

Project Name: Health Risk Appraisal
 Project ID: RSKA0410

Table 1: Invited Peer Reviewer Comments

Reviewer ¹	Section ²	Reviewer Comments	Author Response ³
1	General	A lot of good work went into this review and the analysis is certainly careful and comprehensive. However, the focus is primarily directed at the quality of research (which is not up to clinical trial standards) rather than on a global interpretation of a large body of evidence.	We address the quality of the extracted as per standard systematic review methodology. In our responses to the key questions (results and discussion), we provide a summary or global interpretation of the evidence.
1	General	I am that the overall tone of the review is negative and a great deal of skepticism is voiced related to the likelihood that "HRA Programs" can achieve demonstrable and long-lasting benefits for individuals insured by Medicare. We do point out that many studies did not conduct long-term follow-ups of their participants. This fact makes the assessment of long-term HRA benefits difficult to conduct.	Our review is not negative in this sense. Rather, we conclude that the evidence is too heterogeneous to answer key question # 2. Further, we conclude that the results of HRA studies done in persons under age 65 years cannot readily be generalized to persons aged 65 years or over, which is the only segment of the Medicare population that was included in any of the extracted articles (key question # 3).
1	General	I believe this review does not hit the mark for a number of reasons stated below: The reviewers do not summarize the conclusions of prior reviews on this topic including those by David Anderson et al., Robin Soler et al., Ron Goetzel and Catherine Heaney, and the Rand review for CMS. These should be cited at the beginning as a foundation for this review.	We added a brief summary of each review to Chapter 1: Introduction. The summaries are contained in a new section called 'Earlier Literature Reviews'
1	General	The focus of the review is primarily on the use of an HRA in isolation when the focus should be on the entire process of administering an HRA, providing feedback, and following up with behavior change and risk reduction support programs. It is unreasonable to expect that merely administering a survey instrument, with feedback, will change anyone's behavior, let alone prevent the occurrence	We agree and actually patterned our definition of an HRA after the RAND Report (RAND Santa Monica CAL. Evidence Report and Evidence-based Recommendations: Health Risk Appraisals and Medicare. Contract no.: 500-98-0281. Baltimore, MD: Centers for Medicare and Medicaid Services, 2003). Our definition of an HRA (p. 3, lines 19-24) reads as follows: "Our definition of an HRA contained three components: participants provided self-reported information to identify individual risk factors for disease; participants

of a disease. Previous literature reviews, including those by Rand, Anderson, and Community Guide, came to the same conclusion.

- The focus of the review is primarily on the use of an HRA in isolation when the focus should be on the entire process of administering an HRA, providing feedback, and following up with behavior change and risk reduction support programs. The Community Guide to Preventive Services, in its review of the worksite literature (which is referenced in this review) differentiated between two types of HRA applications: 1) an assessment of health risks with feedback, when used alone, (“HRA Alone”), and 2) an assessment of health risks with feedback as a gateway to more intensive and prolonged health promotion and risk reductions interventions (“HRA Plus”). That review concluded that an HRA Alone intervention is largely ineffectual while an HRA Plus intervention is effective in achieving long-term behavior change (such as quitting smoking) and significant reductions in biometric risk factors like cholesterol and blood pressure. At a minimum, the HRA Plus process would involve the administration of the HRA and production of a feedback report that would form the foundation of a personal prevention plan. However, for the HRA Plus to be most effective, it needs to also include the following components that complement the provision of an HRA with a feedback report:
 - Multiple or serial administrations of HRAs, with longitudinal feedback provided to participants on their health risk status,
 - Ongoing health education programs, provided through pamphlets, books, videos, or interactive computer programs,
 - Motivational interviewing, counseling, and coaching provided face-to-face or telephonically to support behavior change and risk reduction,

received individualized health-related feedback based on the information they provided; and the information was used to give participants at least one recommendation or intervention to promote health, sustain function, or prevent disease”. Any HRA, regardless of its delivery mechanism (e.g., single or multiple questionnaire administration, use of written feedback material, counseling, resource referral, etc.), that fulfilled these three criteria was included in the review. To ensure clarity regarding the type of HRAs included in the review, we added the following sentence to the aforementioned definition on p. 3: “Any HRA, regardless of its delivery mechanism (e.g., single or multiple questionnaire administration, use of written feedback material, counseling, resource referral, etc.), that fulfilled these three criteria was included in the review.” We also introduced the definition earlier in the TA.

		<ul style="list-style-type: none"> • Referral to community resources such as fitness facilities, self-help support groups, or neighborhood volunteer programs, and • Referral to local or national health promotion vendors and services such as smoking quit lines and wellness coaches. <p>Without the above follow-through” activities, the HRA alone program will not succeed. It is unreasonable to expect that merely administering a survey instrument, with feedback, will change anyone’s behavior, let alone prevent the occurrence of a disease. Previous literature reviews, including those by Rand, Anderson, and Community Guide, came to the same conclusion.</p>	
1	General	<ul style="list-style-type: none"> • I agree with the statement: “We believe the process following HRA questionnaire administration, namely feedback and recommendations, provides participants with a sense of engagement that encourages behavioral change. 	Thank you.
1	General	<ul style="list-style-type: none"> • The criteria applied for the evidence review (e.g., the Jadad scale) are applicable to clinical trial studies but not to “real world” evaluation studies conducted at workplace or community-based settings. Of the 115 studies reviewed, 53 were done at worksites and 23 in communities. The requirements for inclusion of studies in the “good” category are overly stringent given the realities of conducting this type of research in applied settings. For example, studies lost points if there was no double blinding, tracking of withdrawals, and reporting of adverse events. Again, these criteria are very relevant for drug or clinical trials but not to workplace based programs when employees at an entire worksite are encouraged to become more physically active or eat a healthier diet. The worksite interventions often involve changes in the physical and social environment, mass communication efforts, policy changes, and creation of support groups for workers wishing to improve their health. In fact, introducing environmental and policy interventions to 	We agree the quality scores were biased downward and we commented on this fact in the discussion.

		<p>improve lifestyle may be more potent than individual counseling, as has been shown to be the case with tobacco use. Quality scores were biased downward because the Jadad scale's adverse effects question was largely inapplicable to evaluating Health Risk Appraisals (HRAs). As the reviewers admit: "Low quality ratings could reflect poor reporting (perhaps prompted by journal word restrictions) instead of poor research. "</p>	
1	General	<p>It is unrealistic to expect "hard" health outcomes such as new cases of heart disease and diabetes from short-term health promotion programs. The reviewers acknowledge that. Health promotion interventions are intended to get people to quit smoking, manage stress, get appropriate preventive screening, drink alcohol responsibly, be physically active, and eat a healthy diet. An expectation of a 12-24 month improvement on these metrics is appropriate and realistic. However, expecting disease incidence to be affected is unrealistic. To assess the impact of smoking prevention/cessation or obesity prevention education and counseling, for example, would require a 10-20 year time horizon where individuals would be tracked longitudinally. The Framingham study provides the evidence for how behavior and biometric risks can cause disease. That long-term study has demonstrated the association between smoking, obesity, high cholesterol, high blood glucose, stress, etc. and disease incidence. Worksite and community-based studies are not designed to replicate multi-decade long research showing similar cause-effect or correlations among risk factors and disease outcomes. It is inappropriate, therefore, to conclude that the review was unable to determine whether HRA programs produced "tangible" health benefits over the medium to long run – when "intermediate" outcomes such as blood pressure, cholesterol, physical activity, and fat intake were affected by these programs. These latter outcomes are indeed</p>	<p>We believe our conclusions are valid. We agree that smoking, obesity, etc. are associated with disease incidence; however, little evidence exists to link the short-term improvements on intermediate HRA outcomes with long-term reductions in disease incidence.</p> <p>Our discussion of short- versus long-term outcomes occupies a very small portion of the TA. To avoid drawing disproportionate attention to this issue, we deleted the word 'tangible' from the TA to prevent the formation of an implied hierarchy where intermediate outcomes are given less importance than disease incidence.</p>

		“tangible.”	
1	General	By design, workplace health promotion program offer a comprehensive array of interventions that may include administering HRAs, delivering feedback in person, in print or face-to-face, engaging individuals in follow-up counseling and coaching sessions, facilitating use of on-site facilities such as fitness centers, reducing barriers to preventive screenings, connecting people to community resources such as Weight Watchers or Smoke enders, etc. These are not one-dimensional interventions where only one variable is manipulated. The analyses generally focus on multi-component programs rather than single focus ones. It is unrealistic to expect one part of this array of interventions to be singled out as effective and another not. It is, however, probable that certain combinations of programs are more effective than others.	Thank you for the observation.
1	General	The analysis states that there is no evidence that one type of feedback is any better than another. The reviewers state: “The feedback and recommendation components of HRA programs appeared to be the primary factors producing encouragement and motivation among participants to modify behaviors, certainly more so than any other component considered in Question 1 a-e. However, the evidence did not suggest a specific feedback or recommendation protocol that was better able than others to lead to behavior modification.” There is good evidence in studies conducted by Strecher, Prochaska, Lorig, Bandura and others showing the individualized, risk reduction, and tailored feedback that uses principles from the Trastheoretical Model of Change and Self-Efficacy Theory are much more powerful in eliciting behavior change than simple untailored feedback (e.g., newsletters).	‘Behavior change’, unless defined as a health outcome such as change in diet or physical activity, did not fall within the scope of key question # 2, which specifically directed us to examine health outcomes. To clarify the point of the two quoted sentences in the reviewer’s comment, we added the following underlined phrase to the text: “...a specific feedback or recommendation protocol that was better able than others to lead to behavior modification <u>that would produce better health outcomes</u> ”.
1	General	The focus of health promotion interventions is on improving population health not only the health of patients with specific diseases and disorders. As such, the baseline values for a number of measures	Thank you for the observation.

		<p>comprise the aggregation of risk factors from both healthier and less healthy people. Consequently, improvement on baseline measures is tempered by the fact that data on individuals with little need for improvement are combined with data from individuals with a great need for improvement. Thus, the effects of the intervention are diluted in such an analysis. Studies where only high-risk individuals are administered an HRA and offered follow up programs show a greater rate of improvement. (See Soler et al's review.) This has significant implications on the ways in which HRA programs are run with fewer time and resources directed at "well" populations vs. "at risk" populations.</p>	
1	General	<p>Some articles excluded in the review, which should have been included, are listed below:</p> <ul style="list-style-type: none"> • Fielding JE, Mason T, Knight K, et al. A randomized trial of the IMPACT worksite cholesterol reduction program. Am J Prev Med 1995;11(2):120-3. PMID:7632447 OVID-Medline. Exclude: Does not report health outcomes – yes it does • Goetzel RZ, Ozminkowski RJ, Bruno JA, et al. The long-term impact of Johnson & Johnson's Health & Wellness Program on employee health risks. Journal of Occupational & Environmental Medicine 2002;44(5):417-24. PMID:12024687 OVID-Medline. Exclude: No comparison group – yes there is • Heaney C A; Goetzel R Z., A review of health-related outcomes of multi-component worksite health promotion programs. American journal of health promotion : AJHP 1997;11(4):290-307 • Can health promotion programs save Medicare money? Ron Z Goetzel, David Shechter, Ronald J Ozminkowski, David C Stapleton, Pauline J Lapin, J Michael McGinnis, Catherine R Gordon, 	<p>We added this article to the TA.</p> <p>We added this article to the TA.</p> <p>We added a summary of this article to the introduction, as per a previous comment from the reviewer.</p> <p>This article is a review, not primary research, and would therefore not be included in a systematic review.</p>

Lester Breslow..Am J Health Promot. ;22
(1):suppl 1-7, iii

- Short Meghan E; Goetzel Ron Z; Young Jared S; Kowlessar Niranjana M; Liss-Levinson Rivka C; Tabrizi Maryam J; Roemer Enid Chung; Sabatelli Adriano A; Winick Keith; Montes Myrtho; Crighton K Andrew, Measuring changes in lipid and blood glucose values in the health and wellness program of Prudential Financial, Inc. Journal of occupational and environmental medicine / American College of Occupational and Environmental Medicine 2010;52(8):797-806.
- Goetzel Ron Z; Baker Kristin M; Short Meghan E; Pei Xiaofei; Ozminkowski Ronald J; Wang Shaohung; Bowen Jennie D; Roemer Enid C; Craun Beth A; Tully Karen J; Baase Catherine M; DeJoy David M; Wilson Mark G. First-year results of an obesity prevention program at The Dow Chemical Company. Journal of occupational and environmental medicine / American College of Occupational and Environmental Medicine 2009;51(2):125-38.
- Goetzel Ron Z; Roemer Enid C; Pei Xiaofei; Short Meghan E; Tabrizi Maryam J; Wilson Mark G; Dejoy David M; Craun Beth A; Tully Karen J; White John M; Baase Catherine M. Second-year results of an obesity prevention program at the Dow Chemical Company. Journal of occupational and environmental medicine / American College of Occupational and Environmental Medicine 2010;52(3):291-302.
- Goetzel, R.Z., Sepulveda, M., Knight, K., Eisen, M., Wade, S., Wong, J., Fielding, J. "Association of IBM's 'A Plan For Life' Health Promotion Program with Changes in Employees' Health Risk Status." Journal of Occupational Medicine,

This article was excluded because it was published after our literature search cut-off date of June 2010.

This article was excluded because treatment and control subjects all received HRAs; the authors are really comparing environmental interventions independent of HRAs.

This article was excluded because it is a companion paper to the previous publication, which was also excluded.

We added this article to the TA.

		36:9, September 1994, 1005.	
1	General	Former Surgeon General C. Everett Koop's name is misspelled	We corrected the instances where the name was misspelled.
2	General	Some thirty years ago, the National Center for Health Services Research (AHRQ's predecessor) contracted for a review of health hazard / health risk appraisal (HHA/HRA). That review defined HRA as "a health promotion technique in which an individual's health related behaviors and personal characteristics are compared to mortality statistics and epidemiologic data in order to estimate his or her risk of dying by some specified future time along with the amount of that risk which could be eliminated by making appropriate behavioral changes." (Wagner et al., 1982 – see below) That review concluded that the attention being devoted to HRA was excessive and that there was a paucity of evidence of effectiveness.	Thank you for this information.
2	General	<p>Since then, the term "HRA" has taken on a broader meaning, which this review defines as encompassing any assessment that includes self-reported information to identify risk factors, individualized feedback, and at least one recommendation. That very broad definition encompasses a huge variety of programs and studies, few if any of which test whether the basic procedure of providing individualized feedback makes any difference. In many such programs that feedback is a relatively minor part of the program being tested, so that if a program has effects on behavior there is little basis for attributing those effects to the HRA component. The present report contains numerous statements, especially in the discussion, that state or imply that HRA itself (i.e., the provision of individualized feedback) has effects, but evidence for this proposition has not been presented.</p>	<p>Many of the included studies compared HRA risk factor questionnaires (or clinical tests such as measures of blood pressure) to broader HRA programs that involved questionnaires (or clinical tests), feedback, and recommendations. Our conclusions regarding feedback relate to the fact that many studies found positive benefits for HRAs with all three components, rather than HRAs with questionnaires or clinical tests alone.</p> <p>Our definition of an HRA, namely that it included questionnaires or clinical tests, feedback, and recommendations, was based on the RAND report (RAND Santa Monica CAL. Evidence Report and Evidence-based Recommendations: Health Risk Appraisals and Medicare. Contract no.: 500-98-0281. Baltimore, MD: Centers for Medicare and Medicaid Services, 2003). We agree the notion of what constitutes an HRA is subject to debate and we have mentioned this issue in the introduction. However, to guide the review, we had to select one specific definition and felt a broad-based definition was valid given that the consensus in health promotion appears to be that HRAs are more than questionnaires or clinical tests undertaken to collect information on risk factors for disease.</p>

2	General	<p>My understanding is that HRA is largely a “packaging” or framework for organizing a behavioral change intervention, rather than itself expected to lead to behavioral risk reduction. When HRA (also called “health hazard appraisal”) was being promoted by Lewis Robbins during the 1960s-1970s, there was a belief that the procedure itself could stimulate behavioral change. In fact it may well have done so, since by using it a physician was providing an occasion to review the patient’s health risks and make recommendations, an activity that was not common in the era before preventive medicine had received the attention it does now. The feedback may have had an impact on patients as well, since many would not have received any physician recommendations before their doctor gave them an HRA.</p> <p>Once preventive counseling became a standard part of medical practice, however, the HRA would not be expected to have any particular impact. Similarly, in the context of health promotion programs, HRA is not expected to accomplish anything unless paired with an effective behavioral program. Has it ever been demonstrated that an effective behavioral program is enhanced in any way by adding HRA?</p>	We were not tasked with addressing this question.
2	General	<p>The draft technology assessment report, perhaps because of the way the key questions were framed, proceeds on the assumption that HRA is a tool of particular significance. It is not obvious to me that it is or that its inclusion in a program somehow distinguishes that program from similar ones that do not include HRA. Although the Affordable Care Act authorizes an annual “health risk assessment”, I have not seen an interpretation indicating whether the language of the Act envisions the HRA procedure (questionnaires, feedback, recommendations, with or without quantification) or a more general assessment of health risks as is typically carried out by a physician during an annual preventive exam that includes inquiry about risk factors. What is</p>	Thank you for this observation.

		commonly called HRA in the health promotion field remains to be demonstrated as having any particular significance.	
2	Page ES-4	The Executive Summary provides a good digest of the report. Page ES-4, Key Question 1c: “we believe all personnel received some orientation or guidance”. What is the basis for this belief? If the basis is simply that we expect it to be the case, perhaps that is not worth including, unless the authors feel that they have more insight into the question than the average reader	We deleted this phrase
2	Page ES-5	Key Question 1e: “In the senior population, workplace cost reduction largely does not apply” – particularly at the younger end of the senior citizen range, a significant fraction of seniors are employed. The fact that many seniors have chronic medical conditions does not imply that preventive behaviors (including blood pressure control, weight loss or maintenance, physical activity, etc.) will not reduce their morbidity, absenteeism, and medical care costs. The HRA studies reviewed may have few data in this area, but the topic should not be dismissed. (This comment also applies to page 75 of the discussion.)	To prevent being overly dismissive, we changed the sentence to read “In the senior population, workplace cost reduction may not apply...”. We made this change in the discussion as well.
2	Intro	History: This section would benefit from by the inclusion of several landmarks from prior efforts to evaluate health risk appraisal programs. For example, some thirty years ago the National Center for Health Services Research, ARHQ’s predecessor, contracted for a “Description, Analysis, and Assessment of Health Hazard/Health Risk Appraisal Programs” (contract no. 233-79-3008). (See “An Assessment of Health Hazard/Health Risk Appraisal”. Edward H. Wagner, William L. Beery, Victor J. Schoenbach, Robin M. Graham. <i>Am J Public Health</i> 1982; 72:347-352, http://ajph.aphapublications.org/cgi/reprint/72/4/347.pdf) Have there been other such reviews prior to the present one? That review defined HRA as a	We added a reference to the Wagner et al. paper in the introduction. The contents of the Wagner et al. paper are consistent with the current text, so we did not modify the text following this paper’s inclusion. Other reviews of HRAs are in the literature. However, the general form of a TA is to provide a brief introduction and background before listing the key questions and turning to the methods. The TA was commissioned to address specific key questions, so our focus was directed to these questions.

		technique that incorporated quantitative risk feedback.	
2	Intro	“Health Risk Appraisal and the Elderly”: “In fact, the Affordable Care Act authorizes Medicare to cover an annual HRA . . .”. Does “health risk assessment” in the ACA really refer to HRA as defined in this report or a more general preventive visit?	We defer comment on legal definitions in the ACA to appropriate legislative experts at CMS.
2	Methods	The methods are clearly described. The review as conducted provides a useful overview of the landscape from a phenomenological perspective – how many studies of what kinds, with what designs, with what length of follow-up, etc. However, since HRA has been defined so broadly and these programs vary to greatly in regard to the interventions tested, outcomes measured, and other critical aspects, that a summary cannot offer much guidance concerning what contributes to an effective program. A useful addition would be for the authors to select a handful of exemplary studies/programs and present brief summaries, indicating what can be learned from these programs. The selection could include studies that did not observe an effect if the study was sufficiently large and rigorous that its findings provide evidence about approaches that are unlikely to be effective.	AHRQ TAs are based on systematic review methodology, which requires identification of articles based on objective, a priori criteria guided by the key questions to be addressed in the research. The selection of some “exemplary” studies is not systematic in that normative values may dictate the definition of ‘exemplary’ and the number of studies to summarize. Nonsystematic selection of studies could provide only a partial picture of the evidence related to a key question.
2	Results	The authors explain that since some studies generated multiple articles, there were more articles (115) than studies (111). Although the report does differentiate between articles and studies, it sometimes refers to articles when it would be more logical and informative to refer to studies. For example, on page 6 bottom, “Samples sizes ranged from less than 100 participants in 16 articles” and “Forty-one articles had between 100 and 500 participants . . .” These statistics are more relevant for studies than for articles, unless the different articles report study components with different numbers of subjects (e.g., an article on subjects younger than 65 years and another article on subjects older than 65 years). Similarly, method of	Generally, we reported counts by article rather than by study since the number of companion papers was minimal. Also, when companion papers report additional results, reporting necessarily requires us to count by article and not by study. The figures pertain to articles. We replaced the word ‘study’ with ‘article’ (and vice versa) in many places throughout the TA to better differentiate between articles and studies.

		HRA (p48 and Figure 6), training reported (p49, Figure 7), methods of follow-up (p50, Figure 8), frequency of follow-up (p50, Figure 9, Table 4), characteristics of patient populations (p63 and related figures), and the Medicare population (page 66, question 3) appear to pertain to studies.	
2	Results Page 14 top	My concern about whether HRA is itself an intervention or not is relevant to statements like “Lengths of followup, though, were 1 year or less in 65 RCTs, which indicates that most trials contained inadequate evidence to evaluate the long-term effects of HRAs.” (p14 top). Would one expect a long-term effect from simply reporting of risk behaviors, the provision of feedback, and the provision of recommendations? If there is to be a longterm effect that must come from other interventions used alongside HRA. Thus it seems odd to speak about a “long-term effects of HRAs”.	The articles included in the review considered HRAs (assessment, feedback, recommendations) to be interventions. Thus, we were concerned about reporting of long-term effects for these interventions.
2	Results Page 16	The discussion of validity of HRA instruments (page 16) provides kappas and correlation coefficients, but this information tells us little except that the instruments vary in their reliability. Depending upon what use is made of the questionnaire results and the way in which they are administered, kappas and correlations in the ranges cited (e.g., $r=0.52$ to 0.90) could result in little distortion or a great deal of distortion of the results. I do not know what more the authors could have done, but unfortunately not much can be made of the information.	We agree with the reviewer that the interpretation of specific kappas and correlations could differ depending on the nature of HRA administration. The key question asked us to describe the characteristics of HRAs and the take home message is indeed that HRAs vary in their reliability/validity.
2	Results Page 65	“Rather, positive benefits from HRAs tended to occur . . .” – this language seems to imply that HRA itself has an effect, but my reading of the report was that there was not evidence indicating that HRA per se had an impact, although interventions that included HRA as one component may have had an impact.	Yes, the statement does imply that HRAs had an effect; however, we wrote the statement with our definition of HRAs in mind (questionnaire/clinical test + feedback + recommendation), rather than a more restrictive definition of HRAs that might be limited to questionnaire/clinical test + feedback or questionnaire/clinical test alone.
2	Results	The report goes on to state: “Notwithstanding the items discussed in Question 1 a-e above, HRA programs involving elicitation of risk factors, individualized feedback, and recommendations appeared to provide participants with motivational	All included articles had comparison groups where the comparator intervention was something less than a ‘full’ HRA with questionnaire/clinical test + feedback + recommendation. We point this out in several locations, e.g., p. 69, last paragraph. Several included articles showed positive

		<p>boosts to alter their behaviors in a positive manner. We believe the process following HRA questionnaire administration, namely feedback and recommendations, provides participants with a sense of engagement that encourages behavior change.” Since all of the studies reviewed had the components of elicitation of risk factors, individualized feedback, and recommendations”, what indication was there that these components had such an impact?</p>	<p>benefits for full HRAs versus comparators, so we concluded that feedback and recommendations (or the added presence of recommendations when comparators included questionnaire/clinical test + feedback) had a positive impact on behavior.</p>
2	Results	<p>On what do the authors base their belief that the process provides participants with a sense of engagement? That is certainly the intuitive belief, but was there any evidence to warrant such a statement? Were there any studies that compared HRA by itself to a control without HRA and found greater engagement and motivation? Were there any studies that compared a behavioral intervention plus HRA to the same behavioral intervention without HRA? If so, those studies should be cited in support of the statement about effects of HRA.</p>	<p>As we state in the results for key question # 2, the reviewed evidence did not suggest specific characteristics of HRAs that were associated with better health outcomes. However, several included articles showed positive benefits for full HRAs versus comparators, so we concluded that feedback and recommendations (or the added presence of recommendations when comparators included questionnaire/clinical test + feedback) had a positive impact on behavior. We were simply unable to determine what types of feedback or recommendations were best able to produce a positive impact.</p> <p>We added a brief summary of 10 studies with non-HRA controls to the results for key question # 2. These studies confirm our assertions regarding the effects of feedback and recommendations.</p>
2	Results	<p>If there are several studies that the authors regard as providing evidence that supports their apparent view of the effects of HRA, it would be helpful if those studies could be summarized in the Results and provided as exemplars.</p>	<p>Please see above comment.</p>
2	Discussion / Conclusion Page 73	<p>“we believe all personnel received some orientation or guidance regarding HRA delivery”. What is the basis for the authors’ belief, other than that surely the studies must have oriented their personnel? If the basis is only that expectation, perhaps it is better to let the reader make that supposition unaided by a statement of the authors’ belief.</p>	<p>We deleted this statement from the TA.</p>
2	Discussion / Conclusion	<p>“HRAs involve multiple contacts with participants and standard followup methods . . .”. I suggest that “typical followup methods” might be a preferable</p>	<p>We changed the phrasing as per the reviewer’s suggestion.</p>

	Page 73	phrasing, so as not to imply that the studies were adhering to any recognized standard – unless that was the case.	
2	Discussion / Conclusion Page 73	“HRAs administered entirely online might not reach groups at highest risk for chronic disease . . . who could potentially reap large benefits from HRAs.” – here again I have to ask is there any indication that HRAs can provide large benefits?	We deleted the phrase “but they are the ones who could potentially reap large benefits from HRAs’ from the text.
2	Discussion / Conclusion Page 73-76	If the authors mean programs that use HRA, at least use this more nearly precise wording. I have the same concern about the following paragraph about frequency of follow-up. Does any of that paragraph refer to the HRA procedure or components in themselves as opposed to intervention programs that happen to use HRA?	Our definition of an HRA is written on p. 3, fourth paragraph: “Our definition of an HRA contained three components: participants provided self-reported information to identify individual risk factors for disease; participants received individualized health-related feedback based on the information they provided; and the information was used to give participants at least one recommendation or intervention to promote health, sustain function, or prevent disease. We excluded studies reporting HRAs without all three components.” We did not waiver from this definition in the TA. We are referring to the number of follow-up contacts undertaken as part of the HRA itself. To clarify this point, we added the word ‘contacts’ or ‘contact’ to create the phrases “followup contacts” or “followup contact”. We also deleted the last sentence of p. 76, paragraph 6 (i.e., “However, the true impact of HRAs cannot be adequately assessed over such a small number of contacts and future research should involve more frequent followups.”).
2	Discussion / Conclusion Page 74	The variety of feedback itself could be enough to keep participants interested in a study.” et suite. Presumably the concern is to keep participants engaged in the intervention program rather than in the study. But do the authors have any evidence for this speculation?	No, this is speculation on our part.
2	Discussion / Conclusion Page 74	(I can accept that HRA may be an effective component of health promotion programs and even that HRA by itself can have any impact. But my understanding is that those possibilities remain to be demonstrated, and using “HRA” as a generic term for a health behavior intervention that happens to have	We point to the fact that no standard HRA definition exists and we used a definition from an earlier report conducted by RAND (RAND Santa Monica CAL. Evidence Report and Evidence-based Recommendations: Health Risk Appraisals and Medicare. Contract no.: 500-98-0281. Baltimore, MD: Centers for Medicare and Medicaid Services, 2003.).

		the components of elicitation, feedback, and recommendations may confuse rather than clarify.	
2	Discussion / Conclusion Page 74 question 1e	P74 question 1e – “The feedback and recommendation components of HRA programs appeared to be the primary factors producing encouragement and motivation among participants to modify behaviors, . . .” Have the authors done any analysis or found evidence in the studies to support this statement? If so that support should be indicated. If not, then what is the basis for the conclusion?	As we state in the results for key question # 2, the reviewed evidence did not suggest specific characteristics of HRAs that were associated with better health outcomes. However, several included articles showed positive benefits for full HRAs versus comparators, so we concluded that feedback and recommendations (or the added presence of recommendations when comparators included questionnaire/clinical test + feedback) had a positive impact on behavior. We were simply unable to determine what types of feedback or recommendations were best able to produce a positive impact
2	Discussion / Conclusion Page 74 question 2	“We consider training to be the specific teaching and instruction given to staff to run HRA programs.” Here again, the term “HRA programs” is used as if there were such a class of entities for which some definition of training is possible.	We point to the fact that no standard HRA definition exists and we used a definition from an earlier report conducted by RAND (RAND Santa Monica CAL. Evidence Report and Evidence-based Recommendations: Health Risk Appraisals and Medicare. Contract no.: 500-98-0281. Baltimore, MD: Centers for Medicare and Medicaid Services; 2003).
2	Discussion / Conclusion Page 75	“We could not conclude whether the followup periods were too short to detect between-group differences, keeping in mind the substantive benefits of HRAs are likely to accrue over the medium or long term . . .” Is there any reason to expect such a “sleeping effect” of HRA? Or are the authors referring to programs that use HRA? Health benefits accrue over time following risk factor changes, but do the interventions conducted in programs that use HRA tend to have late effects on behavioral change? Have any such intervention effects been reported?	To avoid leading readers in unintended directions, we removed this sentence from the text.
2	Discussion / Conclusion Page 76 Top	is it really the case that seniors will be more receptive to HRAs that use “technologically appropriate methods such as paper and pencil”? Would it not at least make a difference whether one is talking about seniors in their fifties, sixties, seventies, or eighties?	We already provide a citation to help support this assertion (Administration on Aging - Department of Health and Human Services. Internet Usage and Online Activities of Older Adults. http://www.aoa.gov/AoAroot/Press_Room/Social_Media/Widget/Statistical_Profile/2010/6.aspx) We agree there could be a difference between age groups and we are referring to seniors as being persons aged 65 years and over. To clarify, we added this definition to the

			text.
2	Discussion / Conclusion Page 77	Lags in feedback could lead to outdated recommendations – if the lag is measured in weeks, rather than months or years, what recommendations would become outdated?	In the absence of a rule of thumb to distinguish appropriate from inappropriate lags, we added the word 'long' to the last sentence on p. 79: "Long lags in feedback provision could promote participant disinterest or lead to outdated recommendations."
2	Discussion / Conclusion Page 77	"We can only conclude intuitively" – what is an intuitive conclusion?	We deleted the word 'intuitively'.
2	Discussion / Conclusion Page 77	"Perhaps most programs that supplement feedback and recommendations with contact such as counseling or telephone followup will motivate participants to adhere to recommendations out of a desire to please project staff, regardless of the type of contacts." If only behavioral change motivation were so easy! Perhaps the authors could provide a relevant citation for their supposition	We already provide a reference (McCarney R, Warner J, Iliffe S, et al. The Hawthorne Effect: a randomised, controlled trial. BMC Medical Research Methodology 2007;7(1):30). The sentence quoted by the reviewer may attribute too much behavioral change to the Hawthorne Effect. Consequently, we modified the sentence to read as follows: "Programs that supplement feedback and recommendations with contact such as counseling or telephone followup may motivate participants to adhere to recommendations out of a desire to please project staff, regardless of the types of contacts".
2	Conclusions	However, few articles considered hard health outcomes such as the incidence of specific chronic diseases" – did any studies look at incidence of chronic conditions?	Only one study, Charlson ME, Peterson JC, Boutin-Foster C, et al. Changing health behaviors to improve health outcomes after angioplasty: a randomized trial of net present value versus future value risk communication. Health Educ Res 2008;23(5):826-39, looked at the incidence of chronic disease and we added mention of this study to the executive summary and discussion.
2	Conclusions Page 78	"We raised several issues that researchers should consider . . . " – this paragraph presents relatively minor specific questions rather than asking researchers to devote attention to the numerous methodological limitations and reporting omissions that the authors found characterized the literature they reviewed. Methodological concerns that limit the ability to say whether HRA has any effect take priority over research to try to optimize an effect that has yet to be demonstrated.	We agree. In the Conclusion section, we summarize the lack of evidence to address the key questions and provide some additional issues to consider. We do not suggest that these issues should be paramount to other concerns.
2	Tables	I suggest bolding the rows in Table 5 for the larger studies (e.g., those with more than 1,000 participants). Also, what is meant by "General	We cannot bold the rows because doing so would be inconsistent with the AHRQ style for TAs.

		health"? Notes at the bottom of the first page of the table should appear (or be repeated) at the end of the table.	<p>We added the following definition of general health to the text: "HRAs targeting general health collect data on an assortment of risk factors without a specific interest in any one disease (e.g., CVD) or behavioral area (e.g., smoking cessation, physical activity)".</p> <p>The notes to Table 5 have been moved to the end of the Table.</p>
2	Figures	Figure 6 has a column labeled "Combination", which is not particularly informative since that could represent almost anything. The method of presentation in Figure 8, where each article (study?) is counted once for each feedback method may be more useful. (Again, I would have thought that studies would be more relevant for tabulating than articles, but perhaps the articles are so different that that is not the case.)	<p>We added a footnote to Figure 6 explaining what we mean by 'combination'.</p> <p>Specific articles may be counted in more than one column in Figure 8. We revised the footnote to Figure 8 to make this fact clearer.</p>
2	Refs	<p>For purposes of continuity, if nothing else, I suggest citing:</p> <p>An Assessment of Health Hazard/Health Risk Appraisal. Edward H. Wagner, William L. Beery, Victor J. Schoenbach, Robin M. Graham. Am J Public Health 1982; 72:347-352, http://ajph.aphapublications.org/cgi/reprint/72/4/347.pdf</p> <p>Health risk appraisal: review of evidence for effectiveness. Schoenbach VJ, Wagner EH, Beery WL. Health Serv Res 1987 October; 22(4): 553-580. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1065456/</p>	We added both references to the introduction.
2	Editorial suggestions	<p>"third party variables" is not a commonly employed term in the public health literature. I suggest the term "potential confounders".</p> <p>Page 34 – "The Mediterranean Easting in Scotland Experience" ?</p> <p>Page 49, Figure 8 – "Posted mail" should (at least for a U.S. audience) probably be "Postal mail".</p> <p>Page 63 – "forth article", "Females composed 30 percent" – should that be "comprised 30 percent"?</p> <p>Page 63, bottom – why does the next-to-last</p>	<p>We made the requested change to the text.</p>

		<p>paragraph end with a mention of a study of seniors recruited from primary care practices but the following paragraph begins by referring to 18 articles (studies?) recruited from physicians' offices. The organization seems confused.</p> <p>Page 66 – “females composed 100 percent” – should that be “comprised”?</p> <p>Page 66 – I would put “random digit dialing in a rural population” as the last item in the sentence, since currently it's ambiguous about whether “health councils, hospital clinics, . . .” were part of the target for random digit dialing.</p> <p>Page 74 – “might better suite participants' preferences” – “suit”?</p> <p>P75 – “two or less followup contacts” – “two or fewer”</p> <p>P77 – “A forth issue”</p> <p>B-2 – “Representatives of the cases” – would that be “Representativeness”?</p>	<p>We made the requested change to the text.</p>
3	General	<p>This is just awful. It must be redone in its entirety. The fatal problems are general, and trace in large part to whoever wrote the charge for this effort. The central issue is whether Medicare could expect to have healthier beneficiaries requiring less medical care if specific HRA-based interventions were encouraged.</p> <p>To address such issues: only studies reporting on seniors should be studied, only studies reporting quantitative outcomes (morbidity, mortality, health risk changes, medical utilization) with comparison groups can be analyzed. Here, only 16 studies were eligible on the age criterion; workplace programs cannot be generalized to seniors. Only a few of these reported outcomes, and those outcomes are not reported in meta-analytic fashion, they seem to be only checkmarks that the program collected the data!</p>	<p>CMS developed the key questions for the TA. The Medicare population was the specific subject of one key question; the other key questions were not limited to the Medicare population.</p> <p>To address key question # 3, we included studies reporting 'health outcomes'. Most health outcomes (in HRA studies of seniors and non-seniors alike) were intermediate markers such as blood pressure. All of these health outcomes were “quantitative”. We discussed the limitations of generalizing workplace studies to seniors. We did not conduct a meta-analysis because the study data were far too heterogeneous to combine statistically, and we pointed this out in the methods chapter.</p> <p>The Evidence Table in Appendix D reports specific quantitative outcomes.</p> <p>The scope of the TA (key questions 1 and 2) encompassed</p>

Thus, by studying seniors for whom outcomes were measured in quantitative detail you could approach the charge with far fewer studies to evaluate in far greater depth. Here, clearly abstracts were used instead of detailed study review. No quantitative data are presented. It's all checklists. Many good studies are not included orl were excluded. Some, easily accessed, were said to be unfindable.

Hypotheses for HRA characteristics likely to improve health include at the top: (1) theory-based, as with self-efficacy or readiness to change;(2)"tailored", with computer triage giving every subject a specific feedback on initial hra and a progress report, in detail, on subsequent hra's, (3) brief, to increase completion and to focus on major health risks, not marginal ones, to decrease questionnaire burden.(4) Positive, self-efficacy prose is more effective that scary threats. None of these appear to have been examined.

The authors admit failure in their discussions of key items 2 and 3 on page es-5

An irony here is that this charge was previously given by CMS to RAND, "Evidence Report and Evidence-based Recommendations: Health Risk Appraisals and Medicare, contract # 500-98-0281; (1998-2003). This study, while far from perfect, was infinitely better than this one, and found generally positive results.

all groups, seniors and non-seniors. We included studies reporting numerical (i.e., quantitative) outcomes, which are shown in the Evidence Table in Appendix D.

We employed standard systematic review methods to screen articles for inclusion in the TA: the first two screening levels involved reading titles and abstracts; the third level involved reading the full article text (see Chapter 2: Methods for further explanation).

In the absence of a list of specific articles from the reviewer, we are unable to assess whether certain excluded articles should have been included in the TA.

We excluded 'readiness-to-change' outcomes because they are not health outcomes. HRAs had to involve individualized feedback to be included in the TA. We were not asked to look at length or 'positivity' of HRAs.

We state for key question # 2 that the evidence was essentially too heterogeneous for us to uncover any patterns that would indicate whether specific characteristics of HRAs were associated with better health outcomes. For key question # 3, we conclude that the results of studies in the 'under age 65 years' group cannot be generalized to the '65 years and over' group.

We are aware of the RAND study and patterned our definition of HRAs after the definition in the RAND study.

	<p>As a result, CMS is funding the Senior Risk Reduction Study, the most rigorous attempt yet to answer these questions with a randomized design and adequate numbers of randomly chosen persons aged 67-74. Did no one associated with this study know of this work?</p> <p>Additionally there are thoughtful reviews of this subject by Chapman, Pelletier, Aldana, Goetzel, and others, critical but generally positive. They describe theory and practice as well as citing the interventions judged most likely to succeed. Did nobody think to synthesize prior work?</p> <p>I looked up about 20 articles on seniors in the report that I knew well and felt to be among the best. Generally, they were not found, were excluded on some technicality of a check list, weren't read carefully, or showed reviewer flaws such as not considering claims data endpoints blinded and objective.</p> <p>I'm out of time and energy so will stop here. This is a sad case in an important area. Sloppy, sloppy, sloppy.</p>	<p>The TA is a synthesis of prior work, provided the prior work addressed any of the three key questions and met our inclusion/exclusion criteria.</p> <p>In the absence of references for these specific articles, we cannot address the reviewer's comment.</p> <p>In the absence of specific details, we cannot address the reviewer's concern.</p>
--	---	--

¹ Peer reviewers are not listed in alphabetical order.

² If listed, page number, line number, or section refers to the draft report.

³ If listed, page number, line number, or section refers to the final report.

Project Name: Health Risk Appraisal
 Project ID: RSKA0410

Table 2: Public Review Comments

Reviewer Name ¹	Reviewer Affiliation ²	Section ³	Reviewer Comments	Author Response ⁴
Anonymous Reviewer 1	StayWell Health Management	General	<p>StayWell Health Management has been a recognized market leader for over 30 years in the development of evidence-based Health Risk Appraisal tools (HRA) and the implementation of worksite-based Health Risk Assessment processes including the administration of HRA tools to populations of employees/dependents and the triage of individual participants into targeted and tailored follow-up health behavior change programs to assist them in reducing health risks identified in the HRA tool.</p> <p>In our continuing efforts to be a market leader and monitor emerging trends, we appreciate the opportunity to provide comments on the 'Health Risk Appraisal Technology Appraisal Report' draft dated January 19, 2011. StayWell understands this is a draft document and the goals of this report were to describe key HRA features, the features associated with a successful HRA and the applicability HRA to the Medicare population as a whole. StayWell appreciates this opportunity to</p>	<p>Thank you for this information</p> <p>Thank you for reviewing and commenting on the draft report.</p>

provide comments and looks forward to providing further input if opportunities arise.

We believe this report will have the greatest value to policy makers if the relevant literature is viewed within the context of the typical role of HRA and the fact that past research has been conducted with limited financial and technical support and often in environments where optimal experimental controls to assure internal validity are not feasible. While the evidence based on this body of research is inherently imperfect, we believe ample evidence is available to guide policy makers and generally supports the use of HRA as a core component of a comprehensive wellness program.

Our comments focus on distinguishing the impact of HRA as a stand-alone tool versus the impact of health risk assessment within the context of a comprehensive wellness program. We believe the evidence, though far from ideal, is substantial enough to recommend against implementing a stand-alone HRA tool and to recommend for implementing a Medicare HRA tool within the context of a broader wellness strategy focused on assisting Medicare recipients in making changes in their daily lifestyle and health practices identified by the HRA.

Thank you for this comment.

Policy-making is beyond the mandate of the TA.

		<p>The following comments draw heavily on a review of the literature on the impact of worksite-based HRA on health-related outcomes, which was completed by StayWell expert staff at the invitation of the Centers for Disease Control and Prevention and published in the American Journal of Health Promotion(1). Although a considerable number of studies on the impact of health risk assessment within the context of comprehensive wellness programs have been published since that time (2,3) the findings and conclusions we drew in 1996 remain largely unchanged.</p> <p>Taken as a whole, the studies reported in the literature reviews by Aldana (2) and Baicker et al.(3) provide imperfect but relatively consistent evidence supporting the effectiveness of HRA in the worksite setting when used as a component of a more comprehensive wellness program. This research also generally provides evidence that 'more is better' when it comes to providing support for individuals attempting to change unhealthy behaviors, but it is not clear whether this refers to the number or types of intervention strategies or to the number of personal contacts with program staff or outreach tools.</p> <p>Based on published literature, there is little evidence that the use of stand-alone HRA is</p>	<p>Comment noted.</p> <p>We are unable to address this comment because it does not relate specifically to the TA.</p> <p>We are unable to address this comment because it does not relate specifically to the TA.</p>
--	--	--	---

sufficient to produce long-term changes in health-related behaviors or health-risk status. While several of the studies report positive changes in seat belt use and self-reported physical activity associated with HRA participation, as well as positive changes in other health-related outcomes (e.g., systolic blood pressure, body mass index), evidence supporting causal inference is lacking due to the many threats to internal validity in these studies. Probably the best conclusion at this time is that evidence for the impact of HRA per se on health-related behaviors is weak. Few studies have been completed, and almost none have been well controlled. It is important to recognize, however, that behavior change is not the typical objective of HRA within the context of a comprehensive wellness process. Rather, HRA objectives focus on health-related outcomes in the early 'prebehavioral' aspects of the change process, such as increased awareness of risks and commitment to change. This suggests that research should focus on assessing the impact of HRA on movement through the initial steps in the change process.(4,5) It is unfortunate that most published studies of HRA effectiveness focus almost exclusively on behavioral outcomes and risk reduction, which only reflect movement through the later, more visible behavioral stages

of change.(4,5) As was also noted in previous reviews of HRA effectiveness,(6,7) if the objectives of HRA focus on 'prebehavioral' health outcomes, then evaluating the effects of HRA on health behavior outcomes is misguided.

Another limitation of the research to date is the general lack of relevance between published HRA research and the evolving nature and role of HRA in practice. Rather than the science of HRA informing its definition and application, practitioners have drawn upon relevant research in related fields (8,4,5) and their own experience to push the definition and use of HRA well beyond its origins and far ahead of its scientific literature. This limitation is a direct consequence of inadequate funding for basic HRA research. As Schoenbach and colleagues recognized in their review of HRA research,(6) ?most of the studies have been carried out with minimal funding, by persons outside the mainstream of scientific research, and without benefit of external review.? This observation from more than 20 years in the past could as well be made today. For HRA research to be credible or helpful to policy makers and to wellness practitioners, more relevant, better controlled studies focused on appropriate health-related measures are needed. Accordingly, if a Medicare HRA

We are unable to address this comment because it does not relate specifically to the TA.

process including follow-up wellness interventions is implemented within the mandated annual wellness visit, we strongly recommend including a strong research component including pilot studies that systematically vary key elements of the program design to assess their relative effectiveness.

Future research should move beyond the general question of the impact of HRA on health-related outcomes, such as changes in participants' knowledge, motivation and behavior, to specific issues of HRA design and delivery approaches. For example, it is important to know not only whether HRA is effective, but also how each characteristic of HRA influences its effectiveness. What are the best risk-projection methods with respect to user comprehension and motivation? Should there be more or less emphasis on mortality and morbidity? Does presentation of historical data enhance HRA effectiveness? What types of design format for the HRA participant report -- text, charts, statistical information, etc. -- are most effective? How should HRA feedback be tailored to different target populations -- varying in age, sex, education, ethnicity, culture, and socioeconomic status -- to maximize effectiveness within each? Many generally agreed-upon assumptions by practitioners regarding answers to these questions have evolved over time,

We are unable to address this comment because it does not relate specifically to the TA.

but these assumptions have been informed almost solely by practice because rigorous research on these questions is virtually non-existent. Again, we recommend research be conducted to address these questions, particularly as they apply to the Medicare population for the purposes of this program.

Biometric screening has played an important historical role in the use of HRA in assessing health risks. There is a trend toward simplified screening protocols, however, as HRA instruments have evolved from mortality-based to habit-based in nature and as screening recommendations have become more targeted.(9) Screening is sometimes even eliminated from the assessment process if it is either not feasible or cost-prohibitive. For these reasons, the role of screening in the effectiveness of the health risk assessment process should be systematically explored. The reliability, validity, and effectiveness of HRA in the absence of screening must be better understood. When screening is included in the health risk assessment process, a better understanding is needed of how to integrate screening measurements into HRA feedback to maximize comprehension and effectiveness.(8) Again, relevant research is sorely needed and there may be some opportunities to test certain aspects

We are unable to address this comment because it does not relate specifically to the TA.

of how best to integrate screening data in this program.

HRA has played an important role in worksite wellness for 30 years and will play an increasingly important role in the future as HRA technology increasingly integrates the emerging knowledge of behavior-change processes, HRA-generated data become an increasingly valuable vehicle for targeting and tailoring follow-up behavioral interventions, and organizations demand greater accountability for measuring the impact of wellness programs on health risks. If HRA research is to support the future development of this core wellness technology, a more relevant and sophisticated HRA science base is essential. As the subject review makes clear, research with stronger internal validity that can still be generalized to the Medicare population and other populations is a priority need as HRA and wellness play a greater role in mainstream health care delivery.

References:

(1)Anderson DR, Staufacker MJ. The impact of worksite-based health risk appraisal on health-related outcomes: A review of the literature. American Journal of Health Promotion, 1996; 10:499-508.

(2)Aldana SG. Financial impact of health promotion programs: A comprehensive review of the literature. American Journal of

We are unable to address this comment because it does not relate specifically to the TA.

			<p>Health Promotion 2001;15:296-320.</p> <p>(3)Baicker K, Cutler D, Song Z. Workplace wellness programs can generate savings. Health Affairs. 2010;29(2):1-10.</p> <p>(4)Prochaska J, DiClemente C. Stages of change in the modification of problem behaviors. In: Hersen M, Eisler R, Miller P, eds. Progress in Behavior Modification (Volume 28, pp. 183-213). New York: Sycamore, 1992.</p> <p>(5)Prochaska J, Norcross J, DeClemente C. Changing for Good. New York: William Morrow, 1994.</p> <p>(6)Schroenbach V, Wagner E, Beery W. Health risk appraisal: Review of evidence of effectiveness. Health Services Research 1987; 22(4):553-580.</p> <p>(7)Edington D, Yen L. Reliability, validity, and effectiveness of health risk appraisals. In: Peterson K, Hilles S, eds. The Society of Prospective Medicine Handbook of Health Risk Appraisals, Second Edition. Indianapolis: Society of Prospective Medicine, 1994.</p> <p>(8)Terry P. The role of health risk appraisal in the workplace: Assessment versus behavior change. American Journal of Health Promotion 1987; 2(2):18-22,36.</p> <p>(9)U.S. Preventive Services Task Force. Guidelines for Clinical Preventive Services. http://www.ahrq.gov/clinic/prevenix.htm</p>	
Don Denmark	Carondelet St. Joseph's	General	The results derived from this comprehensive article review	Thank you for the comment.

	Hospital		addresses the potential short term benefits in limited circumstances. The paucity of longer term follow-up and credible results would make HRA suspect for driving benefits, P4P, andor health policy	
Don Denmark	Carondelet St. Joseph's Hospital	Executive Summary	Representative of content	Thank you.
Don Denmark	Carondelet St. Joseph's Hospital	Introduction /Background	Well complied	Thank you.
Don Denmark	Carondelet St. Joseph's Hospital	Methods	credible	Thank you.
Don Denmark	Carondelet St. Joseph's Hospital	Results	clear in the summary and presentation	Thank you.
Don Denmark	Carondelet St. Joseph's Hospital	Discussion / Conclusion	The poor - fair quality of the studies and lack of standardization gives little support for extrapolating long term health benefits to the use of HRA tools	Thank you for the comment.
Don Denmark	Carondelet St. Joseph's Hospital	Tables	readable and understandable	Thank you.
Don Denmark	Carondelet St. Joseph's Hospital	Figures	For the most part well formatted and easy to understand	Thank you.
John Harris	Healthways	General	<p>We read with interest the draft Health Technology Assessment (HTA) of Health Risk Appraisals and appreciate the opportunity to comment on it.</p> <p>As the industry leader in well-being improvement, we currently provide our health promotion, chronic care management, wellness and</p>	<p>Thank you for reviewing and commenting on the draft report.</p> <p>Thank you for this information.</p>

prevention services, both domestically and internationally, to approximately 40 million people on behalf of more than 1,000 employers and 100 health plans. Our mission is to create a healthier world by delivering solutions that:

1. Keep healthy people healthy
2. Reduce health-related risks
3. Assure the provision of evidence-based care to those who are ill

Healthways has provided health risk assessments (HRA) for more than 25 years. Our current instrument is the Healthways Well-Being Assessment? (WBA). The WBA is based on the latest evidence-based behavioral science available, including the work of Dee Edington, PhD, University of Michigan, James Prochaska, PhD, University of Rhode Island, and Janice Prochaska PhD, Pro-Change Behavior Systems, Inc.; Ron Kessler, PhD, Health Management Research Center; and Gallup, the worldwide leader in research of human nature and behavior. The WBA is configurable for different populations, is certified by the National Committee for Quality Assurance (NCQA), and is currently being implemented by commercial health plans, private sector employers and the Federal Government's Office of Personnel Management, Department of Interior and the General Services

		<p>Administration.</p> <p>Despite the HTA authors' findings of relatively consistent improvement among short term health measures for HRA participants, we are not surprised to learn that the literature review did not lead to conclusive evidence supporting a specific feedback or recommendation protocol. In our experience, developing and implementing interventions designed to modify behavior to achieve improved health is a complex, multi-component process. In a 2008 JAMA article titled, 'The Science of Improvement,' Berwick discusses the limitations of traditional study designs in evaluating effectiveness of these types of interventions. We concur with Berwick's assessment that the effectiveness of interventions geared toward behavior change is sensitive to many influences, environmental and otherwise.</p> <p>Similarly, we have found that taking into account a more comprehensive view of the environmental and other influences that affect individuals substantially improves the effectiveness of interventions intended to drive behavior change. We have invested heavily in the study and measure of the factors that drive individual behavior and the effectiveness of interventions intended to change it.</p>	<p>Thank you for the comment.</p> <p>Thank you for this information.</p>
--	--	---	--

		<p>In partnership with Gallup, Healthways created the Gallup-Healthways Well-Being Index (WBI). With over 1 million surveys of individual well-being completed, the WBI is a rich database of benchmark information on the well-being of the people across the United States. Using the WBI, we created the aforementioned Healthways Well-Being Assessment (WBA), a NCQA-certified, comprehensive tool that captures both 'traditional' HRA information and relevant information about an individual's home and/or other environments and their access to basic essentials needed for achieving and maintaining health and well-being.</p> <p>Through our work with the WBA we have found that understanding the context of an individual's daily life, health and lifestyle behaviors that otherwise may seem irrational, may actually be rooted in specific circumstances which may not be related to physical or emotional health per se. By arraying personal information over a broad range of domains that incorporate many environmental factors, the WBA enables a comprehensive view of an individual's well-being and provides important context for developing interventions to improve both health and overall well-being.</p> <p>Based on our success in developing interventions that drive behavior</p>	<p>Thank you for this information.</p> <p>Thank you for this information.</p> <p>This TA does not discuss coverage policy-related issues. Please direct coverage-related comments or questions to the</p>
--	--	---	---

change to improve well-being, we strongly recommend that the CDC and CMS require the use of HRA tools that are sufficiently robust to improve the likelihood of their effective use. For example, we recommend that HRA tools meet the standards of a recognized accreditation organization such as NCQA, URAC or JCAHO, and have established benchmarks for measuring well-being. Notwithstanding the HTA findings of inconclusive evidence, our own experience demonstrates that by setting an appropriately high bar in the forthcoming HRA guidance, CMS can help ensure that providers have access to high-quality tools that will help them, and CMS, succeed in achieving the triple aim of improving the patient experience, improving health and reducing cost.

In summary, in our experience, HRAs can and do succeed in their primary function of collecting data against which actionable intervention plans can be developed. In addition, well-designed HRAs that are integrated with a comprehensive and targeted intervention program are often successful in achieving their broader goal of improving health outcomes and overall well-being. We would welcome the opportunity to discuss this issue further with CMS and to share our research on the evidence of effectiveness with the CDC as it moves forward on the

Center for Medicare and Medicaid Services (CMS).

Please contact CMS and CDC directly.

			development of guidance on HRAs.	
Dr. K.R. Pelletier	NA	General	<p>1. Fundamentally, the question being answered is not clear to me. It appears they are attempting to determine if there is adequate science to determine if personalized information presented to a person completing an HRA produces and health outcome. I don't think the data are strong in that area ... but it is clear based on the number of studies they discarded that something is amiss with the science or the analytic.</p> <p>2. The definition of HRA as they posit in the article I might agree with; however, many would argue that "the HRA" is the sampling instrument and "the report" and "information given to inform and change behavior" generated by "the HRA" is a fundamentally different question for consideration;</p> <p>3. There are many HRAs tools in existence and most do not have strong published information about the multi-variate correlation between the quality and reliability of the "sampling instrument" and "the report" and "the health outcome";</p> <p>a. I think this absence of scientific information creates a vacuum that could lead the research team in the wrong direction if this paper is the only source of</p>	<p>The key questions answered in the TA are listed in the Executive Summary and at the end of Chapter 1: Introduction. The reviewer's observation about absence of strong data is consistent with findings presented in the TA.</p> <p>We excluded studies based on clearly enumerated, a priori inclusion/exclusion criteria (see Chapter 2: Methods).</p> <p>The reviewer did not specify what he thought was "amiss", so we could not address any concerns he might have with the TA.</p> <p>Thank you for this observation.</p> <p>Thank you for this observation.</p> <p>Thank you for this observation.</p>

awareness about HRAs;

b. There are many innovative sampling tools and methods with valuable health information reports attached that are not "in the published science" at the "level of evidence quality" that is preferred by these reviewers. HOWEVER, if they make fundamental policy decisions based on what is in this selected body of literature they may be missing some great opportunities to guide the market and researchers in what needs to be done to further research the value of HRA instruments and how HRAs could be best used. In my editorial opinion; I would clearly argue there is absence of evidence to make sound decisions (especially based on what I know about the market and seeing how few studies actually passed "muster" for inclusion in this review). I hope the review results in a decision that more research is needed; I think that is about all that can be concluded based on the research that was evaluated.

4. I didn't see information about internal validity assessment (other than one small reference in one study) related to the sampling instruments they assessed. I think this is a concern; if the studies they selected did not use valid instruments and they were only looking for the health outcome of the "intervention" post sampling; there is a fundamental weakness in the

Thank you for this observation.

We present reliability and validity data in our response to key question 1a (Chapter 3: Results).

			assessment. There was some modest assessment of Inter-rater reliability.	
Steve Phillips	Johnson & Johnson	General	<p>The analyses assumed that HRAs are first and foremost standalone behavior change interventions. We would argue that they are better viewed as triaging tools designed to promote population health management by referring people to the appropriate services which could be delivered in various modalities (e.g., high-touch or digital coaching). We would argue that the impact of HRAs cannot be assessed in isolation, apart from the subsequent services participants receive. We would also argue that an HRA is only as effective as the services that result from it.</p> <p>The HRA studies reviewed looked at outcomes across all different risk factors, including smoking cessation, weight loss, physical activity, cardiovascular health, and so on. To compare outcomes across these categories is comparing apples to oranges. On the one hand, we may not expect to be equally successful in all these areas, for reasons such as variation in the effectiveness of the therapy options to address different conditions. On the other hand, the potential overall health effect of the risk factors observed may be different across these risk factors as well.</p>	<p>We clarified our definition of HRAs in the final version of the TA to read as follows (p. 1): “For the purpose of this technology assessment, our definition of an HRA contained three components: participants provided self-reported information to identify individual risk factors for disease; participants received individualized health-related feedback based on the information they provided; and the information was used to give participants at least one recommendation or intervention to promote health, sustain function, or prevent disease. Any HRA, regardless of its delivery mechanism (e.g., single or multiple questionnaire administration, use of written feedback material, counseling, resource referral, etc.), that fulfilled these three criteria was included in the review.” We believe our conception of an HRA includes subsequent services that participants receive on account of the feedback.</p> <p>We included Table 5 in the TA to delineate outcome types (e.g., obesity, cardiovascular health) by article and give readers a sense of the variety of outcomes used in the literature. Statistically significant outcomes are identified in the table, on a per article basis, to provide the type of individual-level detail suggested by the comment.</p>
Steve	Johnson &	Methods	The authors applied a methodology	We did not exclude any articles from the TA based on quality

Phillips	Johnson		<p>to evaluating the quality and reliability of published studies, rightfully excluding those which failed to meet basic standards of scientific rigor. They then made systematic ratings of scientific quality of the 115 studies remaining. Of the 80 randomized controlled trials, 55% were rated as 'poor', 45% as 'fair' and none as 'good'. Ratings of the 35 cohort studies were better, but only 30% of those studies were rated as 'good'. Such low ratings ought to make one very hesitant about coming to any general conclusions about HRAs based on the results of those studies.</p>	<p>score.</p> <p>Readers must consider article quality in their appraisal of the evidence.</p>
Steve Phillips	Johnson & Johnson	Results	<p>The authors point out other problems in the studies sampled, for example, many studies failed to control for confounding variables such as gender.</p> <p>Moreover, they point out that the vast majority of the participants in these studies were not senior citizens. Since some of the critical risk factors don't emerge until later in the life span (e.g., cognitive impairment), it's necessary to ask different questions of the elderly. So how can we generalize from the results of younger cohorts to a Medicare population?</p> <p>The HRA studies reviewed were published over the last 30 years. Technology has gone through rapid changes during that time. While the authors indicated that many of the</p>	<p>Thank you.</p> <p>We agree with comment about generalizability (see key question # 3 in Results and Discussion).</p> <p>Thank you for this observation.</p>

			<p>instruments studied provide 'personalized' feedback, that doesn't mean they were tailored, as we understand that term. This would especially apply to instruments that were not computer based. Research has shown again and again that the degree/depth of tailoring determines the effectiveness of health-related messaging, as demonstrated by a variety of measures (e.g., eye movements, fMRI responses, behavior change, and so on). We would argue that HRAs without tailored feedback are going to be inherently less effective than their more tailored counterparts.</p>	
Louis Tze-ching Yen	University of Michigan	General	<p>This draft report is intended to be a technology assessment where the study subject is the Health Risk Appraisal (HRA). Although this report is solely based on literature reviews primarily from peer-reviewed research publications, those studies did not evaluate the HRA itself as a technology and its effectiveness in identifying health risks or facilitating health interventions. Most of the selected publications chosen for this review used the health questionnaire as tool of measurement in a research or intervention setting and therefore could not support the primary goal as claimed by the authors—that is, to “describe the key features of HRAs and examine which features were associated with successful HRAs and discuss the application of HRAs to a Medicare population.”</p>	<p>We defined 'HRA' to include more than just instruments developed to collect data on risk factors for disease. We based our definition of HRA on the RAND report (RAND Santa Monica CAL. Evidence Report and Evidence-based Recommendations: Health Risk Appraisals and Medicare. Contract no.: 500-98-0281. Baltimore, MD: Centers for Medicare and Medicaid Services, 2003). Our definition is as follows: “For the purpose of this technology assessment, our definition of an HRA contained three components: participants provided self-reported information to identify individual risk factors for disease; participants received individualized health-related feedback based on the information they provided; and the information was used to give participants at least one recommendation or intervention to promote health, sustain function, or prevent disease. Any HRA, regardless of its delivery mechanism (e.g., single or multiple questionnaire administration, use of written feedback material, counseling, resource referral, etc.), that fulfilled these three criteria was included in the review”.</p> <p>Our definition is consistent with the Centres for Disease Control and Prevention (CDC) definition of an HRA. CDC defines HRA as “...an assessment tool used to evaluate an</p>

			<p>This study served more as a review of various research studies which used health questionnaire (not necessary a HRA) as a measurement tool—describing key features of those selected research studies and examining which features were associated with successful research design or implementation. In this way, the HRA itself is no longer the subject of major study; thus, the authors failed to answer the research questions that they originally proposed. There are three major weaknesses in this report as to the research design and implementation:</p> <p>1) <i>a lack of understanding, learning, description, introduction and review of the origin, development and current status of the HRA</i></p>	<p>individual's health. An HRA could include a health survey or questionnaire (see Employee Health Survey); physical examination, or laboratory tests resulting in a profile of individual health risks often with accompanying advice or strategies to reduce the risks." (Centres for Disease Control and Prevention. Workplace Health Promotion – Glossary Terms. Atlanta: Centres for Disease Control and Prevention. Available at: http://www.cdc.gov/workplacehealthpromotion/glossary/#H. Accessed on: March 9, 2011.) We added the CDC definition to the text of the TA.</p> <p>The TA was designed to answer the key questions using the aforementioned definition of HRA.</p> <p>Please refer to our previous comment above.</p> <p>Our mandate was not to provide an in-depth review of the origins of HRAs.</p> <p>The reviewer provides insufficient detail regarding the HRA</p>
--	--	--	--	--

		<p><i>technology</i>—HRA features were not thoroughly and rigorously studied throughout this proposal;</p> <p>2) <i>no efforts given for the direct assessment of the HRA technology</i>—No surveys and no research were conducted to directly evaluate the HRA technology via HRA providers or HRA users; instead, the study used literature searches to evaluate how various researchers have used the health assessment questionnaire as a tool.</p> <p>Theoretically, a technology assessment should focus on the components of the HRAs: the quality of its questionnaire and the health risk appraisal algorithm, the totality and the usefulness of the profile report, the value of the follow-up strategy and the possible implementation and outcomes. Instead the current report expanded the assessment much beyond the HRA technology.</p> <p>3) <i>use of “better health outcome” as the only outcome measure for the success of the HRA technology.</i> The HRA has been used as a tool for health education and awareness. A person who takes an HRA and receives a profile or even several follow-ups would not necessarily receive a “better health outcome.” This reviewer does not believe there is a causal relationship between simply taking an HRA and a “better</p>	<p>features that he claims were not “thoroughly and vigorously studied” in the TA. Without further detail, we cannot address the comment</p> <p>AHRQ-commissioned technology assessments are systematic reviews and do not involve primary data collection.</p> <p>The TA responded to the key questions in the task order from AHRQ. We used an expansive definition of HRA that was consistent with previous reviews in the area (RAND Santa Monica CAL. Evidence Report and Evidence-based Recommendations: Health Risk Appraisals and Medicare. Contract no.: 500-98-0281. Baltimore, MD: Centers for Medicare and Medicaid Services, 2003).</p> <p>Key question # 2 directed us to examine “better health outcomes”. AHRQ, CMS, and CDC approved our use of the HRA definition quoted in a previous comment above.</p>
--	--	--	---

health outcome.” In worksite health promotion and disease prevention, the HRA is positioned as a gateway program to help individuals improve health awareness, attitude, and belief. “Better health outcome” is often refers to health behaviors and health status. Programs for behavioral change - such as smoking cessation, weight control, and increase in physical activity - are not part of the HRA technology, but parts of the comprehensive health promotion and disease prevention program that may complement the HRA. In this proposal, the authors were asked and attempted to evaluate the HRA technology; however, they actually selected and evaluated comprehensive health programs that used some HRA technology as a gateway program by using “better health outcome” as a criterion in their literature search. The authors expanded the definition of “HRA technology” too broadly. This particular limitation is related to the first weakness that the authors showed in the current report—that is, a lack of basic knowledge about the HRA technology, including both the HRA science and its applications.

Overall, the study design is poor and does not fulfill the purposes of this RFP. The authors’ definition of the HRA was incomplete, and their report assessed the research conducted or published on using health assessments (not all of which

See previous comment.

			<p>could rightfully be classified as an HRA), which deviated from the stated purpose of the study (to assess the HRA technology). Because the study's inclusion/exclusion criteria lack a basic understanding of the applications of HRA technology in the field of health promotion and disease prevention, it provides very limited information to fulfill one of the major purposes of the grant—providing useful references for the implementation of the Affordable Care Act to offer an annual HRA to the Medicare population.</p>	
Louis Tze-ching Yen	University of Michigan	Executive Summary	<p>The authors did a good job in summarizing this report. This reviewer prefers to comment on the report section by section but believes that overall, the study was not on target.</p>	<p>Thank you for reviewing the TA.</p>
Louis Tze-ching Yen	University of Michigan	Introduction / Background	<p>The author claimed “no consensus definition exists for HRAs.” This statement is debatable. Apparently, the authors have not rigorously examined the HRA and its background. As the authors stated, the HRA was initially proposed in 1940 and was developed by Lewis C. Robbins, M.D. as a health education tool. In 1964, when the “Reports of the Surgeon General: Smoking and Health,” was released, the Society of Prospective Medicine (SPM) was founded, and six years later Dr. Robbins and Dr. Jack Hall published the first book on the HRA, How to Practice Prospective Medicine. Under Dr. Robbin’s</p>	<p>The fact that the reviewer argues for a more limited definition of HRA (i.e., the instrument only), while the TA utilized a more expansive definition (i.e., instrument + feedback + recommendation), highlights the lack of consensus on an HRA definition.</p> <p>We added a reference to the Hyners et al. (eds.) handbook in the introduction to the TA.</p>

leadership, the SPM played a major role in the development of the HRA into a health education and health assessment technology. The CDC released the first adult HRA nationwide in the late 1970s and by 1981, 12 HRA instruments were generally available. In the late 1980s, the Carter Center at Emory University worked with the CDC to release the 2nd version of an HRA for the US adult population. For over thirty years, until the late 1990s, the SPM was the academic society dedicated to the study of the HRA. The Annual Proceedings of the SPM (published for over 30 years) provide valuable resources and information for the authors and other researchers interested in the history, background and development of the HRA as a technology. Although the SPM has been dissolved, its last publication (edited by Gerald C Hyners, Kent W. Peterson, John W, Travis, James E Dewey, Janet J. Foerster, and Edward M. Framer) entitled The Society of Prospective Medicine Handbook of Health Assessment Tools was published by The Society of Prospective Medicine and The Institute for Health and Productivity in 1999. This 800 page document collects the salient research publications on the origin, development and the status of the HRA for more than 50 years. This particular resource should be included and highlighted in this current study; in fact, the authors did

have access to the handbook, since their reference 133 was a publication selected from this handbook.

Therefore, this reviewer believes the HRA technology was well-defined in the field of health promotion and disease prevention until the early 2000s, when the SPM disbanded. Between the late 1970s and the late 1980s, the CDC and university research groups became the major forces behind HRA promotion and research. The SPM was the main academic agent contributing significantly to the development of HRA technology. Since the 1990s, the rapid development of computer technologies and the commercialization of the HRA by the private sectors expanded HRA technology and widened its functions, academic areas, and health focuses. As reviewed by the SPM in 1999, there were at that time 46 HRA instruments, 28 Quality of Life/Health Status Assessments and 32 Lifestyle-Specific Health Assessments, where the later two categories of assessment were expanded from the HRA but used similar logic as the HRA. Thus, this reviewer agrees with authors' definition of the HRA that the "HRA are techniques or processes of gathering information to develop health profiles, using the profiles to estimate further risks of adverse health outcomes, and providing

Thank you for the HRA overview.

Key question # 2 required that we focus on health outcomes. Estimating future risk is not a health outcome per se and was therefore excluded from the TA.

		<p>persons with feedback on means of reducing their health risks.” However, this reviewer did not see the authors fully integrate the complete definition above into the study inclusion criteria, especially those functions “using the profiles to estimate further risks of adverse health outcomes”.</p> <p>The section on reviews of the history of HRA was not adequate. This shortcome consequently guided this report into the direction of how the health assessment was used as a tool in research studies and away from the stated direction of the features of the HRA technology.</p> <p>The authors also reviewed the importance of implementing an HRA program among the elderly. This is the important background information.</p> <p>Finally, they formulated three important key research questions as follows:</p> <ol style="list-style-type: none"> 1. What are the characteristics of the provision of HRAs? 2. What characteristics of HRAs are associated with better health outcomes? 3. What is the generalizability of the data in Questions1 and 2 to the Medicare population or subpopulation? 	<p>Please refer to our previous comments above.</p> <p>Noted.</p> <p>CMS formulated the questions, not the authors.</p>
--	--	---	---

			<p>The first key question was on target.</p> <p>Another question, concerning the utilization (and repeated utilization) rates and the numbers of individuals who have completed the HRA, should be the direct outcome measures of any technology assessment but were omitted by the authors. Instead the authors put “better health outcomes” as the successful HRA technology outcome measure, which caused the error in the research design.</p> <p>HRA technology is the only one of the many programs to associate with “better health outcomes” either in a research or program setting. As a result, the authors changed the HRA technology assessment to an assessment of a research study or a program implementation where an HRA might be included as a measurement tool or as a program component to improve personal health.</p>	<p>Noted.</p> <p>The TA addressed CMS’s questions, which did not pertain to utilization rates, but concerned health outcomes.</p> <p>Please refer to our previous comments above.</p>
Louis Tze-ching Yen	University of Michigan	Methods	<p>The methods were based entirely through review of the literature, mainly from published journal articles where the authors changed the technology assessment objective from the HRA itself to research studies that used the health assessment as a tool.</p> <p>Unfortunately, the literature review (using an inclusion criterion of RCT</p>	<p>AHRQ TAs are systematic reviews of primary literature; please refer to our previous comments above.</p> <p>The reviewer’s comment is welded in his definition of an HRA, which is different from the definition in the TA. Please</p>

or cohort and case studies) would not be sufficient to adequately answer the first Key Question, especially with all the sub-questions listed beneath it. Thus, the authors' search criteria on research articles focused primarily on the 2nd Key Research Question for "better health outcome." This focus resulted in the following shortcomings: 1) some valuable HRA-related studies were excluded due to "no health outcome measured"; 2) some studies that did not include a true HRA program component (just a health assessment questionnaire) were included; and 3) some studies had "better health outcomes" that did not directly result from the HRA program.

Using literature review to respond the 3rd research question was a questionable strategy. If the authors spent enough time researching the HRA history, developments, and its current status, this question would have been proposed differently. Originally, the HRA technology was developed from and used among healthy adult populations for health education. Both the CDC and Carter Center's HRA stated clearly that HRA instruments were suitable for people aged 19 to 64 and those free of diseases. Because of this qualification, the authors would have difficulty finding existing publications that used an HRA only within the

refer to our previous comments on this issue above.

We included studies with comparison groups to evaluate health outcomes in HRAs with the three components of our definition (data collection + feedback + recommendation) versus HRAs with less than these three components (or versus some other non-HRA program).

We addressed the key question as proposed by CMS.

We found few publications in the Medicare population; all of these publications related to persons aged 65 years or over.

Given the interest in using HRAs in the Medicare population, a search of the literature to find information in this group, as well as an assessment of whether studies done in persons aged 19-64 years can be generalized to the Medicare population, are valid lines of inquiry.

Medicare population. In fact, there have been studies on retirees using an HRA.

In addition to literature reviews using publication databases, this reviewer believes three research approaches are needed: 1) extension of the literature search to those HRA *users* who were selected by the authors with a telephone interview, in order to collect detailed information on the features of HRAs they used; 2) a survey of HRA *providers* for the features of their HRA instruments and their HRA usage status; and, 3) a search through the Internet to focus on HRA providers and their HRA features.

The authors used selected keywords to search titles and abstracts through publication databases and also "hand searched" the American Journal of Health Promotion. However, this reviewer believes that running a literature search based on author - rather than on keywords - would be a more focused strategy. In general, most HRA technologies have been developed by certain research groups, i.e. Dee Edington and The University of Michigan Health Management Research Center; Nicolas Pronk and his colleagues; and so on. All of these groups have extensive research publications on the use of the HRA

Collection of primary data (points 1 and 2) is not a component of a systematic review and is therefore beyond the scope of the TA. Point 3 would be relevant if we were asked to develop a compendium of HRA risk factor assessment questionnaires, which was not part of our mandate.

Search strategies based on well-known authors or research groups can introduce bias into systematic reviews, e.g., only a handful of relevant studies, focused on limited interventions, with a narrow range of conclusions, would likely be included in such reviews. Objective search strategies use keywords linked to the specific research questions guiding the review. A professional medical librarian, with expertise in systematic reviews, developed our search strategy.

We extracted data on articles that passed three levels of screening based on our inclusion/exclusion criteria. If studies from the researchers cited in the reviewer's comment (e.g., Edington, Pronk) were excluded from the TA, then these studies did not meet our specified criteria, which were geared to answering our three key questions.

as a tool for worksite health promotion programs. However, few studies from these researchers were selected in this current proposal.

The reason that these publications that used an HRA as a measurement or program tool were not selected for the review might be due to the authors' inclusion/exclusion criteria.

The authors only included the studies of randomized controlled trials (RTC) or observational studies with comparison groups (e.g. cohort, case control). However, as the authors acknowledged, most HRAs have been used at worksites among voluntary, convenient samples. Since the authors used this research design/sample selection criterion, they missed a majority of the research publications related to the HRA, especially those publications with a focus on the HRA features listed by the authors as sub-questions of the Research Question 1. For example, only three of the 100+ journal articles using HRA data published by Edington and his associates were selected by the authors in the current study. Using these exclusion criteria made the main purpose of this study more difficult to achieve — that is, the discussion of the applicability of HRAs to the Medicare population, because the Affordable Care Act

We agree.

We were tasked with assessing whether certain characteristics of HRAs produced better health outcomes. A basic tenet of evidence-based medicine/practice is to evaluate health outcomes in comparative studies. Consequently, we included studies with comparison groups and excluded case series (no comparison groups).

authorizes Medicare to cover an annual HRA. This Act would result in HRA distributions to voluntary, convenient samples among the Medicare population, a similar distribution to most of the HRA-related research studies at worksites.

The authors were not only biased in their exclusion criteria but also in the inclusion criteria. As the authors acknowledged in their introduction, one of the key characteristics of the Health Risk Appraisal was its appraisal function (or, to use the authors' words, "using the profiles to estimate further risks of adverse health outcomes"). However, this inclusion criterion was not used by the authors.

The authors stated that "our definition of an HRA contained three components: participants provided self-reported information to identify individual risk factors for diseases; participants received individualized health-related feedback based on information they provided; and the information was used to give participants at least one recommendation or intervention to promote health, sustain function, or prevent disease." Thus, by changing the definition, the authors switched the research target from an HRA technology with a program component to any health

Key question # 2 directed us to examine "better health outcomes". Therefore, we followed standard systematic review methodology and designed our inclusion/exclusion criteria to meet the demand of this question. While HRAs can be used to estimate future adverse health risks, this specific issue was not within the scope of the TA.

Please refer to our previous comments above.

			assessment without an appraisal process during the questionnaire to profile procedure. Of the 115 studies the authors selected to study, some of those studies inadvertently used a non-HRA assessment as a tool due to this excessively-wide inclusion criterion.	
Louis Tze-ching Yen	University of Michigan	Results	<p>The authors excluded 5,434 studies from 5,972 unique citations followed two levels of title and abstract screening. No details described what criteria were used for the 1st and 2nd round screening to give this 91% elimination rate.</p> <p>Of the 538 citations promoted to full text screening, 423 were excluded and 115 proceeded to full data extraction and quality assessment. One of this reviewer's publications was selected as the extracted study. However, the review does not appear to be correct and the subsequent conclusion was completely wrong. As shown in Table 3 of the Results, in the "HRA Instruments Used in the Extracted Studies" column, the "Instrument Reference for this study (reference 26)" was listed as "No details on HRA or GM program –This HRA is a product of the StayWell Company, which has won the C. Everett Koop Award in the past. See reference for Gold, 2000 in this chart" and at the "Tools/Instruments" column was listed "HRA in tandem with General Motors LifeSteps program." Actually, the HRA used in this study</p>	<p>We applied the inclusion and exclusion criteria listed in Chapter 2: Methods to remove the citations.</p> <p>We agree with the reviewer that some detail of the LifeSteps Program is contained in the article and we deleted the phrase "No details on HRA or GM program" from the table.</p> <p>The LifeSteps Program won the Koop Award in 2004 and we mentioned this in the table as having won the award "in the past". To clarify, we changed the table cell entry to read "The LifeSteps Program won a C. Everett Koop National Health Award in 2004".</p> <p>When we looked into the LifeSteps Program, we visited the www.lifesteps.com webpage and were directed to a website copyrighted by The StayWell Company, LLC. This prompted us to write "This HRA is a product of the StayWell Company." Of course, our definition of HRA includes the full LifeSteps Program, not just a risk-factor questionnaire. Consequently, our comment in the table appears consistent with the current state of affairs; i.e., LifeSteps is a product of StayWell. To clarify this issue, we added the following underlined word to the text in the table cell: "This HRA is <u>currently</u> a product of the StayWell Company."</p> <p>We deleted the reference to Gold in the table cell.</p>

was not a product of the StayWell Company, and the GM LifeSteps program was described in detail in this article. Furthermore, at the time of study, General Motors had not even applied for the Koop Award and the LifeSteps program have never used StayWell Company's HRA until today. Based on these misleading statements from the authors, this reviewer questions the quality of the authors' work, although the authors claimed they had "reviewed the extracted data to confirm the accuracy of the work."

However, the poor quality of this particular review and its erroneous conclusion were not the primary problems with this proposal.

Rather, the main problem was the authors' focus on studies that might not be involved in HRA technology.

The authors should focus their study on HRA technology, and this reviewer sees no value in Table 1 and Table 2 of the results section, where the authors tried to assign a quality score for the studies published as traditional academic research. The assessment of the quality of the extracted publications had very little relationship to the assessment of HRA technology and the HRA features, such as its

The reviewer is unclear about what he considers the "erroneous conclusion".

Please refer to our previous comments about differences in defining HRAs.

Quality assessment is an integral part of summarizing a body of evidence in a systematic review and forms a basic tenet of systematic review methodology.

		<p>questionnaire, algorithm, profile, and follow-up capacity.</p> <p>Again, relying on literature searches as a research tool may not achieve the main research objectives indicated in the study. Five key sub-questions of Research Question 1 on the characteristics and the provision of HRAs could not be adequately answered due to insufficient data, mainly because the researchers under review did not present those HRA characteristics in their publication (given that these publications used the HRA chiefly as a personal health measurement tool or part of overall health promotion programs).</p> <p>The Table 3 in the study was necessary and informative. However, the structure of the table was dictated by the different studies, not by the HRA provider or the HRA itself. For example: If this table were categorized by the HRA (i.e. HRA for research purpose; HRA from commercial use, such as HealthTrac HRA; StayWell HRA; WebMD HRA, etc), it would be more focused and more useful.</p> <p>In Table 4, a drop-out rate was presented as a major outcome measure on success of the HRA without any indication of a participation rate. This drop-out rate</p>	<p>Please refer to our previous comments.</p> <p>The standard method of presenting extracted data in systematic reviews is by study. Since we did not limit our definition of HRAs to data collection instruments alone, we would not have been able to communicate any practical information using a table structure patterned after the reviewer's suggestion.</p> <p>The drop-out rate is the number of participants who completed a study (data available at the last follow-up time point) divided by the number of participants who began the study at baseline. We added this definition to the text.</p>
--	--	---	--

			<p>was confusing and not clearly defined. If an HRA participant completed an HRA questionnaire and then received a profile, at what point could a drop-out rate be measured? If dropping out occurs before referral to the follow-up programs, another measure – perhaps follow-up rate of the HRA participants - had to be presented before this dropout rate to make the measurement meaningful. However, the dropout rate of the HRA follow-up program or a comprehensive health promotion program with a health assessment component was not really the outcome of HRA and could not lead to a “better health outcome.”</p> <p>Table 5 described the HRA participants’ mean age and percentages of gender in place of the HRA targeting population—those people who received HRA in the authors’ selected studies. The contents of the table did not really correspond to the purpose of the table.</p>	<p>Table 5 lists six outcome domains used in the extracted HRA studies. The row for each study indicates whether the specific outcome was included in the study in question. The asterisk indicates whether the outcome was statistically significant at the 5% level in the study in question. Table 5 is introduced in the text as follows: “Many articles reported benefits for intervention groups in domains such as general health, lowered cholesterol, reductions in blood pressure, reduced fat intake, or improved physical activity”. Table 5 does not pertain to age and gender.</p>
Louis Tze-ching Yen	University of Michigan	Discussion/Conclusion	<p>The section on Quality Assessment in the study is irrelevant to the purpose of study. The HRA was neither designed nor intended for RCT use. Therefore, it was not surprising that none of the 99 RCTs associated with the HRA in this report were of good quality, since the HRA was never designed for this</p>	<p>Quality assessment is a standard component of a systematic review of comparative studies. We address the limitations of the quality assessment in Chapter 4: Discussion.</p> <p>One purpose of this TA is to provide an evidence-based perspective on the impact of HRAs on health outcomes. Assessment of evidence requires consideration of quality (Agency for Healthcare Research and Quality. Methods Reference Guide for Effectiveness and Comparative</p>

		<p>purpose. The authors gave poor quality ratings of the studies/publications from a clinically evaluative point of view, even though the HRA has been largely implemented as a health education activity based on voluntary participation among targeted populations.</p> <p>Comments on the authors' discussions on the five sub-questions regarding Key Research Question 1 are as following:</p> <p>a. Specific HRAs The authors stated, "most articles were concerned with general or cardiovascular health assessment... Articles using questionnaires only often asked participants to self-report previous diagnoses or risk factor for disease (e.g. smoking). HRAs designed for specific objectives such as improving diet rather than improving general health often utilized questionnaires to elicit information on items like participants' food intake." This reviewer does not agree with the authors on their inclusion criteria. Those "specific" HRAs classified by the authors were not true HRs; they were a health assessment questionnaire and should be excluded from this study.</p> <p>b. HRA administration methods.</p>	<p>Effectiveness Reviews, Version 1.0 [Draft posted Oct. 2007]. Rockville, MD. Available at: http://effectivehealthcare.ahrq.gov/repFiles/2007_10DraftMethodsGuide.pdf)</p> <p>The conclusions are based on the articles included in the TA.</p> <p>HRAs as per our definition were not associated with one specific place of administration versus another place of</p>
--	--	---	--

		<p>The authors were able to classify of 115 articles based place of HRA distribution - at the workplace (53 articles) vs. the non-workplace. This observation was the only major finding they could report for this sub-question. The authors stated that “no specific type of HRA was associated with one place of administration versus another.” This inadequate finding arose primarily from a poor research strategy (described earlier) and a lack of knowledge and understanding of HRA technology.</p> <p>c. The training of personnel who administered HRAs. The authors found “21 different descriptions in 78 articles” from the 115 articles they selected. They stated, “we relied on authors’ descriptions of training... although most authors did not detail the specific training regiments required of staff.” If the authors focused on HRA users or providers as the major data resources for this report, this problem would be avoided.</p> <p>d. The methods and frequencies of follow-up for HRA. The authors were able to provide some non-systemic answers regarding this question. They stated: “In workplace HRAs, meeting were a favorite means of contact... Some community-based HRAs also</p>	<p>administration</p> <p>The key question asked us to delineate the training of personnel who delivered HRAs. The many descriptions are not a “problem”, merely a statement of the contents of the extracted studies.</p> <p>We explain what we mean by “follow-up” in Chapter 4: Discussion, under the section for Question 1d. ‘Follow-up’ is the number of contacts between the HRA program and participants over time (relevant for us since we defined HRAs as being more than the collection of risk factors). The reviewer’s comment arises since he defines HRAs differently than us.</p>
--	--	---	---

involved in-person meetings... The increasing popularity of the Internet...however, HRAs administered entirely online might not reach groups at highest risk for chronic disease. Persons in these groups may be less likely to have Internet access... This issue was largely unstudied in the extracted articles, so more work is needed to assess the efficacy of online HRA delivery... Future research should involve more follow-up. ...This prevented us from assessing the durability of HRA over time. ...Further research is required to elicit the specific factors that influence participation in HRAs.” In this sub-question, the authors confused two different concepts: methods of HRA delivery and follow-up of the persons who completed an HRA. The authors were unable to reach clear conclusions in either area and simply suggested further research on the topics.

e. The characteristics of the patient population who received HRAs. This research sub-question, in principle, was not correct. The HRA is a technology that was developed and has been used mainly for healthy populations, not patient populations. The question asks who received HRAs and the authors’ answers were limited to “typical participants,” which is not an

CMS determined the wording of the question. The important word in the question is “populations”, which included any individual enrolled in an extracted study. This particular question was not restricted to the Medicare population.

informative or useful answer. Conceivably, this question could be important for Research Question 3 and the Affordable Care Act, which authorizes Medicare to cover an annual HRA for the Medicare population and not just Medicare patients. Since a blanket distribution of an annual HRA to the Medicare population will be forthcoming, it will be vital to learn the demographic characteristics of the people aged 65 or over who received an HRA or who were more likely to respond an HRA. Such information about demographics and response characteristics among the elderly could be very instructive but was not addressed in this proposal.

The authors concluded Key Research Question 2: “the evidence did not suggest a clear set of characteristics (of HRAs) that were associated with better health outcomes.” Their conclusion was correct and resulted from their inadequate determination of the research question. In actuality, Research Question 2 should focus on the fundamental measurement of a technology or product such as the HRA on user numbers and rates. The discussion determined that these “better health outcomes” were beyond the scope of the study and the data that the authors collected for this report.

Key question # 2 specifically mandated us to obtain evidence on health outcomes, not to collect data on user numbers or counts.

Based on their inclusion/exclusion criteria, the authors “found 16 (of 115) articles included the members of the Medicare population” for Key Research Question 3 - the generalizability of the data (HRA related) to the Medicare population or subpopulation. However, the authors “cannot readily generalize results from HRA studies in persons aged less than 65 years to persons aged 65 or over.” However, there were at least two published studies from Edington’s group that did not make the first two rounds of eliminations but focused on Medicare population; there is also another article that addresses a retiree population that may be relevant to the current study. The three articles are listed below:

- Musich, Shirley A., Aartee Phatak, Timothy McDonald, David Hirschland, Dee W. Edington. Self-Reported Utilization of Preventive health Services Among Retired Employees 65 Years and Older. Journal of American Geriatrics Society. 49(12): 1665-1672, 2001.
- Ronald J. Ozminkowski, Ron Z. Goetzel, Feifei Wang, Teresa B. Gibson, David Shechter, Shirley Musich, Joel Bender, Dee W. Edington. The Savings Gained From Participation in Health Promotion Programs For Medicare Beneficiaries. Journal of Occupational and Environmental

This article does not have a comparison group (all participants received the same intervention). The outcome measure is the percentage of persons who availed themselves of preventive services or health screenings within clinically recommended timeframes (not a health outcome). This article does not meet the criteria for inclusion in the TA.

The outcome in this article is cost savings. Costs are not a health outcome as defined in the TA, so this article does not meet the criteria for inclusion in the TA.

			<p>Medicine. 48(11):1125-1132, 2006.</p> <ul style="list-style-type: none"> • Yen, Louis, Alyssa B. Schultz, Timothy McDonald, Laura Champagne, Dee W. Edington. Participation in Employer-Sponsored Wellness Programs Before and After Retirement. American Journal of Health Behavior. 30(1):27-38, 2006. 	<p>The outcome in this article is participation rate in an HRA, which is not a health outcome, so this article does not meet the criteria for inclusion in the TA</p>
Louis Tze-ching Yen	University of Michigan	Tables	<p>Comments on Tables 1-3 were mainly presented in the Results section.</p> <p>The organization of Table 4 (Methods and Frequencies of Follow-up) was inadequate but its contents were necessary. The table combined HRA questionnaire distribution, HRA delivery method, HRA profile contents, recruiting for HRA participation, and HRA follow-ups. This table should be organized according to HRA features and broken into several tables in a conceptually similar manner as that of the HRA process procedure: i.e. HRA questionnaire distribution and delivery; HRA profile contents and HRA feedback delivery; and, HRA follow-up programs.</p> <p>The contents of Table 5 attempted to document “better health outcome” evidences based on the extracted articles selected by the authors. The authors called Table 5 an “Evidence Table.” However, these “evidences” documented the intervention programs that were associated with the HRA programs and were</p>	<p>The standard method of presenting extracted data in systematic reviews is by study.</p> <p>We defined an HRA as being inclusive of these programs.</p>

			<p>essentially independent of the HRA technology used in the HRA programs.</p> <p>In addition, the categories used in this table were not exclusive of one another. Table 5 provided the outcomes resulted from health promotion programs or activities presented in those selected articles, but these were the evidences of the comprehensive programs with a health assessment component. These evidences, if any, were not the direct results of HRA technology but were, rather, the result of an broad health management program that included an HRA as part of its armamentarium. In other words, the authors incorrectly attributed those “evidence” to HRA technology.</p> <p>In conclusion, three of the five tables did not provide valid information for HRA technology assessment and the other two tables were useful but poorly organized. In addition, this reviewer questions the content validity of these tables based on the authors’ inaccurate descriptions on an article written by the reviewer (Reference 26).</p>	<p>Please see our previous comment, as well as other comments related to our definition of an HRA.</p> <p>The reviewer’s comments on the tables stem from his divergent view of the definition of an HRA. As we mentioned earlier, a consensus definition of an HRA is lacking and the tables are consistent with the definition of an HRA that was used in this report.</p> <p>Regarding reference 26, the information in the table was fundamentally correct: LifeSteps did win the ‘Koop Award’ and it is currently attached to StayWell.</p>
Louis Tze-ching Yen	University of Michigan	Figures	<p>Figure 1 was useful, but the reasons for the first two rounds of elimination were not presented and explained. This reviewer believes that many significant articles regarding HRA technology were excluded, including the last handbook published by the SPM in 1999.</p>	<p>Please see previous comments.</p>

- Hyners GC, Peterson KW, Travis JW, Dewey JE, Foerster JJ, Frammer EM, Eds. *The Society of Prospective Medicine Handbook of Health Assessment Tools*. Indianapolis, IN: The Society of Prospective Medicine. 1999:135-144.

Figures 2 & 3 did not provided valid information for HRA technology assessment, the same as for Tables 1 and 2.

Figure 4 supposedly summarized the objectives of the HRA. The intention was good but the execution faltered. Theoretically, if the objective of an HRA is based on general health issues, it should also include objectives for reducing cardiovascular risks, improvement of smoking, weight, and physical activity behaviors, and the other objectives listed in the same axis in the figure. This reviewer believe that only those articles with an objectives on general health (60) or cardiovascular diseases (49) could be defined as using an HRA, while the others used assessments based on single behaviors—smoking (10); weight (14); physical activity (31); or other (34). Strictly speaking, these single-focus assessments should not be considered HRAs.

Thus, out of the 115 articles that

Please see previous comments regarding different definitions of HRAs.

The stated objectives of several studies involved single behaviors. Based on our definition of an HRA, single-behavior studies could be included in the TA, provided they met our other inclusion/exclusion criteria. Please see previous comments regarding our definition of an HRA.

We reviewed the numbers in Figure 4 and Table 5; we

		<p>made the authors' final cut, there would be maximally 109 articles that could be considered to contain an HRA (with no overlap on 60 general health and 49 cardiovascular disease articles). If Figure 4 presents the same data as Table 5, there may be another 10-20 or more articles that would be excluded—since the authors classified some HRAs with the objectives on both general health and cardiovascular disease.</p> <p>Figure 5 was a good for the type of HRAs.</p> <p>Figure 6 was a good for the method of HRA administration.</p> <p>Figure 7 had a good concept but it was not germane to the study, since it showed the number of articles selected that reported training on HRA administration. The figure could not generate an outcome to designate the quality of an HRA and also had little to do with HRA technology.</p> <p>Figure 8 suffered from the same problems as Table 4, in that it mixed two different concepts together: HRA profile delivery method and follow-up on HRA participants.</p> <p>Figure 9 contained useful information</p>	<p>updated the numbers in Figure 4 to match the numbers in Table 5. Please see our previous comment regarding the number of articles defined as including an HRA.</p> <p>Thank you.</p> <p>Thank you.</p> <p>We included this figure to summarize the data in Question 1c. This figure is unrelated to quality and was not connected to quality in the text. The figure itself was not meant to “generate an outcome”.</p> <p>Please see previous comments regarding different definitions of HRAs.</p> <p>Please see previous comments regarding different definitions</p>
--	--	--	---

		<p>but again, it counted HRA profile delivery as a 1st time follow-up.</p> <p>Figures 10-11 showed the scatter plot between “Numbers of Methods of Follow-up/Feedback” and “Frequency of Follow-up/Feedback.” Both figures should be deleted from the report since the data “failed to suggest possible linkages between tenacity of follow up and dropout rates.” With the mixed concept in HRA follow-up and HRA profile delivery and the loosely defined “drop out” rate from the extracted articles, there were no way to find any correlations between these confusing measurements created by the authors.</p> <p>Figures 12-13 attempted to answered Research Question 1-e on the mean age and gender distributions of the people who received an HRA. However, the authors presented both average values based the HRA participants’ information from the extracted articles, which is probably the only way they could derive the numbers from the publication.</p> <p>In sum, six of 13 figures (figures 2, 3, 10-13) provided invalid information to answer the research questions on HRA technology assessment. In the reviewer’s opinion, further clarifications are needed for the other 5 of the 7 figures.</p>	<p>of HRAs.</p> <p>Please see previous comments regarding different definitions of HRAs, as well as our clarification of the drop-out rate.</p> <p>Information should not be deleted from systematic reviews because it fails to show correlations (linkages, associations etc.). No correlation (linkage, association, etc.) is a relevant result.</p> <p>Thank you for the comment.</p> <p>Please see above comments on the figures</p>
--	--	---	---

<p>Louis Tze-ching Yen</p>	<p>University of Michigan</p>	<p>Appendix</p>	<p>Appendix A contains the computer codes for the literature search. As discussed previously, this reviewer believes the search using author names would be a better choice if the study were to focus on HRA technology through its providers and users. The current search strategy actually focused on health programs using health assessments, where it should focus on the HRA instrument primarily.</p> <p>Appendix B contains the authors' screening forms for the research articles. The authors used the JADAD scale and Newcastle-Ottawa Scale as screening criteria to select articles; as mentioned previously, those exclusion criteria or score methods set by the authors might exclude the majority of publications focused on the HRA.</p> <p>In addition, the Screening Questions for the HRA developed by the authors included significant numbers of the health promotion/wellness program studies without a true HRA program component. The Question 2 reflected this problem: Does this paper refer to health risk appraisal (sometimes called health risk assessment) OR focus on a health promotion/wellness program targeted to a specific, individually identifiable, population (such as employees at particular worksites)? It was clear that the phrase with the "OR" drew those program studies</p>	<p>Please see previous comments regarding different definitions of HRAs.</p> <p>Please see previous comments.</p> <p>Please see previous comments regarding different definitions of HRAs.</p>
----------------------------	-------------------------------	-----------------	--	--

		<p>without a true HRA program component into the current report.</p> <p>Appendix C listed excluded studies. In fact, the first two studies suggested by the reviewer for the authors to review for Medicare population were in this list of exclusions. The one study authored by Musich et al was rejected due to “no comparison group”---actually, the comparison group data on active employees aged 65 or younger than 65 was published in the same year.</p> <p>Another paper written by Ozminkowski et al was rejected because it “does not report health outcomes”---however, this article reported savings gained from participation in health promotion programs for Medicare Beneficiaries. The savings in the study were the medical expenditures which should be considered as “health outcomes.” The authors’ of this proposal appear to define health outcomes very narrowly—perhaps limited only to health behaviors.</p>	<p>The articles report data on the same program in two age groups, which means the articles taken together form a stratified analysis, but not a comparison of two programs. We defined comparison groups to be based on different HRA programs.</p> <p>We did not consider costs or expenditures to be health outcomes (see p.4 – ‘Literature Search Strategy’). Indeed, health economics separates health outcomes from costs in the calculation of the incremental cost-effectiveness ratio, which is the primary means of comparing programs in cost-effectiveness and cost-utility analysis.</p>
--	--	--	---

¹ Names are alphabetized by last name. Those who did not disclose name are labeled "Anonymous Reviewer 1," "Anonymous Reviewer 2," etc.

² Affiliation is labeled "NA" for those who did not disclose affiliation.

³ If listed, page number, line number, or section refers to the draft report.

⁴ If listed, page number, line number, or section refers to the final report.