the cost report. As detailed in the Final Report, there are differences between freestanding and rehabilitation units with respect to what type of information is available on intern and resident counts on the cost report. Data on the
ratio of residents to average daily census are not always consistent with the RRC accreditation requirements for programs in physical medicine and rehabilitation; for example, the expected maximum resident to average daily census should be about 0.125, but there were 15 facilities with a resident to average daily census of 0.20 or higher (Carter et al., 2002).

**ADDITIONAL REFINEMENT ACTIVITIES (YEARS 2 AND 3) ONCE POST-IRF PPS DATA BECOME AVAILABLE**

We will refine our facility adjustments once a sufficient amount of data from the IRF PPS become available for analysis and refinement, which should happen in Years 2 and 3. In light of the data on the new IRF PPS, we will explore the validity of the statistical models that we will develop in Year 1 for facility adjustments, present additional analyses and make recommendations to CMS, and further refine these as necessary. Updated wage index data will be examined when it becomes available from CMS. Issues that were identified in Carter et al. (2001) as requiring further study once post-implementation data become available will be examined, such as the costliness of freestanding facilities relative to units.

Certain facility-level characteristics are worthy of further study. For instance, size and differences in capital costs appear to explain why freestanding hospitals are more costly than rehabilitation units of acute care hospitals. In the fully specified regressions, freestanding hospitals were about 5 percent more costly than units. The higher costs are reflected in the payment simulation results, where the overall payment-to-cost ratio for freestanding hospitals is 0.98. However, the payment-to-cost ratio increases as the hospital’s average daily census increases and is greater than 1.02 for hospitals with an average daily census of 50 or more patients. Average capital costs per discharge for freestanding hospitals are $1,542 compared to $1,145 for units. The issue of whether patterns of care contribute to the higher costs of freestanding hospitals is clouded by uncertainty over coding reliability. Freestanding hospitals currently report fewer comorbidities in each tier and have a length of stay that is about 16 percent longer than the expected length of stay for the patient mix. This issue should
be examined after IRF PPS is implemented and hospitals have responded to
the payment incentives to improve coding practices and to deliver care
more efficiently.

The decision to use the hospital wage index in the IRF PPS is
intertwined with decisions on other CMS payment systems and, therefore,
cannot be fully assessed within the context of the IRF PPS alone.
Nevertheless, some information about the performance of the wage index
in the IRF PPS may be useful to CMS. Just as we had explored wage index
data in the last phase of our research, we will examine updated wage
index data once it becomes available from CMS. In particular, we will
examine whether the wage index is still compressed and how it varies
across facilities and geographic regions. Compression implies that the
wage index overstates the resources required in low wage areas and
understates the resources required in high wage areas. We believe that
the compression might be attributable to a small number of unduly
influential IRFs. When these hospitals are controlled for in the
regression, the wage index is no longer statistically distinguishable
from 1 (Carter et al., 2001).

Further Assessment of Modeling Assumptions

More explicit handling of cases with missing data will be explored;
for example, in the fully specified regression (Carter et al., 2001,
Table 7.8), certification date was missing for four facilities; these
were effectively treated as having an intermediate certification date in
the fully-specified regression, as indicator variables were included for
whether a facility had an old or new certification date. While one
would not expect the results to be sensitive to just four facilities
(especialy since these had a modest number of cases), it is important
to explore the implications of this and similar treatment of missing
data. Alternative methods include logical imputation, when additional
certification data are available, for example, or stochastic imputation,
when the missing data are modeled as functions of related variables
(Little and Rubin, 1987).

Transformation of predictors for which log is taken (teaching, low-
income) was done in the Final Report. Because of the presence of zero-
valued observations, the data were transformed by adding a constant (1 or 0.0001) to their values before the log transformation was taken. There was some concern that the choice of additive constant undesirably influenced the results. We will explore whether alternative specifications that would not involve adding in a constant prior to log transformation, such as taking the square root instead of the log, might be better.
4. MONITORING

A major focus of our ongoing work will be the development of a system to monitor the impact and performance of the inpatient rehabilitation facility prospective payment system (IRF PPS). This system will be used to evaluate the positive and negative effects of the IRF PPS and to determine if the PPS is meeting its intended goals. It will do so in two ways. First, it will track changes in patient access to IRF care and in the costs, quality, and outcomes of IRF care. Second, the system will monitor changes in the care delivered across post acute care settings. This is important because the financial incentives created by the IRF PPS may affect the number and mix of patients using other types of post acute care (PAC). Changes in the payment systems for other types of care may also affect the patients admitted to IRFs. These two components of the monitoring plan are described in more detail below. We conclude this section with a discussion of key analytic issues and the structure of the monitoring system.

MONITORING IRF-SPECIFIC EFFECTS

Our first goal will be to answer a series of questions about the direct effects of the IRF PPS; these include what are the effects of the IRF PPS on beneficiary access to rehabilitation care, the costs of and payments for that care, and the outcomes of that care, and what are the effects of the IRF PPS on individual IRFs? We will also seek to answer questions about how well the components of the IRF prospective payment system are performing, such as how well do the case mix weights predict costs post-implementation; are comorbidities still important drivers of costs; are the outlier targets appropriate; and do the facility adjustments reflect the extra costs of operating in rural areas and serving low-income patients? Baseline measures of cost, access, and quality will be developed over the next year. Post-implementation data will be analyzed and evaluated when it becomes available.
Below, we describe our basic approach to developing the monitoring plan for facility-level characteristics and outline the measures that we believe should be incorporated into the monitoring system. We focus on two major areas of facility-level monitoring of the IRF PPS: how hospital-level activities are affected by payment elements and whether there are changes in coding post-implementation.

**How Hospital-level Activities are Affected by Payment Elements**

Previously, IRF reimbursement depended on a per-case target amount based on the facility’s historical costs and on the facility’s actual cost per case. The Balanced Budget Act of 1997 introduced maximum payment limits that affected some facilities. Now that facilities will receive a fixed payment that varies by patient severity, facilities will have increased incentives to contain costs since their profit will be equal to the difference between the fixed payment and their actual current costs. Thus, it is possible that facilities will have an incentive to treat less expensive patients rather than those who are more expensive, or, more accurately, to treat patients whose conditions fall into more profitable CMGs; conversely, facilities that were at their spending limits pre-IRF PPS might now admit greater numbers of more expensive cases. Because of this uncertainty, it is important to examine the effect of the IRF PPS on beneficiary access to care, costs and payments. Hypotheses to be examined include:

1) Expensive (or, less profitable) patients might experience reduced access to IRF care.

2) Less expensive (or, more profitable) patients might be admitted at increased rates.

3) Facilities that were at TEFRA spending limits pre-IRF PPS will admit more patients in higher weighted CMGs.

4) CMGs with higher variation in cost will exhibit decreased rates of admission relative to CMGs with more stable expected costs.

To address these hypotheses the facility-level monitoring system will include indicators that will measure facility level activities before and after the implementation of the IRF PPS. These will include
measures of payments for IRF care, beneficiary access to IRF care, quality and outcomes of care delivered by IRFs, and the effect of the IRF PPS on various types of IRFs. When developing this set of measures we will keep in mind the key concerns of policymakers about the new financial incentives under the IRF PPS. We will also keep in mind that a challenge to examining these hypotheses is that the data we have may not allow us to exactly distinguish between more versus less profitable patients. Some patient characteristics will be unobserved; for example, we will not have clinical physiological and laboratory measurements for patients, which would allow us to more fully ascertain the severity of their conditions. Below we describe the particular questions we want to focus on related to the four domains that could be affected by the IRF PPS.

Beneficiary Access To IRF Care. Many high-cost and/or low-profitability types of conditions correspond to complex cases -- i.e., patients that are more complex than the average patient in their payment group, or for whom there is more uncertainty about their care costs. It is a concern that these patients may experience reduced access to care once the IRF PPS is implemented. To examine whether this is indeed the case, we will first focus in Year 1 on developing measures and identifying variables that are indicative of access. We will use MEDPAR-FIM data starting with data from 1996. Candidate variables of beneficiary access include discharge rates by DRG, composition of a facility's cases by CMG, and case mix index of a facility. First, we will establish baseline rates and measure underlying trends in the characteristics of patients receiving care at IRFs. In particular, we will look at the fraction of acute care patients discharged to IRFs by DRGs, and assess how this varies over time. We will also track trends in the CMG mix of IRF patients and in comorbidity rates among IRF patients by facility and facility type (e.g., hospitals that were close to or at their payment limit versus those that were not).

Case mix may vary across IRFs in part because medical coverage decisions vary across fiscal intermediaries (FIs); for example, hip replacement occurs less frequently in hospitals belonging to some FIs rather than to others. We will examine whether variation due to FI
significantly affects case mix in IRFs and what this means in terms of beneficiary access to inpatient rehabilitation care.

In Year 1 of this project, our priority will be to establish baseline rates of access to IRF care using the 1999 pre-IRF PPS data. We will subsequently augment these analyses with pre-implementation data from 1996-1998 and 2000 as warranted. Once it becomes available in Years 2 and 3, additional post-implementation data, will be examined to determine whether and/or how access has changed after the implementation of the IRF PPS.

**Costs and Payments for IRF Care.** These potential changes in access to care have implications for costs of and payments for IRF care. In Year 1, we will begin to develop facility-level baseline measures of cost and payments and their relationships to case mix (e.g., CMGs) and other measures of patient severity by examining prior years of data (1996-2001); IRF data for 1996 through 1999 are already in-house, and data for 2000-2001 will be examined when it becomes available to us. We will examine average cost, and payment, and profit (cost minus payment) per case as a function of the percent of patients admitted in the various CMGs at each facility. To assess how the IRF PPS will impact costs and payments for facilities, we will compare these results to data from the IRF PPS once post-implementation data become available in years 2 and 3.

In Years 2 and 3, once post-implementation data become available, we will track the percent of cases that are outliers. We will determine if the loss threshold is actually producing the target amount of outlier payment, and how to change it if not. One of the assumptions of outlier policy is that outliers occur randomly across all hospitals. If, instead, we find outliers are highly concentrated in specific facilities, it might point to a need to either refine the CMGs and comorbidity payments, refine the facility adjustment formula, or increase the outlier payment target (e.g. to the statutory limitation of 5 percent.).

We will also examine the percentage of cases that are coded with comorbidities and/or specific higher-cost FRGs using post-implementation
data once it becomes available; more will be said about this topic in an upcoming section.

**Intensity, Quality, and Outcomes of IRF Care.** IRFs can also try to increase their margins by incurring lower costs in treating patients. Indeed, greater incentives for efficiency are a goal of the IRF PPS. However, IRFs might attempt to reduce the intensity of care they provide (e.g., lengths of stay, amount of therapy provided) down to levels that could adversely affect patients. IRFs could also try to reduce length of stay by discharging patients earlier, sending them home with home health care or to nursing facilities to finish their recovery. This would work against the intent of the payment system to the extent that total Medicare payments are increased by this strategy. Such effects have been found post-implementation of other PPSs; for example, Neu and Harrison (1992) found that the percent of patients who went to SNFs and home health following acute hospital stays increased following implementation of the acute care hospital PPS. Other discharges may have reflected lower quality care, as Kahn et al. (1988) found an increase in the number of patients who were not medically stable at discharge following implementation of the acute care hospital PPS. We will measure and track changes in the intensity of care delivered by IRFs and in the outcomes of that care. We will look at functional improvement achieved by different types of patients using FIM discharge scores and information from the new IRF PAI. In addition, we will track length of stay in IRFs, controlling for changes in the mix of patients seen by those facilities. Candidate measures we will consider are:

- mortality rates
- FIM scores
- cost of care
- acute stay readmission rates
- discharge destination
- length of stay in rehabilitation care
- length of stay in the acute stay preceding rehabilitation stay *(i.e., does "pushback" occur?)*
- discharge destination rates
- readmission rates
- percent transfer cases, including long-stay transfers
- percent patients coded with comorbidities

In Year 1, we will begin to develop appropriate measures of quality and outcomes of care, and examine historical rates using data that are available to us. In Years 2 and 3, we will continue to develop baseline measures, and begin to examine trends.

**Effect of the IRF PPS on Various Types of Facilities.** For the above analyses, we will focus on how these outcomes vary by facility type. We will focus on the following facility characteristics: teaching intensity, geographic location (urban, rural), wage index, freestanding/unit designation, and the low-income measure. We will assess in the pre-IRF PPS data whether and how the outcomes of interest vary by facility type. These analyses will shed light on the adequacy of facility adjustments or highlight areas where changes might be required.

**Identifying Changes in Coding Procedure**

IRFs may change their coding practices. New incentives exist for IRFs to code comorbidities and to assign patients to higher-weight CMGs. Previously, facilities had no incentive to code patient diagnoses accurately or completely. Now, comorbidities will bring extra payments so facilities will likely step up their efforts to code them. IRFs might code beneficiaries into higher weight CMGs and tiers once the IRF PPS is implemented, as higher weight cases will receive higher payments. Further, the training received at the introduction of the IRF PPS may improve the coding of facilities that were unaware of some of the formal rules and improve the uniformity of coding in instances that were previously unclear.

We will examine whether there are other differences in coding under the IRF PPS relative to TEFRA. Acute care hospital data will help us examine changes in coding; we can compare the coding during the acute stay to that in the rehabilitation stay. This will help us disentangle changes in coding from changes in case mix, as well as to distinguish changes in coding due to more accurate practice (e.g., due to clarified
training manual) and upcoding. We will use methods such as that developed by Carter et al. (2002, Chapter 9) for comparing predicted RIC based on acute care stay data to assigned RIC which is based on the IRF PAI data, and compare our predictions to those made previously when the IRF PPS was being designed; of interest is whether a large percentage of predicted RIC items match the RICs assigned using data from the IRF PAI and whether this differs from pre-IRF PPS results. Also of interest will be comparing similar types of patients pre- and post-IRF PPS to see whether coding changes meaningfully for them -- e.g., do their motor and cognitive scores differ.

In Year 1, we will begin to formulate analytical approaches to addressing these issues, such as exploring statistical approaches to handling missing data under a variety of assumptions and methods for comparing coding practices pre- and post-IRF PPS. Most of the work on this issue will begin once we receive patient data from the IRF; much of the work will involve extensive data analysis that will guide our statistical modeling.

MONITORING EFFECTS ACROSS POST ACUTE CARE

Our monitoring system will also be designed to track changes in post acute care delivered across the spectrum of post acute providers, which include home health agencies (HHAs), skilled nursing facilities (SNFs), outpatient departments (OPDs), comprehensive outpatient rehabilitation facilities (CORFs), long term care hospitals (LTCHs), and office-based providers. This is important because the change in financial incentives caused by the IRF PPS might have "spillover" effects on other types of post acute care. For example, if IRFs were to discharge their patients earlier and send them home with home health care at greater rates, the mix of patients using home health care would change. Also, the IRF PPS should function in conjunction with other payment systems to create incentives for beneficiaries to get services in the most appropriate PAC setting.

The key questions we will answer with the post acute monitoring system include what are the effects of prospective payment on post-acute
care before and after the implementation of the IRF PPS? What is their impact on beneficiary access to PAC, Medicare payments for PAC, and PAC outcomes? In addition, we will investigate the causes and outcomes of the wide variation in the use of PAC across markets.

As with the system for monitoring IRF care, our analyses related to monitoring the post acute wide effects of the IRF PPS will have three parts. In year 1 our priority will be to develop baseline models of PAC access. We will also begin to develop measures of PAC costs and outcomes in Year 1. These models (described below) will employ patient, facility, market, and regional-level variables and we will use them to explore the determinants and outcomes of variations in use in the pre-IRF PPS period. The models will then be used to generate a set of key indicators for monitoring PAC that are in line with the concerns of policymakers about the effects of prospective payment on post acute care. When post-IRF implementation data become available in Years 2 and 3 of the project we will track trends in those indicators and follow-up with analyses of problem areas that we identify.

Below we present a review of previous research on post acute care and use it to motivate some hypotheses about the effects of prospective payment on that care. (We will continue to expand this review to ensure that we incorporate relevant findings into our analysis and interpretation of PAC use, costs, and quality.) We also draw on the literature in describing how we will construct our baseline models and measures of PAC access, payments, and outcomes.

**Understanding Post Acute Wide Effects of New Payment Systems**

Other post acute care prospective payment systems, like the IRF PPS, give facilities incentives to contain costs. They may, therefore, limit access to care for patients that are expected to incur more costs than the fixed payment the facility will receive from Medicare for their treatment. Because each of the post acute PPSs has different units of payment and levels of payment for different but overlapping categories of patients, however, one type of provider might be relatively well compensated for a certain patient while another is not. For example, it could be that a hip fracture patient with multiple ADL limitations, good
cognitive functioning, and comorbid breast cancer is a profitable patient for IRFs and home health care agencies but an unprofitable one for skilled nursing facilities. In this case, financial considerations could override a clinical decision that a SNF was the most appropriate post acute care site. In addition, were the patient admitted to an IRF or a SNF that facility could have an incentive to discharge the patient as soon as possible—the SNF because each day of providing expensive cancer drugs causes them a loss and the IRF because the facility receives no extra payments for extra days of care. Thus, the different post acute prospective payment systems may create complex sets of financial incentives. This could be particularly problematic since there is currently little clinical consensus about and wide variation in where patients are sent for post acute care. Financial incentives may, therefore, have undue influence on patient referrals.

Effects of prospective payment systems on post-acute care have been documented. In a study of the effects of the original acute care PPS, Neu and Harrison (1988) and Steiner and Neu (1993) compared PAC use before and after the hospital PPS implementation using a 20 percent random sample of all Medicare discharges from acute care hospitals. They found that the rate of SNF utilization increased rapidly immediately following implementation, but then decreased 3 years later. Home health care use continued its pre-PPS growth trend so that by 1988 14.7 percent of all Medicare discharges went to home health, a 10 percent increase from 1985. The number of visits per home health user also increased by about 30 percent and covered charges for home health care per hospital discharge rose 370 percent. The authors also found discharges to IRFs nearly doubled between 1985 and 1988 (Steiner and Neu, 1993).

The Office of Inspector General for the Department of Health and Human Services has studied the more recently implemented HHA and SNF prospective payment systems. According to the Inspector General, the HHA PPS does not appear to limit access to home health care. In 2001, 89 percent of discharge planners report that under PPS they can place all of their Medicare patients who need HHA and 90 percent report no change in number of patients whom they cannot place since PPS was
implemented (OIG, 2001). These results are consistent with the prior two OIG inspections (1999 and 2000) on access to home health care. There have been no substantial decreases in the 13 most common DRGs discharged to HHAs between 1997-2001 (OIG, 2001). The SNF PPS also does not appear to limit access to skilled nursing facilities. In 2001, 73 percent of discharge planners report that under PPS they can place all of their Medicare patients who need care in SNFs. (An additional 20 percent were able to place all but 1-5 percent of patients.) Eighty six percent report no change in number of patients whom they cannot place since payment increases were implemented. There were no substantial decreases in the 10 most common DRGs discharged to SNFs between 1997-2001 (OIG, 2001). These analyses do, however, raise concerns about the small set of patients who are difficult to place. If these were the most seriously ill or vulnerable patients, then even though they constitute only a small fraction of the population it would be a serious policy matter.

In addition to changes in access and use at a given PAC site, it is fairly clear from prior research that there is the potential for practical substitution among PAC settings. (Practical substitution is the term Neu et al. (1989) used to describe the observed dynamic that areas with relatively little of one kind of post acute care are characterized by relatively heavy utilization of other kinds of post acute care. Practical substitution does not imply medical substitution - i.e., that care in one setting is as effective as care in another setting for any particular condition. For example, IRFs are required to provide three hours of therapy per day while SNFs generally provide much less.) Neu et al. (1989) observed several instances of practical substitution. For example, they found that at an MSA level, as use of SNF care increases, the use of home health care and rehabilitation care decreases. In a later related study, Steiner and Neu (1993) found that SNF and IRF care act as practical substitutes for one another. They also found that rates of rehabilitation use are fairly independent of rates of home health care use, except for the more home health care there is in an area, the higher the probability of discharge to an IRF. Cohen and Tumlinson (1997) observed that there is an inverse
relationship between the number of long-term care hospitals and SNFs in a state and the use of home health care. Home health thus serves as a substitute for long-term care and for SNF care for some. Swan and Benjamin’s study (1990) shows that as the number of nursing home beds increases in an area, Medicare home health use decreases. Kenney and Dubay (1992) observe that higher proportions of Medicare enrollees use home health care in areas with fewer nursing beds, suggesting practical substitution.

In addition, previous research has shown that policy changes in addition to payment reform can have substantial effects on service use. In 1986, for example, fiscal intermediaries were forbidden from using “rule of thumb” rules to deny SNF coverage. As a result denial rates for SNF claims dropped and Medicare expenditures for and utilization of SNF care increased (Liu et al., 1999). The clarification of Medicare benefit language to include home health care that was either part-time or intermittent (instead of part-time and intermittent) also increased home care use. The dollar limit on outpatient therapy in place during 1999 is one example of a policy change that could have caused a shift in PAC use during our baseline period.

Given these precedents, our hypotheses about the effects of prospective payment systems on post acute care include:

1. Prospective payments systems for post-acute care might limit the care available to complex cases.
2. Financial considerations might override clinical considerations in choices among post acute care sites.
3. Patients might be shuttled among PAC sites to increase Medicare payments.

In addition, in order to understand the implications of these hypotheses for patient outcomes we will undertake analyses that will shed light on the degree of medical substitution among PAC sites.

In order to develop additional hypotheses about the wider effects of the IRF PPS, we will analyze each of the post acute payment systems and further synthesize the published literature about PAC care (more of which is discussed below). We will create a comprehensive database that documents the payment systems used for each type of post acute care over
the period 1996 through the end of our data collection period in 2003. Using this information, we will develop specific hypotheses about how changes in the number and types of patients receiving care in each post acute setting will respond to payment system changes.

**Post-Acute Care Access**

A series of hypotheses about the post-acute care market will set the stage for the development of a set of indicators for monitoring access to PAC care after the implementation of the IRF PPS. It will also help us to create a set of baseline models that can generate appropriately case-mix adjusted and benchmarked indicators for PAC access at the market and national level. Prior research on PAC access provides us with a rich set of predictors of use that we can include in our baseline models.

**Patient level predictors.** Many researchers have fit patient-level models of the use of post-acute care and choice of PAC site. They have found that age, gender, race, marital status, functioning, medical condition (DRG), and comorbidities influence the site to which patients are discharged. Liu et al. (1998) found that use of post-acute care was highest among people with Alzheimer’s disease and Parkinson’s disease and that 70 percent of people from DRGs involving hip fracture and 50 percent of the stroke subgroup received post-acute care. Use of post-acute care was positively associated with age and negatively associated with being married. Likelihood of SNF use rose with increasing functional disability; but being married or nonwhite was associated with lower SNF use relative to home health care use. Being on Medicaid was not significantly associated with SNF rather than home health agency use. Lee et al. (1997) observed that patient variables explain 48 percent of the variation in SNF utilization across MSAs but account for only 25 percent of practice variation across MSAs in IRF, home health agency, and ambulatory rehabilitation. Kane et al. (1996) found that living alone and functional dependency at discharge were the most common significant predictors of which PAC setting a patient would be discharged to. In studying the impact of hospital ownership of PAC providers, Blewett et al. (1995) observed that patient characteristics
had a more consistent and significant relationship to whether a patient received PAC upon discharge than did hospital ownership. ADL dependencies assessed just prior to discharge were a significant predictor for stroke, hip, procedure, and hip fracture cases; and functional dependencies as measured by physical demonstration of ADLs just prior to discharge were significant for hip fracture and hip procedure patients. Other significant predictors of PAC use in several DRGs include whether a patient lived alone or had a caregiver prior to hospitalization. Neu et al. (1989) and Steiner and Neu (1993) found that patterns of utilization differed across post-acute care settings. SNF users are older than users of post-acute care in other settings and older than patients who do not use post-acute care are. For rehabilitation care, age based-patterns seemed to depend on DRG. Patients with secondary diagnoses were more likely to use post-acute care than were patients without recorded secondary diagnoses. Whites were significantly more likely to use SNF care than blacks, while blacks were significantly more likely to use home health care than whites. (For IRF care there were no significant differences by race.) Longer lengths of stay in acute care were predictive of a greater probability of post-acute care utilization. Medicaid eligibility was consistently associated with a lower likelihood of using IRF care.

Facility level predictors. Researchers have also found that some characteristics of the discharging acute care hospital affect PAC use and choice. Blewett et al. (1995) found that hospital ownership of PAC was a significant predictor of PAC use. Patients discharged from hospitals that owned PAC providers were more likely to use PAC than patients discharged from hospitals that did not own PAC. Neu et al. (1989) found that patients discharged from teaching hospitals, as opposed to nonteaching hospitals, were no more likely to use SNF, somewhat more likely to use home health, and markedly more likely to use rehabilitation care. Hospital size in their study had some relationship to utilization of SNF and rehabilitation care, but little relationship to home health care use. They also found some support for the proposition that differing lengths of hospital stays reflect differing practice patterns rather than different patient populations. Longer
average hospital lengths of stay led to a lower probability of SNF care, home health care, and/or IRF care for selected DRGs. Medicare patients discharged from urban hospitals were more likely, overall and within each selected DRG, to use post-acute care of all three types than were patients discharged from rural hospitals. Ownership of the discharging hospital had some effect on whether and where patients receive post-acute care: government hospitals are less likely to refer patients to all kinds of post-acute care than are privately controlled hospitals. Nonprofit hospitals were somewhat more likely to refer patients to rehabilitation care than are for-profit or government-owned hospitals. Hospitals with "disproportionate shares" of low-income patients seem slightly less likely to refer patients to SNF care than are other hospitals. In a later related study, however, Steiner and Neu (1993) found that being discharged from a teaching hospital had only marginal effects on the patient's probabilities of using PAC and that the effects of hospital size were mixed.

**Market and regional level predictors.** There are also a number of market and region-level predictors of PAC use. Swan and Benjamin (1990) found that the number of nursing home beds was a negative predictor of Medicare home health use and that the number of home health care agencies per population was positively related to home health utilization. Cohen and Tumlinson (1997) found the percentage of a state's budget allocated to Medicaid was significantly and positively related to the number of Medicare home health users per enrollee and the number of Medicare visits. They also found that when Medicaid home health expenditures decline, the number of Medicare home health users and the number of visits increases. Lee et al. (1997) found that market variables explained 19 percent of geographic practice variability for home health care use. Liu et al. (1998), however, found that neither the numbers of HHA nurses per 1000 enrollees nor the numbers of Medicare SNF beds per 1000 elderly people, nor IRF beds per 1000 enrollees were significantly associated with PAC use—but the level of availability of Medicaid home health was a predictor of use. Kenney and Dubay (1992) found that higher proportions of Medicare enrollees used HHA care in areas with fewer nursing beds, higher hospital discharge rates, shorter
mean lengths of stay, more HHAs per enrollee and in areas located in the New England and Mid-Atlantic regions and/or that are urban. The average number of home health visits was higher in areas where there were more HHAs per enrollee and where a higher share of agencies was proprietary.

Neu et al. (1989) found negative relationships between average hospital length of stay in an MSA and propensity to use PAC. They also found that higher income communities had higher utilization rates of SNF and home health care. SNF bed supply per 1000 elderly population and an increase in hospital-based SNF beds had significant positive relationship with SNF utilization. They also found the propensity to use post-acute care in different settings varies dramatically from state to state. Steiner and Neu (1993) found an even stronger negative relationship between the health market area average length of stay and PAC use; that is, the longer the market area’s average length of stay, the lower the probability of a patient using PAC. However, the number of Medicaid-funded SNF days per 1000 elderly in a particular state showed inconsistent associations with the probability of SNF use: this result differed from Neu et al., who found a consistently negative relationship between Medicaid funded covered days and Medicare covered SNF services across DRGs.

**Access and Utilization Trends and Indicators.** In Year 1, we will develop models of access to PAC care that take into account individual, facility, and market effects as well as the effects of payment and policy changes. These models will be used to generate indicators and track trends in access to care in Years 2 and 3 of the project. Indicators will likely include the fraction of acute care patients receiving care at each type of post acute care setting (for selected DRGs) by market and for the nation as a whole. We know that there is currently wide variation in the use of post-acute care across the country. We will thus use market characteristics to try to understand the causes of this variance in the pre-IRF PPS market and why markets differ in their response to the IRF PPS. We will also monitor trends in discharge settings after IRF care. Specifically, we will track national and regional changes in use of other types of post acute care after IRF discharge, including the number and intensity of services used.
PAC Payments and Outcomes

The models we develop of PAC access and utilization will tell us what drives use of care, but not whether the care patterns or changes in care patterns we observe have positive or negative effects for beneficiaries or the Medicare program as a whole. To evaluate the changes in utilization we will develop and use two types of outcomes metrics: 1) patient care outcomes; and 2) overall costs to the Medicare program. The literature on the costs and outcomes of PAC care is more limited than that on PAC access. In addition, it is generally confined to observational studies and thus the results are confounded by the non-random assignment of beneficiaries to sites of care and/or treatments.

Patient Outcomes. Few studies have examined patient outcomes in post-acute care settings (Keith et al., 1995; Liu et al., 1999; MedPAC, 2001). What little research there is in this area has focused on outcome differences across post-acute care settings. In comparing post-acute care settings, Kane et al. found that rehabilitation facilities, relative to SNFs and home health care, demonstrated the greatest potential for functional improvement for stroke patients who had high ADL dependency scores at discharge (Kane et al., 1997a). In a related study, Kane et al. compared patients’ actual discharge locations with those that would have produced “optimal outcomes.” The optimal discharge location was determined as that which produced the greatest improvement in function after adjusting for patient’s baseline characteristics. They found that a very low percentage of patients (23 to 50 percent) were placed in the post-acute care setting that would provide the most functional improvement (Kane et al., 1997b). In studying whether outcomes and costs differ for elderly patients admitted to rehabilitation hospital, subacute nursing homes, and traditional nursing homes, Kramer et al. concluded that the more comprehensive therapy services provided by rehabilitation facilities can lead to better functional recovery and community placement for stroke patients but not hip fracture patients. Keith et al. compared acute and subacute rehabilitation for stroke patients, and found that acute rehabilitation patients showed substantially greater gains in functional measures (Keith et al., 1995). Another study found that rehabilitation patients
had lower death rates and rehospitalizations than those treated in subacute SNFs (Liu et al., 1999) but most outcomes were not significantly different for patients treated in subacute SNF and inpatient rehabilitation after controlling for patient characteristics. Retchin et al., (1997) found that while HMO and FFS members had different discharge rates to SNF and IRF settings, no significant differences in relative risk of dying were found between HMO and FFS groups. Munin et al. performed the single major study focusing solely on inpatient rehabilitation facilities. The researchers compared effects of time lapse between surgery and admission to rehab. Patients undergoing elective hip or knee arthroplasty who began inpatient rehabilitation 3 days after surgery attained short-term functional milestones in fewer days using fewer total resources compared to patients who began rehabilitation 7 days after surgery (Munin et al., 1997).

Costs of Care. Shifts in the use of PAC care have cost implications for the Medicare program, facilities, and beneficiaries. In the Munin et al. (1998) study earlier rehabilitation yielded faster attainment of short-term functional milestones in fewer days using less total cost. Keith et al. (1995) found that acute rehabilitation patients had twice as much treatment during a stay, twice the daily treatment hours, and twice the average charge per day. Kramer et al. (1997) concluded that the more comprehensive therapy services provided by rehabilitation facilities come at much higher costs that are justified in the case of stroke. Kane et al. (2000) concluded that nursing homes were associated with poor outcomes and high costs compared to home health and inpatient rehabilitation. We will thus need to track the costs associated with patterns of PAC use over time using Medicare claims data, and examine the relationship between cost and outcomes.

Models and Measures of Cost and Outcomes of Post Acute Care. We will attempt to measure improvements in health status associated with care in each post acute setting both at baseline and over time. The measures of health status that we will consider include mortality rates, acute hospital readmission rates, and length of stay in Medicare-covered care. These measures will be tracked across markets and over
time to see if the rate of use of types of care is associated with better outcomes. We realize, however, that these measures of outcomes are very crude. We will, therefore, seek to obtain data, that will allow us to measure outcomes, such as institutional status, at fixed intervals following an admission (e.g. 3 months). We will also explore what would be required to undertake the collection of these types of outcomes data.

We will assess the cost implications of these differences for IRFs, for post acute care providers generally, for beneficiaries, and for the Medicare program as a whole. IRF costs will be examined using their cost reports. We do not plan to estimate changes in costs of other post acute care providers directly, but we will examine how their case mix might be changing using information from patients’ acute care stays. Total Medicare program payments and to beneficiaries for “episodes” of post acute care will be examined using Medicare claims data.

In year 1, we will begin our work on developing measures of costs and outcomes. We will explore the utility of the candidate outcomes measures mentioned above. We will also work to develop measures of costs for PAC episodes involving rehabilitation. For example, we will need to develop measures of office-based PAC therapy. To do this we will review a sample of outpatient claims to determine whether a measure of office costs based only on therapy HCPCS/CPT codes is appropriate, or whether a clinical review of the non-therapy CPTs billed during the period that the office-based therapy is delivered reveals that additional services are part of a comprehensive rehabilitation management plan. We will then refine our list of codes considered PAC therapy for our 2nd stage data request.

**ANALYTIC ISSUES**

In each of these analyses we will have to account for differences in patient severity and circumstances so that we correctly attribute changes to the IRF PPS versus other factors. This means that we must model the factors that affect facilities, costs, and outcomes of care and undertake analyses that account for the fact that patient severity
is not entirely observable. We will also have to exercise care in choosing appropriate baselines given the level of change in the market and evolving medical technology. Finally, we will have to use rigorous methods to determine “problem” levels from statistical and policy perspectives.

For all of the analyses above, we want to establish baseline estimates based on historical, pre-IRF PPS data that reflect current practices. These baselines will allow us to measure changes post-IRF PPS. However, we will also want to establish standards of comparison that reflect clinically meaningful levels of quality of care. Where appropriate, we will consult the literature, clinical consultants, and CMS for clinical, rather than empirical, standards of care.

In developing baseline estimates for facilities and markets and for the purposes of fairly comparing facilities by accounting for differences in case mix etc., regression models will play an important role. Using the example of cost, regression models can help one obtain a measure of average cost per case at the facility level, while accounting for case mix and other characteristics. Hierarchical models go a step further; in addition to obtaining mean costs and payments across all facilities, they can be used to obtain average costs and payments per facility (Goldstein and Spiegelhalter, 1996), while accounting for variation due to unique, unobservable characteristics of facilities.

We will explore Bayesian hierarchical models for monitoring. Advantages are flexibility -- one can readily develop meaningful measures of hospital behavior and outcomes using the results of the fitted model rather than developing separate sub-analyses and tests -- and the ability to correctly model multiple sources of variation (e.g., patient-level, facility-level, MSA-level). Several authors have demonstrated the strengths and benefits of using such models for monitoring a variety of outcomes among health care providers. Normand et al. (1997) use hierarchical Bayesian logistic regression models to develop performance indices to profile medical care providers with respect to an indicator outcome variable of 30-day mortality; their two-stage model accounts for patient characteristics at one level and
facility characteristics at the second level of the model in one of two ways: 1) patient characteristics vary as a function of facility characteristics and 2) baseline rates are assumed to differ across facilities but the relationship between outcome and patient characteristics is the same for all facilities. Further, they develop measures of absolute and relative performance of providers by using the posterior distributions of model parameters; some of these measures assume the current patient population does not change, while others compare provider performance for particular types of patients (e.g., severely ill patients, rural patients, etc.). Christiansen and Morris (1997a,b) use a hierarchical model to estimate true mortality rates while adjusting for case mix at facilities and describe how to interpret the resulting posterior distributions of true, underlying rates to determine the probability that a facility is sub-standard, relative to a pre-established, clinically acceptable standard for mortality.

Our analyses of choice of PAC site will be implemented using conditional logit regression models (McPadden, 1973). These models will be used to estimate the probability that an individual chooses a given PAC site of care (or no PAC.) This probability will depend on patient characteristics and the characteristics of the PAC choice set available. In addition, we will explore adding characteristics of the discharging hospital and market area to the model since there is evidence that PAC choice reflects a combination of health system and individual preferences. We will test whether nesting is necessary in these models (e.g., whether the decision to access PAC care is separable from the choice of a particular PAC site). Estimation of these models requires variation across individuals in their choice sets; this will come from both time series and geographic variation.

In an effort to understand the outcomes of the existing variation in PAC use we will fit models that exploit the change in the PAC sector and the geographic variation in practice patterns. In cross-sectional studies of PAC outcomes it is impossible to disentangle the effects of patient selection from the effects of treatment. For example, home health care patients might have lower readmission rates than SNF users but this would be more plausibly due to healthier patients returning to
their homes than the effectiveness of home health care versus SNF care. We will exploit two sources of natural variation that will allow us to control for such unobserved dimensions of selection. First, each of the previous PPS implementations likely caused shifts in the distribution of patients across PAC sites. This gives us the opportunity to look at costs and outcomes before and after those shifts. In addition, geographic variation in practice patterns allows us to assess the effects of greater or lesser proportions of the beneficiary population using PAC of the various types. We will exploit such differences in treatment patterns to draw conclusions about the added benefit of certain types of post-acute care for various conditions. For example, by comparing the outcomes of joint replacement patients in the west where IRF care is rare, to the outcomes of joint replacement patients in the east where it is more common, we can draw conclusions about the added benefit of more intensive rehabilitation. These conclusions will be valid to the extent that we have measured the important outcomes of joint replacement and that patients are similar across regions on unobservable dimensions of patient severity; we will adjust for observable differences as appropriate for the analysis. Finally, we will use both sources of variation to estimate "difference-in-differences" models that can control for selection biases across areas and unobserved time effects. This work will depend heavily on our ability to develop suitable claims-based outcomes measures.

We will ultimately have six years of baseline, pre-IRF PPS data (1996-2001); in addition, each year of IRF PPS implementation will yield yet another year of data to analyze under the monitoring system. Year-to-year variation will be particularly important to capture, especially when assessing the change in the IRFs pre- and post-implementation and in the early years of the IRF PPS; facilities' behaviors might change as the system is phased in (which will occur in a staggered manner during CY 2002) or shortly after implementation. For this reason, we will explore whether time series models should be used as part of the monitoring system (West and Harrison, 1997). Bayesian hierarchical time series models have been implemented to monitor outcomes in the Veterans Administration system, namely 30-day return rates for three
types of conditions at 136 hospitals over 10 years (Burgess et al., 2000). These models explicitly account for year-to-year dependence of facilities while adjusting for case mix.

A variety of software exists with which we might fit these models. Christiansen and Morris (1997a) have made available S-Plus code freely and publicly for fitting hierarchical models to rate outcomes; Burgess, Lourdes, and West, 2000 and Normand et al. (1997) fitted many of their models using the Bayesian Inference using Gibbs Sampling (BUGS) software package (Spiegelhalter et al., 1999); and other hierarchical modeling programs may be useful, such as SAS PROC MIXED or the GLIMMIX macro (Littell et al., 1996), HLM (Bryk et al., 1996), or MLwiN (Rasbash et al., 2000).

Finally, when fitting any of these models and computing adjusted mean estimates for facilities, it will be important to adjust only for those characteristics that are outside of a facility’s control. We will be certain to address this question in the upcoming year as we identify outcomes for monitoring and characteristics for which we would adjust observed outcomes, such as patient severity of illness.

In Year 1, we will focus on building models of access to PAC care and post-IRF care. We will also explore hierarchical models for monitoring outcomes by fitting models to data we currently have. This will be part of the process of determining baseline measures for the monitoring system and adjusting appropriately for patient, hospital, and market characteristics, as described in the previous sections. If there is time in Year 1, we will also explore whether time series models will be useful for long-term monitoring of the IRF PPS and post-acute care, though it is likely that most of this work will occur in Years 2 and 3. The same is true of our analyses of geographic variation in PAC use and outcomes as they will build on the access models and outcomes measures developed in Year 1.

THE MONITORING SYSTEM

The monitoring system will provide indicators of performance at the level of the individual rehabilitation facility, the rehabilitation market, the post acute care market, and the Medicare program as a whole.
The recommendations for the design and implementation of the monitoring system will include specific rules for each indicator including the calculation formula(s), the source of each data element, and the level of the indicator (e.g., hospital vs. MSA). For hospital statistics, we will include a method to determine if changes in the indicator are statistically significant. As discussed above, we will monitor changes by IRFs in coding comorbidities or assigning patients to higher-weight CMGs and track the percent of cases that are outliers. We will tabulate (or estimate) indicators for years between 1996 and 2001 as appropriate. This will provide baseline measures, as well as providing information on trends and year-to-year variability.

Our design for the monitoring system will not only describe the indicators, it will also show how they can be used together to obtain a clear description of access, outcomes, and costs at inpatient rehabilitation and other post acute care facilities.

The results of these inquiries will be fed back into our efforts to refine the IRF PPS and to understand the far-reaching effects of payment system changes. The system of indicators that we develop will hopefully form the basis for a monitoring system that will give policymakers an understanding of how the system is performing on an ongoing basis and whether it is furthering the goals of the Medicare program.
5. SPECIAL STUDIES

In this section we discuss our plan for several special studies that we will conduct during the first year of phase II. Each study should increase understanding about particular aspects of IRF care and/or the resources that it requires. We will examine how staff time varies by patient characteristics, the ability of MDS-PAC questions to predict costs, the measurement of function, and the role of technology in changing case mix and costs in IRFs. Although some of these activities may affect either refinement or monitoring activities, the connection depends strongly on the findings from the studies.

We expect to conduct additional studies in subsequent years, and to continue with additional work on at least one of the topics studied in the first year of Phase II. The topics of special studies in subsequent years will be chosen during discussions with CMS and our TEP. For example, we might want to examine the costs and payments provided to sole community inpatient rehabilitation providers. We will update our workplan to include these studies as they are sponsored.

STAFF TIME MEASUREMENT STUDY

The Aspen Systems Corporation, as part of a HCFA-funded study, is completing preparation of staff time data that they have gathered from inpatient rehabilitation facilities. The data come from 8 units in 12 hospitals, and cover about 600 patients -- approximately 75 to 80 percent of whom are Medicare beneficiaries. In each unit, they have 24 hours of staff time measurement for 4 weeks. Other data on these patients include admission and discharge date, interrupted stay information, and ancillary charges. The MDS-PAC and FIM were both completed for patients in the study.

We propose to analyze this data to examine whether it explains compression and whether using information from this study could increase the accuracy of our relative weights. One of the potential problems with the departmental method that we have used to estimate the cost of each case is that it assumes that all routine costs, including nursing
costs, are incurred at the same per diem rate. However, it is plausible that medically complex patients require more nursing staff time and that patients with lower function require more nursing staff time than more independent patients.

The effects will be important if predictable variation in nursing costs is similar in magnitude to that found within nursing homes. The original weights calculated for the RUGs system showed that, in the ultra high rehabilitation use categories, the cost of nursing services in the group with most ADL limitations was two-thirds higher than in the group with the fewest limitations. The ultra high rehabilitation groups are for patients receiving the greatest amount of therapy and therefore most like IRF patients. Further, patients with extensive clinical needs such as ventilator patients and tracheostomy patients had exceptionally high nursing costs.

The magnitude of the distortion in relative costs that is introduced by this assumption will be estimated from a model that predicts nursing time for each patient from the patient characteristics. Should the distortion be significant, as we believe it may, we will evaluate ways to improve the weights. Finally, to the extent feasible given the data, we will compare the staff time measurement data on therapy and ancillary services to charges and costs for the same services.

Compression

We show in our final report for phase I (Chapter 6, Carter et al., 2001) that the CMG relative weights for typical cases exhibit CMI compression—i.e., hospitals with high CMIs have higher standardized costs relative to their CMI than hospitals with low CMIs. A possible explanation for the CMI compression is that the resources needed by CMGs with high relative weights are underestimated by our relative weights and those with low relative weights are overestimated. One of the possible reasons for such weight compression is that each hospital is

---

3 Calculated from Table 2E in DHHS (1998) from the nursing index for group RUC (1.30) and RUA (0.78). 2/3 = 1.30/0.78 - 1.
assigned just one per diem for routine cost, yet per diem nursing costs may vary by FRGC.

In the same report, we recommended an ad-hoc method for "decompressing" the weights. Because average hospital cost per case is measured more accurately than average cost per case in a CMG, it is likely that this method results in improved weights. However, it would be much more satisfactory to measure cost per case more accurately. Understanding how nursing cost varies with case mix may allow us to do so.

Research Plan

We will study how nursing costs vary with the patient characteristics related to CMG assignment: primary impairment, FIM motor score, FIM cognitive score, the presence of a relevant comorbidity, and age. The FIM scores will be estimated from the FIM data. In addition, we will examine whether staff time is correlated with the response to items that appear on both the MDS-PAC and the new IRF PAI but not the FIM, such as pressure ulcer stage and degree of pain. If sample size allows, we will also examine whether staff time is related to the clinical characteristics that were found to greatly increase staff time in calculation of the original RUGs weights such as ventilator status, tracheostomy status, and I.V. feeding and medications.

If we can build a model which predicts nursing time as a function of patient characteristics that are found on the FIM (or IRF PAI) and MEDPAR, we may be able to improve our weight calculation. We would estimate relative nursing cost for each patient in the FIM/MEDPAR database using this model. Then we would go from this estimate of nursing cost to an estimate of the routine cost per diems for each patient. The details will depend on the model for nursing time.

Once we have relative routine care per diem costs per case, we can allocate the observed routine costs at each hospital across the patients at that hospital and thus re-estimate the cost of each case. Then we

---

A potential problem would arise if the sample data suggests that the allocation of total nursing cost per day depends substantially on the composition of non-Medicare patients at the particular hospital.
can recalculate the average cost of a typical case in each CMG and the resulting weights. We can compare these improved average costs to the "decompressed weights" or use them to directly calculate improved weights. Final application of the model would wait until data from the IRF PAI are available.

**ASSESSING THE PREDICTIVE VALUE OF MDS-PAC ITEMS**

As has been demonstrated repeatedly, the FIM items provide scales that are highly correlated with LOS and subsequently with cost (Stineman et al, 1994, Carter et al., 1997, Carter et al., 2001). Nevertheless, as we will discuss below, there are concepts that are not measured on the FIM that might add to our ability to classify cases into groups with homogenous resources. Further, there may be more powerful ways to measure either FIM concepts or conditions indicated by ICD-9-CM codes.

The MDS-PAC included both more items and more concepts than the FIM. Thus, it has the potential to provide enhanced explanatory power beyond what can be achieved with the FRGCs. As all data collection is costly, it is important to understand the potential contribution of additional concepts. Thus our approach will be to examine the additional explanatory information provided by items from the MDS-PAC.

This study may identify a limited set of MDS-PAC items that could be either added to the IRF-PAI or that could replace existing mandatory items on the IRF-PAI. Some of these items are now actually on the IRF-PAI, but are voluntary and might be made mandatory. (See section on refining CMG structure in Chapter 2 for how we plan to study these items after the IRF-PAI data become available.) An alternative outcome of the study might be that we find no such items. Such a finding would increase confidence in the existing payment system.

**Background**

A number of experts have expressed concern over the adequacy of the FIM cognitive measures. The MDS-PAC offers a significantly expanded set of cognitive items that may outperform the FIM cognitive items. Morris and his colleagues (Morris, Fries, Mehr, et al 1994) have developed a cognitive performance scale, the MDS-CPS, using these data. The MDS-CPS levels correspond closely to Mini Mental Status Exam and Test for Severe
Impairment scores. This may be important, as work by Eilertsen, Kramer, Schlenker, et al. (1998) indicated that the Mini Mental Status Exam together with FIM motor score and a measure of depression behaviors has substantially greater explanatory power for stroke cases than age and motor score alone as used in the original FIM-PRGs. Work by Hartmaier and her colleagues indicates that the MDS-COGs, an alternative MDS cognition measure, may be an even better predictor of the Mini Mental Status Exam and Global Deterioration Scores. Unfortunately, this measure cannot be scored directly from the MDS-PAC and will need to be rescaled.

In an earlier Rand project, (Carter et al 1997), an expert panel assessed the adequacy of the FIM as the basic tool to provide information for patient classification. That panel concluded that better measures for the following concepts were needed: 1) problem solving; 2) attention/concentration; 3) oral comprehension; 4) endurance/deconditioning; and 5) social support. With respect to the first three cognitive concepts, the MDS-PAC items differ from the FIM, and we will need to test whether they are better for classification purposes. The MDS-PAC includes an item on stamina and change in stamina since the time immediately prior to the precipitating event. Measures of social support are more direct and precise in the MDS-PAC. While it was recognized that measures such as social support may not be appropriate for classification for payment purposes, it may still be important to understand their importance as predictive variables.

Other concepts mentioned but with less consensus by the same Rand expert panel were: 1) producing written language; 2) orientation and short term memory; 3) long term memory; 4) producing spoken language; 5) social interaction, 6) ability to transfer into and out of an automobile; 7) living situation; 8) knowledge of medications; 9) degree of motor loss; 10) spasticity/involuntary movement; 11) joint limitations; and 12) sensory deficit. The MDS-PAC addresses these concepts to some extent.

The expert panel also suggested that measures of severity of some clinical conditions should be studied. These included a measure of
depression and decubiti/skin condition\textsuperscript{5}. Our latest analyses used only ICD-9-CM codes for these conditions. Depression is viewed by many as not well coded and, in any case, did not meet our significance test for a cost increasing condition. CMS did not want to pay extra for cases with ulcers because these may be preventable. The MDS-PAC data allows the creation of scales for both of these items that distinguish mild from severe cases. Such scales may be more predictive and also better coded.

Some of the items on the MDS are quite similar to the voluntary items on the IRF PAI. Of particular interest are the swallowing item and the skin condition items. The skin condition items in our MDS data and on the IRF PAI allow analysis of conditions that were known at admission. Studying these in an existing sample that is not confounded by hospital decisions to complete these items will get us an independent measure of the effect of these conditions on cost.

Data and Methods

The primary data for this study will be the FIM and MDS-PAC data collected during the comparative assessment of these two instruments for use in the IRF PPS (Buchanan et al, 2001). We have already merged these data with the MEDPAR using a probabilistic matching process such as that described in Section 2. We have also estimated the cost of each case using the latest available cost report. (Cost reports for most study hospitals were available for PPS 16; we used PPS 15 cost reports for the remaining hospitals.)

Factor analysis and the guidance that we received from both Technical Expert Panels will shape our approach to selecting items and scales to study. Items with a substantial amount of missing data, or that occur only rarely and are not clinically important, and are not correlated with length of stay will be eliminated. Once the concepts to be considered have been identified, we will undertake some preliminary

\textsuperscript{5} The panel also mentioned recent acute myocardial infarction. This is included as a tier 3 comorbidity and the MDS-PAC does not appear to offer substantially greater information on this issue.
regression analysis to better understand their potential to contribute to the explanatory power of our classification models.

Our analyses will use both the logarithm of length of stay and the logarithm of accounting cost as the dependent variables. For each item and scale that covers a concept that is not in the CMGs, we will ask whether it affects resource use after controlling for CMG weight. Thus we measure the additional value of the item/scale when we already have the FRGs.

For the cognitive concepts, we will also test whether a better scale than the FIM cognitive scale can be fashioned from the MDS PAC items or subsets of both MDS-PAC and FIM cognitive items. We will use CART analysis for this purpose. The analysis will be patterned as much as possible (given the limited amount of data that we have) after the original work limiting both the number of splits and the resultant minimum classification cell size. Our criterion for determining whether a new set of splits are better than a previous set will be the amount of variance explained.

As part of this effort, we will try to replicate the work of Eilertsen and her colleagues (1998) on stroke and hip fracture using data from the MDS-PAC. To do this, we will need to consider the best mapping from items collected in their study to the MDS-PAC. Following this, we will explore whether enhanced MDS-PAC items can contribute even more to the explanatory power of these models. The depression behaviors indicator that Eilertsen and her team used in the stroke analysis was triggered if the patient had crying spells or a withdrawal from social interaction or sleep disturbances, or a suicide attempt. The MDS-PAC has a more extensive set of 11 indicators of depression, anxiety, and sad mood along with measures of intensity and persistence. While it contains an indicator for negative statements such as "Nothing matters; Would rather be dead than live this way; What’s the use; Let me die", it does not contain information on suicide attempts. For the hip fracture group, Eilertsen et al. include an indicator for mobility impairment from arthritis or other joint damage. The MDS-PAC includes indicators for diagnoses of osteoarthritis and rheumatoid arthritis and another variable for impaired joint mobility.
We recognize that this work will, at best, be exploratory, as our sample size will limit the number of impairment groups that we can analyze. We will be able to analyze the stroke impairment group and the MDS-PAC contains substantially more information to better characterize stroke severity and its impact on functionality, both cognitive and motor, so this group is an ideal testing ground. We should also have enough sample in the lower extremity fracture and lower extremity joint replacement impairment groups and a small number of others as well. While, we clearly will not have the sample for a definitive analysis or to analyze all groups. Nonetheless, we believe that this work is important, as it should demonstrate the value of expanding or modifying the information on the IRF PAI.

FUNCTIONAL ASSESSMENT

Our third special study will extend and synthesize what we have learned about measuring functional status from Phase I, the MDS-PAC special study described above, the literature, and the several TEP panels. We will produce a paper that will address challenges in, and special considerations needed to, routinely obtain functional measures for purposes such as determining payment and measuring quality.

Our work in phase I contained two separate threads that are relevant to measuring functional status and Activities of Daily Living (ADLs). First, while examining options for classifying patients, we estimated the marginal effect of the response to each FIM item on case cost. We found that 2 of the 18 items were not correlated with cost as expected in a burden of care model. Our Generalized Additive Model (GAM) provides much more detail than is now available in the literature about how marginal cost varies with each individual response and across the range of motor and cognitive scores.

The second part of our Phase I work that is highly relevant here is our comparison of the FIM and MDS-PAC instruments. That study showed how the reliability of various FIM items related to the complexity of the implied comparison between patient and the norm for the response. The study addressed the comparability of indices that were designed to measure the same concept but were developed from different data sources,
which is a frequent theme in the literature (e.g., Williams et al. 1997). By comparing CMGs assigned with the FIM and MDS-PAC it showed how important the purpose of the comparison is to a judgment of whether the indices are comparable. Other relevant findings from this study emphasize the importance of considering burden on respondents and the ease with which correct codes can be retrieved (or not as in the case of impairment group in this study.)

We will summarize and synthesize our functional measurement research during Phase I and the MDS-PAC special study and examine its potential contribution to the literature on functional assessment. Then we will extend this research through further study of the comparative MDS-PAC/FIM data base. This database can provide even more information about measuring functional status for administrative purposes. For example, we could test the extent to which the FIM instructions to treat cases where an activity was not performed as ‘most dependent’ affects the ability to predict cost. Finally, we will reflect on the practical lessons we have learned from our studies about using functional measures in payment systems.

NEW TECHNOLOGY

The structure of IRF PPS patient classification and payment levels can impact the diffusion of appropriate technologies and beneficiary access to quality-enhancing care. The challenge is to establish an appropriate methodology to assess the cost implications of quality-enhancing technologies and to assure that payments are adequate to provide access to those technologies without encouraging inefficiency. If a cost increasing technology applies to a small subset of patients within a CMG, the average cost on which the CMG relative weight is based may not be adequate to provide access to the technology. However, explicit recognition of the additional cost through patient classification changes or an additional payment could encourage unnecessary utilization. If the new technology applies to a substantial number of patients, there may also be implications for overall costs of IRF care.
Emerging technologies can include:

- Pharmaceuticals, devices, and procedures that are specific to patients with particular conditions;
- Diagnostic equipment or other items and services used in the care of a broad spectrum of conditions; and,
- Infrastructure changes such as new computerized systems to avert drug errors.

Emerging technologies may be either cost increasing or cost decreasing. They can affect the array of services offered by rehabilitation hospitals. Technological advances in acute care as well as rehabilitative care can affect the conditions and complexity of patients that benefit from inpatient hospital rehabilitation and those that can be treated in alternative settings. These changes can affect the case mix and complexity of patients treated in IRFs and the resources required to care for them. Within IRFs, the changes can be detected through:

- Changes in the number of patients treated with specific conditions, including new conditions such as organ transplants that have not traditionally received inpatient rehabilitation;
- Changes in the patterns of care and average resources required within a given RIC; and,
- Changes in within-CMG severity.

IRF-PPS will automatically compensate for some cost impacts of technological change though the case-mix adjusted payment for individual patients and CMG reclassification changes and recalibration. However, changes in case complexity will result in higher aggregate payments only to the extent they are accounted for by the patient classification system. In sum:

- The system will account for increases in case mix through changes in the number of patients in various CMGs and in the presence of co-morbidities. An explicit policy is needed if aggregate payments are to reflect the overall cost impact of technological changes,
including treatment of high-cost new conditions and within-CMG severity.

- Emerging technologies may result in substantial cost differences within a given CMG. Payment based on the average cost for the CMG may not be adequate to assure access to the technology.

We plan to focus our attention on two areas: 1) quality enhancing technologies that are specific to IRF patients with particular conditions; and 2) specific conditions which are rarely seen in IRF today but which are likely to increase in the future with technology changes in acute care. Under the CMGs, patient classification is based on the RIC, FIM scores, and co-morbidities. The use of specific technologies or procedures within the RIC does not affect patient classification. Hence, there may be certain CMGs in which there is substantial variation in either existing or emerging technologies used to treat patients assigned to the CMG. Our focus will not be on the cost of a specific device or pharmaceutical, but rather on whether patients benefiting from the technology have substantially higher cost per case than other patients in the same CMG.

We plan to conduct the analysis of specific technologies in three steps:

1. We will use literature review to identify emerging technologies that are quality enhancing and might have significant cost implications for IRF services furnished to Medicare beneficiaries;

2. We will use MEDPAR data to identify the population of rehabilitation patients potentially impacted by the implementation of these emerging technologies and examine the potential cost implications of this implementation.

3. We will use payment simulation to evaluate the effect of payment policies that specifically recognize the additional costs associated with those technologies that have substantial cost implications.

In addition to examining specific technologies, we will investigate the cost implications of technology changes in acute care that are
likely to change the case mix and complexity of Medicare patients that would benefit from IRF care.

**Identifying New Technologies With Cost Implications**

The first step in our analysis will be to conduct a literature review to identify potential quality-enhancing emerging technologies that are being used in IRF inpatient care that may have cost implications. We will also consider whether any emerging technologies in the acute care setting might impact rehabilitation case-mix.

To structure this literature review, we will begin by surveying a subgroup of our technical expert panel. We will ask them to identify existing quality enhancing technologies not widely implemented at the time that we defined the CMGs, and to identify newly emerging medical technologies that impact rehabilitation outcomes. We will use these responses to target those technologies for which we will conduct a comprehensive literature search.

**Investigating Dispersion Of New Technologies And Cost Impacts**

In the second step of our analysis, we will examine to the extent feasible two basic questions related to emerging technology:

- What proportion of rehabilitation patients are potentially eligible for or impacted by the technology? How dispersed is the technology? How frequently is it being used? Are there systematic differences across types of hospitals (e.g., teaching vs. non-teaching, urban vs. rural) that are using the new technology?
- Are there substantial cost per case differences within the same CMG between patients receiving the new technology and other patients after controlling for other patient and facility characteristics?

The adequacy of the payment rates is an issue only if 1) there is significant use of the technology, 2) there are systematic differences in the use of a technology across hospitals, and 3) use of the technology has substantial cost implications.

Initially, we will rely on MEDPAR data to examine these questions using descriptive and multivariate analyses. However, there may be some potentially significant technologies that we cannot identify through
procedure codes. If so, we will explore using additional information such as data obtained from individual facilities on patients who have received the technology.

We will use a comparable approach to investigate the cost implications of technology changes in acute care that are likely to change the case mix and complexity of Medicare patients that would benefit from IRF care. We will use MEDPAR data to examine the following basic questions:

- To what extent are these conditions currently being treated in IRFs?
- How are the cases distributed across IRFs?
- How do the costs of patients with these conditions compare to the costs of other patients assigned to the same CMG?
- What percentage of the patients with these conditions qualify for outlier payments?

The adequacy of the payment rates for these conditions is an issue if 1) there is likely to be a significant increase in the number of Medicare patients with the condition, 2) the patients with the conditions are concentrated in particular facilities, and 3) the patients have substantially higher IRF costs that other patients assigned to the same CMG.

**Evaluating Alternatives To Recognize New Technology In Payment System**

Ideally, the payment system should be neutral with respect to the use of specific technologies. The cost impact of new technologies have traditionally been taken into account through:

- Changes in the patient classification system
- Revisions in the relative weights; and/or
- The update factor.

With respect to IRF-PPS, the implementing regulations provide for periodic revision of the CMGs and relative weights to take into account changes in treatment patterns and technology among other factors.\(^6\) This

\(^6\) Unlike the statutory provision governing the update factor for acute care PPS, there is no explicit provision for recognition of new
approach builds the cost of technology into the relative weights for the CMG. It may not be adequate if there are significant differences in the diffusion of emerging technology that have substantial cost implications. Concerns over the processes used to recognize new technology under the acute care PPS and hospital outpatient PPS led to statutory changes requiring a temporary pass-through for certain technologies until they are reflected in the claims data and can be evaluated for appropriate payment.

Assuming that we find quality-enhancing emerging technologies with substantial cost implications, we will simulate policies for recognizing the higher costs in the payment system. In doing so, we will consider the precedents set by the acute PPS and outpatient PPS (DHHS, 2001a; DHHS, 2001b). In addition, we will take into consideration the recommendations adopted by the Medicare Prospective Payment Advisory Commission in its March 2001 report. The recommendations dealing with recognition of new technology in Medicare prospective payment systems included the following principles:

- Additional payments for new technology should be limited to new or substantially improved technology that add significantly to the cost of care; and,
- Additional payments should be budget neutral with the costs of new or substantially improved technologies factored into the update factor.

The payment simulations will compare payment-to-cost ratios under current policy with those resulting from a policy that provides additional payment for selected technologies. The comparisons will be at the case level between the patients receiving the new technology and those that do not and at the facility level.

We will also use payment simulations to examine the impact of technology changes in acute care on aggregate IRF costs and on specific technology in the update factor. The annual increase factor is to be based on appropriate increase in a market basket of goods and services representing the services paid under IRF-PPS.
hospitals (to the extent any high cost new conditions that we identify are concentrated in certain types of hospitals). We will investigate the payment-to-cost ratios for these cases relative to other cases in the CMG.
6. SCHEDULE AND DELIVERABLES

In 2002 the IRF PPS research team will undertake two types of activities -- one set will involve the analysis of data through the end of CY 2000 and the other will draw on 2001 administrative data and/or preliminary data from the first two quarters of the IRF PPS.

The analyses of existing data will include investigations of methodological improvements to the classification system, the treatment of complex cases, compression, and facility adjustments. It will also include our investigation of the value of using MDS-PAC items in the payment model and a preliminary set of indicators for monitoring based on analyses of the 1996-2000 standard analytic files. We will report on these analyses in September of 2002. A report on the staff time measurement study will be produced within 5 months of the receipt of the staff time data.

Assuming the subsequent year’s activities are funded, we will meet with our project officer in Baltimore in September 2002. Within 6 weeks of the new fiscal year we will submit a work plan for our next year of activities. This will be repeated annually as funding is extended.

We will have a TEP meeting in November of 2002 to review our methodological refinements, monitoring indicators, special studies, and updated workplan. Prior to this meeting a review of technological change in rehabilitation care will be drafted and circulated to the TEP.

The IRF PPS will be implemented effective January 1, 2002 and will apply to IRFs for cost reporting periods beginning on or after that date. Data from the first several months of the IRF PPS will, therefore, not become available before the fall of 2002. The analyses that require either post-implementation data or administrative data from 2001 (both of which we expect to be available as of October 2002) will thus be presented to CMS in December 2002. This will include our preliminary findings regarding coding changes that affected payment or any other post-IRF changes that suggest the need for immediate refinements. We will report on baseline measures for the monitoring system using data through 2001 in late 2002 as well.
In December we will also provide CMS with a draft summary report covering research conducted during 2002. This report will reflect comments received from CMS on the reports of individual activities. Thus its review and revision can be done quickly. A revised report, for wider distribution, will be available within at most 30 days after receipt of comments.

In addition to formal reports, we will continue to communicate preliminary findings via e-mail and telephone conference call. These will include manuscript that we develop regarding methodological improvements and aspects of the monitoring system.

<table>
<thead>
<tr>
<th>Schedule of 2002 Activities</th>
<th>Completion Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>Activity</td>
<td></td>
</tr>
<tr>
<td>Evaluation of use of MDS-PAC items in payment models</td>
<td>June 2002</td>
</tr>
<tr>
<td>Staff time measurement study</td>
<td>5 months after receipt of data</td>
</tr>
<tr>
<td>Draft paper on functional measures</td>
<td>August 2002</td>
</tr>
<tr>
<td>Methodological improvements report</td>
<td>September 2002</td>
</tr>
<tr>
<td>Version 1 of indicators for monitoring</td>
<td>September 2002</td>
</tr>
<tr>
<td>Final paper on functional measures</td>
<td>September 2002</td>
</tr>
<tr>
<td>Year 2 project meeting in Baltimore</td>
<td>September 2002</td>
</tr>
<tr>
<td>Review of changes in technology</td>
<td>October 2002</td>
</tr>
<tr>
<td>Updated workplan</td>
<td>October 2002</td>
</tr>
<tr>
<td>TEP panel meeting</td>
<td>November 2002</td>
</tr>
<tr>
<td>Preliminary report on baseline measures for monitoring</td>
<td>December 2002</td>
</tr>
<tr>
<td>Preliminary report on refinements to payment elements and coding changes</td>
<td>December 2002</td>
</tr>
<tr>
<td>Interim report on activities to date</td>
<td>December 2002</td>
</tr>
</tbody>
</table>
REFERENCES


Kahn, Katherine L. et al. 1992. The Effects of the DRG-Based Prospective Payment System on Quality of Care for Hospitalized Medicare Patients. RAND. R-3931-HCFA.


Liu, K., et al. 1999. Medicare’s Post-Acute Care Benefit: Background, Trends, and Issues to be Faced. The Urban Institute. (G, Q2)


MEDPAC, Report to the Congress: Medicare payment policy, Washington, DC, March 2001


