Comparative Effectiveness Research Review Disposition of Comments Report

Research Review Title: *Closing the Quality Gap Series: Revisiting the State of the Science Bundled Payment: Effects on Health Care Spending and Quality*

Draft review available for public comment from June 23, 2011to July 21, 2012.

Research Review Citation: Hussey PS, Mulcahy AW, Schnyer C, Schneider EC. Closing the Quality Gap Series: Revisiting the State of the Science. Bundled Payment: Effects on Health Care Spending and Quality. Evidence Report 208 Part 1. (Prepared by the RAND Evidence-based Practice Center under Contract No. 290-2007-10062-I.) AHRQ Publication No. 12-E007-EF. Rockville, MD: Agency for Healthcare Research and Quality. August 2012. Available at: www.effectivehealthcare.ahrq.gov/reports/final.cfm.

Comments to Research Review

The Effective Health Care (EHC) Program encourages the public to participate in the development of its research projects. Each comparative effectiveness research review is posted to the EHC Program Web site in draft form for public comment for a 4-week period. Comments can be submitted via the EHC Program Web site, mail or email. At the conclusion of the public comment period, authors use the commentators' submissions and comments to revise the draft comparative effectiveness research review.

Comments on draft reviews and the authors' responses to the comments are posted for public viewing on the EHC Program Web site approximately 3 months after the final research review is published. Comments are not edited for spelling, grammar, or other content errors. Each comment is listed with the name and affiliation of the commentator, if this information is provided. Commentators are not required to provide their names or affiliations in order to submit suggestions or comments.

The tables below include the responses by the authors of the review to each comment that was submitted for this draft review. The responses to comments in this disposition report are those of the authors, who are responsible for its contents, and do not necessarily represent the views of the Agency for Healthcare Research and Quality.

Source: http://effectivehealthcare.ahrq.gov/search-for-guides-reviews-and-reports/?pageaction=displayproduct&productID=1235
Published Online: August 24, 2012

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #1	Executive Summary and Structured Abstract	Pg 11: The copied image is fuzzy.	We have improved the resolution of the figure.
Peer Reviewer #2	Executive Summary and Structured Abstract	p. vi How were rates set? Update? Effects on growth?	We did not make any change in response to this comment because we could not determine how it was relevant to the page number cited.
Peer Reviewer #2	Executive Summary and Structured Abstract	p. ES-1 Is capitation "bundled"?	We did not consider capitation to be a form of bundling. We clarified this point in the introduction text.
Peer Reviewer #2	Executive Summary and Structured Abstract	p. ES-2 comment on Figure E1 in "Key Design Features" - Which of these measures risk sharing?; general comment on Figure E1 - What about baseline waste? Upcoding?	We added risk-sharing and upcoding to the model. We consider baseline waste (i.e. baseline efficiency) to be specific example of what's meant by "general financial environment," which we clarified in the discussion of the conceptual model.
Peer Reviewer #5	Executive Summary and Structured Abstract	I believe the Abstract and Executive Summary should clearly state that there have been NO randomized controlled trials of bundled payments ever published.	We have clarified the study designs of the included sources in the ES adding the following text: "Among the reviewed studies, 48 employed observational designs, while 9 were descriptive. There was only one study randomized at the provider level, and we identified no studies of bundled payment programs randomized at the patient level."
Peer Reviewer #5	Executive Summary and Structured Abstract	Because the main goal of most of the new payment approaches being considered by CMS and other payors is to bundle institutional payments with physician payments, I believe the Abstract and Executive Summary should clearly state that very few of the available studies involved bundles that included physician services. In addition, some of the studies that did include physician services give very limited descriptions of what services are included.	We added the following sentence to the abstract: "Most of the bundled payment interventions studied in reviewed articles were limited to payments to single institutional providers (e.g., hospitals, skilled nursing facilities) and so have limited generalizability to newer programs including multiple care settings and/or multiple providers." We also added this text to the conclusion of the executive summary.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #5	Executive Summary and Structured Abstract	On a related point, consider noting in the abstract, and I believe you should note in the Executive Summary, that many of the interventions themselves are incompletely described (if we don¹t know what the bundle actually is, it¹s hard to evaluate its impact).	We added the following sentence to the abstract: “The interventions studied were often incompletely described in the reviewed articles.” We also added this text to the conclusion of the executive summary.
Peer Reviewer #6	Executive Summary and Structured Abstract	Comments on the Executive Summary (ES):In the last paragraph of the ""Background"" section on undesired effects of bundled payments, in addition to what is mentioned, it would be worth mentioning:--unbundling (mentioned on p. 11)--spillover effects (from p. 74)--upcoding (mentioned on p. 11)It would be useful in the ES to summarize the evidence on these adverse effects. They are critical for payers to evaluate in deciding whether to implement bundled payment.	We added the suggested undesired effects to the background section of the executive summary. We added a summary of the evidence on these effects to the executive summary, reflecting the fact that evidence was sparse.
Peer Reviewer #7	Executive Summary and Structured Abstract	The report states that the main intended audience is payers and providers in the US (ES-6). Policymakers are also a key audience as they frequently make decisions, particularly for Medicare and Medicaid, that other then have to figure out how to implement.	We added policymakers as a key audience in the applicability section of the executive summary and the summary and discussion chapter.
Peer Reviewer #7	Executive Summary and Structured Abstract	A key design feature that is missing from the conceptual model (ES-2; 13) is how the payment for the bundle is determined (e.g. administratively set, negotiated, set through competitive bidding).	We added this to the figure: “Method for determining payment.”
Peer Reviewer #7	Executive Summary and Structured Abstract	The authors never define what is meant by "financial environment" a term that first appears in the conceptual model (ES-2; 13) and is used in various part of the report. Do they mean profit margins experienced on average by different provider types or are they referring to extent of competition in the provider¹s market (and is generally referred to as market characteristics) or something else entirely.	We have explained what is meant by “financial environment” in the text accompanying the conceptual model.
Peer Reviewer #7	Executive Summary and Structured Abstract	In general, one sentence paragraphs should be avoided (e.g. page ES-4, lines 30-33).	This sentence has been combined with the paragraph that follows.
Peer Reviewer #7	Executive Summary and Structured Abstract	Structured abstract should include the magnitude of the identified effect in lines 39-41 on page vi.	We added the magnitude as suggested. The Structured Abstract conclusion now includes the sentence: “Reductions in spending and utilization relative to usual payment were less than 10 percent in many cases.”

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #7	Executive Summary and Structured Abstract	After reading the summary, I would have been able to tell someone explicitly what bundled payment systems were not included in the report, but not what was covered by the report. The summary would benefit from some additional information, which could be provided in tabular form: the specific interventions addressed, the settings included, the number of articles addressing each and the number of articles that included quality measures).	The descriptive table in the Results section has been added to the ES (Table E1).
Peer Reviewer #7	Executive Summary and Structured Abstract	It wasn't clear in the summary (and elsewhere in the report) whether all of the results discussed were statistically significant.	P-values for individual outcomes are now included in the results section when these were reported in reviewed studies. We added the following explanation to the "Methods" section, "Data Abstraction and Data Management" subsection: "We also abstracted the statistical significance of outcomes when directly reported in the reviewed studies. Studies did not always report test statistics or p-values for individual results or for pair-wise differences or comparisons across groups. We note p-values for individual outcomes when this information was available in the source study."
Peer Reviewer #7	Executive Summary and Structured Abstract	Page 13. Same comment as in summary; suggest adding method used to derive payment to conceptual model.	The method used to derive payment was added to the conceptual model.
Peer Reviewer #10	Executive Summary and Structured Abstract	This well written and well researched report has more limitations than a casual reader might realize, and they should be made more explicit especially in the Executive Summary (ES). For example, only 3 of 22 studies reviewed deal with true episodes of care.	We added several sentences on limitations in the first paragraph of the conclusion of the executive summary. We have also highlighted this limitation in other changes throughout the report, including the abstract (under limitations); in the executive summary (results section, applicability section); the results chapter; and the summary and discussion chapter (summary of key question 2, applicability section, limitations, conclusion).
Peer Reviewer #10	Executive Summary and Structured Abstract	A small table describing the studies along the dimensions of US vs outside US, bundle vs episode, one site vs multiple, etc, would clarify the text in the ES.	The descriptive table from the results section has been added to the ES (Table E1).

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #11	Executive Summary and Structured Abstract	There is a statement about the theorized effect of bundled payment on coordination of care that appears twice (once in the executive summary on page ES-1 and later in the body of the report on page 11). It suggests that coordination might ensue because providers under multi-provider bundles will have to figure out how to allocate the dollars. While this may be true it seems like a rather narrow view of the world. I would suggest another reason that coordination may occur as a result of bundled payment is that it is costeffective. That has historically been the aspiration for capitation and other methods of risk sharing – that not only would providers eliminate waste but they would seek out better ways of caring for patients (e.g. coordinating with other providers) that would both reduce costs and improve outcomes.	We have revised the introduction and executive summary and introduction to address this comment as follows: " Providers are typically given discretion over the allocation of the services used to treat the patient's episode most effectively. This flexibility may encourage providers to use resources to coordinate care; often, these services are not reimbursed under fee-for-service payment. If the bundle includes services delivered by multiple providers in multiple settings, providers have to create a mechanism for managing the shared payment for a given treatment or condition, which could also foster coordination."
Peer Reviewer #11	Executive Summary and Structured Abstract	Page ES-7 in the first recommendation you mention randomization. Because most of these policies were only observed when they were rolled out as national policy that seems like a funny and disconnected suggestion – it is not really for researchers but for Congress. In the longer version of these recommendations I think you did a perfect job of capturing this in a way that was situated in the feasible set of what researchers might control. Fine to mention randomization down the list somewhere but first and foremost you should be focused on better quasi-experimental methods, more detailed info on the intervention and context, etc.	We added new introductory text to the Recommendations for Further Research subsection in both the Executive Summary and full report body. The Recommendations for Further Research subsection (in both locations) was edited to highlight the tension between rigorous and timely evaluations and to highlight practical recommendations under each heading.
Public Reviewer: Arnold Chen	Executive Summary and Structured Abstract	Executive summary: A thought came to mind looking at the abundance of great information, I noticed that complete words are substituted for just letters. Not everyone is familar with abbreviation for example "AHRQ" and such. It would be good to have the whole spelling of each listed.	All abbreviations have been included in a glossary of acronyms, and are spelled out for clarification the first time they are introduced in the document.

Commentator & Affiliation	Section	Comment	Response
Public Reviewer: Dolph Chianchiano	Executive Summary and Structured Abstract	Executive Summary: This comment relates to Recommendations for Further Research. We believe that a complete analysis of the effect of bundled payment systems on quality of care requires an examination of the impact of bundled prospective payment systems on innovation in health care delivery (e.g. new diagnostic and /or therapeutic interventions). It should be possible to compare and contrast different payment models and their impact on innovation. We are concerned that unless there is an adjustor recognizing the additional cost associated with adopting new technology, those who might develop and utilize new items and procedures may not find sufficient incentive to move forward with such work. As a result, opportunities to enhance patient outcomes and, ultimately, reduce demands on health care resources could be lost.	We added the following sentence to the Recommendations for Further Research subsection [emphasis added]: "Important design features to be addressed include the definition of the bundle (how many providers are included, what length of time, which services are included and excluded from the bundle); … *methods to update payment rates to reflect new technologies*; and methods for distributing payment among participating providers." We believe the broader recommendation to integrate an "innovation adjustor" in bundled payment systems is outside the scope of the current report.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #1	Introduction	I would suggest adding the motivation for this review, perhaps something along the following lines: Policy makers are considering the effects of having payers switch from straight fee-for-service to a bundled payment. [None of the studies really look at the impact of these issues with respect to consumer payments.] The hope is that bundling would give providers incentives to choose those inputs that can most efficiently achieve the outcomes desired by the bundle, which is likely to use fewer resources than the situation under reimbursed FFS in which each additional billable service generates revenue. Thus, part of this review will assess the evidence on whether bundled payments appear to result in lower cost to payers or lower use of services. If one assumes, however, that under FFS providers use all the services that could benefit the patient, then a reduction in the use of services could result in a reduction in quality. On the other hand, if FFS leads to excessive use of services, or the failure to compensate for the time for appropriately coordinating care, or the failure to offer services that are not billable, then bundling might improve the quality of care. Thus, the important policy-relevant combinations of findings are: If there was no convincing evidence that bundling resulted in savings to payers, then the effort to make the change would probably not be worth it, UNLESS there was clear evidence that bundling IMPROVED quality. If the evidence was reasonably convincing there were savings to be had, but quality was WORSE, then policy makers would be in a quandary. If the cost reduction evidence was reasonably convincing and quality was convincingly IMPROVED, then it would be a ²double win² Etc. The bottom line of this review is that, while the underlying studies of only fair to moderate methodological quality, (a) the results indicating lower costs or resource use are quite consistent, although the magnitudes of the savings vary and (b) the results with respect to quality are inconsistent, offering no clear evidence of either worsened or improved quality.	We have added a paragraph with the motivation for this review as the second paragraph of the introduction chapter (p. 11). The paragraph draws heavily on the text suggested by the reviewer.
Peer Reviewer #1	Introduction	Pg 19: Bundled payment creates incentives to do more bundles, but so does FFS	We added the following parenthetical statement: "(an incentive to increase service volume also exists under fee-for-service payment)."
Peer Reviewer #2	Introduction	It is fine.	No change in response to this comment.
Peer Reviewer #3	Introduction	The introduction is pretty clear, and lays out the key questions well (although I don't think the figure actually clarifies anything that isn't already put pretty clearly in the text).	No change in response to this comment.
Peer Reviewer #4	Introduction	See comments below for issues that should also be covered in the introduction.	No change in response to this comment; responses to related comments included below.
Peer Reviewer #5	Introduction	The introduction is clearly written and presents a good conceptual model.	No change in response to this comment.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #7	Introduction	Page 10, line 20. Not all fee-for-service payments are [3]defined by a fee schedule[2]. Recommend striking that phrase.	In the revised report, we differentiate more clearly between fee-for-service payment, which we do define as using a fee schedule to define payment rates, and cost-based reimbursement. See Box 1 in the introduction.
Peer Reviewer #7	Introduction	Page 13. Same comment as in summary; suggest adding method used to derive payment to conceptual model.	This has been added to the model.
Peer Reviewer #8	Introduction	Generally, the introduction was well done, but I have 2 suggestions. The first relates to Figure E1 and later (on p. 13) , Figure 1. This figure was taken from other work but parts are not entirely clear. For instance, in Figure E1, is "Organizations' Response" a heading and should the next 2 lines be bullets like in the other entries? This would seem right to me, as care redesign and selection of low risk patients both seem like organization responses to me.	This was a formatting issue, bullets were added to the figure as suggested by the reviewer.
Peer Reviewer #8	Introduction	Also, I might add "investment in infrastructure" as another type of organization response.	We added this to the figure.
Peer Reviewer #8	Introduction	Also with respect to the same figures, I would add "level of payment" as a key design feature. Possibly it could be argued that this is subsumed under payment methodology, but I think it deserves to be broken out and highlighted as a separate bullet point. In the later discussion of findings, the authors point out that different methods were used to establish the level of payment, and that the level possibly can affect the organizational response, although no evaluation was able to address this explicitly. The importance of "level of payment" often gets overlooked as people focus on incentives, but I would argue that level potentially can make a huge difference in provider responses.	We added this to the figure.
Peer Reviewer #8	Introduction	I also think that in the introduction on p. 10 the authors should include formal definitions of fee-for-service, capitation and bundled payment (possibly in a box or cut-out). One could argue that the readers will understand the differences, but that has not always been my experience. The authors then can provide an example of how the formal definitions play out in practice. Also, at some point in the introduction, the authors should define both "spending" and "costs" as these terms will be used subsequently in the review. In reading the summaries of findings, my impression is that most relate to purchaser spending rather than resource value (costs).	We added a box listing features of alternative approaches to bundled payment in the "Background and Objectives for Systematic Review" subsection of the Introduction. We also define our use of the terms "spending" and "costs" in the "The Key Questions" subsection. These definitions state: "By spending we refer to the amount paid to providers in exchange for health care services, i.e., payments to providers. By costs we refer to the value of resources used to provide health care services by providers, e.g., hospitals." We revised the report throughout to conform to these definitions.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #10	Introduction	Limitations: The report states that its "main intended audience" is payers and providers in the US that are considering implementation of a bundled payment program[2] (draft p. ES-6, lines 31-32 and p. 71, line 44). This audience[1]s biggest interest however is in bundles in the form of episodes of care across multiple sites and providers. Despite thorough research, only 3 studies (of 18) address that type of bundle (the rest being one-site and generally one facility provider e.g. Medicare DRG payments and equivalents). Only 11 of the studies (including those 3) were done in the US. Therefore the paper[1]s limitations should be expanded to address this limitation more explicitly, in the Executive Summary, Introduction, Discussion and Conclusions. Otherwise casual readers could be mislead as to the reports main conclusion, [3]In summary, this review found evidence that bundled payment was associated with reductions in health care costs and utilization with inconsistent and generally small effects on quality measures[2] (ES-7, 52-53).	In order to emphasize this limitation, we have referenced this issue in the abstract (under limitations); in the executive summary (results section, applicability section); the results chapter; and the summary and discussion chapter (summary of key question 2, applicability section, limitations, conclusion).
Peer Reviewer #10	Introduction	The conclusions on cost savings should be tempered by the applicability of the findings to episodes of care. The paper only found 5-15% savings in general. Not all procedures are appropriate for episodes of care, and many have to further limited by removing unusual cases. The result is that the proportion of health care to which the savings could be appliedis currently limited. Overall savings to the health care system wouldbe a fraction of 5-15% that bundled payments could achieve.	We added a new paragraph to the "Applicability" subsection of the "Summary and Discussion" section which clarifies that episode-based payment is not feasible in every health care context as the review mentions.
Peer Reviewer #10	Introduction	The perceived need in the past to engage providers by introducing bundles at generous rates (for example, with Medicare DRG payments) is another potential limit to cost savings.	We added a sentence in the "Key Question 2" subsection in the "Summary and Discussion" section to make this point: "At the same time generous initial payments may mitigate the impact of bundled payment on reductions in spending."
Peer Reviewer #10	Introduction	A further technical limitation especially of the episodes of care would be sample size. The report alludes to this in the section on Geisinger ProvenCare, where the N was 117 intervention and 137 comparison patients. These samples are almost assuredly too small to address mortality and readmission rates (unless the baseline was unusually high).	We did not make a change to the introduction in response to this comment as strength of evidence is not explicitly discussed in the Introduction. The reviewer's point is addressed in the discussion section where "small and/or convenience samples of providers" is cited as a main methodological concern. Statistical power was also considered in rating the quality of individual reviewed studies.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #10	Introduction	Many of the pilots involved self-selected institutions, further limiting the generalizability of the resutls. This could be discussed in more detail.	This point is made in the "Key Question 2" subsection in the "Summary and Discussion" section. We added the following sentence at the end of the paragraph describing the Medicare Heart Bypass and Cataracts Demonstrations: "The effect on spending would also likely be lower if hospitals participating in the demonstration were not selected in part on their ability to negotiate low payment rates." We did not make a change to the Introduction section in response to this comment.
Peer Reviewer #11	Introduction	Page 10, second to last para. You write "incentive to improve the services" – I think you mean compress or reduce	We have edited this sentence as follows: "Global payment and capitation create an incentive to reduce the need for services (bundled or not)."
Peer Reviewer #11	Introduction	Page 12, typo in the last para "the impact of these design features IS addressed"	The word impact has been changed to read "impacts."
Peer Reviewer #12	Introduction	The report is well motivated and well written. My only main suggestion, per below, is that the authors may want to discuss in the introduction, the differences between currently discussed policy/payment models (ACOs, cross-continuum efforts) compared to what the past literature focuses on (mostly inpatient pps efforts that excludes physician payments). This will set the stage up front that the past literature really looks at things that are different than is what is currently being considered in many policy circles.	We added this discussion to the introduction as suggested. Following the bullets outlining different levels of aggregation of services possible in bundles, we contrast previous systems with relatively little aggregation (using the example of the Medicare Inpatient Prospective Payment System) to newer programs with relatively greater aggregation (using the example of the Medicare Bundled Payments for Care Improvement initiative).
Peer Reviewer #1	Methods	All these are fine.	No change made in response to this comment.
Peer Reviewer #1	Methods	Pg 27: Use caps: GRADE = Grading of Recommendations Assessment, Development and Evaluation (short GRADE) Working Group www.gradeworkinggroup.org/	This has been changed from "Grade" to "GRADE" wherever it appears, and the acronym has been added to the table of acronyms.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #2	Methods	The criteria are reasonable. The categorization of papers by quality of methods is important, but I am not sure what we learn from the poor quality studies. I also care why the study is poor. Some reasons relate to bias, others to power. I am much more concerned about bias. It is not clear how criteria for quality are weighted. I would like the paper better if it was shorter and focused on only those studies where we really learn something even if it is not necessarily generalizable.	We explain the rationale for all studies rated "poor" or "good" in Appendix A. There was no explicit weighting of criteria, following guidance in the AHRQ EPC Methods Guide. We did not exclude poor studies from the synthesis because several of the poor studies contributed information not included in other studies (e.g., unique study outcomes measured). In response to this comment, we conducted a sensitivity analysis that determined that our conclusions were not sensitive to the inclusion of studies rated "poor." We added a brief discussion of this issue to the summary and discussion chapter.
Peer Reviewer #2	Methods	p. 26, lines 8-18: How would IV or OD studies be rated?	Study ratings are based on several criteria in addition to study design, therefore specific designs are not associated with specific ratings.
Peer Reviewer #2	Methods	pg. 26, line 12: *KB note: reviewer made a question mark on line 12 after text*	We believe the question was about the reference to "Grade" – we replaced this with "GRADE."
Peer Reviewer #2	Methods	pg. 26, lines 25-33: How are these criteria weighted? Emphasis should be on confounding.	There was no explicit weighting of criteria, following guidance in the AHRQ EPC Methods Guide.
Peer Reviewer #3	Methods	The methods section systematically lays out how the study was done, but although it was interesting to see the detail on how the reviews were conducted, I think it would be an easier read if that detail were put in an appendix so the reader could refer to it if interested. More discussion of the ratings of the studies as "good", "fair", or "poor" would be useful. It was hard to understand the meaning of those designations, especially when their results were not distinguished much in the presentation of the results; the designations were referred to throughout, but they were treated pretty similarly in the presentation--analogous to when authors discuss the implications of the direction and magnitude of estimates that are statistically insignificant (which, by the way, the authors of this report occasionally do).	Appendix A includes explanations for all "good" or "poor" ratings. In the summary and discussion chapter, we added a discussion that we did not exclude any studies based on their quality rating, and that we determined that our conclusions were not sensitive to the inclusion of "poor" studies.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #4	Methods	The study completely missed a couple of important studies that have been done on the topic, despite the fact that they are frequently cited in the current literature about payment reform, and given the paucity of studies on bundled payments, this is a serious omission. For example, a project in Michigan that bundled hospital and physician payments for orthopedic procedures and included a 2-year warranty isn[1]t described at all, nor is a project in Texas that accepted a bundled payment for heart procedures and preceded the Medicare demonstration. Both of these projects were initiated by providers, not payers, which makes them particularly important to include to balance the primarily payer-driven projects reviewed in this report. (The report states that [3]new bundled payment systems faced significant initial resistance from providers[2] without ever noting that the studies examined focused almost exclusively on payer-imposed systems.)	We identified the Michigan and Texas programs referenced by the reviewer. We have added the Michigan study (Johnson and Becker, 1994) to the review. The Texas study did not meet inclusion criteria as it did not report an effect of bundled payment on the study outcomes of interest. In our limitations, we note that our search strategy may have excluded relevant studies.
Peer Reviewer #4	Methods	Also, although the report states that it examined the grey literature, the methodology described for finding studies does not appear to have systematically looked for them.	We clarified how we addressed grey literature in the methods section: " These were identified through *ad hoc* searches of research sponsors' Web sites, reference mining, and recommendations from the Technical Expert Panel, as these are generally not included in literature databases."
Peer Reviewer #5	Methods	The inclusion and exclusion criteria for studies are clearly stated and appropriate, as are the definitions of the outcomes measures.	No change in response to this comment.
Peer Reviewer #6	Methods	Are theoretical models/studies of bundled payments within the purview of this report? It should be made clear whether they are or aren't. A section on theories/theoretical/conceptual models of bundled payments would help set up the anticipated effects.	We did not include conceptual papers, and added a note to the methods.
Peer Reviewer #6	Methods	At several points the report mentions the Medicare Acute Care Episode Demonstration and Prometheus Payment. But these programs are not described. I realize they have not been evaluated yet, but it might add to the report to describe these initiatives, at least at a high level.	We have added brief descriptions of these and selected other programs to the results chapter, immediately preceding Table 1.
Peer Reviewer #6	Methods	It might be helpful to clearly distinguish between "payments" and "costs" throughout the report, or at least include a statement distinguishing them somewhere near the beginning.	We added definitions of the terms "spending" and "costs" in the "The Key Questions" subsection of the Introduction. These definitions state: "By spending we refer to the amount paid to providers in exchange for health care services, i.e., payments to providers. By costs we refer to the value of resources used to provide health care services by providers, e.g., hospitals." We revised the report throughout to conform to these definitions.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #6	Methods	The effect of bundled payment on provider risk is mentioned, but rarely discussed. Probably few of the studies considered it, but provider risk is a critical aspect of bundling payments.	We did not explicitly include provider risk as a study outcome of interest for our three key questions. We agree with the reviewer that few or none of the studies would have explicitly reported on it. However, we did abstract and include data on all financial effects of bundled payment reported in reviewed studies.
Peer Reviewer #7	Methods	Page 17, lines 37-39. It states that two reviewers assessed articles and differences were resolved by consensus. A statement about the extent of agreement would be useful. Same comment for page 18, lines 33-35.	We did not calculate inter-rater agreement for this report because most data fields abstracted were free text and almost always differed in some way between reviewers. In general, the reviewers did not encounter difficulty in resolving these differences by consensus.
		Page 17: there is inconsistent use of punctuation in the bulleted lists.	The bulleted lists are now consistently punctuated. Semicolons, commas, and periods are now used uniformly across the lists.
		Page 17, reference to Appendix A (line 56). Appendix A would be easier to consume if the articles were ordered alphabetically within program.	Appendix A is organized alphabetically by primary author.
		Page 18, line 40. Policymakers are also a relevant audience for the report.	Policymakers have been added as a relevant audience in the applicability sections in the executive summary and summary and discussion chapter, as well as in the methods paragraph on applicability.
Peer Reviewer #8	Methods	The methods used to identify relevant studies, assess the strength of the evidence and draw conclusions are described clearly and appropriately.	No change in response to this comment.
Peer Reviewer #10	Methods	These were appropriate.	No change in response to this comment.
Peer Reviewer #11	Methods	The methods were clearly stated and justifiable. It strikes me that if Peter thinks my papers belong in the review (see below) that the terms "case rate" perhaps should be searched to be sure there were no other papers that used that terminology. No statistical methods used. Paper quality designations were based on standard criteria.	We tested the use of the term “Case rate” as a search term, but other than the articles referenced by the reviewer (which were included in the review) the revised search did not yield any additional articles for inclusion.
Peer Reviewer #12	Methods	The search methods seem appropriate and fine to me and are very well and clearly described.	No change in response to this comment.
Peer Reviewer #12	Methods	Yes - methods seem appropriate.	No change in response to this comment.

Commentator & Affiliation	Section	Comment	Response
Public Reviewer: Vinita Ollapally, American College of Surgeons	Methods	The American College of Surgeons (ACS) appreciates the work of the Agency for Healthcare Research and Quality (AHRQ) in preparing this Draft Comparative Effectiveness Review titled, "The Effects of Bundled Payment Systems on Health Care Spending and Quality of Care" (Review). Although this Review provides a literature review of some of what is known about bundled payment currently, we believe that there are important issues related to bundled payment that would be useful for AHRQ to study going forward. As such, our comments are focused on methodology and areas for future research. For the purposes of bundled payment, we question the relevance to today's environment and culture of literature going back almost two decades. Instead of a literature review, it would be more productive to study areas in healthcare where robust bundled payment models have already been in place. One example is transplant surgery, where bundled payments have been used to pay for organ transplantation for two decades. In transplant, the payment for the hospital and physicians are joined, and outcomes are transparent due to the posting of all outcomes on a public website. For these and other reasons, transplant is an ideal example of successfully implemented bundled payment. Relevant questions in the study of transplant could be: What led to this paradigm? What was/is the culture that makes it work? Has it impacted cost? What can we learn about a proven paradigm? What key success factors of this paradigm are applicable to other disciplines?	No change in response to this comment because it is a recommendation for future research priorities for AHRQ.
Peer Reviewer #1	Results	In general, these are fine. The studies are necessarily drawn from a variety of settings. The authors do a good job of making this clear.	No change in response to this comment.
Peer Reviewer #1	Results	Pg 40: "The second reported discharged to home (rather than to community or death)." What distingushes home from community, I assume this was not discharged to the "street"	We corrected the typo pointed out by the reviewer in the Medicare SNF PPS "Effect on health care quality" section. The discharge outcomes are: 1) home/community, 2) hospital; 3) death.
Peer Reviewer #1	Results	I'm not sure if there is a simple alternative to always saying, "in this fair study Jones found" and "in this poor study Smith" Given funding and other constraints, the studies may have been fine for the purposes for which they were written. (At the next research conference I picture the authors being attacked by Smith, Jones, and dozens of other aggrieved researchers.) Perhaps just use A-D ratings to convey the review specific quality scores, "Jones (22C) found... and Smith (31D) found."	The rating system and the use of the ratings in discussing results follows methodology outlined in the AHRQ EPC Methods Guide. We added additional information in the "Assessment of Methodological Quality of Individual Studies" subsection of the "Methods" section to describe our use of these terms and the AHRQ methodology.
Peer Reviewer #2	Results	I think that the results are well presented. I think that there are certain questions, like the importance of context, that I am not sure the literature could support answering at this stage.	No change in response to this comment.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #2	Results	p. 36 lines 40-44: What were the study designs?	Study designs were not noted in the reviewed articles. As we mention in several places, one of the key limitations is the lack of detail provided on the underlying studies.
Peer Reviewer #2	Results	p. 37 lines 31-32: So what should we conclude	We agree that this is an ambiguous result and deleted the passage.
Peer Reviewer #2	Results	p. 39 lines 26-28: Emphasis should be on good study.	No change was made in response to this comment. We describe evidence from all studies meeting the selection criteria and note the strength of evidence ratings as described in the "Assessment of Methodological Quality of Individual Studies" section. We did not exclude poor studies from the synthesis because several of the poor studies contributed information not included in other studies (e.g., unique study outcomes measured). In response to this comment, we conducted a sensitivity analysis that determined that our conclusions were not sensitive to the inclusion of studies rated "poor." We added a brief discussion of this issue to the summary and discussion chapter.
Peer Reviewer #2	Results	p. 40 lines 31-37: Focus on good.	See the previous response.
Peer Reviewer #2	Results	p. 41 lines 3-6: Emphasize those with good design.	See the previous response.
Peer Reviewer #2	Results	p.44 lines 45-46: Why did reimbursement rise?	No change made in response to this comment as this point is not specifically addressed in a reviewed study. The authors hint that the design and implementation of PPS was the cause of the increase. The PPS also includes annual reimbursement rate revisions.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #2	Results	p. 46: Focus on results of good stdies. Supplement with others.	No change was made in response to this comment. We describe evidence from all studies meeting the selection criteria and note the strength of evidence ratings as described in the "Assessment of Methodological Quality of Individual Studies" section. We did not exclude poor studies from the synthesis because several of the poor studies contributed information not included in other studies (e.g., unique study outcomes measured). In response to this comment, we conducted a sensitivity analysis that determined that our conclusions were not sensitive to the inclusion of studies rated "poor." We added a brief discussion of this issue to the summary and discussion chapter.
Peer Reviewer #2	Results	p. 49: An overall summary in each section would be helpful.	No change was made in response to this comment. Summary sentences or paragraphs are included in some but not all results sections. The "Summary and Discussion" section further synthesizes evidence across payment systems.
Peer Reviewer #2	Results	p. 59 lines 24-26: Why fair if DD	Study quality ratings are based on multiple criteria, so that a particular evaluation design is not always associated with a particular quality rating.
Peer Reviewer #2	Results	p. 60 lines 44-50: Why include at all.	Although this study was rated as poor, it did meet our inclusion criteria.
Peer Reviewer #2	Results	p. 62 line 5: Was it pre/post?	The reviewer is correct. We clarified that the Hasegawa et al. result is pre/post.
Peer Reviewer #2	Results	p. 65: lines 3-8: Does this make sense?	Yes, we believe that this makes sense and therefore did not make a change in response to this comment.
Peer Reviewer #2	Results	p. 66: line 32: Which way would bias go? How was trend dealt with	We were not able to address these questions using information from the reviewed study.
Peer Reviewer #2	Results	p. 68, line 28: Why? Sounds like decent control group. Was there evidence of bias.	We added our rationale for the fair rating: "Reviewers rated the report as "fair." due to the small number of participating hospitals and the risk of bias from self-selected participants."

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #2	Results	p. 69, lines 38-39: Shift. See next page.	We clarified the result listed under "other effects." This subsection now begins: "Two of the originally participating hospitals realized statistically significant gains in Medicare bypass market share while a third realized a statistically significant decline in share. All three hospitals added later in the demonstration realized statistically significant declines in Medicare bypass market share." Individual hospitals both gained and lost share. Overall, net changes in market share contributed a small fraction of total savings to Medicare.
Peer Reviewer #2	Results	p. 70, line 5: I thought you said there was savings due to shift.	See the previous response.
Peer Reviewer #3	Results	As I mention above, I would move the results section into an appendix--there is too much detail for the reader to be able to go through, and it would be more useful to have the information as a reference for readers who are interested in that detail. The splitting up of the section by system is logical, but results in a very fragmented text that is really difficult to read through--also, if so much attention is going to be paid to each system, then the descriptions of how the system was designed and what it was intended to do should be expanded. I think Figures 2 and 3 ought to be included in the methods section (they portray the results of the methods in the study, rather than the results of the systems the study reviewed).	We did not reorganize the report as suggested following AHRQ guidance on EPC report structure.
Peer Reviewer #3	Results	I really liked Table 1, but I have one quibble with the contents: it's really confusing to say that the intervention date for the Medicare Inpatient PPS was January 1989. I know the footnote kind of explains why the authors chose to use that date (although I believe that the transition was complete by federal fiscal year 1988), but I would use October 1983 as the date and use the same footnote.	We changed the implementation date to October 1983 as suggested.
Peer Reviewer #3	Results	Also, the paragraph on page 25 on the implementation of the Medicare Inpatient PPS is kind of muddled: it should make clear that the transition was from a blend of hospital-specific and regional average payment rates to national rates.	We have clarified this paragraph as suggested by the reviewer.
Peer Reviewer #3	Results	Also, on line 22 of page 25, I would drop the word 'later': private payers and a number of Medicaid programs adopted systems similar to the Medicare inpatient PPS pretty quickly (and many later dropped those systems).	We dropped the word "later" as suggested by the reviewer.
Peer Reviewer #4	Results	There are some specific aspects of the structure and impacts of bundled payment systems which are ignored or given short shrift in the report:	Specific related comments addressed below.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #4	Results	How the price of a bundle is determined and by whom is a critical issue, but this is typically mentioned in passing in the report rather than given detailed attention, and this is a very important issue for future bundled payment initiatives. Indeed, the report seems to define the issue away by defining a bundled payment as "related to the predetermined expected costs of a grouping of services," when, in fact,the payme not be a clear sense of what costs would be expected under a bundled payment arrangement. (This weakness is exacerbated by the lack of a clear distinction between "cost" and "spending.")	We have addressed this comment in several places in the report. The method for setting the price of a bundle is included as a key design feature of interest in the review. We discuss the issue of how the price of a bundle is determined in the descriptions of reviewed programs (where available from reviewed studies). In the summary and discussion chapter, we include a paragraph discussing the potential impact of the price-setting method on bundled payment effects.
Peer Reviewer #4	Results	Whether the different providers included in a bundle are independently or jointly employed makes a major difference in the ease and impact of implementing a bundle, but this point is given only passing attention in the report.	We included this as a key design feature in our conceptual model and abstracted information on this feature from every reviewed study. However, very few studies reported information on this design feature, so it is not featured in the results section. We raise this as a limitation in the summary and discussion chapter.
Peer Reviewer #4	Results	The report describes bundling "different providers" and "different settings" as though they were equivalent concepts, when they are very different. (For example, paying a doctor and hospital jointly for an inpatient procedure is very different from paying a hospital and post-acute care providers jointly for an entire episode of care.)	We have edited throughout the report to replace the providers/settings distinction with the phrase “multiple providers and/or provider types.”
Peer Reviewer #4	Results	The report does not distinguish between combining multiple services during the same short window of time vs. services delivered over a longer window of time, or between combining services that are almost always provided together vs. services that are only occasionally combined or are actually undesirable (e.g., treatment of a preventable infection).	In the introduction, we discuss the distinction between aggregation of services over time, across providers, and as warranties for complications.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #4	Results	The report does not adequately focus on the need for, effectiveness of, and implications of risk-adjusting bundled payments.	We agree with the reviewer about the importance of risk-adjustment in bundled payment programs and include a discussion of the need for robust risk adjustment in the introduction. We abstracted data on the use of risk-adjustment and any study outcomes related to risk selection or related to risk adjustment. Reviewed studies included little relevant information, which is why this is not a strong focus of this report. We include a paragraph discussing the use of risk adjustment and outlier payments in the summary of evidence related to key question 2 (design features) in the summary and discussion chapter. We discuss the importance of studying the effect of risk adjustment in the suggestions for future research section.
Peer Reviewer #4	Results	The report does not discuss whether all patient conditions or treatments were paid through a bundle or only some of them were in the studies reviewed. The Medicare PPS programs were comprehensive bundling systems, in the sense that a provider could not shift a service or cost from a bundled payment to an unbundled payment, but most other bundled payment projects have focused on only a subset of services and patients, so the ability of a provider to reclassify a patient or a service into anunbundled category is much higher.	In the results section, we include a description of each reviewed bundled payment program, including a description of the bundle definition. Our descriptions were limited by the information available in reviewed studies, as noted in the revised report.
Peer Reviewer #5	REsults	There is excellent detail about the studies. The limitation here is that the original published studies often do not have adequate detail. Figures and tables are excellent.	No change made in response to this comment.
Peer Reviewer #6	Results	It would be helpful in describing results to clearly distinguish short-term or transitional effects from longer-run effects. The longer-run effects would seem to depend critically on the update of the payment rate over time. The discussion of time horizon on p. 72 in the "Recommendations for Further Research" section is good.	We agree the distinction between short and long-term effects is important. As the reviewer notes we make this point in the Summary/Discussion and Recommendations for Future Research sections. We did not note whether each outcome is long or short-term. The “Overview of design of relevant studies” subsection describes the distribution of study time horizons for many payment systems with multiple reviewed studies.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #6	Results	Bundled payment is virtually always compared to "usual", i.e., FFS payment. Is this the correct comparison? What about comparing to other alternative payment systems, such as capitation or global payment? The ultimate goal of policymakers may be to slow public health care spending to e.g. GDP growth plus 1%. What about comparing bundled payment to that standard?	We added a paragraph in the "Applicability" subsection of the "Summary and Discussion" section which states all reviewed studies compare bundled payment to either fee-for-service or cost-based reimbursement.
Peer Reviewer #7	Results	It is not clear whether the descriptions for the Medicare PPS for various settings of care were supposed to reflect when the systems were initially implemented or current day as they were written in a mix of present and past tense (sometimes both in the same description). If they are supposed to reflect present day, there are some inaccuracies and/or inconsistency in the completeness of the descriptions. I encourage the authors to verify their descriptions against MedPAC[1]s payment basics series. The authors in some places state that there are no quality incentives, and in other places are mute on the topic. It should be consistently stated if a setting has a pay for reporting or pay for performance program in place (though this is only relevant if the descriptions are supposed to reflect current policy). Also, cost-sharing and limits on coverage are discussed for some settings, but not others.	We changed the tense of the program descriptions consistently to the past tense. We noted that the review is limited by the completeness of intervention descriptions in the reviewed studies. We included information on the use of quality incentives in the program descriptions where information was available.
Peer Reviewer #7	Results	Page 29, lines 9-13. Please explain what is meant by [3]high degree of financial pressure[2] and [3]lower or negligible pressure[2].	Added definition from source: "Feinglass defined financial pressure as a measure of the 'gap between (hospitals') actual costs and their anticipated or actual PPS revenues.'"
Peer Reviewer #7	Results	Page 29, lines 28-30. The authors state that SNF spending was reduced by$9.5 billion from 1998-2003 and that spending declined to $9.5 billion in 1999. Please verify that 9.5 billion is correct in both places.	No change made in response to this comment. Both $9.5 billion statistics confirmed.
Peer Reviewer #7	Results	Page 29, lines 34-40. There is a long complex sentence that is difficult to follow. Please split into 2 sentences or otherwise simplify.	We made an edit for clarity as suggested.
Peer Reviewer #7	Results	Page 30, line 20/21. How is efficiency defined?	No change made in response to this comment. The term as used by the authors is defined on the next page.
Peer Reviewer #7	Results	Page 32, lines 7-16. The baseline numbers for the three studies seem completely at odds with one another. Are there differences in the population not mentioned in the report?	We corrected an abstraction error. The first two outcomes are use of rehabilitation services, not SNF.
Peer Reviewer #7	Results	Page 32, line 33. [3]Šdid not significantly affect quality.[2] Is this not statistically significant or substantially significant?	We substituted "consistently" for "significantly." This sentence is now consistent with usage elsewhere in the report.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #7	Results	Page 32, line 45. How does discharge home differ from discharge to community?	We corrected the typo pointed out by the reviewer in the Medicare SNF PPS "Effect on health care quality" section. The discharge outcomes are: 1) home/community, 2) hospital; 3) death.
Peer Reviewer #7	Results	Page 34, lines 37-39. I didn¹t understand this sentence.	Reordered the "Payment Method" section in the IRF PPS description to transition immediately from the introduction of Case Mix Groups (CMGs) to a description of CMGs.
Peer Reviewer #7	Results	Page 35, lines 46-47. Please provide more detail on the difference between spending per IRF pt versus reimbursement per IRF pt. Or are you saying there were increases in reimbursement during the implementation period followed by decreases after implementation?	Sood et al. distinguish between marginal and average reimbursement. This description was significantly revised to reflect newly added definitions of "cost" and "spending." See the "The Key Questions" subsection of the Introduction for the definitions, which state: "By spending we refer to the amount paid to providers in exchange for health care services, i.e., payments to providers. By costs we refer to the value of resources used to provide health care services by providers, e.g., hospitals." We revised the report throughout to conform to these definitions.
Peer Reviewer #7	Results	Page 38, lines 4-9. Were the differences between groups significant?	Paddock et al. did not report p-values or test statistics for comparisons across groups. We added the following sentence: "While individual results were significant with $p<0.01$ except for the joint replacement spending result, the study does not report whether the comparisons across groups are statistically significant."
Peer Reviewer #7	Results	Page 41, lines 3-6. ³Hip fracture patients² is used twice. Is one supposed to be the elective surgery group?	We corrected a typo. The reviewer's assumption was correct: The other group is the elective joint replacement surgery group.
Peer Reviewer #7	Results	Page 41, lines 27-30. The effect described sounds like regression to the mean rather than an intervention effect.	No change made in response to this comment. We agree this is a possible interpretation.
Peer Reviewer #7	Results	Page 56, line 12. Inpatient mortality decreased but so did length of stay. Did the study control for LOS in the mortality estimates at all? If not, inpatient mortality is somewhat meaningless and its limitations should be pointed out.	We deleted the mention of the result. This result was unadjusted, appeared only in a table and was not addressed in the narrative in the original source. We agree with the reviewer that this particular result is not meaningful, and we do report mortality one-year post-fracture, which is more meaningful.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #7	Results	Page 63, line 5. What does ³nd² mean?	Added the full definition to the acronym table and text. Nd:YAG refers to “neodymium-doped yttrium aluminum garnet.”
Peer Reviewer #7	Results	Page 63, line 44. Implementation is misspelled.	This has been corrected.
Peer Reviewer #8	Results	The summary of results is well done. With respect to the tables summarizing the results for each study, I would have preferred that spending and costs were not combined in one heading or, if they must be combined to save space, that the tables have a footnote that defines and contrasts the two concepts.	We have revised the results section to more clearly delineated findings on spending and costs and added definitions of these concepts.
Peer Reviewer #10	Results	High level of detail but some implications were not considered. See a and b above as well as the following: 2. Administrative costs: Administration of bundled payments can be costly. The report in fact gives examples of potential for increased administrative costs among providers. The LTACH project noted an increase of one FTE per 1000 inpatient days, possibly due to ³the need for additional administrative staffing² (page 43 line 13). Conversely, payer administration of episodes of care may require significant investments in infrastructure. Both these factors further erode potential cost savings from bundled payments.	We have added a paragraph discussing evidence related to administrative costs in the summary and discussion section.
Peer Reviewer #10	Results	Unintended consequences: Any discussion of unintended consequences (e.g. p. 11 lines 33 and following) should acknowledge that the fee for service system has the same problems, for example the potential for performing unnecessary procedures (instead of bundles) and the risk of cherry picking lower risk patients. Conversely, bundles have a new risk of ³unbundling² and the report mentoins the example of possible delay of YAG capsulotomies to get added reimbursement on expiration of cataract episodes in that pilot (top of page 63). Incidentally, delay of capsulotomy would be a quality defect, in my opinion.	Additional text has been added in several places. E.g., “ an incentive to increase service volume also exists under fee-for-service payment, but not other alternatives such as capitation” (added to introduction).
Peer Reviewer #11	Results	The amount of detail on the individual studies was sufficient to understand what the authors did and found. This sounds self-serving but I think I have a study (my dissertation actuallyŠ) that should be in the review as I understand the criteria. I believe it didn¹t turn up because it uses the words ³case rate² and ³risk sharing² instead of any of the search terms. There are two papers that overlap somewhat but together give the full results.Rosenthal MB. Risk sharing in managed behavioral health care. Health Affairs. 1999 Sep-Oct;18(5):204-13.Rosenthal MB. Risk sharing in mental health care. Journal of Health Economics. 2000 Nov;19(6):1047-65.	We included the two papers referenced in the review.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #11	Results	Page 27, first full para you describe a bunch of findings about Medicare spending and then conclude that "Medicare policy may have driven spending patterns throughout the U.S. healthcare system" – I believe that is true but it doesn't follow from the evidence you presented.	We revised this sentence to follow from the evidence presented earlier in the paragraph: "These findings suggest that the Medicare IPPS resulted in care shifting from inpatient to outpatient settings, and had a modest impact on overall Medicare spending."
Peer Reviewer #11	Results	Page 27, last para and sentence. You say "In particular, other policies" – I think you mean "factors" because the example you give is really a technological rather than a policy change.	The reviewer is correct. We have replaced "policies" with "factors" as suggested.
Peer Reviewer #11	Results	Page 29, first para on SNF PPS implementation. You give the predicted 5 year savings and the first year decrease in spending. These are in line no? You might say that since the apples to oranges comparison makes it non-obvious on a quick read.	The reviewer is correct to point this out. We now note this point: "After implementation, aggregate Medicare SNF spending declined 15 percent from $11.3 billion in 1998 to $9.5 billion in 1999, roughly in line with CBO projections."
Peer Reviewer #11	Results	Page 57, in overview of design you first refer to 2 retrospective obs. Studies and one observational study. Later in the para you call the latter "the descriptive study" – if that is what it is, perhaps you should just call it that from the start?	We have corrected "one observational study" to read "one descriptive study."
Peer Reviewer #12	Results	This is really well written I think. One thing I thought was missing, though it is mentioned at the end, is a discussion about how current discussions about bundled payments, shared savings, ACOs, etc. are focused a bit differently than historical payment model changes such as the 1983 change to PPS for inpatient Medicare patients. For example, the PPS payments excluded phy payments while current discussions focus on inclusion of physician payments as well as care provided across the continuum. My point is that the past literature does not focus on the current policy scenarios exactly. The authors do make this point at the end, but this may be something to address up front in the motivation; specifically the point that past efforts are different from current policy discussions.	We have addressed this comment in several ways. In the introduction, we introduce the distinction between newer and older programs. In the results, we describe several of the newer programs that have not yet been evaluated in order to provide contrast with the older, reviewed programs. In the summary and discussion, we have highlighted the limitations that the review's findings are based mainly on published studies of older programs that differ from newer programs.
Peer Reviewer #13	Results	Selection of studies and level of detail is appropriate.	No change in response to this comment.
Public Reviewer #1	Results	FYI--results on Medicare home health prospective payment (pp. 38-40 in the Draft Research Review, as well as Appendix A):In the late 1990's and early 2000's Mathematica Policy Research conducted a large cluster randomized trial (home health care agencies were randomized) of a prospective payment demonstration that studied effects of a prospective payment system on Medicare utilization and costs, quality of care, and Medicaid and informal care utilization. The final summary report is here--http://www.mathematica-mpr.com/publications/PDFs/hhp-final.pdf	This study has been added to the review.
Public Reviewer:	Results	The American Academy of Ophthalmology (the Academy) appreciates	In the results chapter, section on the Medicare

Commentator & Affiliation	Section	Comment	Response
American Academy of Opthamology		the opportunity to comment on the draft evidence report entitled Closing the Quality Gap Series: The Effects of Bundled Payment Strategies on Health Care Spending and Quality of Care. The Academy is the world's largest association of eye physicians and surgeons—Eye M.D.s—with 19,000 members in the U.S. Comments on the review of the Medicare Cataract Surgery Alternative Payment Demonstration: The Academy has reviewed the draft report and would like to correct inaccurate information cited by AHRQ regarding activities of the Academy opposing the Medicare Cataract Surgery Alternative Payment demonstration implemented from April, 1993 through April, 1996. In the report, AHRQ states that a lawsuit initiated by the American Academy of Ophthalmology helped lead to a low participation rate in the demonstration. While the Academy did seek judicial relief from this poorly designed demonstration, we disagree that the lawsuit was the cause of the poor participation in this project. Instead, the blame should be placed squarely on the design and implementation of the project. Many of our members located in areas considered for these sites expressed significant concerns to our organization when details of the demonstration became known. A primary concern was that the Health Care Financing Administration (HCFA), predecessor to the Center for Medicare and Medicaid Services (CMS), was in fact creating inducements for fraud and anti-kickback violations. The demonstration's design appeared to put ophthalmologists at risk for criminal prosecution and sanctions regarding Medicare participation. An overview document prepared by HCFA regarding the demonstration indicated that participants would be able to provide incentives to Medicare beneficiaries and were permitted to waive deductibles and co-insurance. Further, incentives were outlined for providing referrals, one of the most basic violations of the anti-kickback provisions. Yet there were no exemptions or waivers outlined from the anti-kickback or fraud and abuse requirements of the time. The Academy and its members were concerned that this project was setting out on a course of creating so-called "cataract mills" and incentivizing short cuts to providing quality care. Only months prior to the demonstration, Congress had held a hearing that investigated and raised concerns about such high volume facilities and in response the American Society of Cataract and Refractive Surgeons had written guidelines regarding ethical considerations for cataract surgery. Additionally, the Academy had just undertaken an extensive study partnering with Johns Hopkins University that showed there was no difference in infection rates or other quality measurements in facilities that performed fewer numbers of cataract surgery than those that performed a high volume of procedures. These results were counter to HCFA's purported rationale in the project for funneling surgeries to a	Cataract Surgery Alternative Payment Demonstration, we have added a statement that demonstration participation was low, but removed the statement that this was a result of the lawsuit. We also added a sentence summarizing the grounds on which the Academy criticized the demonstration design: "The Academy criticized the demonstration design on the grounds that it lacked patient protections, violated federal and state anti-trust laws and several state medical practice laws, and did not significantly increase the amount of bundling relative to status quo Medicare payments for cataract surgery."

Commentator & Affiliation	Section	Comment	Response
		few facilities, in which they hypothesized more procedures performed meant higher quality. Unlike some of the other proposed bundling demonstrations, there were few quality requirements outlined in the cataract demonstration. In a side by side comparison of the cataract and heart-bypass demonstration, the emphases on quality were quite different. In the objectives for the heart bypass demo it was clearly stated that the first objective was to increase quality. There was no mention of patient protections or increasing quality in the cataract project objectives. In fact, the only objective discussed was assessing the fiscal benefits of bundled payments. There were also no volume limits for the cataract demonstration which we believed could lead to inappropriate referrals and patient selection. The Academy soundly criticized the initial design because there was nothing to bundle with cataract surgery. The intraocular implants and facility fee were already bundled. Further, performance of the procedure was already mandated to be performed in the hospital outpatient or ASC so there was little variation in facility charges, and unlike the other demos, the unit costs for cataract were a fraction of those for CABG. Ophthalmologists didn't sign up because there was no additional savings that could be made without compromising patient care.The demonstration, in short, failed because it was economically unsound, lacked patient protections, violated federal and state anti-trust laws, and would have violated several state medical practice laws. Several members of Congress also opposed the demonstration.	

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #1	Summary and Discussion	This was the most valuable piece because the authors' expertise and interpretations begins to show through. Before commenting in the stylistic issues in the next section, I offer a couple of other factors that should be included. I think the response of providers to the payment changes needs to be placed in context. To the extent that the shift to bundled payment affects a small fraction of a provider's business, or is expected to be temporary, little change in behavior should be expected, especially in the short run and if there are internal incentives or workflows that need to be changed. Most, but not all, the payment changes involved government programs, and hence could be expected to be reasonably permanent, but they varied in the proportion of patients subject to the new payment scheme. A few studies, were of demonstration projects in which long-term payment changes were far from guaranteed, and generally focused on volunteer sites. Some, e.g., Geisinger, were quite interesting, but involved probably quite unusual sites.	We added the following paragraph to the "Applicability" subsection in the "Summary and Discussion" section: "Some reviewed bundled payment systems, including Geisinger ProvenCare (SM) and several international systems, involved settings where care for most or all patients was reimbursed by a single payer. Many other reviewed payment systems were implemented by a single payer (e.g., the US Medicare program) but applied to providers with diverse patient populations and multiple payers. The reported effects of bundled payment may be blunted if interventions affect only a small portion of providers' overall business. Similarly, providers may not significantly alter behavior if bundled payment interventions are perceived as temporary." We address the issue of self-selected participants in the "Key Question 2" subsection in the "Summary and Discussion" section. We added the following sentence at the end of the paragraph describing the Medicare Heart Bypass and Cataracts Demonstrations: "The effect on spending would also likely be lower if hospitals participating in the demonstration were not selected in part on their ability to negotiate low payment rates."
Peer Reviewer #1	Summary and Discussion	From a policy perspective, the newer bundled payment models have explicit quality measures included, and this differentiates them from most of the older models, which even if the evaluations assessed processes, the direct rewards were not based on quality measures. It is speculative, of course, whether these new models will work better than the old, but the fact that the old were essentially quality neutral is reassuring. With respect to quality, it is worth noting that the power to detect cost differences is probably far greater than it is to detect quality differences. Furthermore, I would expect most of the quality impacts (positive and negative) to play out over longer periods of time than the evaluation periods. The future research section hints at this, and could be strengthened a bit.	In the summary of findings for key question 2, we discuss that newer bundled payment models may explicitly include quality measures, whereas most reviewed programs did not. We have added a sentence in the limitations section noting that small sample size and limited followup time are likely particularly important for detecting quality effects. We also strengthened the discussion of this point in the future research section.
Peer Reviewer #1	Summary and Discussion	In the discussion section(s) it would be helpful to set apart the summary of various findings, the interpretations of those findings, and the speculations on applicability. All three are valuable, but should be separated by headings or a similar convention.	We have included headings including "summary of findings across payment systems," "strength of evidence," and "applicability."

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #3		p. 74, lines 10-18: Some bundled payment models have P4P component. E.g. proven care.	In the summary of findings for key question 2, we discuss that Geisinger ProvenCare(SM) and newer bundled payment models may explicitly include quality measures, whereas most reviewed programs did not.
Peer Reviewer #3		p. 75 lines 36-43: Diff 8 off studies should be better. RCT's will be difficult if effects extend to all patients served by a provider. Would need to randomize providers.	We added new introductory text to the Recommendations for Further Research subsection in both the Executive Summary and full report body. The Recommendations for Further Research subsection (in both locations) was edited to highlight the tension between rigorous and timely evaluations and to highlight practical recommendations under each heading.
Peer Reviewer #3		p. 75 lines 58-60: Who captures savings?	We added definitions of the terms "spending" and "costs" in the "The Key Questions" subsection of the Introduction. These definitions state: "By spending we refer to the amount paid to providers in exchange for health care services, i.e., payments to providers. By costs we refer to the value of resources used to provide health care services by providers, e.g., hospitals." We revised the report throughout to conform to these definitions. In the same subsection we also note that: "The difference between payments (spending) and costs is the provider's margin. The hope is that bundled payment will decrease spending by payers and costs to providers relative to usual, typically fee-for-service, reimbursement. Under this scenario, profits may either increase or decrease depending on the relative magnitude of changes in spending and costs. We also note the distinction between aggregate spending and costs and per-episode spending and costs. If the introduction of bundled payment increases the number of episodes provided, aggregate spending may increase even if per-episode spending decreases.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #3		p. 75 lines 29-31: Why?	We edited the two sentences in the "Key Question 2" subsection to elaborate on when per diem bundled payment might be appropriate: "The Belgian non-medical inpatient PPS and the Medicare SNF PPS used a per diem unit of payment which may be appropriate when length of stay is highly variable. While this bundle definition does not constrain utilization in terms of length of stay, it does affect incentives to provide services in a given day, and studies on both systems reported either declining or steady length of stay post-PPS."
Peer Reviewer #3		p. 78 lines 1-3: Update determines program spending growth	We added the following text to the Introduction section to emphasize the importance of updates (emphasis added): "The design of bundled payment interventions may differ in other key ways including the type of conditions or procedures used as the basis for the bundle, risk adjustment, *methods used to establish and update payment rates, etc.*"
Peer Reviewer #3	Summary and Discussion	This section is very useful and, as I mention above, should be essentially the body of the report. The discussion of Recommendations for Further Research is particularly important--but although the authors do acknowledge the potential for conflict between considerations of design and implementation, discussion of this point should be much more explicit and substantially expanded. Also, although all their recommendations are worthy of consideration from a research perspective, some are more important than others and some (not the same ones, alas) are more feasible to address than others. The discussion of these points should include more recognition of this fact.	We added new introductory text to the Recommendations for Further Research subsection in both the Executive Summary and full report body. The Recommendations for Further Research subsection (in both locations) was edited to highlight the tension between rigorous and timely evaluations and to highlight practical recommendations under each heading.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #4	Summary and Discussion	The report does a poor job of distinguishing the different kinds of bundled payments and the relevance of the evidence to each of them. The vast majority of the research reports summarized are based on Medicare, Medicaid, and international facility-based prospective payment systems that are already comprehensively in use in the U.S. and that likely have little direct relevance to the types of bundling approaches being considered by either public or private payers for future implementation. The technical advisory panel recommended inclusion of those reports because of the paucity of studies on more relevant bundling projects, but the researchers provide far more detail than is necessary on the results of the facility-based prospective payment system changes without devoting enough attention to whether those results are applicable to other forms of bundling. For example, nowhere do the researchers emphasize the fact that the PPS systems were a change from cost-based reimbursement, not from fee-for-service payment, nor do they discuss whether the results of such changes would be applicable to changes from fee-for-service payment systems.	We have made changes throughout the report to address this comment. In the introduction, we introduce the distinction between newer programs and the older programs that were reviewed. In the results section, we describe several newer programs that have not yet been evaluated in order to differentiate them from older, reviewed programs. In multiple places in the summary and discussion including the summary of findings by key question, the applicability section, the limitations section, and the conclusion we highlight this limitation. In the introduction, we added a description of cost-based and fee-for-service payment methods and continued this distinction throughout the report.
Peer Reviewer #5	Summary and Discussion	Unfortunately, the available evidence about the impact on cost simply confirms our expectations and the available evidence about the impact on quality is quite weak. In addition, the published literature is quite flawed, from a methodological standpoint. That said, the authors recognize these weaknesses and mostly make them clear to the reader (for a few additional points, see the next section of my review).	We have strengthened our discussion of these limitations throughout the report as described in responses to related comments.
Peer Reviewer #5	Summary and Discussion	However, there is one important addition I might suggest: Consider listing the types of quality measures that should be included in bundled payment studies (the list of ³undesired effects² in the Introduction does not cover this fully or explicitly). For instance, while bundled payment is likely to lead to improvement in measures of inappropriate utilization (hence, these are less interesting, unless one is studying the mechanisms by which cost savings are achieved), there is more risk that errors of omission will arise (such as incomplete diagnostic testing or missing key components of chronic disease management, such as measuring blood sugar among diabetics). In addition, the impact on long-term outcomes is unclear. By providing a list of topics to be included, the report may improve the quality of future research.	In the future research section, we have added several categories of potentially important quality measures, and go on to note that evaluators' collaboratives could identify priority measurement areas as well as work towards standardization of measurement approaches.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #7	Summary and Discussion	I don¹t completely agree with the recommendation to use standardized measures of impact on costs and quality, especially as it relates to measures of quality. The authors themselves state that evaluations frequently used quality measures that were available, even if they were not necessarily most relevant for the setting/population. The most appropriate quality metrics are going to be setting and population dependent. To the extent that a bundled payment applies to a narrow set of services or procedures, this needs to be taken into consideration as well. The quality metrics one would want to track for cataract surgery bundled payments are quite different than what one would want to track for nursing home PPS.	We added and modified text under the "Use standardized measures of impact on spending and quality" heading in the "Recommendations for Further Research" subsection to address this concern.
Peer Reviewer #7	Summary and Discussion	Page 65, lines 52-54. There seems to be a word missing somewhere.	Added the missing word ("as").
Peer Reviewer #7	Summary and Discussion	Pages 65-66. All articles are treated equally in the summary of findings regardless of their quality. Please include a statement about whether anything changes substantially if ³poor² studies are excluded. Also, the authors should comment on the appropriateness of the measures used, particularly the quality measures. How likely is it that no quality effect was observed because the wrong measures were examined?	We added a statement that no studies were excluded on the basis of their quality rating but that our conclusions are not sensitive to the inclusion of studies rated "poor" in the strength of evidence section under the summary for key question 1. We added a statement to the summary of findings for key question 1 that although the availability of quality measures has increased over time, we don't believe that the conclusions about quality effects are a function of use of the wrong measures.
Peer Reviewer #7	Summary and Discussion	Page 68, lines 11-13. Authors mention ACA includes quality incentive for IPPS. Should also mention an incentive currently exists for reporting quality information.	Made this addition.
Peer Reviewer #7	Summary and Discussion	Page 74, 44-50. The inability to address question 2 is more related to the intervention designs than any problems with the studies.	We added a paragraph in the "Limitations" subsection in the "Summary and Discussion" section, and additional discussion in the "Assessment of Methodological Quality of Individual Studies" subsection in the "Methods" section to stress that intervention design rather than study quality lead to several of the limitations we cite.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #8	Summary and Discussion	I liked the conclusion as stated in the Conclusions section of the structured abstract--"There is weak but consistent evidence..." Given the limitations in the study designs as noted by the authors in their evaluation, the term "weak" should be part of the evidence summary. Unfortunately, in the other places in the review where the authors state their conclusions it is omitted. For instance in the executive summary they say that "the evidence suggests that" bundled payment was associated with a decrease in utilization and costs of services and a decline in spending. (34-36). In the Conclusion of the executive summary (ES7) the authors say that the evidence provides support that the programs are likely to be an effective strategy. This seems too strong given their conclusion that the study designs were weak and effects small. At the end of the review (p. 74) the authors say that "..the review found evidence that bundled payment was associated with reductions in health care spending and utilization. Again, I would think it is more appropriate to emphasize here that there is "weak evidence" given the shortcomings of the study designs, etc.	We have revised the summary of the findings for consistency throughout the report.
Public Reviewer: Ann-Marie Lynch, Advanced Medical Technology Association	Executive Summary and Structured Abstract	Executive Summary: The Advanced Medical Technology Association is pleased to submit comments on The Effects of Bundled Payment Systems on Health Care Spending and Quality of Care. AdvaMed member companies produce the medical devices, diagnostic products, and health information systems that are transforming health care through earlier disease detection, less invasive procedures, and more effective treatments. AdvaMed members range from the largest to the smallest medical technology innovators and companies. AdvaMed concurs with the general findings of the study's review of studies examining the effects of bundled payment on health care costs and quality. In terms of additional specific recommendations for further research, AdvaMed strongly urges that study designs and evaluations incorporate in-depth medical reviews/clinical audits of persons in bundled payment arrangements, comparing their care and health outcomes to professionally recognized standards of care. Evaluation should include analysis of patient medical records, not simply claims data, and a comparison of persons in bundled payment arrangements with persons outside such arrangements, looking at their utilization of specific services, including a review of referrals to medical specialists and their access to medical treatments and technologies. Study designs and evaluations should also include surveys of participating patients and their providers. Provider surveys should include their assessment of the availability of products and services and changes in practice that have been implemented under the bundled payment arrangement. Similarly patients should be independently surveyed to determine what they understand	We added the following text in the Recommendations for Further Research subsection which addresses the reviewer's point regarding the impact of bundled payment on appropriate treatment: "In particular, the use of quality measures was relatively rare and the measures used were inconsistent across studies. Important potential quality effects were often unmeasured, including measures of underuse of appropriate services within bundles, indications of the appropriateness of bundles, measures of the patient experience of care, measures of coordination of care within and across bundles, and health outcomes of bundles of care." We also recommend comparisons to control groups outside bundled payment systems. We agree with the reviewer that five-year time horizons will provide information on many currently unobserved outcomes. This point is also made in the Recommendations for Further Research subsection. Finally, we added the following sentence to address the importance of incorporating changes related to the introduction of new technologies into payment rates: "Important design features to be

Commentator & Affiliation	Section	Comment	Response
		about the potential impact incentives under bundled payment arrangements might have on their care and access to innovative treatments and technologies, as well as their assessment of the care they received. These results should be compared to survey results from persons not in bundled payment arrangements. AdvaMed strongly agrees with the study's recommendation that at least some of the evaluations of bundled payment arrangements incorporate a longer term time horizon of at least 5 years post-implementation in order to capture changes in evolving practice patterns, supplyside market conditions, or clinical outcomes that cannot be assessed in such limited time frames. A critical dimension to include in longer term evaluations of bundled payment impact is patient access to new treatments and medical technologies. As part of this particular evaluation, studies should assess the impact of bundled payments on overall medical progress. Such studies could examine changes in hospital new technology take-up rates, changes in research and development investment, and changes in venture capital funding for new treatments and technologies. Payment systems have to reward the benefits of innovation if a high rate of innovation and corresponding social welfare benefits are to continue to be an important attribute of the U.S. health care system. It is critical that payment and delivery reforms include appropriate metrics for measuring and valuing the benefits of new technology and incorporating those metrics in payment systems. Theses metrics should be included in health care delivery reform efforts. Moreover, study designs should include evaluations of the impact of bundling on access to innovative technologies and on continued medical progress.	addressed include the definition of the bundle (how many providers are included, what length of time, which services are included and excluded from the bundle); … *methods to update payment rates to reflect new technologies*; and methods for distributing payment among participating providers."
Peer Reviewer #10	Summary and Discussion	Further research needed, as described, was clearly stated and could be translated into new research. There are further limitations (see a, b, and d above) There are important further areas for research:a. Items described above: Items 1 through 3 above suggest obvious additional areas of research.	No change made in response to this comment.
Peer Reviewer #10	Summary and Discussion	b. Appropriateness indicators: The health care system urgently need indications for appropriateness of bundles, both to help address the unintended consequence above and also as a measure of quality of care.	We have added indications for appropriateness of bundles in a new description of priority quality measurement areas in the future research section.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #10	Summary and Discussion	c. Determining the best practice contents of (services within) anepisode: Unless the services covered by a bundle are spelled out, purchasers do not know what they are getting. This is especially important in regards to episodes of care. It is concerning that the new CMS pilot intends to allow programs to submit bids on episodes without CMS[1]s specifying the services to be provided. This could result in a race to the bottom. A proliferation of differently constructed episodes at different prices will hinder research into costs and results, let alone predictably cause confusion among patients.	We discuss this important point under the "Incorporate quantitative and qualitative measures of program design and contextual factors" heading in the Recommendations for Further Research subsection of the original draft.
Peer Reviewer #10	Summary and Discussion	d. Quality measures for outcomes, including functional outcomes: Similarly, there is an urgent need for outcome measures, especially those that are meaningful to patients. Otherwise we do not know if the bundle has accomplished the goal of better care and better quality, as well as lower costs, which together make up the Triple Aim.	We have added health outcomes in a new description of priority quality measurement areas in the future research section.
Peer Reviewer #11	Summary and Discussion	The results synthesis is very balanced and does an excellent job of weighing the quality of the evidence and drawing out defensible conclusions about what we know and what researchers need to do to address what we don't know.	No change made in response to this comment.
Peer Reviewer #12	Summary and Discussion	No additional comments other than what I just described above. The report is well written.	No change made in response to this comment.
Peer Reviewer #13	Summary and Discussion	The future challenge of bundled payments is around coordination of clinical care and finances across multiple settings requiring greater provider collaboration. Bundled payments applied to a single care setting are fundamentally different than bundled payments across care settings. Rather than concluding that the evidence supports that bundled payments save money I suggests the authors be more explicit about the fundamental distinction -- and that the conclusion about cost savings is based on single settings but that the evidence on bundling across multiple settings is extremely scares.	We have highlighted the distinction between single-setting programs and multiple-setting programs in the applicability, limitations, and conclusions sections. We also made relevant changes to the introduction and results sections in response to other comments above.
Public Reviewer: Vinita Ollapally, American College of Surgeons	Summary and Discussion	We discuss below areas for future research related to bundled payment namely, inclusion of quality measures, study of bundle design, and study of bundles that incorporate more than one care setting. The Review analyzes studies that rarely integrated quality measures into their bundled payment systems because many of these studies were implemented prior to the recent proliferation of pay-for-performance programs. If quality measures were used, many of the measures were process or intermediate health outcomes measures, or measures that were selected based on availability rather than because they represented the aspects of quality most likely to be impacted by the bundle. In these instances, the effect of the measures is only indirectly related to health outcomes. Today, quality is very much an integral part of any bundled payment	We have made several changes in response to this comment. We have added a statement about the priority of health outcomes measures to the future research section of the summary and discussion chapter. We discuss the importance of studying the differential effects by design features in the future research section. We have highlighted the importance of bundled payment programs including multiple settings and the distinction between these programs and single-setting programs throughout the report as detailed in responses to other comments above.

Commentator & Affiliation	Section	Comment	Response
		model, so we recommend further study of appropriate quality measurement and reporting metrics for bundled payment systems. The ACS also urges AHRQ to utilize outcomes rather than process measures. For example, there is a poor correlation between process-related performance measures, such as the Surgical Care Improvement Project measures, when compared to ACS National Surgical Quality Improvement Program (NSQIP) audited and reliably validated, risk-adjusted outcomes. As such, we support the inclusion of risk-adjusted outcomes measures using clinical, audited data, and we believe that ACS NSQIPderived outcomes-based measures are an ideal example of such measures. Due to lack of evidence, the Review was not able to answer Key Question 2, which relates to the effects on bundled payment systems based on differences in bundled payment design features. We urge AHRQ to continue to study the differential effects by design features. Examples of design features include: length of time of bundle, services included and excluded, methods for limiting financial risk, use of quality measures, methods for distributing payment among providers, whether bundled payment is more effective in highly integrated settings, the role of the general financial environment, and the differential effect between subgroups of patients. It would also be useful to compare the effects of bundles implemented in very progressive institutes and those that are resisting changes to the current way that providers are reimbursed. As part of this comparative work, AHRQ should assess the barriers and the drivers of successful bundled payment systems. A third area that we suggest for further research is the study of bundles that include more than a single site of care. All but three of the 18 bundled payment systems studied in the Review involved a single provider such as a hospital, skilled nursing facility or home health agency. Going forward, we believe it will be important to study the effects of bundled payment involving more than one care setting. Future bundled payment studies should also examine arrangements where physicians and hospitals are integrated. Many private and public initiatives that are currently being developed and implemented (for example the Center for Medicare and Medicaid Innovation Bundled Payment for Care Improvement Initiative) are focused on bundling services provided by different providers over the course of an episode defined by a condition, diagnosis, or procedure. We offer these as a few suggestions of ways to enhance further study of the effects of bundled payment. The concept of bundled payment has received increased attention from policy makers and both public and private payers. To that end, we encourage AHRQ to continue to study these and other important concepts that can improve health care quality and can help physicians better care for their patients.	

Commentator & Affiliation	Section	Comment	Response
Public Reviewer: American Academy of Opthamology	Summary and Discussion	Comments on the Effects of Bundled Payment Systems on Health CareSince we have seen few if any details outlining how the current Center for Medicare and Medicaid Innovations (CMMI) plans on implementing bundled payments, the Academy has not taken an official position on this newest effort at bundling payments. Looking at innovative ways to provide and pay for appropriate episodes of care by seeking input and ideas from stakeholders is certainly an improvement from the previous, forced effort of artificially creating so called "Centers of Excellence" and the cataract bundling demonstration. Because of the outreach by CMMI, we are aware of several members examining ways in which some chronic eye conditions could be appropriately bundled and have had exploratory conversations with staff at CMMI about such possibilities.Given the numerous negatives and weaknesses pointed out by AHRQ in this report and the paucity and poor quality and design of projects from within the U.S., we have difficulty seeing how the conclusions AHRQ researches came to could be justified. The conclusion states: *There is weak but consistent evidence that new bundled payment programs have been effective in cost containment without major effects on quality. Bundled payment is a promising strategy for reducing health spending. However, future programs may differ from those included in this review.* As pointed out, at least one of the three studies, the cataract demonstration had little if any focus on quality and the statement about having no effect on quality was not backed up by the data.As far as the key questions examined, the reviewers clearly indicate that they were not applied consistently throughout the report. In particular, Key Question 2: Differential effects by key design features was not discussed when reviewing the three U.S. demonstration projects reviewed and the reviewers clearly state that this question was not considered within the U.S. studies and therefore could not be included in this review. Reviewers graded the evidence for this Key Question as insufficient to permit an estimation of effects due to the lack of evidence. Yet, the design of any demonstration is key to its success or even more determinate of its failure. And AHRQ verified the Academy's concerns about the cataract study when they found that quality metrics or incentives were rarely integrated into bundled payment systems.The Academy hopes that the current effort to foster local projects that are designed and implemented by the individual facilities and providers will provide a reasoned look at bundled payments that are tailored to the localities and the populations served. These projects could better inform future adoption of these payment mechanisms.	Based on the concerns about the quality and design of reviewed studies and the lack of studies reporting results relevant to key questions 2 and 3, we have rated the strength of evidence as weak for key question 1 and insufficient for key questions 2 and 3. The justification for these ratings and conclusions about effects is presented in the sections on strength of evidence in the summary and discussion chapter and in Appendix D.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #7	Appendices	Appendix A: Additional information would be useful in the evidence table, including setting, clinical area of focus (if relevant), and method used to set bundled payment.	The current draft of this report includes revised appendices with additional information abstracted from studies.
Peer Reviewer #7	Appendices	Appendix B: There are many inconsistencies in the formatting of the references, including handling of page numbers, placement of publication year, and use of journal abbreviations. Also, reference 21 is missing page numbers.	We have addressed formatting issues.
Peer Reviewer #7	Appendices	Appendix C: This appendix also has numerous inconsistencies in the formatting of the references, including handling of page numbers, placement of publication year, and use italics.	We have addressed formatting issues.
Peer Reviewer #7	Appendices	Appendix D is missing from the document.	Appendix D is included in the revised report.
Peer Reviewer #1	General	Overall, this is a well-written and well-balanced review of the evidence. My concerns are minor and focus mostly on wanting to encourage the authors to say a bit more up front about why the review is limited and what they think can be gleaned from it for the purposes of policy makers.	We have added emphasis on the report's limitations to the abstract, executive summary, and introduction, as well as throughout the rest of the report.
	General	There are also places where the wording could be clarified. For example, the key summary point, "this review found evidence that bundled payment was associated with reductions in health care spending and utilization with inconsistent and generally small effects on quality measures" is frequently repeated. It takes a bit too much thinking to see the import in this. On page 83 the sentence is followed by: "These findings were consistent across different bundled payment programs across settings. For policymakers considering implementation of bundled payment programs, this evidence provides support that the programs are likely to be an effective strategy for reducing health care spending. While the effects on health care quality are less certain, the available evidence doesn't support the worst concerns about potential adverse effects of bundled payment." This addendum makes the findings more clear. For the single sentence version, I would suggest: This review found evidence that bundled payment was associated (a) with reductions in health care spending and utilization and (b) with inconsistent and generally small effects on quality measures.(See also comments re Intro and Discussion/Conclusion)	We adapted the reviewer's recommended formulation which we agree is clear and concise.
Peer Reviewer #2	General	The report is meaningful. The key questions are appropriate. The restriction to episode based payment (as opposed to fully bundled payment) is a limitation, but I think that was the charge.	No change made in response to this comment.
Peer Reviewer #2	General	The literature itself is pretty limited so a lot of space is devoted to the experiences from Medicare bundling. Some of that is dated and it avoids the key question of bundling across providers. This last topic is addressed but due the more limited literature, it get less space.	This limitation has been emphasized in the revised report.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #2	General	The authors claim no published results from Prometheus, but there is a related paper about difficulty of implementing it in health affairs. That seems relevant.	We added a description of PROMETHEUS and several other current programs in the results section. However, the referenced paper did not report on spending or quality study outcomes and therefore did not meet inclusion criteria for the evidence review.
Peer Reviewer #3	General	This is an apparently comprehensive, systematic, and encyclopedic review of the findings on the impacts of various bundled payment approaches. The objectives of the report are appropriate and the methodology used to identify candidate studies and select those that would be reviewed are clearly described.	No change based on this comment.
Peer Reviewer #3	General	The topic-by-topic description of study findings, however, is extremely tedious to read through. I have two major suggestions for improving the report's usefulness to potential readers: in the Executive Summary, I would shorten the sections on objectives, conceptual framework, and methods and get to the results more quickly; in the report, I would move the results section into an appendix, so the reader could get to the summary and discussion more quickly and refer to the detail on each system and each topic as desired. I have more specific comments below. *(KB note: specific comments are included in the Executive Summary and Results tabs.)*	We did not make this change in following AHRQ guidance on the structure of EPC reports.
Peer Reviewer #3	General	As I mention above, if I had my druthers I would reorganize this report to focus much more on the discussion and recommendations, while preserving the detailed results in an appendix for useful reference. I would also recommend clearer distinctions among the many issues raised here, so the reader's attention can be called to the most important points that are raised.	We did not make this change in following AHRQ guidance on the structure of EPC reports.
Peer Reviewer #4	General	The evaluation in the report is based on an overly narrow and misleading statement of the goal of bundled payment systems. The report states the goal as "decreasing health spending while improving or maintaining the quality of care." While decreasing spending is certainly desirable and a goal for payers such as Medicare that have imposed such systems, providers have often supported them because of the greater flexibility they give the provider (relative to fee-for-service payment) to determine the most appropriate combination of services for the patient. This is a critical distinction, because the cost-based reimbursement systems of the past typically did not limit the types of services for which a provider could be reimbursed, but fee-based payment systems often do.	We have added the goal of increasing providers' flexibility to determine the most appropriate combination of services for the patient in the executive summary and conclusion. The newly added paragraph in the introduction providing motivation for the report includes the statement "The hope is that bundled payment would give providers incentives and flexibility to choose those inputs that can most efficiently achieve good health care outcomes." Later in the introduction, we state: "Providers are typically given discretion over the allocation of the services used to treat the patient's episode most effectively." This sentence was also added in the introduction section of the executive summary.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #4	General	Providers also support bundles involving multiple providers because of the opportunity they provide to achieve coordination; although this is mentioned in the report, it is described as a secondary effect of the providers[1] need to manage the payment, rather than a primary goal that may attract the interest of the providers.	We have revised the introduction and executive summary and introduction to address this comment as follows: "Providers are typically given discretion over the allocation of the services used to treat the patient's episode most effectively. This flexibility may encourage providers to use resources to coordinate care; often, these services are not reimbursed under fee-for-service payment. If the bundle includes services delivered by multiple providers in multiple settings, providers have to create a mechanism for managing the shared payment for a given treatment or condition, which could also foster coordination."
Peer Reviewer #4	General	The authors also do not clearly distinguish between "costs" and "spending." A bundled payment may enable a reduction in a provider[1]s costs without a comparable reduction in the payer's spending, and conversely, the bundled payment could be lower than the combined amount of previously unbundled payments without enabling or resulting in any reduction in costs for the involved providers. A bundled payment could result in lower spending per case or episode, but higher spending overall if the number of episodes increase, or it could result in lower total spending if the number of episodes decrease or if the increase in episodes is due to shifts from other types of more expensive care.	We added definitions of the terms "spending" and "costs" in the "The Key Questions" subsection of the Introduction. These definitions state: "By spending we refer to the amount paid to providers in exchange for health care services, i.e., payments to providers. By costs we refer to the value of resources used to provide health care services by providers, e.g., hospitals." We revised the report throughout to conform to these definitions. We note the distinction between aggregate spending/costs and per-episode spending/costs in the same location.
Peer Reviewer #4	General	And finally, a bundled payment structure could enable a slowing of the increase in spending without necessarily reducing it, and this distinction is not clearly made in the report. These many different potential goals and outcomes are of different relevance in different settings, but the report attempts to draw one inappropriately broad conclusion, i.e., does bundling (of any type) reduce spending (by payers) of all types.	We added the following paragraph to the "Applicability" subsection of the "Summary and Discussion" section to make this point: "All reviewed studies assessed the impact of bundled payment relative to either fee-for-service or cost-based payment. The magnitude and sign of effects relative to fee-for-service or cost-based payment may differ from absolute effects. For example, bundled payment might slow an increase in absolute spending relative to usual payment. Transitions to bundled payment from other payment methods, e.g., salary or capitation, may have other effects."

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #4	General	The report fails to realistically discuss the challenges of conducting "good" research on payment reforms in real-world contexts. While the statements in the report about the limitations of the study findings are correct, the implication is that the research studies which have been done have not been "good" ones, rather than being done as well as was likely possible given that the payment changes were typically made as matters of policy rather than explicitly done as formal demonstrations. The suggestions made about how to improve research (e.g., to "randomly select participating site" in order to create a randomized control trial) are naïve and inappropriate.	We added a paragraph in the "Limitations" subsection in the "Summary and Discussion" section, and additional discussion in the "Assessment of Methodological Quality of Individual Studies" subsection in the "Methods" section to stress that intervention design rather than study quality lead to several of the limitations we cite in the report. On the point related to recommendations, we added new introductory text to the Recommendations for Further Research subsection in both the Executive Summary and full report body. The Recommendations for Further Research subsection (in both locations) was edited to highlight the tension between rigorous and timely evaluations and to highlight practical recommendations under each heading.
Peer Reviewer #5	General	This is a well-written and well done evidence review.	No change made in response to this comment.
Peer Reviewer #5	General	The report will be considered timely by both clinicians and policy-makers. The key questions are clearly stated and meaningful.	No change made in response to this comment.
Peer Reviewer #6	General	In general this is a good report. It is a thorough review of the existing empirical literature on bundled payments.	No change made in response to this comment.
Peer Reviewer #7	General	Overall, I found the report to be well organized, clearly presented, and usable.	No change made in response to this comment.
Peer Reviewer #7	General	I found the report to be generally well written and informative. There are a number of areas that it should be clarified or otherwise strengthened, however. I have 5 main concerns.	Remainder of this comment with concerns is listed in the relevant section of this table.
Peer Reviewer #7	General	The treatment of quality measures was somewhat uneven and difficult to follow throughout the report. For each of the interventions, the following should be addressed: a) was a specific incentive for quality included (if the answer is "no" for all, this needs to be more explicit in the summary and conclusion sections); b) was a certain level of quality expected at a minimum as part of the intervention (this was the case for participation in the Medicare Participating Heart Bypass Demonstration) c) what evaluators found to be the effect of the intervention on quality.	We added a description of the possible uses of quality measurement in bundled payment programs to the introduction. We include a description of information abstracted from reviewed studies on the use of quality measurement and effects of bundled payment on quality in the results section. We summarize the use of quality measures in reviewed programs and discuss the evidence of bundled payment effects of quality in the summary and discussion chapter.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #7	General	I was somewhat confused by authors versus reviewers of articles for the report. Were the authors also the reviewers? At one point, the report indicates there were 3 reviewers, I believe, but stated in the abstract that there were 2 authors, but only Dr. Hussey is listed as an author on the report's cover page.	In the methods section, the initials of reviewers were added. The final version of this report also includes a complete author list and suggested citation, which were omitted from the blinded review copy of the report.
Peer Reviewer #7	General	All abbreviations need to be spelled out, for example, nd.	All abbreviations have been included in a glossary of acronyms, and are spelled out for clarification the first time they are introduced in the document.
Peer Reviewer #8	General	This is an outstanding review of the literature on the impact of bundled payments on a variety of outcomes. The discussion of the limitations of the evidence was on target, as were the suggestions for future research. All in all, an excellent addition to the literature on an increasingly important topic.	No change made in response to this comment.
Peer Reviewer #8	General	The report is well-structured and organized.	No change made in response to this comment.
Public Reviewer: Ann-Marie Lynch, Advanced Medical Technology Association	General	The Advanced Medical Technology Association is pleased to submit comments on The Effects of Bundled Payment Systems on Health Care Spending and Quality of Care. AdvaMed member companies produce the medical devices, diagnostic products, and health information systems that are transforming health care through earlier disease detection, less invasive procedures, and more effective treatments. AdvaMed members range from the largest to the smallest medical technology innovators and companies. AdvaMed concurs with the general findings of the study's review of studies examining the effects of bundled payment on health care costs and quality.	No change made in response to this comment.
Peer Reviewer #10	General	Yes well structured and organized, with clear main reports. Conclusions are somewhat overstated given the large limitations. If limitations and needs for added research were addressed the report's usefulness for informing policy decisions would be increased.	The limitations and needs for added research have been revised, as noted in response to other comments.
Peer Reviewer #10	General	Limitations affect the conclusions on cost savings and applicability. Other areas of improvement in the report include acknowledgement of administrative costs, the discussion of unintended consequences, and additional directions for future research including the need to describe the best practice contents of an episode of care, and the urgent need for appropriateness and outcome quality measures. The report could be improved in a more general way by describing the connection of bundled payments and their goals to the Triple Aim.	We have emphasized the limitations of the report as discussed in responses to other comments above. We have included discussions of administrative burden, unintended consequences, and the suggested directions fo future research. We have added a motivation paragraph to the introduction that discusses the overall goals of bundled payment, although the paragraph does not explicitly reference the Triple Aim.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #11	General	Yes. The report is systematically structured to identify the target audience, the questions and scope of the report and the criteria for review and all conclusions. Overall, the review is comprehensive, precise, and provides clear findings. I found it particularly helpful that the results summaries noted the quality of evidence throughout (as cumbersome as that is, it is also critical.)	No change made in response to this comment.
Peer Reviewer #11	General	As noted early, this report is lucid and well-constructed. Despite its length and technical nature it reads quite well.	No change made in response to this comment.
Peer Reviewer #12	General	The 3 key questions are stated well and answered well. The report is clear and easy to follow. It is very well written and the conclusions seem appropriately conservative and consistent with my sense of the literature, thus providing a lot of face validity for me.	No change made in response to this comment.
Peer Reviewer #12	General	The report is well written and clearly points out that past literature is only suggestive for current policy discussions because of differences in the programs, context, and characteristics of the past payment models. I agree with this conclusion. I also agree with the authors' suggestions for the needs for future research. These may be made more strongly as the need for better studies may be the most important and relevant policy suggestion, which of course requires resources to be provided by the policy community.	We have revised and strengthened the discussion of needs for future research.
Peer Reviewer #13	General	This report is thorough, detailed, and appears to be methodologically sound. The key questions are appropriate and explicit.The report is clinically meaningful - but only to the extent that 1) the underlying bundled payment programs; and 2) the underlying studies of those programs are meaningful.	No change made in response to this comment.
Peer Reviewer #13	General	As the researchers note, most of the current interest in bundled payments is focused on programs that bundle payments for services across multiple providers and settings. Only three such studies were present in the review (Geisinger, Medicare CABG, Medicare cataract) of which two were graded methodologically fair and one poor. The report very aptly demonstrates both the paucity of relevant bundled payment programs and of relevant research.	The differences between the majority of programs reviewed and the programs subject to most of the current interest have been highlighted more strongly in the revised report, as noted in responses to other comments above.
Peer Reviewer #13	General	I agree with the author's recommendations for further research. As you are aware, the upcoming CMS bundled payment for care improvement program will create a significant opportunity to incorporate these recommendations. Though it may not be in scope of the report it may be useful for the authors to share their thoughts in the context of this new opportunity for evalation.	We have revised and strengthened the recommendations for future research. We do not focus on recommendations only for the CMS program because they are relevant to other programs in the private sector and elsewhere, but our recommendations clearly apply to the CMS program.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #13	General	Aside from the above comment and the general concern about lack of programs that bundle across settings I believe the report is well structured and the main points clearly presented. Its use for policy and practice is unfortunately more limited as future implementation of bundled payments will focus on more expansive bundles.	No change made in response to this comment.