Comparative Effectiveness Research Review Disposition of Comments Report

Research Review Title: *Closing the Quality Gap: Revisiting the State of the Science Quality Improvement Interventions To Address Health Disparities*

Draft review available for public comment from November 17, 2011 to December 15, 2011.

Research Review Citation: McPheeters ML, Kripalani S, Peterson, NP, Idowu RT, Jerome RN, Potter SA, Andrews JC. Closing the Quality Gap: Revisiting the State of the Science Quality Improvement Interventions To Address Health Disparities. Evidence Report/Technology Assessment No. 208 Part 3 (Prepared by the Vanderbilt Evidence-based Practice Center under Contract No. 290-2007-10065.) AHRQ Publication No. 12-E009-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/index.cfm/search-for-guides-reviews-and-reports/?productid=1242&pageaction=displayproduct
Published Online: August 27, 2012

Commentator & Affiliation	Section	Comment	Response
TEP #1	Executive Summary ES-17, Line 31-32	The target populations are explicitly defined in terms of disparities and disparities indicators. Though the following sentence,".....and the fact that any two populations may represent multiple and overlapping disparities," is important enough that it should be introduced much earlier. Perhaps in sections within the ES and the Full Report where disparities indicators are described. This would explicitly point out that disparities indicators (race/ethnicity, gender, insurance status, etc...) are not mutually exclusive and that disparities populations often share multiple disparities indicators. We tend so much to compartmentalize these in our studies and in policy discussions, so I strongly recommend we make this statement upfront. The report will then come full circle with this point when it points this out as one of the challenges later on.	We have added a sentence to this effect in the first paragraph of the executive summary and in the background of the main report.
TEP #1	Executive Summary ES-9, Line 41	Change IMPROVED to REDUCED in the sentence, ".....with two studies reporting no evidence that the intervention improved the disparity". We don't want the disparity to get better; we want it to go away.	Revised "improved" to "reduced", so the statement reads, "… two studies reporting no evidence that the intervention reduced the disparity. "
TEP #1	Executive Summary ES-16, Line 24	Under Discussion. Sentence starting with "Based on the limited evidence..." is an awkward sentence. It seemed to be missing a verb on first read but then upon reading it again, it seems to just be an awkward sentence.	We have rewritten this sentence
TEP #2	Executive Summary	Generally seemed clear and accurate; however, I wondered about some duplication of information across (a) footnotes in analytic framework, (b) inclusion/exclusion criteria table, and (c) text – some information is here three times.	We have removed the analytic framework from the Executive Summary. This eliminates one source of duplication.

Commentator & Affiliation	Section	Comment	Response
TEP #2	Executive Summary	I think you could/should drop Table ES-1. AHRQ prefers "fewer" tables in the Executive Summary to "more," and it is not needed (you have most of this information in text, and it is unusual to have this particular table in a summary). IF this is the standard approach for ALL reports in this update series, then fine (but then cut your text way back), but otherwise I'd drop.	The Office of Communications and Knowledge Transfer (OCKT) has set a ceiling of three tables and requires that if one or more of those tables is (are) longer than a page, the text should be cut equivalently. If the EPC needs to choose between a higher number of simpler tables and greater complexity of fewer tables, however, OCKT will permit a higher number of tables. We have seven tables, including the inclusion/exclusion criteria and the individual effects on disparities tables. We have one figure (literature flow diagram). The analytic framework has been removed.
TEP #2	Executive Summary	Please look carefully in Executive Summary and then everywhere throughout the report for how you elect to list the conditions. Sometimes they are termed "priority" conditions, which is fine (although that does have a particular meaning for AHRQ), but some readers may find this confusing, because these are not the IOM's priority conditions for quality improvement (from the 2003 report)—some are, but not all, and not in the IOM's order or exact terminology, and some that you have may not be in that list).	These are conditions selected for the focus of this report, and are not intended to mirror a specific set of national priorities. We have revised the text to make this clear.
TEP #2	Executive Summary ES-6	Yours also aren't exactly the 10 priority conditions from MMA (which is where AHRQ's conditions typically have come from). You mention AHRQ and IOM publications as sources; apparently your choices are unique within those sets as agreed to by AHRQ (EPC), your TEP, and the authors. You would do well to make this clearer on Page ES- 6.	We have provided clarification.

Commentator & Affiliation	Section	Comment	Response
TEP #2	Executive Summary	References 6-8 cited on that page are not correct and, in fact, as best I can tell you do not have the citations at all in the ES reference list. (These are items 6- 8 in the main list, but frankly, I think 7 and 8 are fairly odd cites and not really primary cites; 6 is the correct one for the original IOM report.)	We have deleted the references for the Institute of Medicine (IOM) Commitment to Quality of Care Collection of Reports (citation #7) and the AHRQ Partners Guide (citation #8) and left the citation for the IOM 2003 "Priority areas for national action: transforming health care quality" report (citation #6).
TEP #2	Executive Summary	I cannot figure out what logic you used for listing the conditions, given that you have reasons for picking the ones you pick. They are not alphabetic, not in order by obvious organ system, not by prevalence (I don't think), not by severity, and not by the relevant clinical intervention (except that the screening ones come first). The list is the same in the table and in the text, but not for the analytic framework. See below for comments about the eventual order of the analyses/results.	We have reorganized the conditions alphabetically to provide an overarching logic to the organization.
TEP #2	Executive Summary	My suggestion: [1] sort out what you really want to list for the AF, [2] drop the table here (but fix it in the Methods chapter), and [3] look everyplace in main text where you list conditions and get them organized and in the same way. Some kind of obvious logic to the list order would be nice (even alphabetic would help, though keeping the two cancers together might make sense). A little more explanation of why these might be sensible.	We have deleted the Analytic Framework from the Executive Summary. We have revised the table of inclusion and exclusion criteria. We have ordered the conditions alphabetically, keeping cancer and cardiovascular disease together.
TEP #2	Executive Summary	Also go back through text where you first introduce conditions and consider noting this is what you mean by clinical target (or at least that's what I infer). THEN: try to make some logic out of what order you use for results (starts with diabetes – completely unpredictable), and see if you can warn readers about what's in, what's out, and in what order. You could do this on Page ES-7, perhaps, where you describe a whole bunch of other characteristics of the literature, but say nothing about conditions, which were very much cut back once you settled on the articles to include.	We have reorganized the report so that the results are organized by population characteristics associated with disparity, rather than by clinical target. Within disparity, when we do provide information by clinical condition, it is organized alphabetically.

Commentator & Affiliation	Section	Comment	Response
TEP #2	Executive Summary ES-5	List on ES-5 is race and ethnicity, socioeconomic status, insurance status, sex, sexual orientation, health literacy/numeracy, and language barriers. I wondered, however, about the connotation of those being "disparities" (or even indicators of disparities). I guess that's okay, but really these are simply characteristics of certain kinds of populations thought to be a high risk of disparities in health and health care. CLEARLY, one would not try to use QI interventions to change them (except, perhaps literacy/numeracy). Elsewhere the disparities are considered as being measured by utilization, access, or quality (process measures or outcome measures). Just be sensitive to this language so that nobody takes offense (imagine saying that somebody should try to change ethnicity, or sex, via these interventions) or gets confused.	We have edited throughout to make the language more clear. We agree that these are characteristics of populations; they are, however, characteristics that have been shown to be associated with health care disparities.
TEP #2	Executive Summary	Note that your analytic framework does not deal with disparities this way (by personal characteristics) at all. Here the disparities are in health care (but not also health outcomes), per the title. No footnote indicates what characteristics of populations you consider important for looking at these disparities.	We have changed the title of the Analytic Framework to match the report title. We have added a table note listing the disparity indicators/characteristics. We have edited the Analytic Framework table notes to order the conditions as noted above and to order the QI strategies as noted below.
TEP #2	Executive Summary	AF: the footnote about settings isn't quite accurate. Settings don't conduct things. You mean that settings include those in which QI interventions or programs were implemented or tested, such as those in hospitals, physician (or provider?) offices, or clinics.	We have revised the Analytic Framework footnote to "Settings include those in which QI interventions were tested: hospitals, provider offices, and/or health care clinics."
TEP #2	Executive Summary	You have 12 studies; of these, all have QI interventions of at least 2 elements (Barr et al. doesn't, apparently), so I don't think the description on Page ES-7 (Line 22) is quite correct.	We inserted "but one" to read, "All but one of the studies used a multicomponent approach…".

Commentator & Affiliation	Section	Comment	Response
TEP #2	Executive Summary	The two ranking (tied) elements of these interventions are patient education and provider education (10 of 12), by my count; right behind is self-management (9 of 12). All other elements are found in only a minority of these studies. You might consider providing some kind of clearer overview (maybe a new paragraph on ES-7 starting at Line 22) about the QI intervention elements; whatever is in the ES tables ought somehow to be predictable from text.	We have revised this text.
TEP #2	Executive Summary	One question: will readers understand what a "collaborative care" model is (for the depression studies)? Double check your description in the main report (around pp. 16-17). That's where Table 3 shows up.	We have added a brief description of the collaborative care model in which multiple types of providers (e.g., physicians, nurses, case managers) collaborate to provide a coordinated set of interventions in the Executive Summary, and the Results and Discussion chapters of the report.
TEP #2	Executive Summary	I realized, after spending a lot of time trying to figure this all out, that you have picked the clinical conditions as your principal analytic category, evidently followed by the characteristics of the populations (what you call disparity indicators), then the actual QI interventions, and finally the actual disparity outcomes measured. I wasn't expecting that – had assumed that either the QI interventions or the disparity indicators were the critical element – but the focus on clinical conditions is probably okay as an organizing principle and the QI interventions are so complicated and diverse that focusing on them isn't going to work. It's just that you should try to make this organizational strategy clearer at the outset, because the presentation around Page ES-5 doesn't set this up – the conditions come after the other elements.	We have reorganized the report so that clinical conditions are not the focus of the results.
TEP #2	Executive Summary	Literature search: Noting the two interfaces used Is useful here (and in main text), but put them both in parentheses.	We have inserted parentheses for the two search interfaces.
TEP #2	Executive Summary	Don't understand why you refer to the articles as "papers." That is unusual language for a systematic review. Suggest you use articles throughout, especially because you have no gray literature here.	We have opted to use "papers" throughout the report for consistency.

Commentator & Affiliation	Section	Comment	Response
TEP #2	Executive Summary ES-7, Lines 8-24	Saying your 17 papers represented 12 distinct populations is, at best, misleading because, really, you have 12 different studies; those studies do have different populations (on at least some dimensions, though not necessary all), but that's not the main point for the lit search.	We have deleted "…distinct populations addressed in…" so statement reads, "Seventeen papers met criteria, representing 12 studies of cancer, diabetes, heart disease, hypertension, and depression."
TEP #2	Executive Summary ES-7, Lines 8-24	Later you say studies, but then you call out a Table 3 that does not exist in the Executive Summary (and you really shouldn't cite a table in main text without making it abundantly clear that's where it is).	We have deleted "(Table 3)".
TEP #2	Executive Summary ES-7, Line 14	Make a new paragraph on Line 14 about the studies ("Of the 12 studies...").	We have inserted a paragraph break.
TEP #2	Executive Summary ES-7, Line 19	Make another new paragraph on Line 19 ("Included studies..."). That will help break up a too-dense paragraph.	We have inserted a paragraph break.
TEP #2	Executive Summary	PRISMA flow diagram: [1] You can keep the flow diagram for the Executive Summary if you like, but you do not "have" to include it, and I think you probably have enough just with the text.	Thank you. We have deleted the literature flow diagram.
TEP #2	Executive Summary	PRISMA flow diagram: Please try to correct it, however (here and in main text), to ensure that it refers to articles – either you don't specify at all or you call the elements studies, which is incorrect.	We have revised the figure to reference "papers" instead of "studies".
TEP #2	Executive Summary	See note above about preparing readers for only four main conditions (out of a whole bunch). It's not quite clear to me why you start with diabetes – maybe because of hierarchy of study design and number of studies. (That's fine; I just would have expected this to be cancer, then diabetes, heart disease, and finally depression.)	The report has been substantially reorganized as noted above, and we have added text specifying that although we sought information on a wider range of interventions and clinical conditions, studies were not available for all of them.
TEP #2	Executive Summary	Mostly I think this can be edited to make it shorter and easier to read. Consider splitting the disease-specific sections into paragraphs that deal only with study characteristics, then (separately) the results.(Example: make a new paragraph on Page ES-8, Line 52 at "In two...")	We have substantially edited the Executive Summary and the Full Report.

Commentator & Affiliation	Section	Comment	Response
TEP #2	Executive Summary ES-9	In studies of CRC and breast cancer screening: Probably best to refer at first to the Department of Veterans Affairs, and consider not using "firms," which only people really into VA or long-ago quality studies will understand (I suspect).	We have revised "Veterans Affairs firms" to "Department of Veterans Affairs clinics".
TEP #2	Executive Summary ES-9, Line 29	In studies of CRC and breast ca screening: I'm not convinced people will get what you mean about interaction (significant or otherwise) – I don't in this case.	We have revised the text to clarify that although a difference in effect was seen between English-speaking and Spanish-speaking participants, the difference was not significant.
TEP #2	Executive Summary ES-10, Line 37	In studies of depression: Readers expect "intervention" to be the QI effort, but here you invoke a design variable (group-level randomization), which surprised me. However, if this was some kind of cluster RCT (group level = practice), how were individual providers in those practices, or patients, able to pick the treatment? Not exactly sure how the members of the group (practice) were able to do this or, even more, how patients were able to pick their own treatments. You might look at this description again.	That sentence was unclear. We have deleted the phrase referring to the randomization. Although there was randomization at the practice level, providers could break that randomization and select treatments on a patient-by-patient basis.
TEP #2	Executive Summary ES-10, Line 48-51 (or so)	In studies of depression: Is something missing for the dangling phrase about some measures?	Yes, this was an error. We have corrected the statement to read, "For example, one study demonstrated a greater effect on clinical outcomes in the less educated group, and the effect of a second intervention was amplified in minorities on some measures." in the Executive Summary and in the Full Report.
TEP #2	Executive Summary ES-10, Line 48-51 (or so)	Also, why introduce these RCT names here if you don't use them again in the Executive Summary?	We have removed the names of the RCTs; they are described in the Full Report.
TEP #2	Executive Summary	Your language of "disparity target" may be useful to invoke more often. The targets are the various diagnosis-specific measures for health care or health outcomes; they are not the population characteristics.	We have edited and clarified throughout the report.

Commentator & Affiliation	Section	Comment	Response
TEP #2	Executive Summary	For ES-2 and the related tables, consider changing the titles to say: Summary of included studies by clinical condition, disparity target, and race or ethnicity. Next one would be by Summary of included studies by clinical condition, disparity target, and insurance status. And so on – you get the idea. The problem is that right now, as written (with that dash), you are saying that race/ethnicity (or insurance status, or sex, or whatever) *is* the disparity.	Thank you. In reorganizing some of the text of the report, we have elected to rename the table titles to "Summary of effects on disparities in health outcomes associated with xx "
TEP #2	Executive Summary	In these tables, capitalize all the main words in the column heads	This has been corrected throughout the report.
TEP #2	Executive Summary	Change "effect on disparity" to "effect on disparity target." In table column headings	We have revised the table column heading to "Effect on Health Disparity".
TEP #2	Executive Summary	I never did figure out what order or logic was operating in the list of QI intervention characteristics. Consider reorganizing these cells to start always with patient education (patients trump providers), then provider education, then promotion of self-management, and so on in reverse numerical order of appearance (and then order all the ones that appear in only a single study in alphabetical order). If a given study didn't include something, just move to the next possible element in the list. EXAMPLE: Sequist et al. would be Provider education, then Audit and feedback, whereas, say, Mahotiere et al. would be Patient education, Provider education, Promotion of self-management, Audit and feedback, and Community intervention. (I do not understand how/why community intervention is part of a QI intervention, but that's okay.) That might make it easier for readers to find the studies that have a given element in their QI intervention. However, if you have some other conceptualization in mind, can you explain or describe it someplace?	We have revised the order of QI strategies in all tables: • Patient education • Provider education • Promotion of self-management • Audit and feedback • Facilitated relay of clinical data to providers • Patient reminder system • Provider reminder system • Organizational change • Other
TEP #2	Executive Summary Table ES-2	Consider explaining QI-Therapy and QI-Meds someplace. Not entirely sure what QI + [therapy] [meds] means, especially when the other construction uses -. (a minus sign?!) Meds are a form of therapy, so the distinction was lost on me.	These terms are no longer used in the Executive Summary are explained in full in the results and discussion. We have replaced "QI+Therapy" and "QI+Meds" with "QI-Therapy" and "QI-Meds".

Commentator & Affiliation	Section	Comment	Response
TEP #2	Executive Summary ES-14	One of the spots where language concordance is not clear, especially when paired with language barrier in the table. (Language barrier is not listed anyplace as an indicator of disparity anyhow).	We provided text to clarify: "Both of them studied language concordance, in which strategies are provided in the native or preferred language of the participant (e.g., in Spanish for native Spanish speakers)."
TEP #2	Executive Summary ES-14, Line 12	I don't understand "no significant language by intervention interaction if there was a greater effect in the Spanish-speaking group. The explanations in the table seem better.	We have clarified the text as follows: "For breast cancer screening, Spanish speakers were more likely to be up to date at baseline than English speakers (OR 1.46; 95% CI: 1.16 – 1.84). The intervention was associated with increased rates of screening overall, with subgroup analysis indicating a greater effect in the Spanish speaking group (OR 1.85; 95% CI: 1.38, 2.47) than the English speaking group (OR 1.18; 95% CI: 0.82, 1.71). However, the overall multivariate analysis failed to confirm these results, finding no significant language-by-intervention interaction"
TEP #2	Executive Summary ES-16, Lines 22-23	Need "is" before unavailable, and "that" before qualified.	The referenced sentences have been revised to "We identified individual studies that suggest benefits in particular subgroups known to suffer from disparities in health and health care, but evidence is unavailable to guide QI efforts specifically to reduce disparities. Based on limited available evidence, several strategies are worthy of future study, and possibly wider implementation."

Commentator & Affiliation	Section	Comment	Response
TEP #2	Executive Summary ES-16, Line 27	Can you find a spot someplace to clarify what language concordance means – it's sometimes an alternative to language barriers. Do this throughout the Executive Summary AND the main text. Does it mean that patients and providers spoke the same language or that materials were the patients' language, or both, or something else? I assume literacy concordance means that materials were done at a level of literacy commensurate with that of the patients (it's probably not that the providers were of the same literacy level as patients).	We have clarified that these are interventions in which strategies are provided in the native or preferred language of the participant (e.g., in Spanish for native Spanish speakers).
TEP #2	Executive Summary ES-16, Line 27	Say that Data are insufficient to support...	We have revised from, "There do not appear to be sufficient data to support…." to "Data are insufficient to support …"
TEP #2	Executive Summary ES-16, Line 33	Did you REALLY have studies that did something in homes? That's not the implication of the analytic framework, which seems to suggest only hospitals, physician offices, or clinics.	We included studies of interventions initiated in clinical settings. Supporting patients in home-based blood pressure monitoring was one component of an included study intervention. We identified this study as being conducted in a health care setting since the interventions and patient population originated in two primary care sites. We have clarified the sentence to indicate that the intervention includes a home-based component.
TEP #2	Executive Summary ES-16, Line 48	Consider a new paragraph at "A small number of..." Some of these paragraphs are far too long for anybody to follow them easily.	Thank you. As per the comment below, this statement has been deleted. We have inserted paragraph breaks as needed throughout the report.
TEP #2	Executive Summary ES-16, Line 51	You need a period after interpret.	We have made this correction.
TEP #2	Executive Summary ES-16, Lines 52-53	This seemed like a non sequitor or just unnecessary – it's not anything related to heterogeneity of the articles.	We agree. This statement appears to have been misplaced and has been deleted.
TEP #2	Executive Summary ES-16, Lines 52-53	I think you could then have another new paragraph at "Our assessment..."	We inserted a paragraph break before, "Our assessment…"

Commentator & Affiliation	Section	Comment	Response
TEP #2	Executive Summary ES-17, Line 7	Please specify whether this is many, some, several, or only a few – a number is too ambiguous.	We have revised this statement so that it refers to the authors of studies included in this review.
TEP #2	Executive Summary ES-17 Line 19	I thought the NLM librarians added terms to the MeSH tree, not to PubMed.	We revised the text from " For example, the subject term "Quality Improvement" was only added to PubMed in 2011" to "For example, the subject term "Quality Improvement" was only added to the PubMed Medical Subject Heading Database (MeSH) in 2011".
TEP #2	Executive Summary ES-17 Line 19	You might in any case hint at the challenges that the NLM indexers have – for example, it's not THEIR lack of consistency or agreement re what is QI, it's how and what authors write (especially abstracts) and the key words they might suggest. The points are right – just make sure the "culprits" aren't the NLM librarians	We have revised the statement from "Compounding this heterogeneity is poor indexing of quality improvement strategies in the medical literature databases. For example, the subject term "Quality Improvement" was only added to PubMed in 2011; before this time, myriad subject terms were employed to describe the various strategies employed in the QI literature, understandably leading to tremendous variability in how similar studies are categorized in the database.", to, "Compounding this heterogeneity are challenges to indexing of QI strategies in the medical literature databases. For example, the subject term "Quality Improvement" was only added to the National Library of Medicine's Medical Subject Heading Database (MeSH) in 2011; before this time, myriad subject terms were used to index the various strategies described by authors of the QI literature, understandably leading to tremendous variability in how similar studies are categorized in the database."

Commentator & Affiliation	Section	Comment	Response
TEP #2	Executive Summary ES-17 Lines 39-44	Applicability. I think “gaps in” applicability may be more gentle than “risks to”?	We revised the statement from “Despite fairly large studies in diverse areas of the United States, all of the studies had substantial risks to applicability…” to “Despite fairly large studies in diverse areas of the United States, all of the studies had substantial gaps in applicability…”
TEP #2	Executive Summary ES-17 Lines 39-44	Consider rewriting second sentence as: Therefore, health systems or clinicians wishing to replicate ay of these interventions should carefully assess whether the interventions...	We revised the statement from “Therefore, health systems or clinicians wishing to replicate any of these interventions should carefully assess whether they apply or must be modified…” to “Therefore, health systems or clinicians wishing to replicate any of these interventions should carefully assess whether the interventions apply or must be…”
TEP #2	Executive Summary ES-17 Lines 47-55	Conclusions. Can stand some rewriting. Say “The literature on ... health care is large. Whether those interventions are effective at closing... populations remains unclear.”	We have modified the statement as suggested.
TEP #2	Executive Summary ES-17 Lines 47-55	Use comma after rather, and say QI has not been shown specifically to reduce known disparities in health care or health outcomes. [This is the point to keep in mind – you are concerned with health care AND with health, and you are not concerned with reducing disparities in population characteristics that may signal the potential for disparities.]	We have revised from “rather QI has not demonstrated to specifically reduce known disparities in healthcare” to “rather, QI has not been shown specifically to reduce known disparities in health care or health outcomes.”
TEP #2	Executive Summary ES-17 Lines 47-55	I would not use “subpopulations” – nobody wants to be classified as a “sub”...	We have revised all instances of “subpopulation(s)” to “population(s)”.
TEP #4	Executive Summary ES-2 Line 29	You reference TEP members but don't define the acronym until page ES-3, Line 39	This has been corrected; TEP is spelled out in the first instance of the Executive Summary (Page ES-2).
TEP #4	Executive Summary ES-17, Line 22	Insert "is" between "evidence" and "unavailable."	We have revised the statement to read, “…. but evidence is unavailable to guide….”

Commentator & Affiliation	Section	Comment	Response
TEP #2	Abstract Line 14	don't need the interface for PsycINFO here (haven't done for MEDLINE anyhow)	We have deleted "(CSA Illumina Interface)".
TEP #2	Abstract Lines 17-25	Something is wrong with prepositions and then verbs across all this. Turn it into 2-3 separate sentences.	We have revised the statement and created three separate sentences.
TEP #2	Abstract	Here, as elsewhere in the report, the logic of the order of conditions is not clear. Similarly, the logic of the order of the disparity variables (also throughout the report) is not clear or predictable;	We have re-ordered the conditions alphabetically. We ordered the disparity variables by the number of studies available so that those with more possible evidence are described first.
TEP #2	Abstract	Change gender to sex EVERYWHERE in this report (as it is not consistent but gender is not correct).	We have revised "gender" to "sex" throughout the report.
TEP #2	Abstract Lines 30-45	Spell numbers 10 and over.	Consistent with the OCKT publication guidance, we spell out numbers less than 10. We use numerals for time and measurement, and in tables and parentheses.
TEP #2	Abstract Lines 30-45	Why in the second/third sentences aren't the "commonly used" interventions in some logical order (e.g., 10 of 12, then 9 of 12, and so on); 7 of 12 is barely more than half, so why is this considered "common?"	We have revised this.
TEP #2	Abstract	Through the abstract, drop the constructions using "with," which are wordy and potentially confusing; use a semi-colon and write with active verbs.	We have revised accordingly.
TEP #2	Abstract Line 45	What does "confirmatory" mean in Line 45? (Confirmatory is not a concept in most systematic reviews, unless you want to claim the studies confirm an earlier hypothesis, but that's not really the case for most of these studies.)	This sentence was confusing and has been deleted.
TEP #2	Abstract Lines 47-53	Drop the "There is..." construction and write as "The literature on QI interventions is large…" and then sort out what the sentence is supposed to be after that through to the point that says "...is insufficient." Something isn't working or clear here.	These statements have been revised to: "The literature on QI interventions generally and their ability to improve health and health care is large. Whether those interventions are effective at closing the gap between two populations remains unclear."

Commentator & Affiliation	Section	Comment	Response
TEP #2	Abstract Line 52.	Insert “been” between “not” and “demonstrated”.	We have revised the statement from “…QI has not demonstrated to specifically…”, to, “…QI has not been demonstrated to specifically…”.
Peer Reviewer #1	Introduction	I would like to have seen some of the conceptual issues raised earlier addressed in the Introduction. The conceptual model presented is straightforward, but I think misses some of the nuances of the field.	We have expanded the introduction somewhat.
TEP #2	Introduction	Okay in very basic ways.	Thank you.
TEP #2	Introduction	Should first sentence say health care (not just care) and health outcomes?	We inserted “health”, so the sentence reads, “QI is a multidisciplinary, systems-focused, data-driven method of understanding and improving the efficiency, effectiveness, and reliability of health processes and outcomes of health care.”.
TEP #2	Introduction Lines 23-24	Last sentence, say: “... considering whether these interventions.... as well is a logical next step.”	We have revised the sentence from “… it is logical to consider that these interventions could be fruitful in reducing disparities as well.” To “… considering whether these interventions could be fruitful in reducing disparities as well is a logical next step.”
TEP #2	Introduction	The QI history. The first use of much of this dates back 60 years or more (to the end of the WWII), so you might indicate tools first used in industry and then on to other sectors for the past six decades. Or something like that. This stuff is not new, exactly. I’d note that PDSA is one of the earliest approaches, so I think your history is a little off – it’s not simply another model.	We have revised the QI introductory section accordingly.

Commentator & Affiliation	Section	Comment	Response
TEP #2	Introduction	Please consider what you can excise from text if material is in a table. You should keep Table 1, but I don't understand the organization at all. Please go back to the "ordering" you should impose for the tables in Executive Summary and present the QI strategies in that order. Will set things up much better for the remainder of the report. E.g., start w/ patient education, then provider education, then promotion of self-management. Then I think you might go as collaborative care model as a separate entry (not another), then audit and feedback, facilitated relay, patient reminders, provider reminders, then organizational change, and finally other. Frankly, though, I'd put care manager in with organization change, and maybe make guideline adherence a separate category (if you've pulled the other 2 things out).	We have edited the report. We have reordered the QI strategies in Table 1 and in other places in the report as suggested.
TEP #2	Introduction	Your language about disparity, and noting that differences in outcomes or determinants are associated with social conditions and demographic attributes, is the useful way to tackle this. Attributes that describe people is the way to go	Thank you.
TEP #2	Introduction	Change gender to sex.	We have revised "gender" to "sex" throughout the report.
TEP #2	Introduction	Be sure to make clear that the items in ( ) are only examples (e.g., in the first item's example will do that).	We have inserted "e.g.," at the beginning of each item in parentheses.
TEP #2	Introduction	Identify the NHDR as belonging to/produced by AHRQ. Is the 2009 report the right one to be citing (don't know – just that it'll be 2012 by the time this review is out, so 2010 or maybe even 2011 might be the better report to cite)?	We now reference the 2010 NHDR, published in March 2011.
TEP #2	Introduction Lines 30-32	Number these (not bullets) because you say three.	We have revised to Arabic numerals per the publishing guidance.
TEP #2	Introduction	Scope of report: you say health care access and utilization and health outcomes. These are not precisely what has been said above/elsewhere – you may want to equate health care with health care with access and use.	Revised.

Commentator & Affiliation	Section	Comment	Response
TEP #2	Introduction	Here and elsewhere re the KQs: You probably cannot do anything about it, but the use of a wholly unjustified change in the font and font size is unsettling (at best) to readers – there is no obvious reason to do this and it just stops readers cold, wondering what's going on.	The key questions are in 12 pt. Arial regular font per the publishing guidelines.
TEP #2	Introduction	Your analytic framework and discussion belongs in this introductory chapter, not in methods, unless something has really changed. Pull it out of there, but shorten... See comments re Executive Summary, too.	Consistent with AHRQ CER Content Guidance, dated September 2011, we have moved the Analytic Framework to the "Scope and Key Questions" section.
Peer Reviewer #2	Introduction	In the executive summary, the background section was overall very brief. The definition of disparities is very broad and is not truly congruent with IOM report or other widely accepted definition of disparities. The authors accurate define disparities in the Full Report. More development of what disparities in health care truly means in the summary section would be helpful	We have added text from the Full Report into the Executive Summary.
TEP #3	Introduction	Good, solid and clear introduction.	Thank you.
TEP #3	Introduction	The analytic framework diagram (Page ES-3, Figure ES-1) is nicely done; clear and informative.	Thank you.
TEP #4	Introduction	The background is clear. However, given that the disparities cite in the report are population-based (e.g., higher mortality due to coronary heart disease among black men and women compared to whites), it may be helpful to say more regarding how it is envisioned that QI intervention studies that are essentially local and heterogeneous in terms of settings and individuals can be translated to effect population-based change. Certainly, we have examples, as in tobacco policy, that demonstrate population-based change (even if not necessarily differentiated by disparity condition), but how can clinic-based QI interventional studies do that?	We have added some additional text to the introduction: "QI interventions have been successful at improving health outcomes generally and in a number of settings; it is possible that they could be adapted or targeted to narrow a health or health care gap. If they were implemented in either a targeted way (i.e.,, in locations with especially high disparities) or broadly, they could potentially affect disparities at the population level"
TEP #4	Introduction	QUALITY IMPROVEMENT: The explanation of QI is well done and Table 1 provides a good overview.	Thank you.

Commentator & Affiliation	Section	Comment	Response
TEP #4	Introduction	The one thing in this section that I question is the discussion of QI paradigms (TQM, CQI, etc.). Addressed here, they seem never to appear again, except on Page 3, lines 45-47. If they are important to the analysis of the individual studies, perhaps their role in those studies merits mention elsewhere. Otherwise, I don't see the need to include the paradigms.	These are intended as examples of approaches that we anticipated potentially finding in the literature and that serve to help readers newer to the concept understand the report.
TEP #4	Introduction	DISPARITY: This is a good and brief overview.	Thank you.
TEP #4	Introduction	APPROACHES TO REDUCING DISPARITIES: Starting at Page 4, Line 5, the report states, "Thus, for a QI intervention to be effective for reduction of disparity, both intervention effectiveness and disparity reduction effectiveness must be demonstrated; the intervention would be more effective for disadvantaged groups or individuals than for advantaged groups. A judgment of effectiveness is therefore not possible when the intervention is targeted only at disadvantaged individuals or groups." It goes on to note that such research "is thus complicated and rare." While I appreciate the desire for head-to-head comparisons, and their importance in some circumstances, I am not convinced they are the best way to ascertain whether a QI intervention is effective and reduces disparity. The head-to-head approach seems to diminish the impact single population studies can have in addressing a particular disparity. If women's participation in breast screening in an Indian Health Service (IHS) Clinic, is significantly lower than participation by white women seen in a clinic that is majority white, then interventions in the IHS clinic that raise screening participation to approximate that of white women would seem to be an approach to reducing disparities.	As noted in the report, while it is theoretically true that changing health outcomes in a minority population could reduce disparities if rates in the majority population do not change, this is an empirical question for which there is currently inadequate evidence. We agree that studies of the impact of interventions in underserved populations provide important and valuable data for measuring improvements in care in those populations; they do not necessarily demonstrate changes in disparities. Furthermore, while single population studies certainly may be associated with reductions in disparity, it would not be possible to observe a change in disparity without a comparison group (i.e. the majority group). Furthermore, there are known and accepted risks of bias associated with single arm studies and they are unable to provide evidence that changes in outcomes were not due to underlying trends or other interventions. Nonetheless, we agree it is important to acknowledge the value of study designs that were not included in this report. We have added a statement to the report Discussion Chapter indicating that although this report included only studies designed in such a way to measure disparities, other types of studies, such as those including an underserved population only may offer

Commentator & Affiliation	Section	Comment	Response
			valuable information to clinicians and policymakers, We have inserted a statement in the "Gaps in the Literature" section of the Discussion Chapter: "Our review specifically sought studies that could measure a potential change in disparity. Other types of studies, such as those that only included only underserved individuals, might provide valuable information to policymakers and clinicians hoping to improve care in those populations, but would not have been included in this review"
TEP #4	Introduction	KEY QUESTIONS: They seem reasonable, however, I don't see that they necessarily imply the approach discussed in the previous section, as they do not explicitly narrow the pursuit of answers to identical QI interventions applied uniformly to "disparities" and "non-disparities" populations at essentially the same time.	The non-disparity group could have been an external referent population, thus controlling for secular trends. In this case, the non-disparity group would not necessarily have received the intervention. There did, however, need to be some comparison group, even if it was within the disparity group.
TEP #4	Introduction	USES OF REPORT: The audience that will use the report is well described. After reading the report, I think the audiences who could most benefit are funding agencies who develop RFAs and researchers who develop the protocols that are responsive to the criteria laid out in the RFAs.	We have revised this section.
TEP #4	Introduction	TARGET POPULATION AND AUDIENCE: Yes, I think they are explicitly defined to include: policy makers, health systems leaders, funding agencies, and researchers. I would say that the report could be of greatest import for researchers and funding agencies, if it spent more time emphasizing specific research criteria that need to be included in research and in the RFAs that support the research.	We appreciate your comments.
TEP #4	Introduction	Key questions are appropriate and explicit.	Thank you for your comment.
TEP #4	Introduction	The introduction defines the scope of the charge and the problem that they hope to clearly articulate.	Thank you for your comment.

Commentator & Affiliation	Section	Comment	Response
TEP #5	Introduction	Introduction is generally good and clear. I especially want to compliment the writers in their clarity in defining QI interventions. This is a great improvement over earlier reports in the series.	We appreciate your comment.
TEP #5	Introduction	I did find the inclusion of payment strategies as a QI intervention odd, and would have been happier excluding that from the list (I know it was ultimately excluded as it was covered elsewhere). The argument that QI interventions that reduce disparities are inherently difficulty and rare was not fully compelling. IF the QI intervention includes an assessment of the source of disparity then customizes the intervention to the source, the differential impact burden need not be so high. One could argue that the same language intervention described later is such an intervention---if they started with an assessment of the preference for language, they would have found that one group needed a new set of counselors and the English speaking group was fine, then you'd have different interventions based on an assessment.	As noted, payment strategies were not included in our assessment.
TEP #1	Methods	Overall methods are good. Please see my comments above in terms of framing	Thank you.
Peer Reviewer #1	Methods	I would prefer to see interventions included that were rigorously implemented and evaluated if they targeted only minority or other underserved groups, even if there was no white comparison group (per above comment).	Please see prior response.
Peer Reviewer #1	Methods	The authors have not attempted to distinguish the conceptual difference between interventions that "target" minority populations or other underserved populations and their needs versus those interventions that are "generic" and not meant to specifically address disparities. This is a very important conceptual difference for the field -- as many believe we do not need targeted interventions, but rather should implement programs that improve care for all. The literature does have examples of both ("generic" and "targeted" efforts) and it would be helpful to distinguish these in your analyses.	This is true that we have included all studies that could potentially have provided data on changes in disparities. Several of these included studies clearly provide data on a targeted intervention (e.g., language concordance, or a depression care model focused on the elderly) and we have attempted to note in the text where this is the case.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #1	Methods	It is not clear why interventions that focus only on underserved populations would not be included, if they are intended to improve the health of minorities, women, the poor, etc. For example, a program to improve mental health care that was a randomized trial involving only black patients would not be included here (I believe) because it did not have a white comparison group. But if that intervention was successful at improving care for black patients, I think that is important information for the field of disparities.	The particular difficulty in assessing the impact of any health care intervention on disparities is that research must show effectiveness across multiple planes. Evidence of the effectiveness of the intervention needs to be demonstrated using a non-intervention comparison group, and at the same time, a disparity in outcome must be narrowed within the intervention group, but not in the comparison group. Thus, for a QI intervention to be effective for reduction of disparity both intervention effectiveness and disparity reduction effectiveness must be demonstrated; the intervention would be more effective for disadvantaged groups or individuals than for advantaged groups. A judgment of effectiveness in reducing disparities is therefore not possible when the intervention is targeted only at disadvantaged individuals or groups. We agree that studies of the impact of interventions in underserved populations provide important and valuable data for measuring improvements in care in those populations, they do not necessarily demonstrate changes in disparities. While single population studies certainly may be associated with reductions in disparity, it would not be possible to observe a change in disparity without a comparison group (i.e. the majority group). Furthermore, there are known and accepted risks of bias associated with single arm studies and they are unable to provide evidence that changes in outcomes were not due to underlying trends or other interventions

Commentator & Affiliation	Section	Comment	Response
TEP #2	Methods	Topic refinement. This should be shortened to maybe 100 words – readers don't care and it's just a whole page to get past (or skip). See the "guidance" for CERs for some sample language, which ought to help.	The first paragraph of the topic refinement description was provided by AHRQ for the series. A paragraph break has been inserted to break up the general topic refinement description from the TEP description.
TEP #2	Methods	You can keep how you defined QI at the bottom (lines 45-54) or move it someplace else – it's important but lost here.	This has been addressed in editing.
TEP #2	Methods Line 7	For Study Selection: Make a new sentence "Populations could be characterized..." This is impossible to follow otherwise.	The statement has been separated and a new sentence has been added.
TEP #2	Methods	Make a new paragraph at "As an inclusion criterion," because this has nothing to do w/ study populations. You might even consider a new heading for Sample Size.	We have inserted a level two heading for "Sample Size".
TEP #2	Methods Line 26	Don't mention TEP again.	In line with this suggestion and the current AHRQ CER Content Guidance, we have revised and shortened the introductory material in the Methods chapter.
TEP #2	Methods	Geographic limitation, etc. Change the "as" to "because" in both places here.	Revised "as" to "because" so the sentences read, "Because health care systems, disparity indicators, and groups subject to disparities vary geographically, we limited eligible papers to studies of patients within the U.S. health care system."… "Searches were limited to papers published in 1983 or later because seminal work regarding QI strategies began to be published in the early 1980s."
TEP #2	Methods Line 55	Study groups. Suggest a new paragraph at "To be eligible..."	Inserted paragraph break before "To be eligible…"
TEP #2	Methods Line 56	Data are	We revised the statement from "…data that was collected..." to "…data that were collected…"

Commentator & Affiliation	Section	Comment	Response
TEP #2	Methods	On the next page, are you sure that "source... at the state or local level" will be clear (is it needed?) – Evidently no national data sets would be eligible?	National data sets would not have been eligible, no studies were found to use an external referent so the criterion was not invoked.
TEP #2	Methods	So all studies had to have three groups: the target group, the referent group, and some non-QI-intervention comparison group, which I gather wasn't to be the referent group. Not sure that's quite clear, and I didn't come away from the description of studies with the idea that all studies had to have three groups.	We have clarified the language around this.
TEP #2	Methods	Conditions: same problems as described for Executive Summary.	We have ordered the conditions alphabetically throughout the report. Congestive heart failure, hypertension, and coronary artery disease fall under "heart disease". Breast cancer and colorectal cancer (including screening) have been placed under "Cancer".
TEP #2	Methods	Table 2: Probably keep the table but figure out what you can cut from text that duplicates this information. Reconsider the order of the interventions listed in it. I think the footnote belongs in the text along with the description of sample size – it makes little sense buried here.	We have kept the table and the text but have edited the text. The intervention row of the table has been revised to describe a QI strategy. We removed specific examples of QI strategies from the table. We have deleted the table footnote and added a section for "Sample Size".
TEP #2	Methods	Quality assessment: do you need a citation for the risk of bias tool?	We have inserted the citation for the Cochrane Risk of Bias Tool
TEP #2	Methods Page- 12, Line 13	This should be These.	We have revised the statement from "This risk of bias scoring and quality assessment tool ratings were…" to "These risk of bias scoring and quality assessment tool ratings were…"
TEP #2	Methods	I think the font/size for Data Synthesis may not be the right level (looks as if it should be one level higher).	We have changed the font and size using the level one heading.
TEP #2	Methods	Grading SOE: Change the heading to Grading the Strength of Evidence for a Body of Evidence.	We have revised the heading to "Grading the Strength of Evidence for a Body of Evidence"

Commentator & Affiliation	Section	Comment	Response
TEP #2	Methods	Consider adding a sentence to comment on high and moderate SOE, given that you say something about low and insufficient.	We have added a statement to describe high and moderate strength of evidence.
TEP #2	Methods	Not sure if you need a separate section for peer review, but I imagine you can add it for the final if you do need.	We have added a section heading for "Peer Review and Public Commentary" and inserted the following language recommended by the CER Content guidance:"Experts in the fields of QI and health disparities and individuals representing stakeholder and user communities were invited to provide external peer review. AHRQ and an associate editor also provided comments. The draft report was posted on the AHRQ website for 4 weeks to elicit public comment. We addressed all reviewer comments and revised the report as appropriate. A disposition of comments report will be available 3 months after the Agency posts the final report on the AHRQ website."
Peer Reviewer #2	Methods	Yes [the inclusion and exclusion criteria are justifiable]. However, the participant per group exclusion of 50 eliminated over half of the studies retrieved. As noted by the authors, this was an estimate for minimum required sample size. Given the nature of these types of studies, it may be useful to consider expanding the inclusion to allow smaller sample sizes to include more studies. Particularly since the results are presented in a narrative form and not being used for meta-analyses or other pooled data analyses.	At the full text screening level eight papers were excluded solely because the number of participants was less than 50 for one or more study groups. This number is liberal in the sense that it is highly unlikely that studies even with 50 individuals per arm would have the power to assess the differences being measured in this literature. In fact, we selected this number with the hope that we would be able to combine studies quantitatively, but could not do so due to the heterogeneity. Therefore, knowing that the studies would only be presented in narrative form would have suggested that we set the minimum number of participants much higher.
Peer Reviewer #2	Methods	Yes - search strategies are explicitly stated and logical.	Thank you.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #2	Methods	[Definitions or diagnostic criteria for outcome measures] Not explicitly stated.	We identify examples of health outcome measures, process measures, and harms in the table of inclusion and exclusion criteria (table two of the Full Report).
Peer Reviewer #2	Methods	How much of a reduction in a "disparity" was deemed to indicate a successful QI intervention? Clearly, this will vary on the individual study's power to detect a meaningful change. More detail of how this was determined would be useful.	We did not set an a priori level of reduction. Any change was reported.
Peer Reviewer #2	Methods Page 5, Lines 52-53	(Regarding the appropriate use of statistical methods) All analysis is narrative. Yet, the abstract suggests definitive findings of no evidence to suggest QI reduce disparities." …rather QI has not demonstrated to specifically reduce known disparities in health care."	Given the heterogeneity of the literature it would not have been possible to conduct statistical analysis. Furthermore, the individual studies did not provide strong evidence that the QI interventions studied resulted in reductions in disparities.
Peer Reviewer #2	Methods	Also mention of "large data on QI and disparities". 17 studies does not strike me as large. This seems misleading.	We have revised the conclusion to state "The literature on QI interventions generally and their ability to improve health and health care is large."
TEP #3	Methods	The search strategies are explicitly stated and logical.	Thank you.
TEP #3	Methods	The inclusion and exclusion criteria are, overall, clearly justified	Thank you.
TEP #3	Methods Page 8, Lines 10-29	However, the justification for including a minimum sample size of 50 per study intervention group is less than clear. It was created "in the absence of published information...to inform a power calculation, we derived the minimum sample size from expert opinion... "Although we could articulate a rationale for setting the minimum number of study participants in excess of one hundred, we decided to choose a number lower than the lowest possible number that could yield statistically significant results...This rationale and selection criteria was reviewed...." Even with this explanation, the rationale remains unclear. The text seems to argue that the authors made a deliberate decision to include inadequately powered studies---but the basis for that decision is not clear.	We have clarified our reasons with the following text: "Setting a minimum sample size based on quantitative power calculations would have required studies in excess of 100 individuals per arm. We set the sample size liberally at 50 in order to allow for the possibility that we might be able to combine studies quantitatively in a meta-analysis."

Commentator & Affiliation	Section	Comment	Response
TEP #3	Methods Page 3, Line 54; Page 4, Line 12	The paper successfully and clearly describes how to best assess and measure for a true reduction in one or more disparities	Thank you.
TEP #3	Methods ES-5, Line 26 and other	Defines the referent group as "...comprised of individuals for whom the disparity indicator was absent." I find the term "absent" somewhat unclear, particularly when I interpret the term "indicator" as the process or outcome measure of interest. Perhaps I am misinterpreting the term "indicator"? Would it be more clear to say that the referent group had a more positive process or outcome measure compared to the disadvantaged population? Or that it had the most positive indicator compared to the other populations? Or something similar?	This has been edited to be clearer.
TEP #3	Methods Page 5, Lines 16-17	"Applicability to the Evidence-based Practice Center (EPC) program/amenable to systematic review" This consideration is unclear. Should they be two separate considerations?	This text was provided by AHRQ and is standard.
TEP #3	Methods Page 13, Line 5	"...we assessed the number of studies and range of study designs for a given intervention-outcome pair, and downgraded..." Should this instead state "...we assessed the number of studies and range of study designs for specific disease condition and disadvantaged population outcome pairs..."?	Thank you for the suggestion; however, the statement is correct as written. No changes were made to this statement.
TEP #4	Methods	INCLUSION/EXCLUSION CRITERIA: Exclusion criteria are largely implied, except where explicitly laid out in Appendix M. Table 2 is a very helpful distillation of the inclusion criteria as discussed in the Analytic Framework section (Page 6) and the Study Selection section (pages 8-9). The criteria seem reasonable given the last paragraph of the Topic Refinement section (Page 6).	Thank you.
TEP #4	Methods	SEARCH STRATEGIES: The database search process (Appendices B and C) and two phase screening of studies (Appendices D and E) were well-articulated and well-designed for guiding identification of possible QI articles for inclusion in the assessment.	Thank you.

Commentator & Affiliation	Section	Comment	Response
TEP #4	Methods	DEFINITIONS/DIAGNOSTIC CRITERIA for OUTCOME MEASURES: They are appropriate as outlined in Table 2 and consistent with those found the Quality Assessment Tools (Appendices F-I)and in the Evidence Tables (Appendix J) and with Appendix K (Quality of Individual Studies) and demonstrated in the textual review of the various articles.	Thank you.
TEP #4	Methods	STATISTICAL METHODS: They seem reasonable in terms of tests applied in the studies themselves as well as by the report team in those instances where the team had to calculate statistical significance.	Thank you.
TEP #4	Methods	With the criteria used, I don't believe that any relevant literature was excluded. However, you might discuss how elimination of any of the exclusion criteria in Appendix M, might have altered results. Would any relaxation of the exclusion criteria have added to understanding of QI interventions involving disparities populations?	Because we did not extract data from studies that were excluded, we cannot comment on this.
Peer Reviewer #3	Methods	There are a few concerns about the use of inconsistent racial categories (e.g., African American vs. black). Please refer to OMB's directive 15 and subsequent revisions for appropriate categories http://www.census.gov/population/www/socdemo/race/Ombdir15.html	Yes, we agree that the use of racial categories across the literature is inconsistent; however, we tried to reflect the categorizations used by the study authors so as not to infer incorrectly. We have added a statement to the Methods chapter explaining that we use the race and ethnicity categorizations as used by the authors of the primary literature
TEP #4	Methods	The methods are logically outlines. The strategies are clear as are the inclusion and exclusion criteria. The statistical methods are appropriate.	Thank you.
TEP #5	Methods	Seems fine. Assume that papers failed to meet criteria for ESRD and CF? But the methods appeared straightforward.	No papers discussing end stage renal disease or cystic fibrosis were eligible for inclusion.
TEP #1	Results	Good.	Thank you.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #1	Results	The results section is very clear, and the tables are all readily interpretable. Please see my comments above on whether additional studies may have been relevant to include.	Thank you. We have addressed the comment above.
TEP #2	Results	Basic intro: the list of conditions isn't exactly the same as earlier or Executive Summary; you probably should combine hypertension and CAD into heart disease, if you are going to combine them in other places. (See re Table 3 below, too.)	We have addressed this throughout by using an alphabetical order. Hypertension and coronary artery disease are now grouped under cardiovascular disease.
TEP #2	Results	Search Results...No need to use within; in will do.	We have corrected as suggested.
TEP #2	Results	Use articles instead of papers	We have elected to use "papers" throughout.
TEP #2	Results	Keep the same order for conditions, don't list depression first and combine HTP and CAD (and figure out what order the two conditions go in and be consistent).	We have addressed this throughout by using an alphabetical order. Hypertension and coronary artery disease are now grouped under cardiovascular disease.
TEP #2	Results	Note your language about referent group, but what happened about the third comparison group? I suspect there is some confusion way back in methods – see query #4 above.	We have revised the text and substantially simplified the language.
TEP #2	Results	Figure 2: identify everything as articles – these are not studies.	We have corrected the instances (here and elsewhere) where we indicated "studies" instead of "papers". As noted above, we have opted to use "papers" instead of "articles".
TEP #2	Results Page 16, Lines 54-55 and continuing to next page	Overview. Please see some comments in the Executive Summary. I don't think the material is as accurate as it could be.	We have made revisions as noted above.
TEP #2	Results	Disease-specific Materials – only general comments because the results are simply too difficult and poorly written to try to wade through.	All of the text has been edited.

Commentator & Affiliation	Section	Comment	Response
TEP #2	Results	Key Points. I expected to see strength of evidence grades here. In addition, this is supposed to be findings relevant to the key questions, so in my view a bullet about the number of studies is not appropriate; however, if all the reports in the series are doing that, then okay. All in all: think carefully about what should be in the key points section and bullets. This is likely to be all that most users will read, and if you are missing specific findings here, they will never find them.	We have added strength of evidence tables to an appendix. The Results chapter does not include a "Key Points" section. Consistent with the recommendations in the CER Content Guidance, presentation of results as "Key Points" and: Detailed Synthesis" is required in CERs but optional for other types of reports.
TEP #2	Results	I do not believe you need "overview of the literature" as a separate heading, and you must check whether all this is redundant with what might be above (in overview) or in detailed analysis. If you want to keep this paragraph, ensuring you have not already said this anyplace else, then consider simply moving the "detailed analysis" heading above it.	We have reorganized and edited the report.
TEP #2	Results	Detailed analysis. I read these as best I could, but basically in my view all this needs to be cut by at least 50 percent, maybe more. It is impenetrable as written, and people will have to skip over it. Virtually everything in text in the "detailed analysis" should be dropped if it can be found in the relevant tables, and only material that "synthesizes" the findings across studies should be retained.	We have substantially edited the report, included the detailed analyses.
TEP #2	Results	In addition, the headings are completely confusing for readers to follow (i.e., understand the structure across all the parts and subparts) because they are not sufficiently "different" to give any guidance, but I suspect that is an artifact of guidance from AHRQ and you may not be able to do anything about it. Cutting back heavily on text may help.	We have substantially edited and reorganized the report. The headings have been updated accordingly.
TEP #2	Results	In at least one section (e.g., hypertension), you have a completely unexpected subsection of "critically important outcomes." For a section with apparently only two studies, that did seem to be a surprise, given that you offer no hint about any of that in the "key points."	We have reorganized the report and this should no longer be an issue.

Commentator & Affiliation	Section	Comment	Response
TEP #2	Results	In depression, why do you have three tables (I see no reason for not combining that material)? You have departed from the logic of previous sections (e.g., no "detailed analysis" heading), and instead you seem to treat the studies as the detailed analysis sections. This is just confusing to readers, but in any case, much of this is in the tables and these subsections should be drastically cut back. This is another example of having unexpected headings (critically important outcomes, but now also surrogate clinical outcomes) that cannot be tracked easily by readers.	This is not the case in the reorganization.
TEP #2	Results	This chapter is too long by perhaps at least 50%. Nothing documented in tables should be repeated in text, as it is now; much of the chapter is unreadable and likely to be skipped, meaning that important findings may be missed.	The report has been edited.
TEP #2	Results	I did not try to figure out whether any studies were missing or, conversely, should have been excluded; I think the authors have justified their inclusion/exclusion decisions adequately.	We appreciate your comments.
TEP #2	Results	The tables need a good bit of work, but at least they are easier to understand than the text. Some may need some revisions to make them consistent across the entire report (and, possibly, conforming with 508 compliance problems).	Each table has been reviewed and edited to conform to the guidelines in section 508. Furthermore, several tables have been reorganized to reflect a consistent order and organization for the individual studies, clinical conditions, and disparities.
Peer Reviewer #2	Results	Yes, a great amount of detail is contained in the results section. In fact, many of the tables in the results seem redundant with table included again in the discussion section. In addition, the discussion of the included studies for depression is in much greater detail than the other disease states – would condense to make them uniform. For example, would consider having one table with all studies as done in the other sections.	The report has been edited. Thank you.
Peer Reviewer #2	Results	Yes [the characteristics of the studies are clearly described].	We appreciate your comments.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #2	Results	The authors seem to criticize the use of "surrogate" outcomes for the QI interventions instead of "critically important clinical outcomes". This downplays the importance of many of the patient centered outcomes and seems to reflect their biases on what types of outcomes are most important. Further, given the nature of these types of interventions (multi-faceted, time intensive) and resources involved, to show a significant difference in the "important" outcomes they allude to such as mortality would typically be time and cost prohibitive.	We have moved away from this distinction between "surrogate" and "critically important" outcomes. We have edited the report throughout to use more accurate language.
Peer Reviewer #2	Results	Figures, tables and appendices are more than adequate.	We appreciate your comments.
Peer Reviewer #2	Results Page 69	In Table 12 it may be interesting to group findings under the heading of provider, patient and system level interventions. Since the IOM and others break down the causes of disparities into these potential categories, understanding at what level the QI interventions studies target would be useful. Further, breaking down the discussion of results by the target level for the QI intervention (patient, provider, system or combination) would be useful.	This is an interesting idea but we have elected to maintain the current organization.
Peer Reviewer #2	Results	Investigators did not overlook any studies that ought to have been included or include studies that ought to have been excluded.	We appreciate your comments.
TEP #3	Results	The amount of detail presented is appropriate and helpful. The studies are clearly described and the key messages clearly presented in word and visual presentation. Figures and tables contain adequate and useful detail.	Thank you.

Commentator & Affiliation	Section	Comment	Response
TEP #3	Results Page 32, Lines 50-55	Current evidence-based guidelines for the treatment of depression recommend either pharmacotherapy or psychotherapy as the first line of treatment for mild - moderate depression. The decision should be made primarily by the patient and clinician as they weigh potential side-effects, preferences and cost. The guidelines also recommend pharmacotherapy combined with psychotherapy as the first line of treatment for major depression (http://www.guideline.gov/content.aspx?id=9632). The description of this study seems to indicate that these evidence based guidelines were not utilized (because pharmacotherapy was emphasized as the first line of treatment via a predetermined algorithm and only offering psychotherapy as a second line of treatment). This deviation from evidence based practice may have potential implications for how the findings should be interpreted.	As noted in the description of this study, the decision about which treatment line to pursue was made by the clinician and the patient, which is an important element of evidence-based depression care, as the reviewer notes.
TEP #3	Results Page 32-33	I may have missed it, but it seems the description of the PROSPECT study does not include the age inclusion/exclusion criteria.	The age-related selection criterion has been added.
TEP #3	Results Page 34, Lines 13-14	The description of the IMPACT study seems potentially inaccurate when referencing the effort to analyze disparities between older adults of different incomes. Instead of "older low-income adults" should it say "among older adults of varied income levels"?	The language in the sentence you refer to has been adjusted to reflect the income designations of the authors in the IMPACT publication (Arean 2007).
TEP #4	Results	DETAIL PRESENTED/STUDY CHARACTERISTICS/FIGURES, TABLES AND APPENDICES: Appropriate detail is presented in terms of types of included study designs and quality, disparities populations, outcomes, and QI interventions. Tables 3-11 characterize the studies well. Appendices as noted previously are very helpful.	Thank you.

Commentator & Affiliation	Section	Comment	Response
TEP #4	Results	KEY MESSAGES: The Overview of Included Studies was a concise and helpful introduction to the section. Organization of results across and within health conditions was consistent and generally easy to follow with each health condition introduced by a Key Points presentation and Overview of the Literature, followed by a Detailed Analysis section, the appropriate summary table, and then a discussion of Outcomes (Critically Important, Surrogate Clinical, and Process, as applicable to the specific studies).	Thank you for your comments.
TEP #4	Results	OVERLOOKED OR EXCLUDED STUDIES: The review appears complete and thorough under the criteria set for inclusion and exclusion of studies. As noted elsewhere, I have some concern that comparing the same intervention in a health disparity and non-health disparity population may mask the degree to which an intervention actually closes the disparities gap between the two populations. There are, as noted in your examination of the literature, numerous studies that focus solely on one population and apply a tailored intervention that reflects the population and can show improvement between Time 1 and subsequent assessments of the intervention. I can think of studies utilizing black churches and barber shops to promote exercise, cancer screening, blood pressure checks, diabetes monitoring, etc. Collaborative interventions between these community entities and health care facilities and providers seem effective because they are tailored to the culture. Attempting to utilize these interventions in the white population would not be culturally appropriate. Outcomes derived from such comparisons would not be meaningful. Excluding a study from consideration as a QI intervention because the intervention was not applied to both populations, on the other hand, prevents capture of potential reductions of health disparities when examining the intervention against the usual treatment in the white population.	Actually, it is possible that a study targeting one disparity population could have been included if they also provided referent data for a non disparity population, and if they used a comparison group within the disparity population. Therefore, studies such as those you describe could have been included if they included a comparison group (e.g., intervention A in some black churches and no intervention or intervention B in others), and provided some external referent data in the nondisparity population (whites in this example) to show the degree to which outcomes in the disparity population were improved with the intervention AND the degree to which those outcomes approached outcomes in the referent group. No studies met these criteria.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #3	Results Page 22, par. 5	Inappropriate use of percent change, even though you refer to it as "an absolute improvement." The correct formula for percent change = (New Value-Old Value)/(Old Value) *100. An example is you cite a 6.9 percent absolute improvement when in fact it is 8.2 percent. There is a great article from the Planck Institute on statistical literacy at http://library.mpib-berlin.mpg.de/ft/gg/GG_Helping_2008.pdf	You are correct. We made an error and appreciate your attention to detail.
TEP #4	Results	The authors exhaustively reviewed the results with appropriate and detailed descriptions of the included studies. The key messages are explicitly stated. Both the included and excluded studies are justified.	Thank you.
TEP #5	Results	I found the description of the results in the executive summary difficult to follow. I found the description clearer in the actual results section. There were some papers that I am still not sure of from the description. The paper on gender disparities in depression...I think the findings are that women were receiving better care at baseline, and that by the end of the study, men's care had improved more so that care for both groups was better at two years and equal? This was not consistently indicated in the results. The several studies that showed greater improvement in the disparate group, with improvement making a statistical gain for the disparities group and not the majority group, but there not being a positive interaction term were quite confusing...not sure what inference to draw from this. Also I thought the authors were inconsistent in their approach to p values. For one study, p=.06 was dismissed as not significant, but in another study p=.08 was noted as highly suggestive. The paper that appeared to have different findings of an intervention for Black and Latino patients seemed to combine the two groups and indicate no overall effect. Seems to me that if the groups are different, they can't be combined in a single minority group (can't look for a main effect if there are differences in the subgroups).	We have edited the report to achieve consistency. The reporting of changes in effect even in the absence of statistical significance is intended to provide information on approaches that may be promising and therefore useful to readers, but we feel it is important to point out when different effects by group do not reach statistical significance, as would be indicated by a significant interaction.

Commentator & Affiliation	Section	Comment	Response
TEP #1	Discussion/ Conclusion	The conclusion states the following: Despite positive results seen in specific studies on specific clinical outcomes in some or all study populations, the strength of the evidence for QI interventions reviewed n this report to affect disparities is insufficient. The choice of the word AFFECT vs. REDUCE seems weak to me.	The conclusion now uses the word “reduce.”
Peer Reviewer #1	Discussion/ Conclusion	I would like to have seen more insight from the authors regarding the interpretation of their findings. Essentially, the findings are null in terms of interventions to address disparities. That is an entirely reasonable finding. However, what the field needs is an understanding of "what's next" --- there must be some information in those 10 or 15 studies they reviewed carefully to suggest what might represent next steps, etc. As it is, the report does not point the field in specific directions, and I think this is a missed opportunity given the amount of work the authors have accomplished here.	It is difficult to be more specific about what the next steps should be given the overall need for research in this area. We have attempted to provide this information more explicitly in the discussion, including in future research, applicability and the conclusion.
TEP #2	Discussion/ Conclusion	Look again at how you are characterizing “disparities” to be sure the concept and term refers to health, health outcomes, and /or health care. I wonder (Page 40, Line 30) about saying that people “suffer from” disparities (rather than, say, experience them). The disparity targets are not the characteristics of the patients (such as race or age); they are “unequal” (all other things taken into account) process or outcome measures.	We have edited the report and checked our use of these terms.
TEP #2	Discussion/ Conclusion	The disparity-target tables tend to be a little clearer about this, but you might do better to change the titles to, for instance: Table 17, Summary of included studies by clinical condition, QI intervention, and effect on disparity, by race or ethnicity of patients, so that nobody misinterprets “by disparity—race/ethnicity” as meaning that that is the disparity. As with Table 12, consider how you are organizing the table and listing the QI interventions.	Table titles have been revised to “Summary of effects on disparities in health outcomes associated with xx”
TEP #2	Discussion/ Conclusion Page 48, Line 37	Should it be Table 17, not Table 19?	You are correct. That was an error. All tables have been renumbered and reordered.

Commentator & Affiliation	Section	Comment	Response
TEP #2	Discussion/ Conclusion	Go through the entire discussion chapter and rewrite any sentence longer than about 25 words by splitting into two or more sentences. For example, page 40, lines 22-26 are probably about 60 words and cannot be easily tracked as one sentence.	Thank you. This is a helpful reminder. We have split long sentences into two or more sentences where appropriate.
TEP #2	Discussion/ Conclusion	What is a "countercurrent?"	We have deleted the sentence that included the word "countercurrent".
TEP #2	Discussion/ Conclusion	Go through this chapter and split very long paragraphs. For, example, page 40, lines 29-43, would benefit from having a new paragraph at "Thus,...". The aim is to make this easier for readers to get through.	Thank you. This has been addressed throughout the report text.
TEP #2	Discussion/ Conclusion: Page 40, Lines 43-50	Query: I wonder if "evidence ... suggests the benefit of QI..." is rather weak and dismissive. You didn't examine the huge literature about QI, of course, but still...	We have tried to ensure that we are specific about our conclusions as they relate only to reduction in disparities. We have further acknowledged the large QI literature that demonstrates benefit generally.
TEP #2	Discussion/ Conclusion Page 42, top lines	I was looking for the next sections to be those "suggestions" for additional research, but you apparently pick that up later. You might want simply to say you are reviewing findings by clinical target and disparities target and let it go at that.	We have revised the sections to include summary statements on the findings and suggestions for additional research.
TEP #2	Discussion/ Conclusion	As with results, please do not repeat in text anything that is in the various tables. The aim is synthesis and discussion, not yet another recap of material study by study. Between the text and tables you have too much duplication.	The report has been substantially edited.
TEP #2	Discussion/ Conclusion Page 54	Retain the paragraph that begins "in sum" or perhaps move it to earlier in this whole section.	We have retained the paragraph, but it is now under a revised heading, "Effects on Disparities".
TEP #2	Discussion/ Conclusion	I recommend you delete "We identified ... that" everywhere	This has been done in most instances.

Commentator & Affiliation	Section	Comment	Response
TEP #2	Discussion/ Conclusion	I expected to see "Research Gaps" as a main heading, with gaps in literature and limitations (??of the review??) as subsection leading up to something like Future Research. This is all choppy and the gaps, limitations, and future research sections don't quite sum to (or lead to) a clear analysis of what research to do next.	We have retained the section organization, but have edited the text under "Gaps in the Literature" and "Future Research" to improve readability.
TEP #2	Discussion/ Conclusion	Much of the discussion warrants editing by a professional – quite a lot of basic mistakes and a good deal of wordiness.	The report has been edited.
TEP #2	Discussion/ Conclusion	Conclusions: I'd rewrite to say something along the following lines: The literature on QI interventions generally and on their ability to improve health and health care is very large. These interventions are not, however, clearly effective in closing gaps between two populations characterized by various sociodemographic or other variables. Readers should not construe this report as assessing the general effectiveness of QI in health care settings. Rather, the message is that the QI efforts tested to date have not been shown specifically to reduce known disparities in health care or outcomes. In a few instances, disadvantaged populations may have experienced some positive effects. QI experts (? Investigators) should replicate these programs and extend their studies to include other interventions with potential for addressing these disparities.	We have rewritten the conclusions, taking into account your suggestions.
TEP #2	Discussion/ Conclusion	I wasn't entirely sure what distinction you were trying to make between replication and study further, and I wondered whether you ought to pick some to identify specifically the QI interventions you regard as "most promising."	We have noted those we found to be most promising.
TEP #2	Discussion/ Conclusion	The discussion seemed to repeat too much of what was in the results chapter.	We have edited the discussion to include fewer results
TEP #2	Discussion/ Conclusion	I am unclear on what overall guidance, for the ENTIRE SERIES, has been provided, so I had questions about what was presented for strength of evidence.	The strength of evidence tables are included as an appendix and we included summary text of strength of evidence in the section "Review of Main Findings" in the Discussion chapter.

Source: http://effectivehealthcare.ahrq.gov/index.cfm/search-for-guides-reviews-and-reports/?productid=1242&pageaction=displayproduct
Published Online: August 27, 2012

Commentator & Affiliation	Section	Comment	Response
TEP #2	Discussion/ Conclusion	The discussion did not really have any (other) literature, so it did not exactly put the findings into a broader context. For quality improvement, however, that could be difficult.	Agreed.
TEP #2	Discussion/ Conclusion	The heading structure is so strange that limitations, gaps, and future research do not flow well, and I did not come away with a clear view of what research actually might best be done and, as important, what questions do not warrant any further examination because the QI interventions are either "proven" (which is pretty much not the case) or are shown to be so ineffective for the "disparities reduction" purpose as to not offer much useful research opportunities relative to the resources available.	We have reorganized the report and hope it is clearer.
Peer Reviewer #2	Discussion/ Conclusion	The discussion is very well written. It clearly outlines why population based QI interventions may not improve disparities among certain groups as everyone in the population is likely to benefit (rising tide raises all boats). In fact, the most successful QI interventions in this report appear to be those that target interventions specifically aimed at the disadvantaged group (low health literacy, language concordance). In other words, it appears to suggest that wide ranging QI interventions are likely to be less successful than those targeting specific causes of the disparities. In general, this review highlights the importance of needing more research/study to understand what the causes of the disparities in care are in order to know where to target interventions. Taking a blanket approach of QI for all is unlikely to eliminate specific disparities.	Thank you.
Peer Reviewer #2	Discussion/ Conclusion	Yes, the discussion section very nicely summarizing the limitations and obstacles of this type of review. The tone of this discussion seems more appropriate as it implies the evidence is not available to draw strong conclusions as to whether QI interventions truly reduce disparities. A similar message should be convened in the conclusions of each subgroup section in the results portion of the document.	Thank you. We have noted your suggestion as we have edited the report.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #2	Discussion/ Conclusion	The biggest limitation in this review is the small number of studies ultimately included for analyses. Although the authors admit this is a narrative/descriptive piece and not a quantitative comparison of the effectiveness of QI interventions, they still conclude in several areas of the document that QI interventions do not reduce disparities. This is misleading as these conclusions are based on very limited literature. Instead, the message should be that some evidence is supportive, some is not. No strong conclusions on QI can be drawn at this time and more research needs to be done.	We have tried to make the point that no strong conclusions can be drawn throughout. We do note, however, that we did not exclude studies that could have answered the key questions – the small number of studies is less a reflection of our criteria than an indication that there is very little literature available on this topic.
Peer Reviewer #2	Discussion/ Conclusion	Further, given the overall small number of studies, breaking them down further into subgroup discussions is risky. No conclusions can be drawn from 3 studies among diabetic patients or 2 studies among patient with heart disease that used different populations and outcomes. Conclusions in these sections needs to be reflect the extremely limited data in these areas.	We offer the results in two ways- first by type of disparity and then by condition – and in editing the report have attempted to be clear about the limitations.
Peer Reviewer #2	Discussion/ Conclusion	The investigators did not omit any important literature	Thank you for providing comments.
Peer Reviewer #2	Discussion/ Conclusion	The future research section is clear and easily translated into new research.	Thank you for your comment.
TEP #3	Discussion/ Conclusion	Overall, the "review of main findings" on Page 40 is strong.	Thank you.
TEP #3	Discussion/ Conclusion	The paper highlights the possibility of a QI intervention resulting in improved care for all groups or segments of the population (rising tide lifts all boats) such that all disparities remain or are worsened. It is also theoretically possible that a QI intervention will worsen quality of care and/or outcomes for a disadvantaged population (or perhaps the advantaged population). It may, or may not, be worth discussing these potential outcomes as well.	Thank you – we have added a brief note that is also a possible outcome.
TEP #3	Discussion/ Conclusion	The limitations are described adequately.	Thank you for providing comments.
TEP #3	Discussion/ Conclusion	I am not aware of any omitted literature.	Thank you for providing comments.

Commentator & Affiliation	Section	Comment	Response
TEP #3	Discussion/ Conclusion	The future research sections are clear. However, it seems that most investigators do not conduct studies to assess the potential of an intervention to reduce disparities (as rigorously defined in this paper) because the financial costs of such studies are too high. As noted, they are methodologically complex and time-consuming. I don't know a way around this, but it seems that researchers do not have the necessary incentives to conduct such studies. Can the paper address this more explicitly? Perhaps by suggesting that funding organizations (government, for-profit, non-profit) emphasize support of the necessary research methodologies? There may be other ways to encourage the recommended types of research studies.	We hope that the portion of the report on research gaps will provide information for funders on important research that is not being done, likely in part due to lack of funding for conducting these complex studies. Thank you for raising this issue.
TEP #4	Discussion/ Conclusion	Before getting to the responses requested, let me raise some questions. You say on Page 40, starting at line15, "Given the potential for quality improvement (QI) strategies to improve the quality of care across the population, interest has developed in whether they might be used to reduce specific disparities, potentially by having an amplified effect among groups affected by disparities." My understanding of QI, from my former work in education leadership, is that QI is applied within organizations in ways that are distinctly responsive to each individual organization's characteristics, cultures, and internal and external environments. If I am correct in this understanding, doesn't the idea that a QI intervention in one disparities setting with one disparities sub-population could "improve the quality of care across the population" expect more from QI efforts than is intended? If you then attempt to administer and compare the effect of a QI intervention in matched disparities and non-disparities settings, doesn't the expectation become even less realistic?	The original sentence over-reached and has been made more specific.

Commentator & Affiliation	Section	Comment	Response
TEP #4	Discussion/ Conclusion	A second question: Since QI interventions are long-term in their implementation, do any of the studies measure or assess organizational adherence to the intervention over time and document its evolution over that time, or do they only measure the sub-population response overtime without adjusting for changes in the intervention or the degree to which it is persistently and consistently applied overtime? If the latter, how does one know how any changes in the intervention affect the measured outcomes?	This information was not provided in the papers, although it is clearly important. We note the need for capturing organizational adherence in the future research needs portion of the report.
TEP #4	Discussion/ Conclusion	You use the terms “consistent”/”inconsistent” and “precise”/”imprecise,” but nowhere do I find a clear explanation of what they actually characterize. This should be discussed in the text as well as in the appropriate appendix.	We have added definitions for strength of evidence to Appendix J.
TEP #4	Discussion/ Conclusion	A third question to ask and posit answers to: Why are so few QI studies appropriate for inclusion in this review? You have set certain criteria for inclusion in your review, why do few studies incorporate these? The report hints at reasons, but doesn’t make them as explicit as they might be. I think they should be made boldly.	As noted in the report, we sought only those studies that were designed in a way that they could measure change in disparity. The QI literature has not focused specifically on disparities.
TEP #4	Discussion/ Conclusion	A fourth question: In the introduction to this section (pages 40-41); you have a table of the quality improvement strategies used in the studies. Why do you not address the implications of most studies using one or more of three interventions--provider education, patient education, and promotion of self- management)? Is this because they are more effective per the literature, less costly, quicker to implement? What do you make from their pervasive presence and what recommendations do they suggest to you for future work in this area?	Assessment of individual components of the interventions was not possible given the available literature and a separate assessment of each is out of the scope of this review. Therefore, we cannot comment on this issue although we agree that the reasons behind the selection of components in a multi-component intervention are interesting and worth considering.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #3	Discussion/ Conclusion	Quality improvement measures must include the clinician decision making process as well. There is a lot of information regarding patient reminders, etc, and the responsibility is on the patient, absolutely. However, if clinicians are making decisions and suggestions differently based on a population, this is a HUGE contributor to disparities and not mentioned here. I believe that bias and disparity in treatment plays more of a role than we like to believe.	Thank you for your comment. It is difficult to assess this issue in the literature, but the concern is noted.
TEP #4	Discussion/ Conclusion	The implications of the major findings are clearly stated. The bottom line is that there is a dearth of available evidence that demonstrates the value of quality improvement initiatives to address health care disparities. The adage that a rising tide lifts all boats does not seem to apply to QI initiatives when tied to healthcare disparities.	Thank you for reviewing the report.
TEP #5	Discussion/ Conclusion	This section was clearly long. I found the tables on strength of evidence confusing as this isn't something I have previously reviewed/seen. In general, I think the findings were more encouraging for the potential for interventions than the authors concluded, and I think there appears to be some grounds for differentiating across conditions and intervention...again, I think interventions need to be matched to the proposed source of disparity and one can't expect a generic intervention (unless so generic that it simply says--assess source and fix it) to work in all settings. As always, more research, better research, more consistent measures, comparison populations, etc. But what this really cries out for is a theory and a strategy to address the theory. I also thought the authors were consistently too harsh on multi-component interventions. Any QI intervention worth its salt for something as complicated as a disparity has multiple components. I didn't understand the critique of such interventions (e.g., the NY State lipid testing intervention).	We agree and have removed the SOE tables. We have also reviewed our discussion to assess the way in which we judged the literature. We do not understand your comment about our judgment of multicomponent interventions – all of the interventions included multiple components and were judged on their scientific merit.

Commentator & Affiliation	Section	Comment	Response
TEP #1	Clarity and Usability	Generally good structure and organization. However, if the report is supposed to be clinically meaningful and therefore meaningful to clinicians, I recommend pulling out a subsection on clinical relevance of findings. Right now the findings, the challenges, including challenges with broad uptake, etc. are all together. It would be very useful to pull out a brief synthesis of the clinical meaningfulness with a subheading highlighting this.	The section on applicability is intended to be helpful in this way. We have expanded it by including information by clinical condition, which is likely of specific interest to clinicians.
Peer Reviewer #1	Clarity and Usability	The report is reasonably well organized. However, it could benefit from a better setting of context via their conceptual model up front, as well as interpretation of the findings.	As noted throughout, we have substantially revised and edited the report and believe that the organization and context are clearer.
TEP #2	Clarity and Usability	Glad to see [applicability] called out directly.	Thank you for your comments.
TEP #2	Clarity and Usability	In the disease-specific material, however, I think you need to focus only on the bottom line: what is specifically pertinent for clinical practice and what is missing. Repeating again what different studies did masks what you regard as the “conclusions” about applicability.	The report has been substantially reorganized and edited and we hope it is clearer.
TEP #2	Clarity and Usability Page 56, Lines 13-16 or so	In depression, these studies had a range of adult ages, both men and women and were racially diverse. By the time you combine all this, why is this considered representative only of a “small proportion” of individuals with depression. Who is missing, apart perhaps from children and adolescents, or persons with dual diagnoses? Practice setting is entirely different matter.	We have clarified in the text that even with this range of populations, the limited settings likely affected applicability, particularly in not including individuals with poor access to care.
TEP #2	Clarity and Usability	The report is structured fairly predictably. The headings within chapters are not well done -- they do not easily convey (by font or by size) how the sections and subsections fit together, so readers will get lost. Whether this is a consequence of guidance for all the "Quality" series reports or a peculiarity of AHRQ specifications, I don't know, but in that sense the structure is rather unpredictable and inconsistent.	The headings are formatted to AHRQ specifications, but we have reorganized the report and believe they are now clearer

Commentator & Affiliation	Section	Comment	Response
TEP #2	Clarity and Usability	The main points are not always clearly presented for two reasons: (1) many are lost in endless descriptions of studies (especially in results, but to some extent in discussion), and (2) the ways QI interventions are listed, conditions classified, and similar organizing principles presented vary across the report in ways that can be confusing to readers.	In reorganizing the report we have made the main points more prominent.
TEP #2	Clarity and Usability	Cutting much of the text back by 50% to 75%, getting the MS professional edited, and focusing on telling the story clearly will help. Working back from the "main messages" (conclusions) to understand what to emphasize and focus on might help.	We have edited the report.
TEP #2	Clarity and Usability	The conclusions are not precisely relevant for clinical practice decisions (I did not think that the report or the series has this as a goal); I think the basic conclusions can be applied in decisionmaking about starting or continuing QI interventions to help in reducing health and health care disparities, but only to a limited extent (because of the disappointing findings).	We agree.
Peer Reviewer #2	Clarity and Usability	Yes, the report is well structured and organized.	Thank you.
Peer Reviewer #2	Clarity and Usability	Yes, the main points are clearly presented.	Thank you.
Peer Reviewer #2	Clarity and Usability	Yes, the conclusions can be used to inform policy and/or practice decisions.	Thank you.
TEP #3	Clarity and Usability	The Analysis by Clinical Target and Analysis by Disparity Target sections might be easier to read if subheadings are added. Otherwise, the paper is very well organized and main points are clearly presented.	We have substantially reorganized the report.

Commentator & Affiliation	Section	Comment	Response
TEP #4	Clarity and Usability	CLARITY AND TRANSLATABILITY: The Future Research subsection is the most essential part of the entire Discussion section. I think under the limitations imposed on the review it is quite clear and offers some guidance to researchers pursuing QI interventions for disparities populations. Again, I would bold key message points and perhaps incorporate a diagram showing ideally what elements funders need to insist upon and researchers need to incorporate into their designs. Perhaps, give a couple of illustrations of those designs as applied to a research question.	Development of a more specific research agenda is important but not within the scope of this project. We hope that stakeholders in the field will take note of current gaps in the literature and move toward such an agenda.
TEP #4	Clarity and Usability	STRUCTURE AND ORGANIZATION: I have referred to this elsewhere. Within the boundaries set for the review, it is very well structured and consistently well organized. I think it would be helpful for the report to give a bit more explanation of terms and include more discussion of the exclusion criteria in the text. Too much is left to the appendices, and some terms that I have mentioned are not even well explicated there.	We have clarified the inclusion/exclusion criteria in the text and tables within the report, specifically, in the Methods section.
TEP #4	Clarity and Usability	IMPLICATIONS--CLARITY: The major or main findings section spends too much time rehashing the findings for each of the disease and disparity conditions. What you have is clearly written, but you would serve the reader better by speaking directly to the implications of those findings. Reference a table of findings and then say directly what the implications for future research are for similar findings and designs.	Thank you. We have shortened much of the text to make the main findings more clear.

Commentator & Affiliation	Section	Comment	Response
TEP #4	Clarity and Usability	Across all of the discussion segments (whether implications, limitations or future research), be bold about take home messages. Literally bold those sentences that characterize the discussion. Also be blunt in the assessments. On Page 54, beginning line 17, you sum up the studies, acknowledging some benefit, and deficiencies that seem to far outweigh the benefits. The take home message from this report for me is much more like: "At this stage, and with the available evidence, formal QI interventions do not appear sufficiently effective to be a first line option for improving health in health disparities populations, especially if the goal is to decrease disparity between these populations and non-disparity populations."	We are following the formatting guidelines of our funding agency.
TEP #4	Clarity and Usability	LIMITATIONS--ADEQUACY: I would place and discuss Gaps in the Literature in the Limitations section. In other respects this section is adequate.	We have organized the report per the required approach for EPC reports.
TEP #4	Clarity and Usability	The report is exceedingly well structured, organized, and well written. The conclusions can be readily used to inform policy and the allocation of resources to design RFAs and RFPs to address this area of investigation.	Thank you for your comments.
TEP #5	Clarity and Usability	Because the authors conclude that more research is needed and that we can't draw meaningful conclusions with confidence the implications for practice are small and for policy (except for funding decisions about research) are minimal. As noted above, I believe the conclusions for promising interventions can be stated a little more positively and that the call for research and innovation can be shaped along the lines noted above.	Thank you for your comments.

Commentator & Affiliation	Section	Comment	Response
TEP #1	General	*These comments refer the abstract and would also apply to subsequent sections of the report where the issues may be relevant.* Objective: "This review evaluates the effectiveness of QI strategies in reducing disparities in health outcomes." I know we discussed how we were defining "health outcomes" during one of the TEP meetings. Many people think of processes and outcomes as very distinct. Clinicians, in particular, are worried about being evaluated on health outcomes (rather than processes) because many health outcomes are outside the parameters of their control. So I worry that in the abstract, we only refer to reducing disparities in health outcomes. I fully recognize that the report later describes health outcomes to also include processes (and other elements too) but this is the first sentence readers will see and I worry that it sends an inaccurate message, requiring readers to delve further into the report before realizing that the definition adopted is more encompassing than what it appears at face value.	We have revised the objective statement to: "This review evaluates the effectiveness of quality improvement (QI) strategies in reducing disparities in health and health care." We have made similar clarifications in other sections of the report.
TEP #1	General	As far as the clinical meaningfulness of the report goes, it may lose clinicians early on. Why don't we just say, in the objective, "This review evaluates the effectiveness of QI strategies in reducing disparities in health and health care?"	We have revised the objective statement to: "This review evaluates the effectiveness of quality improvement (QI) strategies in reducing disparities in health and health care."
TEP #1	General	Final point on this is a general one, when we say disparities in health outcomes, it implies that we are only examining health disparities rather than health care disparities, but the report is doing both so I would be in favor of the more general language.	We have clarified that we mean heath and health care.

Commentator & Affiliation	Section	Comment	Response
TEP #1	General	The other point I'd like to raise here is the ambiguity of the objective itself. It is unclear to me (and remained so throughout the report) whether the review was evaluating the effectiveness of GENERAL quality improvement strategies in reducing disparities or whether the report was evaluating the effectiveness of TARGETED quality improvement strategies to reduce disparities (with the target being the disparity itself). I know this point is discussed later in the report (e.g. pg ES-5 lines 7-32) but it needs to be underscored early. A looming practice and policy question is whether general QI can reduce disparities or whether we need QI targeted to reduce disparities to actually see a reduction. In reading the upfront material, the report seems to allude to a focus on GENERAL QI, which is not the case.	This is a good point. We included both literature that focused on reducing a disparity and general QI studies that included data to assess any reduction in disparity.
TEP #1	General	Also, when I saw that there were only 17 papers that met the inclusion criteria in the past 28 years, my first response was "really, that is it!?" However, recognizing that the report is including only those papers that focus on targeted QI to reduce disparities then makes more sense in terms of the limited number of papers that met the inclusion criteria.	Correct.
Peer Reviewer #1	General	I enjoyed reviewing this report, the findings will be of help in advancing the field of interventions to reduce health disparities	We appreciate your comments.
Peer Reviewer #1	General	Comments on the opening key questions: The title of the paper is on "health disparities", but it is not clear whether the work here is focused on "health disparities" or "health care disparities", or some mix of the two. This is an important distinction as interventions that affect the delivery of care may very distinct from those that affect broader issues affecting health (e.g. social issues, community issues, etc)	We included disparities in health and in health care and have noted that more clearly.
TEP #2	General	I found this review quite wordy, and I recommend “global” editing changes.	We appreciate your comments. We have made changes throughout to make the report more succinct.

Commentator & Affiliation	Section	Comment	Response
TEP #2	General	Do not start ANY sentence, anywhere in the report, with "it" when the "it" has no immediate referent. This means that all sentences beginning with language such as "it is not clear ..." must be rewritten.	Thank you. This is a good reminder. We have revised sentences that begin with "It…" and changed to active voice.
TEP #2	General	Similarly, do not start any sentence with "There is/there are...that." These can virtually always be rewritten to remove those three words and eliminate the awkwardness of the fact that "there" means, mainly, "not here."	These sentences have been corrected and/or revised throughout.
TEP #2	General	"...in order," in a phrase such as "in order to do something..." is nearly always unnecessary and can be deleted.	We have deleted most all instances of the phrase, "in order".
TEP #2	General	"...the purposes of" can be deleted everywhere.	Done.
TEP #2	General	In many places (chiefly results) the language "we identified.... that..." can be deleted as both unnecessary and repetitious (annoyingly so). Just give the facts of "NN studies did ..."	This has been done in most instances.
TEP #2	General	Acronyms. At some point, please ask a good editor to identify (a) all the places where an acronym (such as TEP) must be spelled out at its first use (in abstract, in executive summary, in each separate chapter of the report itself) and then (b) all the places where the acronym, having been correctly introduced, must be used thereafter. This is especially a problem for quality improvement, when much text, even in the same paragraph, toggles back and forth between QI and quality improvement. Except at first use, EVERY time thereafter should be QI in a given part (e.g., chapter) of the report.	The consistent use of acronyms has been addressed and revised accordingly.
TEP #2	General	Active voice. The manuscript is rife with passive voice. Please try to revise into active everywhere, but particularly for all material that involves actions by the authors in conducting the review. This will help with the wordiness, too.	We have revised the text to active voice, with particular attention to statements referring to the report authors conducting the review.

Commentator & Affiliation	Section	Comment	Response
TEP #2	General	Fix numbers in various places. Time (days, months, years), for instance, is always Arabic numbers unless it starts a sentence. Usually, all numbers 10 and higher are Arabic.	Thank you. We have fixed the numerals throughout the report.
TEP #2	General	Referring to people, populations, etc. I'd urge you NOT to refer to human beings (e.g., patients, people, and populations) by a descriptor such as black, white, nonwhite, or whatever. Those are solely characteristics of these individuals, and I think you should be referring to black patients, or white participants, or whatever. (Same is true for diseases – don't refer to people by their condition, although I don't think you do that in this review, but you might check	Thank you for this reminder. Throughout the report, we have revised the descriptions of participants to focus on the individuals and not on the characteristic or descriptor of the participant or participant group.
TEP #2	General	Say race or ethnicity (none of these studies is a combination, as best I can tell, but in any case a slash is always ambiguous).	We have revised throughout from "race/ethnicity" to "race or ethnicity".
TEP #2	General	Strength of evidence: Suggest you try always saying the SOE was low [insufficient] about X or Y, rather than "there was insufficient evidence," as the issue is really one of the SOE per se.	We have clarified all of our discussion of strength of evidence.
TEP #2	General	Check a variety of little things – ask an editor for help: Black and white should not be capitalized (as race), and many words should not be hyphenated.	We are using guidance provided by AHRQ on issues such as this.
TEP #2	General	You can delete "of the" in constructions such as "All [of the] studies" or "NN [of the] studies..."	We have corrected as suggested throughout.
TEP #2	General	Also: the medication adherence report in the series is not restricted to (nor does it cover) HIV/AIDS, so "in HIV/AIDS" should be deleted.	We have deleted references to "HIV/AIDS" as part of the medication adherence report topic within the series.
TEP #2	General	I was glad to see "compared with" (rather than compared to), but sometimes the construction should be that X was lower/higher in [patient group] THAN IN [other patient group], not compared with.	We have edited the report, including these instances.
TEP #2	General	Ask an editor to catch the spots when more/less [higher/lower; smaller/larger; etc.] should be followed with "than."	We have corrected throughout.

Commentator & Affiliation	Section	Comment	Response
TEP #2	General	You have a large number of errant commas floating around, yet you are missing some before and in a series.	We have corrected throughout.
TEP #2	General	In some places you say you are not doing something in the series (which is fine) but one of the specific things isn't right: RTI is doing "medication adherence" per se, across several chronic conditions, but it is not doing med adherence in HIV/AIDS, and my impression is that NOBODY is doing that because it was just done (or something like that). We're not doing some severe mental illness issues either (re med adherence), but it is a quite broad report. My suggestion: look for those spots and just delete "in HIV/AIDS" and leave med adherence unspecified. Just to say in your current report, excise "in HIV/AIDS" and let it go at that. It's really only in footnotes and at some point is a very easy correction.	We have deleted references to "HIV/AIDS" as part of the medication adherence report topic within the series.
Peer Reviewer #2	General	In general, this is a well written and detailed review of the current state of the evidence regarding quality improvement interventions and health disparities. The authors should be commended for undertaking such a thorough review in an incredibly important area. Overall, the report fairly summarizing the limited, good quality studies in the area.	Thank you for your comments.
Peer Reviewer #2	General	However, I am concerned about the strength of their conclusions in the executive summary stating that QI has not demonstrated to reduce known disparities. The Full Report more accurately details that given the number of difference disparity groups included (race/ethnicity, SES, insurance status, language barrier, health literacy, and gender), very small numbers of studies ultimately included for analysis, the variability allowed in the medical condition as well as the intervention, it is difficult to make any general conclusions about QI on disparities. As many readers of this report may only focus on the executive summary, I would suggest using some of this language in the summary. Further, the summary should refer to potential areas of promise for future study.	We stand by our conclusions but have attempted to ensure that promising data are noted. We also took care to ensure that a lack of evidence was not construed to suggest that there was evidence that the interventions were ineffective.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #2	General	Another overall concern relates to the focus on QI interventions to reduce disparities without a great deal of discussion regarding the lack of evidence in the current literature documenting the true causes of disparities. As the authors are well aware, the majority of disparities literature has focused on defining where disparities in health care exist with much less evidence for the potential causes. The authors do discuss proposed reasons of health care disparities including evidence from the IOM report. However, this should be expanded upon in the discussion. These issues should be expanded upon in the discussion.	This is a good point, but beyond the scope of this particular review. We have added a comment but cannot go into detail having not reviewed the literature completely on this topic.
Peer Reviewer #2	General	The reason for failure of many of these studies to reduce disparities may be that they are targeting the wrong causes.	Absolutely.
Peer Reviewer #2	General	Moreover, many of the studies included for review were not designed to specifically eliminate or reduce a disparity a priori, but instead subgroups were examined retrospectively. If the initial study was not powered to find a difference by race, age, gender etc. it is not surprising no differences were found.	This is correct and noted in our text.
Peer Reviewer #2	General	Overall, the focus on disparities is meaningful. More information about how priority conditions were chosen would be helpful. For example, were these the conditions in which disparities among some groups (i.e. age or race/ethnicity) have been shown to be the largest? The report later refers to the choice based on AHRQ priority areas. It'd be helpful to justify the choice of some of the conditions in a disparity focused report. For example, are large disparities seen in conditions such as cystic fibrosis?	The conditions were chosen through an iterative process of reviewing existing priority lists and working with our technical experts and library scientists to identify a set that provided some representation across the population (e.g. some that affected both adults and children), and were known to be associated with disparities.
Peer Reviewer #2	General	Target population and audience are explicitly defined.	Thank you for your comment.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #2	General Line 11-14, Page 10	Yes, to some extent [KQs are appropriate and explicitly stated]. The key questions focus on “health outcomes” yet the objective includes health care access, utilization and health outcomes. If the authors are considering access and utilization under the umbrella of health outcomes, then these should not be listed separately in the objectives	We have clarified the objective.
Peer Reviewer #2	General Page 9, Line 53	Drop the word “process”?	We have retained the word process, as we sought health outcome measures and process measures to answer key questions.
Peer Reviewer #2	General Page 61, Line 39	drop the “t” in front of The (under Critically Important Outcomes)	We have corrected this.
Peer Reviewer #2	General Page 68, Line 53	Period is missing after first sentence, last paragraph, between “interpret” and “heterogeneity”.	We have corrected this.
TEP #3	General	The report is meaningful and the key questions are appropriate and explicitly defined. The report is clinically meaningful from the perspective of assessing the potential promise of various interventions targeting specific disparity populations.	Thank you for your comments.
TEP #4	General	If the question is: "Do the report results suggest ways to change practice to bring about patient benefit and/or reduce health disparities between a defined health disparity population and a non-disparity population," the answer is “No.” However, if you mean, "Does the report give meaningful guidance to practitioners, policy makers, funders, researchers, etc. regarding constraints, resource investment, and uncertainty of effect," the answer is “Yes.”	Thank you for your comments.

Commentator & Affiliation	Section	Comment	Response
TEP #4	General	Overall, the report seems to say, "We can't tell you much about what to do, but we can point out the problems with what is out there. After reading and thinking about the report, I wonder whether the results and conclusions were not predictable at the outset. If one intervenes in two populations, under most circumstances, is it not likely that the effects will be similar (except in a few cases, such as offering interventions in different languages or accounting for different reading/health literacy levels, cultural characteristics)?	We note your concern but feel that the approach is appropriate, even if it primarily serves to note concerns about the current literature and suggestion directions for future research.
TEP #4	General	Also, these interventions seem to assume that many potential intervening variables are not all that relevant. For example, education level affects individuals and communities in ways more pervasive and long-lasting than can be addressed by any of the analyzed interventions. However, if one looks across generations and sees advances in education levels, one sees general advancement in SES, and improved health status.	These are good points, directed at the researchers of the primary studies rather than our report.
TEP #4	General Page 58, starting Line 23	Perhaps the most effective (only?) QI intervention that can truly address health disparities would be improved education in under-served/disparities communities that extends across and continues for generations. On the other hand, the report notes that many studies focus on interventions to change health behavior or health outcomes within disparities populations or settings, but don't attempt to compare the intervention to a non-disparity population	We note that our review focused specifically on QI interventions that derive from a health care system, so general education as a policy measure would not have been included.
TEP #4	General	It seems reasonable to me that changes within a single population from Time 1 to Time 2, to Time 3, etc. is quite important to those disparities populations and such interventions can bring about meaningful change in those specific contexts, regardless of education level. Still, absolute decrease in disparities maintained over time will likely be a result of educational attainment and differential interventions that are tailored to specific community, cultural, and linguistic characteristics.	See note above.

Commentator & Affiliation	Section	Comment	Response
TEP #4	General	In my work with tribes, it is apparent that bands of the same tribe, even living within 50 miles of each other, respond differently to interventions. In some instances, such differences may be apparent across the districts within the same tribal band. Some may adhere more closely to traditional ways, while others may be more comfortable with Euro-western practices. In some, social networks play a much stronger role than in others. These differences call for interventions that are responsive to very local circumstances.	Certainly to be effective any intervention should be appropriately designed for the given circumstance. Nonetheless, studies such as those included in our review should be able to provide some information on types of interventions that are more or less effective, along with applicability information for providers adapting the interventions to specific populations, as you note.
TEP #4	General	MAIN POINTS: Main points should be in bold print.	The formatting guidelines for these reports do not support the use of bold font to identify main points.
TEP #4	General	CONCLUSIONS: Main points should be in bold. As the conclusions stand, they are of limited value to policy makers or practitioners. I would emphasize more of what you have in the future research section, so that funders are left with what they need to incorporate in future RFAs.	See above.
TEP #4	General Page v, Line 52	"QI has not demonstrated to..." Do you mean, "QI has not been demonstrated to..." or "QI has not demonstrated that it...."?	Revised from “rather QI has not demonstrated to specifically reduce known disparities in healthcare” to “ rather, QI has not been shown specifically to reduce known disparities in health care or health outcomes
TEP #4	General Page 20, Line 11	You first use the acronym "HDZ" in Table 4, but don't fully explain it until Page 21, Line 30.	The description of the acronym for the health disparity zone (HDZ) is listed in the table notes.
TEP #4	General Page 21, Line 30	The explanation of HDZ seems ambiguous. Do you mean that within an HDZ, minorities have diabetes prevalence that is greater than for minorities overall? Or are you saying that minorities have diabetes prevalence that is greater than for the average in the U.S. general population?	We have clarified in the text.
TEP #4	General Page 21, Line 20	Insert "was" between "intervention" and "not shown."	We have revised from “, and the intervention not shown to be effective” to “, and the intervention was not shown to be effective”

Commentator & Affiliation	Section	Comment	Response
TEP #4	General Page 51, Line 14	You write, "was associated in increased..." Change to: "was associated with increased..."	We have revised from “language concordance was associated in increased cancer screening...” to “language concordance was associated with increased cancer screening…”
TEP #4	General M-1, Line 46	The last sentence is a fragment: “Data from external referent groups”	We have deleted the statement, “Data from external referent groups”
TEP #4	General	The report is clinically meaningful and clearly makes about the dearth of available evidence that quality improvement initiatives improve healthcare disparities or narrow the gap across a wide spectrum of target populations and disease processes. The target populations are clearly delineated. The key questions are appropriate and explicitly stated.	Thank you for your comments.
TEP #5	General	I found the report meaningful and useful. As with many systematic reviews, the limitations in the available evidence makes meaningful inference difficult, which is frustrating but not the fault of the writer. The key questions are well stated and the populations well defined.	Thank you.
TEP #5	General	The comments, especially in the executive summary about why papers that had no disparities at baseline were included was/is confusing.	We set our criteria a priori to include any studies that included data on two groups so that we would not miss those cases where a disparity did not exist at baseline but was created or in which the intervention seemed to have greater effectiveness in a particular subpopulation than another.

Commentator & Affiliation	Section	Comment	Response
TEP #5	General	My overarching concern with the report is the absence of a theoretical framework for the creation of/amelioration of disparities, and whether the cause/cure for disparities is the same across conditions, different disparate groups, etc. So, one might argue that some conditions (ESRD--must not have been studies meeting criteria) have a highly technical element, so that a technical QI intervention (e.g., standardizing dwell time) might work to reduce disparities; while a similarly technical, one size fits all approach wouldn't effectively address chronic care management. Such a framework--modeled perhaps on the one that Dr. McDonald wrote for the AHRQ coordination of care measurement atlas--would help in interpreting the studies here.	You're absolutely correct that different types of QI interventions would be most appropriate for different conditions. This is exactly why our definition of QI was broad and would have included a wide range of specific interventions had research on them been available.
TEP #2	Figures Table 3.	Consider whether you should combine CAD and HTN in this table.	Yes- we have elected combine CAD and HTN under the category "Cardiovascular disease".
TEP #2	Figures Table 3.	What is the logic for the order of articles (rows) within diseases? It's not alphabetic, not chronological, and not by a (presumed) hierarchy of study designs. You must have used some organizing principle.	We have organized by alphabetically by condition and then by author last name.
TEP #2	Figures Table 3.	See comments elsewhere as well about the order in which QI intervention chx are listed. "	We have ordered the interventions as described above.
TEP #2	Figures Table 3.	"Outcomes" is a really important column, and I'm wondering whether you want to be sure you have, someplace, noted that these are your disparity targets (if that's the case).	Table 3 has been replaced with a summary table that does not include specific outcomes.
TEP #2	Figures Table 12.	This is probably a useful table, but I do not understand the ordering of the rows. I suspect you meant to go by disease, and if so, consider creating a column (really, the table stub) to show that.	We have organized by alphabetically by condition and then by author last name.
TEP #2	Figures Page 40-41	The text is misleading in that it refers to cancer first, and it also calls hypertension and CAD "heart disease," so readers will get no guidance from the text about understanding the table.	We have organized by alphabetically by condition and then by author last name. We have elected combine CAD and HTN under the category "Cardiovascular disease".

Commentator & Affiliation	Section	Comment	Response
TEP #2	Figures	Please go back to earlier comments and tables, however, and rearrange the remaining columns in the principal order of the interventions – with the "most frequent" coming first (further to the left, such as patient education and provider education) and the "least frequent coming last (further to the right, such as organizational change (which, frankly, I'd combine with other – one whole column for a single study/example just wastes space).	We have ordered the interventions as described above.
TEP #2	Figures SOE Tables	Check to see what other reports in the series may be doing for strength of evidence; at this point I'm not sure whether everybody is putting the domain scores in tables. (If so, fine; if not, perhaps all the authors across the reports can come to some agreement about this.)	We have put the strength of evidence tables in an appendix.
TEP #2	Figures SOE Tables	If you keep, then you may need to explain whether risk of bias as part of the SOE scoring is the same score as you would have done for quality assessment of individual studies (as I recall, you do good/fair/poor);	The strength of evidence tables are in Appendix J. We have added some text about how strength of evidence was done.
TEP #2	Figures SOE Tables	I believe in addition to number of studies (and number of subjects) you should give the type of study (RCTs or observational all lumped together) and the quality grades. Nonetheless, consistency across all the reports in the series may be what people need to shoot for.	The strength of evidence tables are now in Appendix J. We have added the number of participants, number of studies, and the study design.
TEP #2	Figures	Check whether your approach will meet 508 compliance standards. It took me a while to figure out that your grayed rows are the outcome, and the items under it are the intervention (the heading does not quite imply that).	Each table has been reviewed and edited to conform to the guidelines in section 508.
TEP #2	Figures Page 42	Do make certain that what you say is the intervention is, in fact, one that you have identified way back in introduction or results as an intervention. For instance, I question that "cultural competency training" is an intervention (I didn't remember it; sure doesn't exactly sound like QI per se).	Cultural competency training has been categorized under provider education throughout the report table. In the strength of evidence tables we have revised from "cultural competence training to "Provider education (cultural competency)". In the text, where we describe individual study interventions, the strategy is still described as cultural competency training.

Commentator & Affiliation	Section	Comment	Response
TEP #2	Figures	Get an editor to copy edit these tables once you have them revised.	We have taken extra care to review and correct table formatting, spelling, and punctuation.
TEP #2	Figures	Keep punctuation after “insufficient” the same throughout, and decide whether to capitalize the first word after it or not.	Corrected. Capitalized “Insufficient” throughout. The tables have been revised and do not include used a colon and lower case for subsequent word. The SOE tables are now in an Appendix. .
Peer Reviewer #3	Figures	Some of the tables have a black background and are hard to read.	This appears to be a technical artifact due to conversion of the document to PDF. We will be checking the final submission for this or other similar errors.
TEP #2	References	Check references (see comment earlier) They need to be run against your EndNote database separately from the full report	This has been done.