RESPONSES TO OMB QUESTIONS ON PRA RENEWAL APPLICATION
December 30, 2008

Can you just confirm that the EHR demo will be following the protocol laid out in the evaluation design report? The report says the evaluation will be using random assignment: is this true?

Yes. A randomized design will be used for the EHR demonstration.

Also, can you have CMS provide a discussion of what the limitations of the studies (all 3 of them) are, both in terms of their internal validity as well as their external validity? For external validity, for example, can CMS provide a discussion of how the sites selected for the demos may not be representative of sites or patient population groups at large (e.g. demographic factors like patient population characteristics did not factor into how the sites were chosen, so the evaluation may find that the sites are not representative in important ways; some sites were chosen because they already had an HIT program; etc.).

Under each of these demonstrations, participating practices report clinical quality measures relating to the care of diabetes, congestive heart failure, coronary artery disease, and preventive care. The proposed ambulatory care measures being used were developed by CMS in conjunction with the American Medical Association's Physician Consortium for Performance Improvement and the National Committee for Quality Assurance. In addition, CMS worked directly with the industry and participating practices to minimize administrative burden and to align the measures with those used by commercial payers.

The Physician Group Practice (PGP), Medicare Management Performance (MCMP), and Electronic Health Record (EHR) demonstrations were designed to study responses to different types of incentives. Practices were not selected at random under these demonstrations, and selection bias is always an issue in these types of demonstrations. Demonstration evaluations will be conducted, and these will include empirical analyses used to identify demonstration effects. These effects will be estimated impacts of the demonstration per se, estimated by comparing participating 'treatment' practices with 'control' practices. The definition of 'control' practice varies across these demonstrations. In each evaluation, comparisons will be obtained after controlling for a variety of characteristics of practices and beneficiaries using treatment and control practices. While research tools will be used in attempts to deal with selection bias, results of these demonstrations will not be generalizable to populations of practices that treat Medicare patients nationwide.

Ten large physician group practices are participating in the PGP demonstration. These practices were not selected at random from large Medicare practices, but in response to a review panel's assessments of 26 practice applications. Under the demonstration, PGPs

would earn shared savings by reducing PGP beneficiary expenditures and by improving quality measures. The PGP evaluation is assessing access, quality, and expenditures by comparing experiences of 'loyal' PGP beneficiaries with experiences simulated from trends for the average Medicare beneficiary (not necessarily with a 'loyal' practice attachment) who does *not* use PGP services but resides in the PGP's market area.

Under the MCMP demonstration, small-to-medium size (generally up to 10 physicians) practices providing primary care to at least 50 Medicare beneficiaries in each of the four Doctor's Office Quality-Information Technology (DOQ-IT) pilot states were recruited by the states' Quality Improvement Organizations (QIOs). The MCMP evaluation is using a non-randomized comparison group (quasi-experimental) design. Comparison states were selected based on their similarity to demonstration states (census region or sub-region, number of small practices as a percent of all physicians in the state, the ratio of specialists to primary care physicians, Medicare expenditures per beneficiary, and Medicare managed care penetration rate). Practices participating in these comparison states were matched to practices in the demonstration states. Matching variables included practice size, prior experience with HIT, number of Medicare beneficiaries with the targeted conditions, average number of evaluation and management visits and average number of hospitalizations per patient. Demonstration effects may differ by site (pooling of observations across states may be inappropriate due to differences in state regulations, pay-for reporting and pay-for-performance initiatives, and HIT/EHR penetration). Results will not necessarily apply to small-to-medium practices where Medicare beneficiaries are treated, nationwide or within each of the demonstration states (as practices may differ by state and participating practices likely differ from those choosing not to participate).

Under the EHR demonstration, 12 sites were chosen from a number of applicant sites (geographic areas, often states) for the demonstration by a panel led by the Office of the Secretary that included representatives from ASPE, the Office of Policy at CMS, and the Office of the National Coordinator for HIT. In each site, a community partner is recruiting practices to participate in the demonstration. Practices must be small-to-medium size (up to 20 physicians) and they must provide primary care to at least 50 Medicare beneficiaries. Stratification of the sample in each site will ensure that the treatment and control practices are similar. Practices will be stratified by characteristics including practice size, urban v. rural location, and whether or not the practice has an EHR system. Evaluation findings may be representative of small-to-medium size practices in sites where participating control and treatment practices represent a substantial fraction of all small-to-medium size practices in those sites. However, findings of the evaluation will not generalize to practices participating in the Medicare program nationwide because of the nonrandom process used to select the 12 demonstration sites.

It would also be helpful to know other limitations of this study for purposes of roll-out on a more large-scale basis. For example, it does not seem like these demos will allow policy makers to titrate incentive amounts for desired impact (e.g. it will not enable you

to say, for example, that a bonus payment of $50 resulted in 10% improvement, while a payment of $100 resulted in 25% improvement).

The MCMP and EHR demonstrations were not designed to model changes in quality resulting from changes in the bonus, or level of payment for use of EHRs. Under MCMP, demonstration practices are eligible for up to three incentive payments: an incentive of $20 per Medicare beneficiary (up to $1,000 per physician, or $5,000 per practice) for reporting baseline clinical quality measures; a payment for each of the three demonstration years based on whether the composite score meets quality standards for each chronic condition (up to $70 per beneficiary) and for meeting standards in delivery of preventive services (up to $25 per beneficiary); and a payment for using a certified medical records system that can extract and submit performance data to CMS electronically (which would increase incentive payments up to 25 percent). Similar multiple payment incentives exist under EHR. Incentive design of a larger roll-out would depend on policy goals.

It would also be helpful to get a discussion from CMS about some of the unintended consequences that P4P programs may result in and how the data that will be collected —or which will be available to the evaluators through other mechanisms—will or will not be used to monitor for these unintended consequences, such as: cream skimming (the denial of high-risk or non-compliant patients); selection bias (patients electing to receive care from high-performing providers thus skewing the patient caseloads for particular providers); reduced quality of care on those dimensions of health care that are not being measured/rewarded; impeded knowledge and transfer and innovation among providers who now see themselves as competing for bonuses with other providers; cost-shifting (e.g. if Medicare payments fall below a certain level for underperforming providers, do they shift costs to the private sector or to patients?); etc. If these unintended consequences will not be monitored, it would be good to know this as well.

Unintended consequences of changes in incentives are always possible, and some of these may be discernible during analyses conducted as part of these demonstrations. Several unintended consequences would appear to be more likely under our ongoing PGP, MCMP, and EHR P4P demonstrations.

Coding creep – A liberal use of disease coding may be encouraged under the PGP demonstration, as increases in disease severity or co-morbidities increase target expenditures for PGP beneficiaries under the demonstration. Increases in target expenditures may increase incentive payments to PGP sites. Future PGP evaluation work will address this issue.

Increases in disparities – Increased use of HIT/EHR might exacerbate disparities by race, geographic location, or income class, e.g., if practices that adopt HIT/EHR are more likely located in geographic areas that are not racially balanced or treat higher income beneficiaries. Available data on urban v. rural location, race, and eligibility for Medicaid

in addition to Medicare will support some analysis on changes in practice composition of Medicare beneficiaries over the demonstration time period.

Perverse quality incentives – Whether P4P models encourage provision of quality from which bonus payments are generated at the expense of quality that is not rewarded is an issue of concern to CMS. CMS evaluation staff are conferring with the evaluation contractor to determine whether analysis of this consequence can be studied as part of the EHR demonstration evaluation.

It is important to emphasize that any evidence of unintended consequences might be revealed only during the evaluations' planned analysis phases; these analysis phases are not routinely scheduled for purposes of ongoing monitoring of various dimensions of practice performance.