

Response to Questions about the Multi-site Coordinated Evaluation of the Impact of Quality-based payment Strategies for AHRQ

1. Please provide a copy of the articles by Young et al. 2005 and Meterko et al 2006.

Copies of these articles accompany our response.

2. Financial incentives linked to physician performance are but one type of incentive health plans use to influence provider behavior. How will this study “isolate” the effects of these explicit financial incentives, which are the focus of this study, from other financial and non-financial incentives these physicians face?

We will use our semi-structured interviews of senior managers in each study site to identify whether other types of incentive arrangements have been established with respect to physician performance. Should other incentive arrangements exist, we will consider such information as a possible confounder relative to any significant changes that we observed in quality measures following the introduction of the quality-related incentives.

3. A2: Please explain how the quality targets were chosen. For example, why is HbA1c a quality target for “child health” but not for diabetes? Also, why are retinal exams and LDL cholesterol level testing the only quality targets for diabetes?

The selection of quality targets and the decision to select some measures but not others is part of the proposed inquiry. We intend to use the interviews with senior managers to assess why some targets were chosen over others. To clarify the table presented in section A2 of the supporting statement, these are examples of quality targets communicated to us by Montefiore and BMC Health Plan at the time of the proposal. For example, Montefiore has changed its quality measures several times as its program has matured and physicians improved. We have replaced Table 1 with a list of all quality targets for each site as of 2006. Please note that BMC Health Plan program began in January 2006 and has a substantially shorter quality target list than the decade-old Montefiore program.

4. A2: how has the Montefiore financial incentive program changed since its inception in 1996? In conducting pre/post analyses, at what time points will you be conducting analyses for the Montefiore sites?

We will be using the semi-structured interviews and review of program documents to understand how the financial incentive program has changed as it has matured. Over time, Montefiore’s program has introduced both different quality measures and also changed the magnitude of the incentive payments. We will be examining measures that were introduced between 1999 and 2004, allowing us to have several years of baseline and post-intervention of experience. The time points we will use to conduct the pre/post analysis therefore will depend upon their introduction into the program.

5. A2: the literature is fairly clear that what is important to patients (i.e. what they define as “quality care”) is often very different from what is important from the perspectives of physicians, hospital administrators, health planners, and health plans. Will there be any

attempt to capture the impact of these financial incentives on quality of care from the patients' point of view?

While we agree completely that it is important to understand the patients' perspective on financial incentives for healthcare providers, this inquiry exceeds the scope of our study. We do believe that our study findings may be useful to designing future studies that consider the patient point of view.

6. Please provide more information on the study of non safety net settings that will be used as a comparison for this study including information on the sample, survey methods, and response rate.

As described in the supporting statement, Boston University evaluated the National Rewarding Results Demonstration that included seven different pay-for-performance programs. We intend to compare results from the safety-net settings to those obtained from one of the Rewarding Results programs: the Rochester Individual Practice Association. A description of this setting, the sample surveyed, survey methodology, and response rate are presented below.

Our comparison site from the Rewarding Results program is located in Rochester, New York. The site consists of a partnership between Excellus, Inc. and the Rochester Individual Practice Association (RIPA). Excellus, a single health maintenance organization (HMO), has enrolled more than 70 percent of the commercial (non-Medicare and non-Medicaid) population in the nine county region surrounding Rochester, New York. RIPA is a contracting entity comprised of more than 800 primary care physicians who collectively care for more than 420,000 commercial patients, approximately 80 percent of whom are members of Excellus. Excellus and RIPA have collaborated to share administrative data for the purpose of their P4P program. The sample for the survey consisted of physicians who qualified to receive a financial reward from RIPA for achieving quality targets related to care pathways for at least one of four conditions: asthma, diabetes, otitis media, or sinusitis. A total of 597 physicians (152 family practitioners, 290 internists, and 155 pediatricians) met this criterion; all were included in the survey sample.

To administer the survey, cover letters and questionnaire booklets were distributed in envelopes addressed to each physician by name. Each packet also contained a prepaid business reply envelope that respondents were instructed to use to return their completed surveys directly to a third-party data entry service. In Massachusetts, most surveys were sent to a liaison at each of the 32 selected contracting entities, who then distributed the surveys either at a medical group meeting or by placing the packets into office mailboxes. At RIPA, where most physicians were members of solo or small group practices, questionnaires were distributed by mail or by RIPA managers at group events such as grand rounds at affiliated hospitals.

The response rate for RIPA was 43%.

7. Please provide more information on the psychometric properties of the PAI-26 and the development and validation of this instrument.

We conducted an extensive analysis of the psychometric properties of the instrument based on physician survey data that we collected in two of the seven sites in the Rewarding Results Demonstration, specifically Massachusetts and Rochester, New York. The results of this analysis are presented in the previously noted paper published in Health Services Research (Meterko et al., 2006); a copy is included with this set of responses.

We conducted several types of psychometric analyses to evaluate the reliability and validity of the instrument. To assess the reliability of the instrument, we computed a Cronbach's alpha on each proposed scale. We also applied multi-trait analysis. Additionally, to examine validity, we conducted a factor analysis. We used the factor analysis to determine if the number of factors and factor-item loadings were consistent with our proposed conceptual model. We also tested the construct validity of the instrument by testing whether respondents' subscale scores (which conceptually indicate physicians' attitudes toward the incentive programs in which they participate) were statistically related to their scores concerning the perceived impact of the incentive program.

In general, the psychometric analysis provided strong empirical support for both the reliability and validity of the instrument. Details of the results are presented in the above referenced publication.

8. Please describe in more detail what kinds of pre/post-test analyses will be undertaken with the clinical performance data.
 - In some places, the supporting statement seems to imply that the actual performance of physicians/physician groups will not be assessed (p15). Does that mean this study will not assess whether quality of care improved after the introduction of the financial incentives, compared to quality of care provided before the introduction of financial incentives? If not, why not? It would be interesting to know how patients fare when they are under the care of a physician participating in the financial incentive program vs when they are under the care of a non-participating physician.

For each research site, we intend to examine whether improvement occurred on the targeted quality measures following the introduction of the programs. We will compare baseline scores with post-intervention scores by applying the statistical procedure of repeated measures analysis of variance. This procedure is useful for assessing whether the post-intervention scores are significantly different from the pre-intervention pattern in terms of both levels and trends. We applied this statistical procedure to assess the impact of an incentive program in the Rewarding Results demonstration (Young et al., Journal of General Internal Medicine, in press). An advance copy of this unpublished paper accompanies our responses. We may also consider using other procedures that take into accounting nesting of observation as we recognize that physicians working together in the same practice or clinic are not necessarily truly independent observations from a statistical point of view. Thus, we may also examine data using repeated measures analysis but this will depend on whether preliminary tests indicate moderate correlations among physician performance scores within the same practices or clinics.

- What kinds of important but financially unrewarded clinical activities will be examined? How will these activities be chosen?

At each site, we in collaboration with managers and clinicians will identify a set of quality measures that have not been explicitly rewarded but still serve as important indicators of quality of care. One concern about linking financial incentives with quality measures is that it will encourage providers to adopt a “teach to the test” mentality” so that they may pay less attention to clinical activities that are not financially rewarded even though they are though are clinically important.

- Are these the same as the “clinical tracers” mentioned on page 18?

Yes, by tracers we were referring to quality measures that were not explicitly rewarded but are still important indicators of quality of care.

- Will there be some kind of analysis of patient attrition or caseload shifting among physicians? If not, why not?

Yes, we will have data on patient caseload and will use this information in our analyses. In particular, we will want to include in our impact analyses only those physicians who had a sufficient number of relevant cases for a quality measure so that performance scores are reliable.

9. What percentage of physicians choose not to sign up for the financial incentive program and what are their characteristics (and what are the characteristics of their patient caseload)? It would seem to be important to survey this population as well. Will there be an attempt to compare quality of care and patient outcomes between participating and non-participating physicians?

In the two programs that we are studying, physicians do not “sign up” for the program. They are included based on their organizational affiliation. Their performance is measured automatically relative to the selected quality measures and they are eligible for financial rewards. During our structured interviews with program managers, we will inquire about any decisions that were made to exclude certain practices/clinics or individual physicians from the incentive programs. If certain exclusions have been made that potentially bias the study results, we will follow up in the manner you have recommended. However, our preliminary conversations with program managers do not point to any such exclusions.

10. Please clarify how key informants are being selected for interview. Also, in section A.9 of the supporting statement, the proposed honorarium is listed as \$50, but the letter and consent form state that it is \$100; please clarify.

Interviewees will be selected from short lists of key informants developed by the project liaisons at each research site based on their availability for interview during times when the project staff can travel to them. We will attempt to accommodate as many informants as our travel schedule allows. We have revised the honorarium amount in the consent form to the correct \$50 honorarium amount.

11. Please clarify which sites will be picked for the interviews. Will some of the sites not be participating in the financial incentive program? The informed consent form suggests that the interview sites will all be participants, and yet the interview protocol suggests that some of the interview sites will be non-participants.

All sites will be eligible for interviews and we will attempt to interview at least one key informant from each practice site. None of the sites we will interview are non-participants. We have revised the interview protocol to reflect this fact.

12. Please clarify the timing of the interviews in relation to the surveys. It might be useful to undertake a preliminary analysis of the survey data before conducting the interviews. This preliminary analysis might yield questions which researchers can probe for in the interviews.

Because of the tight project schedule, it will not be possible to analyze the survey data prior to conducting interviews. The interviews will be conducted concurrently with survey fielding. The Rewarding Results Evaluation was also conducted in this manner. We will ask each interviewee if we can follow-up with them by phone should any question come up following the survey.

13. Section A.10 of the supporting statement did not provide any statutory authority to promise confidentiality to respondents; however, the letters and informed consent make this promise. Please provide the statutory authority for confidentiality of this information.

Individuals and organizations contacted will be assured of the confidentiality of their replies under Section 924(c) of the Healthcare Research and Quality Act of 1999. We have amended Section A.10 of the supporting statement with this information.

14. At times, the purpose of this study appears to be exploratory (e.g. page 4); at other times, the purpose of this study appears to be evaluative and prescriptive (e.g. page 8). Please clarify. Given that this is the first study to explore financial incentives in a safety net setting, and given the case-study approach, a more exploratory approach would seem to be most appropriate: i.e., one where the results of the study would be used to inform directions for future research rather than used as the basis for making recommendations on how financial incentives might be used in other settings.

We agree with the OMB that the project is a first step toward understanding the implementation of financial incentives in safety net settings. We do see it as exploratory in the sense that little research has been done on financial incentive programs generally and on such programs in safety net settings specifically. We also recognize that our study design, (in terms, for example, of number and geographic representation of sites) is far too limited to provide any definitive results about the impact of financial incentives on quality in safety net settings. Still, we believe the study results can be extended to other settings. We note that our data collection is designed in accordance with a conceptual framework that we have now tested and refined with data from the Rewarding Results demonstration. These data include the results presented in the previously noted publication addressing the psychometrics of the PAI-26. Thus, our study is not exploratory in the sense that we still in the process of developing and explicating key theoretical/conceptual constructs. Much of this work has already been done

and will provide a solid conceptual foundation for generalizing study results to other settings.

15. In case study research—as it appears to be the case in this study—sites are often selected because they have unique characteristics which lend themselves to exploring the research question at hand. However, these same unique characteristics typically also limit the extent to which results from case study research can be generalized. Please describe how the results from the study will be generalized and what limitations of the case study methodology will be acknowledged in reports of these results.

As discussed in our response to comment 14, we are collecting data in accordance with a well developed conceptual framework on financial incentives and quality that we developed and have tested. This framework provides key constructs that we will measure in this study. The use of this framework promotes the generalizability of the results since we are not studying these programs with measures that have already undergone broader testing regarding their reliability and validity. At the same time, we recognize that the study results are still limited in the sense that they will be based on a small number of sites that have not been selected randomly from a universe of such sites and in fact do not offer any meaningful geographic representation. Accordingly, the results from our study will need to be further tested and expanded with large scale studies that include a large number of sites that offer representation along key dimensions such as geographic and demographic characteristics.

16. A16 states that the findings are primarily for internal use and yet other sections of the supporting statement imply that the results of this study will be used to inform future policy. Please clarify.

As noted above, we believe strongly that because our data collection protocol is grounded in a well developed conceptual framework that our study goes well beyond a pilot -type project. We believe the value of the project will be in providing a platform for larger studies to address the impact of financial incentives on quality in safety net settings. We do not believe our results should in any way be viewed as definitive and will qualify them with the appropriate caveats in all publications from the study. However, policy makers and managers are desperate for any information that can help guide them in their efforts to design and implement financial incentive programs and so we expect that there will be interest in our results with policy and practitioner communities.

Response rates

17. Please provide more information on the previous studies surveying this same population where you achieved a 60% response rate, and how that rate was calculated.

The response rate was calculated as the number of individuals included in the sample who returned a questionnaire divided by the number of individuals included in the sample, adjusted for the number of individuals who left the organization between the time the sample was constructed and the time the survey questionnaires were distributed. The sample at the RIPA site (where we achieved a response rate of 43%) consisted of all primary care physicians who

were eligible to receive financial awards under the RIPA quality-related incentive program.

18. In the package submitted for the briefing, the estimated response rate was 30%. This package estimates a 60% response rate, and yet it doesn't seem like the data collection methods have changed very much. Can you please explain how you will obtain the 60% response rate?

Since the briefing, we have had an opportunity to discuss survey logistics with our liaisons at the research sites and based on these discussions we believe that we will be able to achieve a much higher response rate than previously estimated. Our original estimate was intended to be very conservative and was offered without our having the opportunity to have detailed discussions about survey logistics with the site liaisons. Now that we know that we can expect a very high level of commitment from the liaisons in terms of promoting the survey and motivating respondents to complete and return the questionnaires, we believe we will achieve a response rate that exceeds what we achieved in the noted RIPA site.

19. Are "thank you/reminder" postcards being used after the initial distribution and prior to the second questionnaire distribution?

No, we will not be using thank you or reminder postcards after the first round of surveys are distributed, as was the case in prior administrations of the survey for the aforementioned sites in question. As noted above, we will work with a liaison at each site to remind survey respondents to return their survey. The liaison will be either an office manager, practice executive, or other administrative staffperson designated by the project site Principal Investigator.

20. B1: As part of the non-response bias analysis, it might be interesting to analyze whether those physicians who either performed very well or did not perform very well are either disproportionately represented or not represented.

Yes, we agree and plan to conduct this analysis. We did perform such an analysis (presented in the forthcoming Journal of General Internal Medicine paper accompanying our response) at the RIPA site and found no significant differences in performance between respondents and non-respondents.

21. B2: Please clarify how the survey will be distributed because B.2 states both that:

- "Boston University will send surveys to a site liaison who will distribute them individually to each physician participating in their organization's program."

and

- "Questionnaire booklets will be sent by first class mail address to the individual respondent and will be accompanied by a cover letter from the CEO of the relevant medical group and a pre-paid business reply envelope addressed to a third-party data entry vendor."

We have amended the supporting statement, section B2, to reflect the former method—that is, a site liaison will distribute surveys individually to each physician participating in their organization’s program.

22. B2: please clarify whether informed consent for interviews will be obtained verbally or in writing (in the first paragraph of B2, it says “written informed consent” while in the second paragraph, it says “informed consent will be obtained orally”).

Informed consent for the key informant interviews will be obtained in-person in writing, not verbally. We have amended the second paragraph in the supporting statement, section B2, to reflect this fact.

23. Please include a response to any public comments received (for example, the comment from the American College of Cardiology). Was the comment from the ACC the only comment received?

The comment from the American College of Cardiology was the only comment received by AHRQ in response to public notices regarding this project. We have attached a letter with our responses to their three comments.

Physician Survey Instrument

24. Is this survey the PAI-26? If not, has it been pilot-tested?

Yes, the survey instrument is the PAI-26, which was pilot tested and validated, as described in Meterko et al., 2006.

25. The formatting of the survey should be fixed, as the current formatting tends to cut off several of the questions (e.g. question 23, 25, 26, 27, etc).

We have fixed the formatting of the survey so that all questions are clearly presented.

26. Where will the standard PRA blurb, OMB control number, and expiration dates be printed for the survey and the interviews?

The survey has been revised to reflect placement of the standard PRA blurb, OMB control number, and expiration dates.

27. Q15: Delete “have” (I believe the question was supposed to say: “About what percent of your patients are eligible for the procedure...”)

The word “have” has been deleted from this question.

28. Section 3:

- How are respondents supposed to answer these questions if they work in the clinic that serves primarily uninsured patients?

This section of questions was included in the survey used in the Rewarding Results Evaluation where the programs were sponsored by health plans. We

will ask these questions in the BMC version of the survey because the BMC program is sponsored by BMC Health Plan. BMC Health Plan is a Medicaid managed care organization. All the physicians eligible for their survey will be participants in the BMC Health Plan. We have revised this section to specify BMC Health Plan.

Montefiore's financial incentive program is sponsored by the Medical Center and implemented by Montefiore Medical Group. Section 3 will not be included in their survey.

- What is the main purpose behind these questions? Are the quality targets being imposed on the physicians by these health plans? Or is the aim behind these questions to see whether health plans help or hinder physicians from achieving the hospital-based quality targets? Can the questions be revised to be more specific?

The main purpose behind these questions is to determine whether health plan sponsorship affects physicians' attitudes toward pursuing quality targets tied to financial incentives. As stated, BMC Health Plan imposes quality targets on their participating physicians. We have revised the questions to refer specifically to BMC Health Plan.

- Physicians tend to find it very difficult to answer broad questions like these which ask them to generalize about their experiences with health plans because of the wide diversity of third-party purchasers/health plans physicians contract with. Can the questions be revised to be more specific? Otherwise, you may end up getting a lot of "neutral" responses.

The questions have been revised to specifically refer to BMC Health Plan and will not require physicians to generalize their experiences with all the health plans they participate in.

29. Q54: Would it be useful to add a follow-up question to assess professional satisfaction prior to the implementation of the financial incentive program, to see whether satisfaction has gone down or up as a result of this program?

While we agree that it would be interesting to understand the role professional satisfaction plays in the success or failure of pay-for-quality programs, we are reluctant to add a follow-up question about professional satisfaction after the implementation of the program. Ratings of professional satisfaction prior to program implementation may be subject to recall bias. For example, because the Montefiore program began in 1996 some Montefiore physicians may have participated in the financial incentive program for a substantial period. Physician survey respondents may not be able to remark on their professional satisfaction prior to program implementation. The questionnaire does include a question about physician's current professional satisfaction and we will compare responses to that question with their performance on the quality measures.

30. Can a question be added to the survey that explicitly addresses the issue of cream-skimming? For example: "As a result of this financial incentive, physicians are more hesitant to treat non-compliant patients." Or "As a result of this financial incentive, physicians actively seek out those patients who are most compliant and easy to treat."

The current questionnaire does include several items that address such unintended consequences of using financial incentives. We are reluctant to add any additional items as this would add to the time required to complete the questionnaire without any clear benefit to the reliability and validity of the existing scales.

Interview Schedule

31. Has the interview protocol been piloted?

Yes, a very similar semi-structured interview protocol was piloted and used in the aforementioned Rewarding Results Evaluation project. We adapted the protocol to reflect programmatic differences in the two research sites, but the questions are largely the same. Two articles presenting results from data collection using the interview protocol have been published and are attached to our response (Bokhour et al., 2006, and Sautter et al., 2007).

32. Respondents are not likely to have answers to many of these demographic/administrative questions at their fingertips (e.g. how many physicians practice at the clinic, what is the racial/ethnic composition of the patients at the clinic, what percent of patients are low-income, etc.). It may be a better use of researcher and respondent time if these questions were asked in survey prior to the interview. At minimum, respondents should be given a copy of the interview protocol in advance so that they will have the information ready and available.

We agree that this information may be difficult to provide in a semi-structured interview without prompting respondents to gather these data prior to the interview. We will provide each interviewee with a copy of the protocol prior to the interview. We will also highlight the administrative and demographic questions when scheduling the interview so that respondents are aware they may need to collect information prior to being interviewed.

33. It seems like it would be important to understand how the quality targets were developed and *why* each center chose *which* quality targets. They seem to differ, for example, from the NCQA quality targets which are widely-used performance indicators. It would be interesting to know why they chose to adopt some targets and not others.

We agree that it will be important to understand the selection and criteria for the quality measures targeted by each financial incentive program. We intend to probe each interviewee, when appropriate, about this issue.

34. It would also seem to be important to ask Montefiore and BMC how they decided to structure their financial incentives. For example, why did BMC decide to assess the performance of whole clinical groups rather than individual physicians?

We agree that understanding the rationale for structuring each site's financial incentive program is an important issue. We have amended the interview protocol to probe this issue.

35. It would also seem to be very important to explicitly ask respondents how their financial incentive program differs from those implemented in non-safety net settings and why.

Our primary goal in the interviews is to gather specific information about the development and implementation of the financial incentive programs at each site. We recognize that each site may have modeled their program after those financial incentive programs implemented in non-safety net settings and we will probe interviewees about whether the structure of their programs was influenced by features included in non-safety net financial incentive programs.

36. Given the different stages that Montefiore and BMC are at with regard to their respective financial incentive programs, it would seem like two different interview protocols would be needed for each site. For example, since Montefiore is a “mature” program—founded in 1996—it would be interesting to know how the financial incentive program evolved (e.g. why did Montefiore find it necessary to stratify incentives based on physician performance and patient compliance? Were there unintended consequences early on in the program that seemed to necessitate this stratification?). For BMC, on the other hand, it would be more important to understand what their early challenges are.

While we agree that the differential maturity of the programs may require us to ask additional questions, we believe using the same semi-structured protocol across all sites would allow us to make some comparisons that two different protocols would not allow us to do. We will probe interviewees about the development of their respective programs and the challenges each program faced during implementation.

37. Section V: the first question in this section implies that some sites selected for the interview will not be participating in the financial incentive program. Is this the case? If so, it would seem to be a good idea to ask them why they chose not to participate in the program rather than skip to the end of the protocol.

We are interviewing only informants from practice sites that are participating in the financial incentive program.