Collection of Information Employing Statistical Methods

C.1   Respondent Universe and Response Rates

The target population for this project is fee-for-service Medicare beneficiaries receiving Part B-covered ("outpatient") therapy services in hospital outpatient departments (HOPDs), nursing facilities, outpatient rehabilitation facilities (ORFs), comprehensive outpatient rehabilitation facilities (CORFs), home health agencies (HHAs), and by clinicians in private practices. Only patients of providers participating in this study will have assessment data provided by or for them. RTI and CMS estimate that approximately 38,632 assessments will be collected from 19,316 Medicare beneficiaries treated by up to 190 participating providers recruited from across the United States. Each provider will participate for at least four months in the data collection process, though not all simultaneously (Table B-2).

C.2   Procedures for the Collection of Information

Statistical Methodology for Stratification and Sample Selection

The sample design for this study is a stratified clustered design with the following hierarchy of design elements. We will first assemble a set of providers engaged in outpatient therapy using Medicare claims data. Claims will be aggregated by provider ID (using NPIs as well as legacy OSCAR and UPIN identifiers). Claims will also be used to indicate which providers have higher Medicare volume, so that these providers can be targeted for sampling (for greater data collection efficiency), instead of only using provider program participation files.

The next step in the sampling process is stratifying providers by the six provider type strata mentioned above, defined using the provider program participation data. We will use the volume information from claims data to identify the providers within each stratum with Medicare volumes above the median—however, if this criterion results in excluding a disproportionate number of providers in low density population areas, we will adjust the volume threshold. We will develop a sample of these providers and begin the recruitment process.

During the recruitment process we will determine, through discussion with leadership in the sampled organizations, the relevant organizational characteristics of each provider, for example the patient case mix, service specialization, and divisions into units or offices. Stratifying by categories of one or more of these organization characteristics will produce very small cell sizes, potentially unduly increasing statistical error. However, during recruitment we will attempt to balance the sample in these and other (e.g., geography) characteristics to produce a nationally representative sample.

Within each stratum, our sample of patient-episodes[1] is not a simple random sample. The reason is clear. It would be too costly and impractical to select a simple random sample of patient-episodes because of data collection logistics—providers must be recruited. As a result, the patient-episodes within each stratum are clustered into 190 primary sampling units based on the 190 proposed facilities/practices. Since patients may have multiple episodes with a particular provider,

---

[1] Note that a patient-episode may involve therapists from a single discipline or from multiple disciplines.

the sample of episodes can be viewed as clustered by patient; however, we believe the principal source of sampling error to be provider-based clustering.

Statistical Power and Target Number of Outpatient Therapy Episodes

Power gives the probability that a true, real difference from zero (or other specific reference value) will be identified as a significant difference. Because of random variation, a measured difference may in fact be zero, or close to it, despite a true nonzero difference. A high-powered test of a case mix coefficient in a payment model is less sensitive to this random error and is more likely to identify a true nonzero difference of that case mix coefficient from the population average.

Our estimated total sample size is driven by the need for tests of case mix group coefficients in regression models of outpatient therapy episode payments to be reasonably powerful to identify meaningful differences from the average payment for reasonably-sized groups as statistically significant. Conditional on the difference from the mean, or "effect size," of the case mix group (which determines the case mix weight) and the underlying program cost homogeneity of that group, the power of the test of the case mix weight from 1.0 will be determined by the proportion of patients in that group. Alternatively, conditional on the proportion of patients in that group and the underlying program cost homogeneity of that group, the power of the test of the case mix weight from 1.0 will be determined by the effect size of the case mix group.

Power analysis for a regression model is based on an F-test of coefficient estimates (Taylor and Muller, 1995). For this power analysis, we consider the special case of a test of a single regression coefficient (i.e. for a single case-mix group adjustor in a model of log episode payment). The power of the F-test of a coefficient estimate is given by

$$P = 1 - F_N \left[f_{crit}(0.95, 1, N - k), 1, N - k, t^2 \right], \tag{1}$$

where F_N is the cumulative distribution function of the non-central F distribution evaluated at $f_{crit}(0.95, 1, N-k)$, the critical F value for the test; with 1 numerator degree of freedom, corresponding to the single coefficient being tested, and N denominator degrees of freedom, corresponding to the number of degrees of freedom in the model; and a non-centrality parameter t equal to the value of the F test statistic. N represents the sample size, and k is the number of regressors in the model.

When the only restriction tested is for a single coefficient, the F statistic can be simplified to the square of the t statistic for the regression coefficient (Greene, 1993). However, because of the clustered sample in this study, an adjustment must be made to coefficient standard errors to account for the design effect D; the adjusted standard error will be equal to the standard error, assuming simple random sampling, multiplied by the square root of the design effect. Also, it will be useful to express a case mix group coefficient b as a percentage p of mean expenditures m, so that b .

As a result, the F statistic in equation (1) can be expressed as:

$$f_{\text{effect}} = t_{\text{effect}}^2 = \frac{(pE)^2}{D\sigma_b^2},$$ (2)

where b is the estimated demonstration effect regression coefficient, and σ_b is the standard error of the estimate. Substituting (2) into (1), the power of the test is given by

$$P = 1 - F_N \left[f_{\text{crit}}(0.95, 1, N-k), 1, N-k, \frac{(pE)^2}{D\sigma_b^2} \right].$$ (3)

Using equation (3), to estimate the total episode sample size N to achieve a desired power of 80 percent of tests of case mix coefficients, it is necessary to provide several inputs into that equation. For the number of regression coefficients, we assume k will equal 100 to approximate the models included in the Ciolek and Hwang (2004) episode-based payment model report and to provide a conservative (high) estimate for the number of regressors in the models estimated for this study. Consistent with power estimates underlying the sample size estimates for the Post-Acute Care Payment Reform Demonstration, we assume a design effect of 3 for this study. This design effect estimate is based on average design effects encountered in the Psychiatric Inpatient Routine Cost Analysis project (Cromwell, et al., 2003), which collected primary assessment and resource use data on 838 Medicare patients in 40 inpatient psychiatric facilities. Estimates of effect sizes (E) and standard errors are based on selected diagnosis groups in the Ciolek and Hwang (2004) model report. However, we assume the regression model will reduce standard errors by 17 percent, as in the log episode payment models estimated by Ciolek and Hwang (2004). In addition, based on findings from the Cromwell, et al. (2003) study, we assume regression values will be further improved by 25 percent from using assessment characteristics not currently included in administrative data versus a completely claims-based model. Based on these assumptions, we estimate that there will be sufficient power (80 percent) to identify case mix groups equal to 15 percent of the total outpatient therapy population with a 10 percent effect size, or case mix groups equal to three percent of the population with a 20 percent effect size, with a sample of 19,316 outpatient therapy episodes.

We estimate that 84.1 percent of episodes will have a PT component, 26.6 percent will have an OT component, and 15.9 percent will have an SLP component (these percentages sum to over 100 percent because some episodes have clinicians of multiple disciplines contributing to the patient's care). This represents a 50 percent oversampling of SLP services and 20 percent for OT services (and a 10 percent under sampling of PT services). This deviation of our expected sample from the national distribution of outpatient therapy services will improve statistical power and accuracy for OT and SLP services, which are smaller proportions of all outpatient therapy services than are PT services. Sample weights will be computed for computing nationally representative estimates of utilization and cost of outpatient therapy services across disciplines and settings.

Assuming a data collection period of four to six months (averaging five months, with a length of data collection time based on a provider's volume), we calculate the total number of providers required to achieve 19,316 episodes to be 190. This assumption is based on the average

number of episodes per month for providers with above-median volume, but capped at 35 per month assuming that is the maximum number of assessments per month providers will be willing to collect (based on experience in the Post Acute Care Payment Reform Demonstration). This volume count can be achieved with: 29 hospital outpatient departments; 45 skilled nursing facilities; 29 CORFs, ORFs, and home health agencies; and 29 each of practices that are PT-only, PT plus OT, and any with SLPs. Furthermore, to capture more complex patients in day rehabilitation programs, 7 of the 29 targeted hospital outpatient departments will be day rehabilitation programs, identified during recruitment.

During the data collection period, we will request that an admission and discharge assessment be submitted for each Medicare beneficiary beginning an episode of care during the data collection window. Participating providers will be asked to complete the admission assessment within the first two visits of an episode and attempt to complete a discharge assessment within the last two expected visits. Episodes beginning near the end of the four-month data collection period will be included, and we will ask that providers submit discharge assessments for these beneficiaries if it occurs after that data collection period is over.

Survey Eligibility

Individual and institutional providers providing outpatient therapy services will be eligible for this data collection. The relevant patient population is all patients receiving outpatient therapy during the data collection period.

C.3   Maximizing Response Rates

The strategy for provider recruitment needs to consider several competing objectives, including practice setting, provider and geographic variation. Unlike the development of the CARE tool, we are not wedded to markets and as long as we can conduct regional training we can accommodate rural providers in the study. We do need to make sure that we include providers from different practice types, and that the data collection is not overly burdensome so that small provider practices can participate.

We will make every attempt to obtain geographic variation and will map the locations of proposed practices for CMS' consideration and to illustrate the geographic location of proposed provider sites. In considering provider participation we will use the strata illustrated in Exhibit 1 to make sure we have sufficient sample for each practice setting.

We propose to begin provider recruitment by taking advantage of the extensive network available through our subcontractors. We will try to engage the support of the various associations, including AOTA, APTA, and ASHA, to gain their support for participating in the data collection effort. In addition, through stakeholder involvement, particularly through conference calls and the project Web site, we hope to solicit provider interest, particularly for unaffiliated providers. We will make available an e-mail address and a telephone number so that following a planned upcoming Open Door Forum, providers can contact us if they have questions about participation. As with the development of the CARE tool, we found that speaking at national meetings heightens awareness of the project and is an excellent forum for soliciting provider participation.

C.4   Contacts

CMS-10298 Developing Outpatient Therapy Payment Alternatives

David M. Bott, Project Officer, Developing Outpatient Therapy Payment Alternatives
Office of Research Development and Information
Phone: (410) 786-0249
e-mail: dotpa@cms.hhs.gov

Edward M. Drozd, Project Director
RTI International
Phone: (781) 434-1716
e-mail: edrozd@rti.org