Comparative Effectiveness Research Review Disposition of Comments Report

Research Review Title: *Emerging MRI Technologies for Imaging Musculoskeletal Disorders under Loading Stress*

Draft review available for public comment from December 27, 2010 to January 26, 2011.

Research Review Citation: Chung M, Dahabreh IJ, Hadar N, Gaylor J, Raticheck SJ, Trikalinos TA, Lau J. Emerging MRI Technologies for Imaging Musculoskeletal Disorders under Loading Stress. Technical Brief No. 7 (Prepared by the Tufts Evidence-based Practice Center under Contract No. HHSA 290-2007-100551.) AHRQ Publication No. 11-EHC024-EF. Rockville, MD: Agency for Healthcare Research and Quality. November 2011. Available at: www.effectivehealthcare.ahrq.gov/reports/final.cfm.

Comments to Research Review

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

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

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

Source: http://www.effectivehealthcare.ahrq.gov/index.cfm/search-for-guides-reviews-and-reports/?pageaction=displayproduct&productID=578
Published Online: August 2010

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #1	Executive Summary	The abstract does explain what the technology is for, what is important about the technology, and what are the most important issues about that technology that need to be addressed for load bearing MRI (actual/simulated). The positional and kinematic aspects are not mentioned in the purpose.	Please note that kinematic MRI (unless performed under stress loading conditions) was not considered within the scope of the Technical Brief. A clarifying statement has been added in the Methods section.
Dr. Jamie Phelan	Executive Summary	Concise with appropriate information relevant to the topic. Provides intriguing findings that urge readers to continue into the full report.	Thank you. No further response necessary.
Peer Reviewer #1	Introduction (background)	This section adequately describes the clinical problem that stress loading MRI is meant to address and discusses well the current medical practice as it relates to the clinical problem.	Thank you. No further response necessary.
Peer Reviewer #1	Introduction (guiding questions)	This section does note a change in terminology from "positional MRI" to "stress loading MRI" that occurred during the research process in order to more broadly capture the concept. While this makes some sense, it appears that non stress loading kinematic imaging was lost with this change and may be inappropriate (since kinematic might be considered a special form of positional imaging).	The preliminary protocol intended to capture our focus on imaging under loading stress (specifically with upright weight bearing MRI). During further stages of the review, based on Key Informant input and our own understanding of the literature, the scope was extended to include a broader set of devices.
Peer reviewer 2	Introduction	clear and concise	Thank you. No further response necessary.
Peer reviewer 3	Introduction	The Background needs to set the stage for the rest of the report. In order to do this more effectively: 1. There needs to be at least some mention that the magnet strength for some of the systems is less that 1 Tesla (0.2 for some, 0.5-0.6 for others) and that the resolution/quality of images is influenced by this ....although the general physical principles for all are the same, the image quality is not. Although this is described later under findings/other sections, the intro (and Executive Summary) leads one to believe that they are all the same.	We agree that image quality is indeed determined by magnetic field strength, as well as other key technical factors. We have added a sentence indicating that it is likely that between-device differences in field strength may have impact on image quality both in the Executive Summary and the Background section.
Peer reviewer 3	Introduction	2. For the Washington State report, one of the issues raised related to additional views (e.g. flexion/extenion for spine). There are 2 issues, one, do they add to the diagnostic ability and #2 additional cost of these extra views. It may be good to at least mention this as an issue as part of the context.	We have added the following information in the Background section: *"Additionally, the technology assessment could not determine whether technologies that allow positional imaging (for example, flexion and extension views) provide additional diagnostic information, despite the acquisition of non-neutral views being associated with additional costs."*

Commentator & Affiliation	Section	Comment	Response
Peer reviewer 3	Introduction	3. Page 13 lin 9 - Preclude may not be the appropriate word: diminish or alleviate may be more appropriate.	We have used “alleviate.”
Peer reviewer 3	Introduction	4. Somewhere, perhaps in an appendix, it would be nice to have a listing or the types of scanners covered in this report that is a bit more specific.	We agree that this information does not belong in the Background section. In page 20 (“Who are the current (major) manufacturers of stress-loading MRI devices? What is the current FDA clearance status of these MRI devices?”), we list specific devices and manufacturers. Appendix C presents photographs of most of the devices we included. Table 3 (section II. Evidence map) provides exact operational definitions and examples for major device categories.
Peer reviewer 3	Introduction	5. There should be a brief paragraph/overview of types of basic clinical epidemiology considerations that are important to evaluating the role a testing method - i.e. that across the populations for which the test is intended that it measure/evaluate what it is intended to evaluate and classify diseased vs. non-diseased (validity and accuracy) with be reproducible across populations for whom the test is intended. This would help set the stage for the implications and recommendations.	We believe that such an introduction to general epidemiologic principles would be out of place in the Introduction of the Technical Brief. Further, at the outset of this work one would not have been familiar with the specific epidemiological principles that apply (for example, in many papers MRI imaging was obtained from healthy individuals and some studies simply used the test to obtain anatomic measurements in patients with an established disease diagnosis, in both cases diagnostic accuracy is irrelevant). Additionally, other uses of the test could have been evaluated (for example predictive accuracy for assessing response to treatment); we feel that listing all relevant epidemiological concepts would be inappropriate for the Background section. Instead, we have expanded the relevant sections in the Future Steps section where the applicable epidemiologic principles fit better with the flow of the Technology Brief.
Peer Reviewer 3	(Introduction) Background, page 10, line 42	Need to point out that not all MRI imagers in this report aber between 1 and 3 T, but rather 0.5T or less [*sic*]	We have provided relevant information in the Methods and Findings section.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer 3	(Introduction) Background, page 10, lines 56-57	Should also point out here that strength also influences image quality/resolution to set stage for other aspects of report.	We have added a sentence to that effect.
Peer Reviewer 3	(Introduction) Background, page 13, line 10	preclude = prevent, may not best word - diminish or alleviate may be better	We have used “alleviate”.
Peer Reviewer 3	(Introduction) Background, page 13, line 25	need to mention that magnet strength in some of these systems is much less than conventional MRI and may influence resolution	We have added a sentence to that effect.
Peer Reviewer 3	(Introduction) Background, page 13, line 49	? and views taken - part of the issue for uMRI described in the WA report was the use of additional images	Relevant information has been added. Please also see our response to the related comment above.
Dr. Jamie Phelan	(Introduction) Background	Divided well so that the MRI systems are clearly defined and operations of the different machines are stated for complete understanding. Figure 1, however, was difficult to differentiate between the keyed lines. This is most likely due to the report being printed in black and white. May want to consider more significant differences in key lines to allow more of a contrast and ease in reading data outlined in the graph.	Thank you. It is difficult to make a line graph with many lines easily readable (black and white coloring is required). We have tried to make the different shades and line patterns more distinct.
Peer Reviewer #1	Methods	This section clearly and concisely describes how the data for this report was gathered and integrated showing balance and relative thoroughness. One related concept not completely expounded upon was kinematic imaging. The expanded scope for extremity MRI is interesting but the non-relevant aspects/studies with respect to the primary purpose should not be included in this report.	Thank you. Regarding kinematic MRI, please note that it was not considered within the scope of the review. We have added a clarifying note to that effect. We agree that extremity MRI is not directly related to weight bearing MRI. The inclusion of extremity-dedicated devices was based on Key Informant input and a joint decision by AHRQ and Tufts EPC. To address the reviewers concern, we have moved most of the material relevant to extremity dedicated MRI to an Appendix. Please note that some statements regarding these techniques have remained in the report to maintain continuity and for comparison reasons.

Commentator & Affiliation	Section	Comment	Response
Peer reviewer 2	Methods	including extremity dedicated scanners in the report does not seem justified.	Please see above responses regarding dedicated extremity MRI scanners.
Peer reviewer 3	Methods	Overall, the methods appear to be adequate with regard to the search and conceptual framework.	Thank you. No further response necessary.
Peer Reviewer 3	Methods	1. There needs to either be a description/definition of the categories and circles in the "Definitions' and "Data Extraction" portion and/or referenc to the figure and/or Table 3 before the findings so reader knows what these are.[*sic*]	We have provided a reference to Table 1 in the Definitions and Data Extraction sections.
Peer Reviewer 3	Methods	2. No explicit PICOD was presented but was mentioned as a basis for include information for studies for Categorie E equipment (pg 8 line 49, which seems to also contradict statement on pg 12, line 15-18). Exclusion criteria could be better specified.[*sic*]	As described above, by the very nature the Technical Brief, a strict definition of Populations, Interventions, Comparators and Outcomes cannot be adopted throughout. This is because of the exploratory nature of the work presented. For our systematic literature searches, the criteria used to determine study eligibility are listed in pages 9 and 10. The comment regarding dedicated extremity MRI devices probably is due to a misunderstanding. For studies of such devices we recorded PICO elements from the studies – we did not select the studies based on specific PICO elements. The only criterion used for study inclusion was the use of a device of interest. This also explains why the studies on dedicated extremity MRI (now listed in Appendix E) are so heterogeneous. We have revised the relevant Methods section to clarify this point.
Peer Reviewer 3	Methods	3. Search description/method seems adequate.	Thank you.
Peer Reviewer 3	Methods	Limitations of the MAUDE data base should be noted.	We agree that the MAUDE database is an imperfect source of information. We have added a citation to a review discussing some of the databases limitations and a caution regarding the databases limitations.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer 3	Methods	4. There is limited information on diagnostic criteria used for the various anatomic areas or how/if they were assessed.	This information was not collected from the primary studies. We adopted the definitions used in each study since the exact diagnostic criteria were often not reported.
Peer Reviewer 3	Methods	5. A systematic approach for describing the limitations of the studies overall could be described.	We have described our approach for making methodological recommendations for future studies in the "Data Presentation and Analysis" subsection of the Methods.
Peer Reviewer 3	Methods, page 14, line 15	somewhere should be a more specific list current types of equip that are included	Please see our response to the related comment above.
Peer Reviewer 3	Methods, page 16, line 16	There are limitations of MAUDE in general-- there is no denominator information, reports may be unverified, reports can be generated by providers, manufacturers, patients, attorneys.....there is opportunity to reports from multiple sources on the same individual.	Please see above for response related to the MAUDE database.
Peer Reviewer 3	Methods, page 17, lines 20-21	?? Refer to figure? ? circles; This appears to be the first mention of these and therefore unclear to reader	We have provided a reference to the relevant "Findings" sub-section. Please also see our response to related comments above.
Peer Reviewer 3	Methods, page 17, line 42	?? Categories A-D are what? Category E? Some reference to the figure and/or brief explanation of categories needed before the "findings" to allow reader to understand what is going on - before they get to the findings section	We have provided a reference to the relevant "Findings" sub-section. Please also see our response to related comments above.
Peer Reviewer 3	Methods, page 17, lines 48-49	Where is the PICO provided?	As discussed in response to relevant comments above, by the very nature the Technical Brief, a strict definition of Populations, Interventions, Comparators and Outcomes cannot be adopted throughout. This is because of the exploratory nature of the work presented. However, in our systematic literature searches this information was explicitly provided (please see Methods section).
Dr. Jamie Phelan	Methods	It seems that this section contains all of the needed information to understand the direction that draft reporters followed to acquire the final findings of the review. Using Key Informants as a means of data provided a "real" outlook on the information provided in the literature; a nice cross-reference since text information does not provide all perspectives.	Thank you. No further response necessary.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #1	Results (I. Description of stress-loading MRI)	This section covers the technology and its context very well.	Thank you. No further response necessary.
Peer Reviewer #1	Results (I. Description of stress-loading MRI)	The authors are willing to expand their scope to include extremity MRI yet they state that it is infeasible to review all studies with flexion/extension; this seems misguided. Are not the kinematic MRI studies more relevant to this report that a non-kinematic, non-stress loading extremity MRI study?	As noted above, we agree that extremity MRI is not directly related to weight bearing MRI. We have moved most of the material relevant to extremity dedicated MRI to an Appendix.
Peer Reviewer #1	Results (I. Description of stress-loading MRI)	Bore Configuration: wide-bore is also known as open-bore and should include this term as a synonym.	We have added a note to indicate that “wide-bore” scanners are often referred to as “open-bore”. We have also clarified that this term should be distinguished from “open MRI”.
Peer Reviewer #1	Results (II. Evidence Map)	This section clearly summarizes what studies have been done and comparing the evidence to the claims and questions for stress loading MRI as defined.	Thank you. No further response necessary.
Peer Reviewer #1	Results (III. Projected Uptake and Potential Growth)	This section summarizes potential directions predominantly from key informants and it seems that this section could be omitted and incorporated into the “Future Steps” section for organizational simplicity and ease of reference.	The section on “Projected Uptake and Potential Growth” attempts to list information from different sources, including manufacturers, Key Informants and our literature searches. Because the information presented does not represent our opinion (but was instead derived from external sources) we feel that it should remain in the Results section.
Peer Reviewer #1	Results (IV. Dedicated extremity MRI devices)	This section should be omitted since it is not part of the primary goals, appears out of scope, and lengthens the report unnecessarily.	We have moved all information relevant to dedicated extremity MRI to Appendix E.
Peer reviewer 2	Results	The results are well summarized. There appears to be a problem in Table 3 under category D which describes under examples "... or weight-bearing (static) postures in a conventional closed-MRI scanner" except that according to Figure 3 area D represents the intersection of stress-loading and open configuration so it cannot include weight-bearing postures in a conventional closed scanner. perhaps this is a cut & paste vestige from category C above it.	Thank you for your comment. Regarding Table 3, we have corrected the example provided for category D, as suggested.
Peer Reviewer 3	Results	d. Results: This section (Summary and implications) would benefit from additional clarification and detail.	We have revised this section (please see below for specific changes). Because of the limited data extraction performed as part of the Technical Brief there is a limit to the details we can provide.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer 3	Results	1. With regard to "potential benefits" - it seems that it would be logical to at least briefly discuss the validity of the tests (or lack of evidence) first before discussing safety OR related to indications for use. This would set the stage for Box 1. Before "assesing the impact" on outcomes, it seems logical that we need to know if the test can really do what it is supposed to do? [*sic*]	We agree that diagnostic validity needs to be established before assessing impact on clinical outcomes or decision-making. This approach was also suggested in our proposed framework for future research. In the first paragraph of the "Summary and implications" subsection we have added a comment that the evidence on the diagnostic validity of MRI under stress loading is limited.
Peer Reviewer 3	Results	2. Safety: The ramfiicaitons regarding potential need for alternative imaging methods (pg 16, lines 10-13) to rule out specific pathologies is to some extent a safety issue (e.g. ? radiation exposure?) that might be considered in more detail. Unfortnately, ss the validity and role in clinical decision making haven't really been established, it is probably not known how/what types of implications there may be for false + and false - with respect to safety. This could also be pointed out and has potential cost consequences. [*sic*]	We do not think this is absolutely accurate. Both in the FDA documents and the studies we reviewed, the additional imaging is obtained using conventional MRI modalities, thus there is not risk of radiation exposure. We respectfully disagree with the comment regarding the implications of false positive (or negative) results. Although, as we have pointed out in the text, the evidence on diagnostic validity is limited, the implications of false results can still be anticipated: for example, for diagnostic use of the test treatments (potentially including surgical procedures) and increased costs when none is necessary (for false positive results); and no administration of treatment when treatment is needed. What is unknown is the relative frequency of these eventualities (exactly because diagnostic accuracy has not been established).

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer 3	Results	3. The extent to which the stress-loading MRI may be used for replacement, triage or "add-on" may depend on application? For the spine, "additional views" ...i.e. extension/flexion views - do these not constitute "add-on" imaging? Are both "weight bearing " and non wt bearing image taken, again constituting an "add on"? [*sic*]	Here we have considered the use of specific devices, not specific imaging views. If one adopts the approach proposed by the reviewer, then every medical test is an add-on test (since no patient receives only one test). When comparing MRI devices, using weight-bearing MRI in addition to conventional MRI would constitute "use as an add-on test". In contrast, using weight-bearing MRI instead of conventional MRI (which is often proposed by the manufactures) would constitute "use as a replacement test". We hope that our revised answer to the question " To which populations and for what indications might stress-loading MRI apply? Is stress-loading MRI being proposed as a replacement, triage, or add-on test?" (page 20) clearly indicates that we are comparing MRI using devices to obtain images under stress loading versus conventional MRI (and not specific imaging views).
Peer Reviewer 3	Results	4. Costs- (pat 29 line 40) It is not clear what costs are referred to (device, facility, images, all of these? etc.) [*sic*]	We have clarified in the text (pages 20 & 21) that we attempted to address both types of costs. Note that costs of obtaining and operating stress-loading MRI devices are an obvious driver of costs to patients; both types of costs are of interest to healthcare policy decisionmakers.
Peer Reviewer 3	Results	5. There has been marketing of devices to attorneys for spine related cases.	We were able to confirm this information by perusing the websites of several legal firms. We have added a relevant sentence in the "Findings" section.
Peer Reviewer 3	Results	6. Evidence Map - It would be more helpful to have the description (sample sizes etc.) broken down by study type (e.g. comparative, longitudinal, validation, etc.) to get a better sense of the literature - or at least by area (e.g. spine). Although the bubble plot provides a bit a visual for this, overall, it seems misleading to talk about # of cases and # of controls across all study types without knowing how many correspond to comparative studies and how many to non-comparative studies for example. [*sic*]	We have provided the count of the number of studies (and comparative studies) as additional information on the bubble plot. We have also generated another graph that presents information stratified by anatomic location.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer 3	Results	Table 4. The label of "Adverse event" for this column does not seem appropriate- some are more patient limitations, limitations of the procedure, etc. completion rates, etc. not what one would classically call an adverse event.	We believe “Adverse events” to be an accurate description of the information contained in the Table. It may not be appropriate to consider pain or new neurologic deficits induced by axial compression as “patient limitations”. Furthermore, as we have indicated in the table, in cases of “incomplete exams” the reason was always the development of new or the worsening of existing neurological symptoms (pain, neurological deficits) that should qualify as “adverse events”. Note that given the lack of any follow-up information from the majority of studies, other adverse events from the imaging procedures were unlikely to be captured. We have also expanded the description of adverse events in the main text.
Peer Reviewer 3	Results	Table 5: Row titles seem in appropriate: for "study setting" these appear to be more study goals? vs. where (i.e. a physical settting) that the study was done. Exam location - Anatomic location or region is more accurate [*sic*]	We have replaced “setting” with “clinical setting” throughout the report. We have replaced “location” with “anatomic region”, as suggested.
Peer Reviewer 3	Results	Section describing diagnostic outcomes could be enhanced: a) little description of validity/accuracy evidence (or lack of it) is provided or whether diagnostic criteria were used, described, etc. These are important regardless of the role of the test. b) although the goal of this report is not extensive critical appraisal of the studies, additional context on limitations of available literature is needed - e.g. potential for bias is described for axial loading studies but not the other; c) in addition to the validity/accuracy of a test, studies on reliability are also important. The extent to which a-c above are not well addressed is important to describe in order to better understand (and lead to) the "implications and conclusions". [*sic*]	The Technical Brief did not look at specific outcomes from each study; we only provide a general classification of the research goals and study designs of individual studies. Consequently, we cannot provide a more detailed assessment of the potential for bias in each individual study. Instead, based on our reading of the specific studies and general epidemiological principles, we elected to present broad methodological guidance for future studies. We agree that reliability of a test (sometimes referred to as analytic validity) is important, in addition to diagnostic accuracy. However, we note that we did not extract information on the results for studies addressing analytic validity. We have revised the manuscript to indicate the importance of validity in the assessment of novel MRI devices.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer 3	Results	Reliability: For a test to be of use, it has to be reproducible, including the reading of the results - Inter and intra-rater reliability are important and have were not mentioned/addressed.	Please see our response to comment above.
Peer Reviewer 3	(Results) Findings, page 20, line 37	Implications of lower strength need to be briefly discussed some where	The interpretation of the findings from the literature review has been limited to the section entitled “Summary and implications”. Given the lack of quantitative data comparing different diagnostic modalities and the conflicting input from Key Informants, it was not considered appropriate to expand on this issue in the “Findings” section.
Peer Reviewer 3	(Results) Findings, page 23, line 8	This needs to be described earlier. Up to this point it would seem that all systems are comparable because they use the "same physical principles"	Mentions to the importance of low field strength in determining image quality have been added in the Background section. Please note that the (accurate) statement about physical principles does not logically imply “equivalence” in image quality, despite the reviewer’s assertion.
Peer Reviewer 3	(Results) Findings, page 23, lines 26-30	not clear what this means --- implications/correlation between patient limitations and recording of information is not clear	As we have argued above, pain or neurological symptoms developing during axially loaded imaging should not be considered a “patient limitation” but an adverse event. We have clarified in the text that this refers to new-onset pain/symptoms, developing during the procedure.
Peer Reviewer 3	(Results) Findings, page 24, line 48	See previous comments regarding limitation of MAUDE database	Please see above for response related to the MAUDE database. We have added a cautionary statement regarding the database’s limitations.
Peer Reviewer 3	(Results) Findings, page 25, lines 12-13	To what extent should this be considered as a safety issue in more detail. Since validity and role in clinical decision making haven't really been established, it is probably not known how/what types of implications there may be for false + and false -	Please see response above regarding safety issues, as well as our response regarding the association between test diagnostic validity and potential implications for patient safety/outcomes.
Peer Reviewer 3	(Results) Findings, page 27, line 13	For WA report, one of the primary issues was payment for "additional views" ...i.e. extension/flexion views - do these not constitute "add-on" imaging? Are both "weight bearing " and non wt bearing image taken?? again constituting an "add on"?	Please see our response above to relevant comments.
Peer Reviewer 3	(Results) Findings, page 28, line 40	Does this refer to the costs of the device, the images taken or both?	We have clarified this point in the text of the report. Please also see our response to the relevant comment above.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer 3	(Results) Findings, page 29, line 15	At the time of the 2007 report there was also direct marketing to attorneys and suggestions by attorneys that uMRI/positional images were important to whiplash and other cases	We have added this information in the report. Please also see our responses to relevant comments above.
Peer Reviewer 3	(Results) Findings, page 31, lines 23-24	I assume these stats include comparative an non-comparative studies. It may be more meaningful to separate out stats from comparative and non-comparative studies OR by type of study attempted (e.g. longitudinal studies, validation study, reliability study etc.)	The bubble plot in Figure 4 provides relevant information – we added numerical in the graph to facilitate interpretation, following the reviewer's suggestion. We have also provided additional images that present information regarding anatomic location.
Peer Reviewer 3	(Results) Findings, page 33, line 6	Adverse event needs to be defined and/or a different label used - some of these appear to be patient limitations to undergo imaging with the loading, wt bearing etc...? vs. an adverse event related to the safety of the device or procedure? Some are more "patient" characteristics - need for pain control prior to wt bearing	We respectfully disagree with this point. The examples provide indeed constitute adverse events that occurred during the image acquisition process. Please also see our response to the relevant comments above.
Peer Reviewer 3	(Results) Findings, page 35, line 10	need to indicate that number in ( ) is ? percent - ? of studies	We have provided this information. Thank you.
Peer Reviewer 3	(Results) Findings, page 35, line 38	? Better term - study goal or intent (setting to some may imply a physical location, type of center, etc.)	We have used "clinical setting" throughout the manuscript.
Peer Reviewer 3	(Results) Findings, page 35, line 46	Anatomy examined	We have adopted the suggested wording. Thank you.
Peer Reviewer 3	(Results) page 36, lines 13-15	Do studies that describe anatomic measurements also correlate those measurements with specific pain, clinical outcomes, etc. - if they don't this would be an important point to make to support the point about clinical applicability of these measures. Studies of measurement, and those of reliability should be enumerated separately	Studies of anatomic measurements did not report information on clinical outcomes. Unfortunately information on anatomic measurements was not captured separately from information on reliability. We note that is often not possible to separate the two types of studies (for example agreement of MRI measurements with those obtained during surgery provides information on reliability).

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer 3	(Results) Findings, page 37, line 35	hardly seems like an appropriate thing to compare anxiety on ?	The comment refers to data extracted directly from the study – the comparison (upright MRI versus myelography) reflects what was actually done in the study.
Peer Reviewer 3	(Results) Findings, page 37, line 43	Were diagnostic criteria provided? potential for bias is briefly described for axial loading studies, why not for the uMRI studies? OR the others described below	We did not extract data on diagnostic criteria; please see above for our response to relevant comments regarding criteria for diagnosis used in the primary studies. Please refer to Appendix D, where we present author overlap for all studies we considered in full text. Special mention was made to axially loaded studies because the "shared" authors were also inventors of the technology.
Peer Reviewer 3	(Results) Findings, page 39, lines 12-13	Unclear - is the intent to say that there is a need for research or that this is an area where there is increased research interest or enhanced research effort?	We have clarified that Key Informants considered this to be an "actively growing research field".
Peer Reviewer 3	(Results) Findings, page 39, line 42	Do they provide citations to support this?	The Key Informants were not required to provide citations for every statement. Instead, in the report's text we have carefully identified statements made by the Key Informants, distinguished that from the information uncovered by our searches and noted when both sources of information (i.e. the Key Informants and the published literature) were in agreement or disagreement. This approach has been consistently used in the "Findings" section.
Dr. Jamie Phelan	Findings	Excellent. The figures and tables used in this section are a terrific source of quick comparison of the findings listed in the text	Thank you. No further response necessary.
Dr. Jamie Phelan	Summary and Implications	Very much agreeable with the conclusions made by the draft reporters. I was actually surprised that the amount of literature available on the potential benefits and implications of stress-loading MRI is so lacking. However, the authors of this review appear to have extensively searched for data, and made educated conclusions based on the information available across the spectrum.	Thank you. No further response necessary.
Peer Reviewer #1	Discussion (next steps)	A conceptual framework is provided that could help organize future research and policy. Specific recommendations are given based on the summary presented in earlier section.	Thank you. No further response necessary.

Commentator & Affiliation	Section	Comment	Response
Peer reviewer 2	Discussion	Yes this section is well described.	Thank you. No further response necessary.
Peer Reviewer 3	Discussion	e. Discussion/ Conclusion: These sections could be clearer and would benefit from some re-consideration and re-organization of information presented.	We have revised the Future Steps section following the reviewers' comments. Please see our responses to the comments below.
Peer Reviewer 3	Discussion	Box 1: a)Nothing has been said about study reliability. b) Little information is presented in previous sections to support the statement that potential harms from FP may be substantial. c) it doesn't seem that meaningful cost-effectiveness modeling can be done given the available data - the accuracy/validity is unclear (thus false + and false - cannot be adequately modeled), the role (triage, replacement, add-on) is not clear and there are not high quality data available. [*sic*]	a) We agree that study reliability is an important aspect of the evaluation of imaging tests. We added a statement to that effect in Box 1. b) As noted above, although the frequency of False Positive and False Negative results is not well established for many of the devices we evaluated (because the diagnostic accuracy has not been fully assessed), the importance of false results can be considered known, given that MRI images are often used for clinical decision making (including the decisions for invasive surgical procedures – a point emphasized by our key informants). c) We agree that the accuracy of cost-effectiveness modeling would be improved with the availability of additional data. However, modeling, even in the presence of uncertainty can be useful for obtaining best- and worse-case scenario estimates, quantifying uncertainty and formalizing the clinical decision problem. We have clarified that the usefulness of modeling may be improved by acquiring higher quality diagnostic accuracy data.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer 3	Discussion	The reader should be reminded that MRI in general may not adequately discriminate between those who do/do not reqire intervention, that for the spine, there is not always a correlatio between finding on MRI and symptoms, and that there may be "findings" even in asymptomatic patients at least across studies of conventional MRI in the spine. It is unclear to what extent these new MRI techiques will enhance the diagnostic ability. [*sic*]	We have added the following comment: *"Increasing use of conventional MRI has been associated with increased utilization of orthopedic surgical procedures, with unclear effects on patient outcomes. The apparent limited impact of MRI on patient outcomes may be due to the limited ability of MRI to discriminate between patients who require intervention and those who do not, as indicated by the high frequency of positive MRI exams on clinically asymptomatic patients. Currently, the published evidence on stress-loading MRI is inadequate to determine whether use of these devices will improve patient outcomes compared to conventional imaging techniques."*
Peer Reviewer 3	Discussion (NEXT STEPS section)	a) some awkward/poor wording in initial paragraphs	We have revised the section for clarity.
Peer Reviewer 3	Discussion (NEXT STEPS section)	b) the conceptual framework is nice - some definitions of some componentes would be helpful -e.g. are false negatives considered "adverse events associated with testing" [*sic*]	Thank you for your comment. False negatives are not considered as adverse events associated with testing. Negative consequences of false negative results would be captured in our proposed framework as suboptimal outcomes of test directed treatment (as well as an excess of unnecessary treatment-related adverse effects). We have provided a more extensive legend, explicating particular components of the proposed analytic framework.
Peer Reviewer 3	Discussion (NEXT STEPS section)	c) the top of pg 45 "future research" suggestions does not flow logically from what is presented regarding the limitations or the conceptual framework. Has the "feasibility" been reasonably established? What about accuracy? The bullet order leads one to believe that the steps/recommendations are in order of importance. It seems to assume that appropriate, high quality studies validating/defining the accuracy have already been conducted and this is established, which doesn't appear to be the case from the information presented. [*sic*]	We have re-arranged the bullets to follow the proposed sequence. We have also specifically addressed test reliability and diagnostic accuracy in our recommendations.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer 3	Discussion (NEXT STEPS section)	d) From a clinical epidemiology perspective, there are several additional issues that are not mentioned, considered in the list, only a global "methodological issues" concern which is not formative.	It is beyond the scope of the Technical Brief to provide a step-by-step methodological tutorial in biostatistics and epidemiology. We have expanded the Table of methodological issues to include items suggested by the reviewer. Please also note that we argue from first principles, epidemiological or statistical, and would prefer to avoid using the term “clinical epidemiology” (please also see Miettinen O. et al., Journal of Evaluation in Clinical Practice, 2009).
Peer Reviewer 3	Discussion (NEXT STEPS section)	Why is the "most practical research strategy" described on page 48 line 28 and not presented as the first concept described/ most important things to establish first (prior to or with clinic decision making evaluation)?	We have re-arranged the Next Steps section to address the reviewer’s comments. We believe that the new structure, along with the additional recommendations made will prove helpful to those interested in designing further studies.
Peer Reviewer 3	Discussion (NEXT STEPS section)	Methodological Considerartions table: This table needs work. It should logically lay out specific next steps/solutions based on the most important methodological shortcomings. There are additional epi/methods issues in these studies (at least in those related to the spine that I am familiar with) that are not listed. Some proposed solutions are misleading -e.g. it isn't only in case-control studies that a representative patient population is needed or that need description of sampling methods - they all do. [*sic*]	As discussed in regards to previous comments, it is beyond the scope of this Technical Brief to provide a detailed step-by-step tutorial on designing future studies. Instead, we have opted to provide a list of recommendations that we feel are most pertinent (based on our reading of the relevant studies) along with references to detailed methodological sources (epidemiological and statistical) that can provide more detailed guidance for interested readers. We have expanded the original list of recommendations to include the reviewer’s suggestions.
Peer Reviewer 3	Discussion (NEXT STEPS section)	In addition to apprporiate power/sample size, the bottom-line, classic basic epidemiological methods considerations for validation include: •Broad spectrum of persons with the expected condition •Appropriate reference standard used •Adequate description of test and reference for replication •Blinded comparison of tests with appropriate reference standard •Reference standard performed independently of diagnostic test	We have expanded our list of recommendations to include some of those proposed by the reviewer.
Peer Reviewer 3	Discussion (NEXT STEPS section)	The bottom-line basic epi methods for reliability studies include: •Broad spectrum of persons with the expected condition •Adequate description of methods for replication •Blinded performance of tests, measurements or interpretation •Second test/interpretation performed independently of the first	We have expanded our list of recommendations to include some of those proposed by the reviewer. Please also see our response to the comment above.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer 3	Discussion (NEXT STEPS section)	Depending on the role of the test (e.g. triage) parameters in addition to sensitivity and specificity may be as/more appropriate to describe the consequences for testing (e.g. 1- negative predictive value), assuming there are appropriate studies (broad spectrum of patients)	We agree that sensitivity and specificity are not the most clinically interpretable information. At this point however, as the reviewer also pointed out, it is unclear what role the new technologies we evaluated will have in clinical practice (triage, replacement, add-on, etc). Given the many possible clinical settings, we prefer to suggest the reporting of sensitivity and specificity (in accordance with widely accepted reporting guidelines) instead of prevalence-dependent measures such as positive/negative predictive value. Please also note that the dependence between prevalence and predictive values is present regardless of the representativeness of the enrolled patient population. Further, predictive values derived from representative samples drawn from different populations are likely to be heterogeneous (if the prevalence of "true" disease is different).
Peer Reviewer 3	Discussion (NEXT STEPS section)	The text portion is heavily referenced but does it provides adequate direction for those who are looking to improve the quality of evidence. Part of the problem with many studies of diagnostic testing studies (including those that I am familiar with here) is that there is a basic lack of understanding about how to effectively design and execute a methodologically sound study based on epidemiological principles and approaches to reduce bias. More focus/guidance on the basics and logical flow of next steps is would be helpful to potential clinical/industry/other researchers. [*sic*]	As discussed above, it is beyond the scope of the Technical brief to provide a detailed step-by-step tutorial of the relevant biostatistical and epidemiological methods. We have however incorporated some of the reviewer's suggestions in our guidance for the design of future studies.
Dr. Jamie Phelan	Next Steps	Again, excellent use of figures and tables to draw the text together in this section. I agree that this is indeed a relevant topic to continue to gather information, primarily for the stress-loading and open bore MRI systems. There is a tremendous amount of variability in diagnostic testing, and further study conduction with literature review providing comparison of said studies will only enhance the quality and availability of all forms of MRI technology, thus, in turn, providing more comprehensive diagnoses and treatment plans for our patients.	Thank you. No further response necessary.
Peer Reviewer #1	Conclusion	Based on the data presented this section summarizes well the most important issues to be considered for this medical technology. This section also does well at identifying key decisional uncertainties.	Thank you. No further response necessary.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer 3	(Conclusion) Summary and Implications, page 42, line 20	are these limitations described adequately? See comments in previous section - potential limitaionts/biases pointed out for some types of studies/applications, not for others [*sic*]	Please see our responses to relevant comments above regarding strategies to avoid bias.
Peer Reviewer 3	(Conclusion) Summary and Implications, page 42, line 28	The validity (or reproducibility) is currently not well supported either Role in clinical decision making is also unclear (replace, triage, add on)	We have clarified these points in the text. Please also see our response to relevant comments above.
Peer Reviewer 3	(Conclusion) Summary and Implications, page 43, line 13	? in this report OR in the literature?	The Technology Brief did not present results from cost-effectiveness analyses (please consult the Methods section). We also did not identify such results from external studies. We have clarified this point in the "Summary and implications" section.
Peer Reviewer 3	(Conclusion) Summary and Implications, page 43, line 24	Awkwardly worded	We have rephrased the highlighted text.
Peer Reviewer 3	(Conclusion) Summary and Implications, page 43, line 27	Not much information presented in previous sections that support this statement (see previous comment)	Please see above for our responses to relevant comments
Peer Reviewer 3	(Conclusion) Summary and Implications, page 43, line 29	I question that meaningful cost-effectiveness modeling would be possible with available data - the accuracy/validity haven't been well established and the role hasn't been really established. There are not data from high quality studies available. It would seem that any true cost-utility or cost-effectiveness study would likely be based on poor quality data and guesses about the validity, implications of false + and - and role these scanners play in the diagnostic work up.	We have revised the text to indicate that additional data may be necessary to maximize the usefulness of any modeling effort. Please see our response above to the relevant comment.
Peer Reviewer 3	Next Steps, page 44, line 20	Poorly worded	We have rephrased the highlighted text.
Peer Reviewer 3	Next Steps page 44, lines 21-22	not sure what "despite the special features" has to do with assessment of tests following stepwise approach?	We have revised the sentence for clarity.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer 3	Next Steps, page 45, line 6	This bullet list leads one to believe that these are the next steps (in order of importance) and seems to assume that appropriate, high quality studies for levels 1 and 2 in the above diagram have been conducted... Given that other portions of the report indicate that such studies haven't been done and that there is some evidence maybe for feasiblity, but basically very sparse and limited evidence from poor quality studies on accuracy/validity, I suggest rewording and re-evaluation of the list [*sic*] Before these it seems that better information on the accuracy/validity and correlation between findings and symptoms is needed, The role of these test also needs to be better delineated. If it is to replace a test, that has one set of implicaitions (and need for documentation and research) vs. if it is to triage for further testing, that has a different set of implications (and need for documention/study) [*sic*]	We have revised the list and amended the sequence of recommendations in accordance with the reviewer's suggestion. Please also see above for our response to relevant comments.
Peer Reviewer 3	Next Steps, page 45, line 21	Is this really the first research priority? Isn't it as/more important to be sure that the text is valid, accurate and reproducible?	We have revised the order in which research priorities are presented. Please note that the ranking does not denote relative importance or priority. Please also see our response to relevant comments above.
Peer Reviewer 3	Next Steps, page 45, lines 26-27	what does this mean?? Language needs to be simplified and made more understandable.	This is standard terminology for an appropriate method to evaluate the impact of information derived from a diagnostic test on clinical thinking. We have provided a relevant reference for interested readers along with simplified descriptions of the proposed methods. Please note that commonly used approaches such as the percentage of diagnoses affected by the diagnosis tested are less sound methodologically.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer 3	Next Steps, page 45, line 49	Section Needs work there are additional issues from an Epi perspective and the proposed solutions need to be better thought out Reliability is also required for a robust diagnostic test (inter and intra-rater) RE: representative based - not just needed for c-c studies! Look like proposing C-C studies as a solution for not providing inclusion/exclusion criteria?? Doesn't make sense. In addition to blinded interp of test being compared, they need to be performed independently of each other [*sic*]	We have adopted the reviewer's suggestions. Please also see our response to relevant comments above.
Peer Reviewer 3	Next Steps, page 46, lines 7-9	Diagnostic criteria also need to be explicit, any thresholds clearly described and reliability of any thresholds assessed	We have addressed these suggestions in the text of the report. Please see above for our response to relevant comments.
Peer Reviewer 3	Next Steps, page 46, line 20	Correlated data analysis methods may be a better term to use?	We have used both terms.
Peer Reviewer 3	Next Steps, page 47, line 28	This should be stated up front before giving the list of research areas and issues to be dealt with	We have revised the Future Steps section following the reviewer's suggestion.
Peer Reviewer 3	Next Steps, page 47, line 29	Parameters in addition to sensitivity and specificity may be important depending on the role the test is to play. [not sure "predictive" accuracy is an appropriate term?	We agree that "diagnostic accuracy" would be a more appropriate term.
Peer Reviewer #1	Figures	Figure 3 is a Venn diagram: each zone A-G should have caption and description/definition for reference (definitions are provided in Evidence Map incompletely in Table 3).	We have provided a detailed description in the Figure legend.
Peer Reviewer #1	Figures	Figure 4 – should vertical columns reflect the relevant proposed categories A-G?	As detailed in the Methods section, we did not consider devices in all categories during our evidence review. For this reason we have only presented information only from studies of devices that were considered within scope. In the legend of Figure 4 the corresponding device categories from Figure 3 are listed. It was considered more practical to use abbreviations for the specific categories on the horizontal axis (instead of directly using the letters corresponding to those categories).

Commentator & Affiliation	Section	Comment	Response
Dr. Jamie Phelan	Figures	As noted above.	Please see above regarding Figure 1. No further response necessary.
Peer Reviewer #1	Tables	Table 3 – need F, G definitions	As noted above devices belonging to categories F and G were not considered in the Technical Brief. Definitions for the devices classified under these categories and reasons for exclusion are listed in the Methods section. For Table 3, we expanded the legend to provide a brief description of the listed device categories (A-E) and pointed readers to the Methods section for additional details.
Dr. Jamie Phelan	Tables	As noted above.	Thank you. No further response necessary.
Peer Reviewer #1	References	Mostly appropriate given the stated scope. As mentioned previously would remove the nonstress loading non-kinematic extremity MRI references and include studies of kinematic MRI under the rubric of positional imaging.	Kinematic MRI was considered outside the scope of the Topic Brief (please also see response to comments above). We moved all information relevant to dedicated extremity MRI to Appendix E.
Dr. Jamie Phelan	References	Number of references used, spanning across the past two decades, is appropriate and supportive to the findings of this draft.	Thank you. No further response necessary.
Peer Reviewer #1	Appendix C	Figure b: Should make a note that Signa SP is no longer manufactured or sold by GE.	We have added the relevant information. Please also note that all Figures have been updated since we have now used photographs for which we could obtain copyright permission.
Peer Reviewer #1	Appendix C	Figure e: This product is now owned by GE and has been rebranded. Since this device is not used for stress MRI, it may be omitted.	We have removed the photograph of this device from the Appendix.
Peer Reviewer 1	Appendix E	Should be omitted since it not part of the primary goals, appears out of scope, and lengthens the report unnecessarily.	As discussed above, the inclusion of dedicated extremity MRI was based on input from Key Informants, following discussion between AHRQ and the Tufts EPC. To address the reviewers concerns, we have removed all information about dedicated extremity scanners from the main text of the Technical Brief and have moved them to Appendix E. Please also see our responses to other points relevant to dedicated extremity MRI devices.

Commentator & Affiliation	Section	Comment	Response
Dr. Jamie Phelan	Appendices	Appendix A may require a brief explanation for those readers who may not be computer savvy. Otherwise, appendices are an excellent source of additional relevant information.	We have provided guidance in Appendix A. Readers may also consult the extensive methodological literature on search strategies – such a discussion is beyond the scope of the Technical Brief.
Peer Reviewer 1	General	Need to reconcile and define the terminology/nomenclature better and be more consistent in its use throughout the manuscript. The kinematic imaging aspect should also be expounded a bit more since this may be considered a form of positional MRI.	Because of the extremely large number of studies pertaining to kinematic MRI and the different technical aspects of kinematic imaging technologies, it was decided upfront that kinematic MRI imaging was outside the scope of the Technical Brief. We focused on stress-loading applications; extremity specific imaging information was collected based on input by the Key Informants and a joint decision by AHRQ and the Tufts EPC. We have provided some additional information on kinematic MRI and stated more clearly that such technologies were out of the scope of the Technical Brief. We have attempted to harmonize terminology throughout.
Peer Reviewer 1	General	Other modalities such as computed tomography and ultrasound could also be used for kinematic, positional or load bearing imaging. Radiography for spine and lower extremities is comely performed in weight bearing position. Perhaps a brief comment could be made to recognize this and state that it is out of scope of this technical brief but that some of the same issues and questions may exist for these other modalities.	We have added a note that non-MRI imaging modalities may be used to obtain imaging under stress loading (page 4). Kinematic imaging was not considered within the scope of the report (unless imaging was obtained under stress-loading conditions). (Please see response to comment above and as well as to comments on specific sections).

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer 2	General	Overall the report represents a clear statement of the issues and questions and a well done synthesis of the existing evidence base. The inclusion of dedicated extremity units which are not used for weight-bearing or loading stress applications, despite the authors explanation, seems added on and out of place and does not seem to add to the report.	Thank you for your comments. The inclusion of dedicated extremity MRI devices in the Technical Brief was based on Key Informant input and a joint decision by AHRQ and the Tufts EPC. To address the reviewers concerns, we have removed all information about dedicated extremity scanners from the main text of the Technical Brief and have moved them to Appendix F. Please also see our responses to other points relevant to dedicated extremity MRI devices.
Peer Reviewer 2	General (Clarity and Usability)	Well structured except for the add-on bit regarding extremity scanners.	Thank you for your comment. Please see above responses regarding dedicated extremity MRI scanners.
Peer Reviewer 3	General	The target population is wide and varied since the report includes a wide array of devices. An explicit definition is not apparent.	Because this is a Technical Brief, i.e. a broad overview of diverse emerging technologies definitions of Populations, Comparators, Interventions and Outcomes cannot be strictly defined. Even so, for the systematic literature review component our report, we made clear that the population of interest is "patients with musculoskeletal conditions" who are candidates for MRI imaging. This is a broad population description, but well suited to the Technical Brief's purpose.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer 3	General	The order of the key questions doesn't appear to support a logical flow for the practical recommendations for research.	The Key Questions of the report were defined based on extensive input from Key Informants (including stakeholders such as payers – public and private – radiologists, experts in evidence-based medicine and patients), AHRQ and the Tufts EPC. The questions do not follow the norms of comparative effectiveness reviews because of the distinct purposes a Technical Brief is expected to serve (i.e. provide a broad overview of a diverse set of emerging technologies). It is not within the scope of the Technical Brief to provide detailed recommendations for future research. Detailed recommendations for future studies are addressed in different types of documents (e.g., future research needs documents). Nonetheless, because some of the limitations of the current evidence base were apparent, we provided some specific suggestions that can be implemented in future studies.
Peer Reviewer 3	General (Clarity and Usability)	The initial part of the report was difficult to follow: e.g. there was mention of the "circles and categories" before these were defined and ther was not reference to the figure to assist.	We have revised the problematic subsections by providing clarifying references to appropriate sections of the report.
Peer Reviewer 3	General (Clarity and Usability)	Portions are poorly worded and unclear.	We have revised the report for clarity and used more uniform terminology.
Peer Reviewer 3	General (Clarity and Usability)	There are statements made in the summary and conclusions that don't always have suppporting information presented in the previous sections. [*sic*]	It is not clear which statements the Reviewer refers to. We hope that the revisions made throughout the report in structure and content will address this concern.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer 3	General (Clarity and Usability)	The logic and flow of the recommendations for next steps and future research needs work to better inform future research, practice decisions and policy. If there is truly interest in helping to inform clinical and research communities about the next steps to enhance evidence, there needs to be a logical flow and practical, prioritized "next steps" and clearer information (or basic references) for avoiding specific types of bias.	As mentioned above we have revised the report in an effort to improve clarity. It is beyond the scope of the Technical Brief to provide an extensive tutorial on epidemiological or biostatistical methods. We have provided an extensive reference list that interested readers can consult. This list covers topics ranging from general epidemiological principles to specific issues in the analysis of diagnostic test data in the absence of a gold standard. Furthermore, because we did not use explicit methods (qualitative or quantitative) to prioritize research needs, we cannot offer a detailed list of research priorities. Instead, we provide general guidance for future studies and an analytical framework that may facilitate the planning and design of such studies. We have considered the reviewers suggestions in revising our recommendations for future research.