Comparative Effectiveness Review Disposition of Comments Report

Research Review Title:
Lower Limb Prostheses: Measurement Instruments, Comparison of Component Effects by Subgroups, and Long-Term Outcomes

Draft research review was available for public comment from October 24 to November 21, 2017.

Suggested citation: Balk EM, Gazula A, Markozannes G, Kimmel HJ, Saldanha IJ, Resnik LJ, Trikalinos TA. Lower Limb Prostheses: Measurement Instruments, Comparison of Component Effects by Subgroups, and Long-Term Outcomes. Comparative Effectiveness Review No. 213. (Prepared by the Brown Evidence-based Practice Center under Contract No. 290-2015-00002-I.) AHRQ Publication No.18-EHC017-EF. Rockville, MD: Agency for Healthcare Research and Quality; September 2018. Posted final reports are located on the Effective Health Care Program search page. DOI: https://doi.org/10.23970/AHRQEPCCER213.

Response to Peer and Public Comments on this Research Review

The Evidence-based Practice Center (EPC) Program encourages the public to participate in the development of its research projects. A draft form of each research review is posted to the AHRQ Web site for public comment. Comments can be submitted via the Web site, mail or email. At the conclusion of the 3-4-week public comment period, authors use these comments to revise the draft research review.

In addition to public comments, each draft research review is independently evaluated by peer reviewers before it is finalized. Because they are chosen for their expertise in the subject matter and research methods, and freedom from conflict of interest, peer reviewers help to assure that the final report is accurate and free from bias.

The table below includes the original comments by peer reviewers and the public, as well as the authors' response for each comment that was submitted for the draft research review. Comments are not edited for spelling, grammar, or other content errors. Each public comment is listed with the name and affiliation of the commentator, if this information is provided. Peer reviewers are listed by number. 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 AHRQ.

Source: https://effectivehealthcare.ahrq.gov/topics/prosthesis/research
Published Online: September 6, 2018

Research review section	Reviewer, Affiliation	Comment	Response
Abstract	Public Reviewer #2 Andreas Kannenberg, Otto Bock HealthCare LP	Quote: "Currently, there is no evidence to support the selection of specific components for patient subgroups to maximize ambulation, function, and quality of life or to minimize abandonment or limited use." Comment: This conclusion should be added by another statement from the executive summary and/or main report that the systematic review did not attempt to review all evidence about comparisons between different types of components to prevent misinterpretation as no evidence for overall differences between specific types of prosthetic components. Also, it should be added in the results and conclusion sections that no evidence was found that the MFCLs/K-levels were good predictors which lower limb amputees would most benefit from a given component.	We have added in several sections clear statements that KQ 4 does not directly address the comparison of prostheses, per se. We have also added to the discussion that MFCLs have not been evaluated for their psychometric properties.
Abstract	Public Reviewer #4 American Orthotic Prosthetic Association	The structured abstract of the draft report provides general information regarding the background, methods, results, and conclusions of the AHRQ systematic review. AOPA is concerned that the language used in the "conclusion" section of the structured abstract is overly ambiguous and may lead to misinterpretation of what is presented in the literature. The language states, "Currently, there is not evidence to support the selection of specific components for patient subgroups to maximize ambulation, function, and quality of life or to minimize abandonment or limited use." AOPA is concerned that this statement may be interpreted to mean that there is no difference in quality and/or function of prosthetic components. The AHRQ conclusion that the existing evidence does not support the selection of specific components does not mean that there are not significant differences in function and performance of specific prosthetic components. Specific components are indeed selected each day matched to the patient's status, needs and desired (and possible) outcomes, and how the functions of the components match up with those objectives for the patient.	We have added that the evidence is too sparse to address this question. A lack of evidence is not evidence of a lack of effect. The review addresses only the existing study evidence.

Source: https://effectivehealthcare.ahrq.gov/topics/prosthesis/research
Published Online: September 6, 2018

Research review section	Reviewer, Affiliation	Comment	Response
Abstract	Public Reviewer #4 American Orthotic Prosthetic Association	AOPA is concerned that the "results" section of the structured abstract places inappropriate focus on the subject of abandonment of the prosthesis. The structured abstract states that review of the clinical literature contained in eight studies indicate that 11 to 22% of unilateral amputees who receive a prescription for a prosthesis, abandon their prosthesis within one year and that trans femoral amputees are twice as likely to abandon their prosthesis and trans tibial amputees. The decision to place emphasis on this in the draft report based on a limited number of studies is unusual, and must be questioned. A sounder premise might have been to point to recent reports indicating that fewer than 50% of amputees ever receive a prescription for a prosthetic device, that the likelihood of receiving such prescription declines even more markedly with each advancing decade of the patient's age, and the fact that studies in Western Europe show that patients there—Scotland and Scandinavia—have a much higher percentage of amputees receiving such prescriptions. Why is the U.S. behind other countries in terms of patients here being less likely to be prescribed a prosthetic device? This is more important than the rate of abandonment. But since AHRQ has chosen to focus on rates of abandonment, we must note that the better, more advanced (likely K-3) devices often include step monitors which actually have the capability to initiate a communication to the prosthetist if they are not used for a day or multiple days. This can prompt a patient contact and inquiry—is there a reason why the device has not been used? Is there a repair or adjustment of the device that can help? Has the patient been ill? Has the patient lost contact with the location of the device? Unfortunately, data shows that the access of U.S. Medicare patients to the better prosthetic devices, those more likely to include those step monitors has been interdicted in the past four years—down roughly 35% in the 2011-14 period contrasted to the previous few years, a	We have added information about the percentage of amputees who receive a prescription for a prosthesis to the background. We do not further address this issue since it was not among the Key Questions. The structured abstract summarizes the evidence pertaining to the Key Questions. This is also not a narrative review that addresses what devices exist that may improve LLP use or other features. We also do not address policy issues.

Research review section	Reviewer, Affiliation	Comment	Response
		sign that policy has been pointing in the wrong direction, reversing the potential impact of advances.	
Abstract	Public Reviewer #5 Orthotic and Prosthetic Alliance	The search date (i.e., November 30, 2016) should be updated to reflect the date of the final search. (See; p. v.)	Correct. That was the search date for the draft report.
Evidence Summary	Public Reviewer #2 Andreas Kannenberg, Otto Bock HealthCare LP	Quote: "The current standard approach for matching patients to prostheses relies heavily on performance-based assessments, self-assessments, and wearable monitoring technologies that record patient activity (3); although prosthetists often rely on clinical judgment to match patients to prostheses." Comment: In the U.S., clinical decision making is primarily driven by insurance coverage criteria rather than clinical judgment of the prosthetist. Internationally, Ottobock offers recommendations for component selection that are based on clinical considerations, but these often do not match the coverage criteria in the U.S. In order to avoid confusion of U.S. prosthetists, these recommendations for component selection therefore often differ between documents for the U.S. and other countries	We have removed the sentence about the choice of LLP in the US, as it is redundant with prior sentences.
Evidence Summary	Public Reviewer #2 Andreas Kannenberg, Otto Bock HealthCare LP	Quote: "Specifically, it excludes from evaluation biomechanical and other non-patient-centered intermediate outcomes. It also does not attempt to review all evidence about comparisons between specific components. Instead, it largely focuses on those comparisons, which provide within-study data to allow assessment about how components compare in different subpopulations of patients based on their characteristics." Comment: This is a very important statement that should also be made in the structured abstract to prevent misinterpretation of the finding that "Currently, there is no evidence to support the selection of specific components for patient subgroups to maximize ambulation, function, and quality of life or to minimize abandonment or limited use." as no evidence for overall differences in effects between different types of prosthetic components.	We have added this to the abstract, along with a statement that the review does not cover overall comparative effectiveness. These points are highlighted repeatedly throughout the document for clarity, specifically in the abstract, introduction (objectives), relevant results sections, and discussion.

Research review section	Reviewer, Affiliation	Comment	Response
Evidence Summary	Public Reviewer #2 Andreas Kannenberg, Otto Bock HealthCare LP	Quote: "The searches were conducted on November 30, 2016. [The searches will be updated in all databases upon submission of the draft report for peer and public review.] No publication date or language restrictions were applied." Comment: That means that the literature searches were performed before the public comment on the original research plan could be considered (deadline for the submission of comments on the research plan was December 20th, 2016). That does not shed a good light on the systematic review as it raises the question whether or not the public comments on the research plan were to be considered or ignored.	The literature searches were conducted early to help us gain insight on the issues and types of studies we would expect to consider. However, the study selection and all subsequent phases occurred after the protocol was approved. Furthermore, the search was updated in October 2017. We apologize for the poor wording in the draft.
Evidence Summary	Public Reviewer #2 Andreas Kannenberg, Otto Bock HealthCare LP	Also, it is noteworthy that no PICO questions were formulated, which is unusual for a systematic review.	We disagree. The questions follow standard PICO formulation.
Evidence Summary	Public Reviewer #2 Andreas Kannenberg, Otto Bock HealthCare LP	Key Questions 1-3 Comment: It should be mentioned in this section that, although many prosthetic studies reviewed had assessed K-levels as one of the patient characteristics, the K-levels themselves are a non-validated, very vague administrative patient classification that is hard, if not impossible to prove as their definitions leave a lot of room for interpretation and argument. In this context, though I fully agree with your recommendation to use validated measures to characterize patients in future studies, it is important to note that currently neither Medicare nor private insurances accept any of the validated assessment techniques, prediction tools, or outcome measures listed to support the determination of the MFCLs/K-levels. Inasmuch it is quite questionable how a better characterization of patient subgroups with validated measures could be used to direct coverage of prosthetic components in the future. That would require that either certain validated measures would be approved or accepted to inform the K-leveldetermination or that the MFCL system would be replaced by a new patient classification that is based on rehabilitation science and validated measures.	We have added that K levels have not been validated. We make no comment on what Medicare or other insurers use to support their coverage.

Research review section	Reviewer, Affiliation	Comment	Response
Evidence Summary	Public Reviewer #2 Andreas Kannenberg, Otto Bock HealthCare LP	Quote: "Overall, the studies do not provide evidence that any specific subgroup of patients consistently have differentially better outcomes with any specific component than other subgroups of patients." Comment: This statement raises the question what "consistently" means or requires. Does it mean that ALL studies would have had to demonstrate differentially better outcomes of certain components in specific subgroups of patients, including those studies that were underpowered or conducted with mediocre or inferior representatives of a component category? For example, the vast majority of studies with the C-Leg has shown consistent benefits in K2 and K3 patients, whereas most studies with other microprocessor knees have failed to do so. Also, you might want to consider the possibility that certain components deliver consistent benefits in most, if not all patient subgroups. In addition, this finding should be added by a statement re-emphasizing that the systematic review did not attempt and did not review all evidence comparing the overall effects of the different types of prosthetic components.	This sentence has been removed. We have added statements about what KQ 4 does not address; namely, it "does not focus on the relative effect of different components nor does it include the majority of studies that compare components (but do not provide subgroup analyses). Nor does it evaluate components based on biomechanical, nonclinical or other nonpatient-centered intermediate outcomes." This is noted in the abstract, introduction (objectives), relevant results sections, and discussion. For example, we included studies that directly compared K2 and K3 patients, but not studies that reported only that a component is effective among (combined) K2 or K3 patients.
Evidence Summary	Public Reviewer #2 Andreas Kannenberg, Otto Bock HealthCare LP	In this section of the evidence summary and main report, the study of Hahn et al., 2016, is disqualified as "methodologically flawed". I would like to ask you to consider that "flawed" is a very strong, if not derogatory term that should only be used for research of exceptionally poor quality. The study of Hahn et al. has, as any research study, methodological limitations that have been openly discussed by the authors in the original paper. It is my understanding that you will receive a detailed response of the authors on their papers published in 2015 and 2016. However, I would like to emphasize that, despite the methodological discussion that seems to focus to a large extend on less central aspects of the paper and that may at least in part be due to a misunderstanding of the analysis performed, the central point of the observation appears to have been missed by the reviewers. Responders to an intervention may be found in a wide range of demographic and	We agree that the language was too strong and have removed the offending phrase. We believe that the actual findings of the study are fairly narrow in extent, that among people selected by prosthetists to receive a Genium knee (based on their likelihood of benefiting) an analysis with technical methodological concerns failed to discern predictors of an unclear set of unvalidated outcomes. The issue is not that the prosthetists have a secret knowledge, but that the criteria used for selecting patients together with the prosthetists' experience and clinical knowledge result in highly biased sample of people who are likely to benefit. While this is a common issue, it was particularly notable in

Research review section	Reviewer, Affiliation	Comment	Response
		other characterizing parameters such as age, amputation etiology, mobility grade/K-levels and many more variables studied in this paper. It is especially interesting that the studies of Hahn et al. found that a relatively high percentage of K2 patients benefitted from microprocessor knees, which is consistent with the results of a number of clinical trials in that population (1-8). Inasmuch, the justification to generally withhold this technology from K2 patients in the U.S. appears to be highly questionable. Also, both studies have been criticized for "preselection bias" as subjects were deliberately invited to participate by their prosthetists. Given the multitude of patient characteristics analyzed for their predictive power to no avail in both studies, it is quite surprising that the reviewers seem to believe that the attending prosthetists had a kind of "secret knowledge" or even "secret stomach feeling" who of their patients may or may not benefit from the interventions tested but do not recommend research with these prosthetists to identify those assumed characteristics.	this study. The people in the available dataset were likely very different than a more general population of people receiving lower limb prostheses.
Evidence Summary	Public Reviewer #2 Andreas Kannenberg, Otto Bock HealthCare LP	Key Question 7 Quote: "Subgroup analyses in single studies tended to be underpowered to detect differences, mostly leading to determinations that the evidence was insufficient. However, we found a moderate strength of evidence, based on six studies, that about 11 to 22 percent of lower limb amputees who receive a LLP prescription abandon the prosthesis (stop using it) at about 1 year; these studies are generally representative of people with LLP, in particular older adults and those with dysvascular etiologies. Three of these studies provide low strength of evidence that people with unilateral transfemoral amputations are about twice as likely to abandon their LLP than those with unilateral transtibial amputations. Potential differences among other subgroups had insufficient evidence due to conflicting results among three studies or only a single, imprecise study with data. Also based on four, generally representative studies, there is low strength of evidence that 11 to 37 percent of LLP recipients	The evidence is sparse overall. For the outcome with moderate evidence (abandonment) studies were mostly consistent and we discuss differences across countries.

Research review section	Reviewer, Affiliation	Comment	Response
		use their prostheses only indoors; however, these studies are somewhat inconsistent and imprecise. There is low strength of evidence abouthow likely different subgroups of people use their prostheses only indoors, suggesting that people with transfemoral amputations, or who are older, or with bilateral amputations are more likely to be limited to indoor use. There is insufficient evidence about why people abandon their prostheses." Comment: When reviewing research done outside the United States to answer this key question, it should be considered that the primary prosthesis fitting rates after an amputation are much lower in the U.S. compared to other developed countries. In many European countries, specifically in Germany and Scandinavian countries, the health care systems require the attending physician to explain why a patient with an amputation is NOT a candidate for prosthesis fitting, whereas in the U.S. we face the opposite situation that the physician and prosthetist have to explain why a patient should be fit a prosthesis. Therefore, prosthesis abandonment rates and rates of sole indoor use of a prosthesis can be expected to be higher in European settings than in the U.S., as many patients who receive a prosthesis in Germany or Sweden would never get a prosthesis in the U.S	
Evidence Summary	Public Reviewer #2 Andreas Kannenberg, Otto Bock HealthCare LP	Quote: "Suboptimal matching of patients to LLPs may unnecessarily increase health care utilization, prevent attainment of maximal patient function, and defer realization of improved quality of life attainable with an appropriate prosthetic." Comment: Throughout the evidence summary and the main report, there is the repeated notion that inadequate matching of patients to prostheses may only result in increased health care utilization and "over-prescription". Only once in the whole report (page 3 of the main report) there is a more balanced statement that under-utilization or under-prescription may result in suboptimal outcomes either ("Variability and subjectivity in assigning or predicting the K level of prospective LLP recipients may inadvertently lead to	Thank you. This was an inadvertent omission. We have added "or "under-prescription," where people are given LLPs without capabilities they could benefit from" to the sentence in the Discussion where we talk about "over-prescription". This is the only other place than the Introduction that this issue is covered.

Research review section	Reviewer, Affiliation	Comment	Response
		inefficient or inappropriate LLP matching. 13 This can occur if a person receives a LLP allowed for lower K levels when a LLP allowed only for higher K levels would enable better function, or if a person receives a LLP approved for higher K levels, which might be unnecessarily complex for an individual who would have equivalent or better function with a simpler component."). Although I can understand that over-utilization is the main concern of CMS/Medicare, the repeated focus on this aspect may be considered a bias of the research group. I would therefore recommend to use the more balanced statement throughout the report	
Evidence Summary	Public Reviewer #2 Andreas Kannenberg, Otto Bock HealthCare LP	Quote: "Although not a limitation, per se, it should be noted that this review makes no attempt to make conclusions about the overall effects of different LLP components." Comment: This is a very important statement that should be repeated throughout the report more often to prevent misinterpretation of the results of the systematic review.	We have done so.
Evidence Summary	Public Reviewer #2 Andreas Kannenberg, Otto Bock HealthCare LP	Quote: "However, most of the studies that analyzed heterogeneity of treatment effect or provided data to allow subgroup analyses were observational and did not control for underlying differences during use of one component or the other. For example, studies did not describe or control for rehabilitation, training, or acclimation with each of the components. In particular, in the pre-post studies (where everyone switched from an old (simpler) to a new (more complex) LLP, one would expect that patient characteristics such as age, strength, and mobility will also have changed." Comment: I think the prosthetic research community would greatly benefit from an explanation how exactly rehabilitation, training, acclimation, age (?), strength, mobility, etc., should be controlled for in pre-post studies	We have added further language about controlling for differences in potential confounders.
Evidence Summary	Public Reviewer #2 Andreas Kannenberg, Otto Bock HealthCare LP	Quote: "To as great an extent as possible, studies should assess validated, patient-centered outcomes related to ambulation, function, quality of life, and related outcomes. … Ideally, studies should use a core set of validated, patient-	This may be a reasonable policy approach, but it is beyond the scope of this evidence review. Our review found limited evidence about the usefulness of standardized metrics

Research review section	Reviewer, Affiliation	Comment	Response
		centered outcomes (in addition to other study-specific outcomes, as needed)." Comment: Though I basically agree with this recommendation from the research perspective, it also requires alignment with CMS/Medicare/DME MACs as to what validated, patient-centered outcome measures may be used to corroborate the determination of the K-level of a patient. Without that alignment, it will not be possible to translate research results into reasonable coverage criteria that are based on K-levels. Ideally, either certain validated measures would be approved or accepted to inform the K-level determination or the MFCL system would have to be replaced by a new patient classification that is based on rehabilitation science and validated measures. I think it is absolutely necessary that the respective authorities understand that.	for distinguishing between patients by K level. However, importantly, we cannot make clinical or policy recommendations. These comments have been shared with CMS.
Evidence Summary	Public Reviewer #2 Andreas Kannenberg, Otto Bock HealthCare LP	Quote: "Lower limb amputees are clearly a highly heterogeneous group with distinct needs dependent upon age, etiology of limb loss, level of amputation, comorbidities and health status, postoperative stage, and rehabilitation status. Better understanding of which component would be best for which patient could both maximize individual's ambulation, function, and quality of life and minimize waste due to either abandonment or due to "over-prescription," where people are given LLPs with specific capabilities that they cannot benefit from." Comment: The prosthetic research community would greatly benefit from understanding what patient characteristics should be chosen to define "relevant subgroups". Also again, the language neglects "under-prescription" as a relevant problem in the U.S. healthcare system today.	Better studies are needed to determine evidence-based guidance on which patient (or other) characteristics should be chosen to define relevant subgroups. Key Question 4 attempted to answer this question, but the evidence is sparse and problematic. We have added a sentence that the evidence base does not yet answer this question. We have added language about under-prescriptions. This quoted sentence already talks about maximizing patient function and quality of life.
Evidence Summary	Public Reviewer #2 Andreas Kannenberg, Otto Bock HealthCare LP	Quote: "The majority of the evidence addresses the question of which components maximize ambulation and function in the average patient, as opposed to which component would best suit the needs of a given individual." Comment: This statement is true but does not only apply to prosthetic research but to the vast majority of pharmaceutical and	We have added to the quoted sentence that this issue is common across medical research. The suggested approach seems reasonable, but we cannot make policy or treatment recommendations.

Research review section	Reviewer, Affiliation	Comment	Response
		medical device research. While current evidence is still somewhat limited with regards to the number of subjects enrolled in prospective clinical studies, trial fittings may be used to determine whether or not patients benefit from certain components. Trial fittings for microprocessor components are a procedure commonly used in European health care settings (e.g. Germany, France, UK, Italy, Netherlands, Austria, Sweden, Norway) to honor promising but yet-limited evidence by making sure that only patients with proven individual benefits have access to this technology. I think that this approach would be a feasible compromise to bridge the gap between limited evidence and proven individual benefits of a patient with an amputation.	
Evidence Summary	Public Reviewer #4 American Orthotic Prosthetic Association	The "background" section of the evidence summary states the following. "The current standard approach for matching patients to prostheses relies heavily on performance based assessments, self-assessments, and wearable monitoring technologies that record patient activity; although prosthetists often rely on clinical judgment to match patients to prostheses." AOPA believes it is important to recognize that while component selection should always be based on the clinical needs of the patient, reality indicates that insurance coverage considerations remain a significant factor in prosthetic component selection and should be recognized as a contributing, and often a limiting factor.	This is an important point. We have added the sentence: Insurance coverage policies often dictate which prostheses and components are available for a given patient.
Evidence Summary	Public Reviewer #4 American Orthotic Prosthetic Association	The "objectives" section of the evidence summary states that, "Specifically, it excludes from evaluation bio-mechanical and other nonpatient-centered intermediate outcomes. It also does not attempt to review all evidence about comparisons between specific components. Instead, it largely focuses on those comparisons, which provide within study data to allow assessment about how components compare in different subpopulations of patients based on their characteristics." AOPA believes that the systematic review should have considered any evidence that discussed the comparison of	The systematic review addresses the Key Questions as laid out in the protocol. We now make it clearer what topics the review covers and does not cover. We agree that a general review comparing components would be of interest, but that is not the topic of this review.

Research review section	Reviewer, Affiliation	Comment	Response
		specific components, especially as they relate to improved patient outcomes as a result. The failure to consider studies that directly compare the effectiveness of one prosthetic component to another creates a significant issue regarding the validity of the results of the systematic review—the review seems to have systematically eliminated the very comparative effectiveness studies which should be its essence and focal point.	
Evidence Summary	Public Reviewer #4 American Orthotic Prosthetic Association	Another discrepancy that concerns AOPA is discussed in the "methods" section of the evidence summary which states that several databases were searched for clinical literature regarding lower limb prostheses through November 30, 2016. The AHRQ requested and received public comments regarding the key questions that would be used to prepare the systematic review through December 20, 2016. AOPA and several other organizations, such as The Orthotic and Prosthetic Alliance, worked diligently to provide valuable input regarding the key questions by the December 20, 2016 deadline. If the database search criteria was limited to studies available prior to November 30, 2016, AOPA is concerned that perhaps neither the comments that were provided on December 20, 2016 were not considered by the AHRQ in performing the systematic review, nor important new, cutting edge studies that have occurred in the past twelve months. Any systematic review needs first and foremost to be current and comprehensive at the time it is published. It appears this report may well have fallen short on both of those markers. In addition, the search of the databases did not indicate a beginning date for the search which could potentially allow the use of studies that are significantly aged and may have included prosthetic componentry that are outdated and/or no longer commercially available—some dated studies have continued relevance while others do not.	The search date should have been explained better in the draft review. It represented the date that the actual search was done but not the dates of the screening process, which took place in 2017. Subsequently the search (and screening) was updated through October 30, 2017. In discussions with Key Informants and Technical Experts, it was decided to not exclude older studies.
Evidence Summary	Public Reviewer #4	The next section of the evidence summary addresses the key questions that were used to direct the systematic review. In its	We believe we have included all potentially eligible studies. We have screened in full text

Research review section	Reviewer, Affiliation	Comment	Response
	American Orthotic Prosthetic Association	comments submitted on December 20, 2016, AOPA offered detailed responses to each of the key questions along with references to existing clinical literature that should be considered for inclusion in the systematic review. In reviewing the studies that were included in the AHRQ systematic review, it appears that very few, if any of the studies cited by AOPA in its comments were included in the systematic review. AOPA believes that exclusion of these studies represents a significant omission in the systematic review and urges the AHRQ to move proactively now to rectify these omissions and include these studies in its final report.	all references recommended by reviewers. Our eligibility criteria, particularly for Key Question 4, were strict in many instances. As we note the large majority of studies that compare components were not eligible because there was no possible assessment of heterogeneity of treatment effects. Appendix B includes the list of all rejected articles, including from peer reviewers, and reasons for rejection.
Evidence Summary	Public Reviewer #4 American Orthotic Prosthetic Association	In its December 20, 2016 comments on the key questions for inclusion in the AHRQ systematic review, AOPA pointed out the importance and relevance of in-progress research by both the RAND Corporation and the health economics firm Dobson DaVanzo. While both studies were incomplete at the time AOPA's comments were submitted, preliminary reports from both organizations were included as addenda to AOPA's comments. Both studies have now been completed and the final reports are included in these comments as addenda.	As per protocol, we included only published, peer reviewed articles.
Evidence Summary	Public Reviewer #4 American Orthotic Prosthetic Association	The RAND study, entitled Economic Value of Advanced Transfemoral Prosthetics developed a simulation model to assess the differential clinical outcomes and costs of microprocessor controlled prosthetic knees compared with non-microprocessor controlled prosthetic knees. It is based upon, and includes, a comprehensive review of all relevant scientific literature about lower limb prosthetics, and in this sense should be a good companion to the AHRQ report. One would have expected the RAND and AHRQ's contractor, conducting a literature review on the same general topic and at much the same time, would have significant overlap. Sadly, and for whatever reason, there is not such significant overlap, which causes us to question the appropriateness and comprehensiveness of the AHRQ contractor's systematic review. The final RAND report was released on the RAND	The RAND review and this review addressed largely non-overlapping questions, so it is not surprising that the evidence base differs. This review did not address economic issues. This review does not address the overall comparative effectiveness of components, only (as pertains to KQ 4) the heterogeneity of treatment effect.

Research review section	Reviewer, Affiliation	Comment	Response
		website on September 5, 2017, which is a fully citable, Index Medicus source. While it is expected to be published in a traditional print format, there is no reason or impediment to justify this important study not being included in the AHRQ literature review. The RAND study concludes that transfemoral amputees who do not receive a microprocessor knee are approximately 450% more likely to die as a result of a fall—that finding is critically important to AHRQ's area of investigation!	
Evidence Summary	Public Reviewer #4 American Orthotic Prosthetic Association	The Dobson DaVanzo study, entitled Retrospective Cohort Study of the Economic Value of Orthotic and Prosthetic Services Among Medicare Beneficiaries reviewed the extent to which patients who received select orthotic and prosthetic services, including lower limb prostheses, had less healthcare utilization, lower Medicare payments, and/or fewer negative outcomes than patients who did not receive orthotic and prosthetic services. The Dobson DaVanzo study is an update to a previous study that analyzed data from 2007 through 2010 to include data from 2011 through 2014. Sadly again, AHRQ's contractor's search did not include the 2015 report of the 2007-10 Dobson-DaVanzo study in Military Medicine, and also failed to take any cognizance of the updated 2011-14 data.	This review does not address the overall comparative effectiveness of components, only (as pertains to KQ 4) the heterogeneity of treatment effect. We have, however, referenced the article in the Background for the main report regarding remaining in the home and ER visits.
Evidence Summary	Public Reviewer #4 American Orthotic Prosthetic Association	AOPA strongly contends that the studies conducted by the RAND Corporation and Dobson DaVanzo are extremely valuable resources for inclusion in the final AHRQ report and urges the AHRQ to consider both studies for inclusion in its final report.	Based on the key questions of this report, these studies were not eligible.
Evidence Summary	Public Reviewer #5 Orthotic and Prosthetic Alliance	The Evidence Summary is sparsely cited (i.e., only 6 references for 24 pages of text and tables). As the Evidence Summary is likely to be the most widely read section of the review, it is important to summarize the review using appropriate evidence. (See; p. ES-1 – ES-24).	Per AHRQ style, the Evidence Summary can have only a limited number of references. Therefore, it is standard to reference the included studies only in the main report.
Evidence Summary	Public Reviewer #5	As noted above, we request the authors to qualify in the final report the statement "…deemed to be generally applicable to	The Key Questions and eligibility criteria are quite clearly not restricted to the Medicare

Research review section	Reviewer, Affiliation	Comment	Response
	Orthotic and Prosthetic Alliance	the Medicare population...” (p. ES-7). This is an important criterion for this review, as it significantly limited the evidence considered and included. It seems entirely appropriate to indicate in the evidence summary the percentage of participants with dysvascular amputation and participant ages that were considered “adequate” for inclusion in this review	population. We categorize studies for KQ 1-3 (and to a lesser degree other KQ) based on generalizability to the Medicare population, but this is not a criterion for inclusion.
Evidence Summary	Public Reviewer #5 Orthotic and Prosthetic Alliance	The authors have indicated that they “encourage journal editors to require use of validated measures.” (See; p. ES-20.) While we appreciate and agree with this comment, we would encourage the authors to clarify this point so it is clear that use of measures without formal evidence of validity may be appropriate in select instances (e.g., when measures to assess a particular trait or construct do not exist, or when ad hoc measures are combined with validated measures to expand upon the information that can be measured with “validated” measures).	We agree and have added this concept to the paragraph.
Evidence Summary	Public Reviewer #5 Orthotic and Prosthetic Alliance	Table A is confusing and likely to be difficult for readers to interpret. The table raises several important questions. For example, we are not clear what is meant by “(various)” after a measure title? Why are both italics and bold necessary for “population”? (Would plain text and bold suffice?) Under reliability and validity, what is the difference between a blank and “no” evidence? Why is “no” evidence bolded for validity and not reliability? Why does the IPAQ have “adequate” validity, and all other measures are “yes,” “no,” or blank? Why is “mix” usedas a means to indicate validity? Does a “yes” under floor/ceiling effects mean the measure has or does NOT have a floor or ceiling effect? (A “yes” in other columns is a desirable trait, whereas here that would be an undesirable trait of the measure.) What does “(most)” mean in regards to a floor/ceiling effect? We encourage the authors to clarify this table, perhaps by using a system to clarify the strength of evidence (e.g., +, ++, and +++ to indicate low, moderate, or strong evidence of validity).	We have removed the table and have simplified the presentation to lists of validated and reliable instruments. The revised summary tables for each instrument address these issues and we believe are clearer.

Research review section	Reviewer, Affiliation	Comment	Response
Introduction	TEP Reviewer #1	The introduction addresses MFCL and describes it's use in classifying functional level. Yet, it was not used as a criteria for determining if current outcome measures could differentiate between current MFCL levels since this is the current standard of prosthetic for prosthetic prescription in the Medicare Eligible population. Please explain why that was not explored	It is not the case that MFCL were not used as a criterion for evaluating instruments' test validity. However, studies rarely reported this as an outcome measure for validity assessment. K levels were used for predictive validity in a few cases (e.g., 1 leg standing balance) and not uncommonly for divergent validity (e.g., 6MWT). This should be much clearer in the new analysis and presentation of summary results for KQ 1-3.
Introduction	TEP Reviewer #2	The introduction helps delineate this "complicated problem" are highlighting that "patients are heterogeneous" pg 3 line 32.	Thank you
Introduction	TEP Reviewer #2	the definitions for function included as a footnote on g 9 lines 3-12 are to be highlighted: Functional and patient-centered outcomes include: quality of life, disability measures, activities of daily living, mobility measures, including use of prostheses only for transfers, self-care, pain, fatigue after use (e.g., end of day), daily activity, time LLP worn per day, falls, satisfaction with LLP, and others (but not simple preference of one component over another). Ambulatory outcomes include: gait speed, step count, walk distance; uneven or wet surface, low lighting walking; ramps and incline traversing; step/stair climbing function; ambulatory function measured in the community setting (e.g., self-report or activity monitors); achievement of bipedal ambulation; and other patient-centered ambulatory function measures. Adverse events include: skin ulcers and infections, injuries from falls due to mechanical failure, and other problems with prostheses.	These are the footnotes to the Analytic Framework. We do not think it is necessary to repeat the lists in the text.
Introduction	TEP Reviewer #3	Although, dysvascular disease is the largest cause of amputations in the United States. Individuals who lost their limb from other etiologies (e.g., congenital anomalies, cancer, and trauma) are disproportionately represented in the total limb loss community (Ziegler- Graham et al. 2008). This may be largely due to the relatively high mortality rate experienced by persons who lose a limb from dysvascular disease	Thank you. This is a good point. We have added this to the Introduction.

Research review section	Reviewer, Affiliation	Comment	Response
		(Robbins 2008). This is not clearly communicated in the introduction to the report.	
Introduction	Peer Reviewer #1	Minor comments: Introduction, Page 2, Line 12. The reference for "Medicare coverage guidance" is incorrect	Thank you. This has been corrected. It was reference 11.
Introduction	Peer Reviewer #3	The Introduction outlines the purpose for this systematic review however the title of the SR needs to more accurately reflect what the SR entails. Lower Limb Prosthesis (Prosthetics) in Medicare only patients. As is, this document will undoubtedly be used to extrapolate across all patient populations (not just those with Medicare) which will have negative implications on those do not meet this criteria.	This review is not limited to the Medicare population. The title has been change to Lower Limb Prostheses: Measurement Instruments, Comparison of Component Effects by Subgroups, and Long-Term Outcomes
Introduction	Peer Reviewer #4	pp 1-3 - well written, excellent presentation of the key problems with current research and the fundamental problem of trying to assess the relationship between prosthetic componentry and function, the measurement of function and the prediction of future function.	Thank you
Introduction	Peer Reviewer #4	KQ4 could use some changes in wording to enhance clarity. Not sure what you are getting at exactly. Especially 4a	KQ 4 has been reworded.
Introduction	Peer Reviewer #4	Not sure why KQ6 is included it does not mesh with the primary goals of the assessment and is not included in the analytic framework.	All KQ's were agreed upon with the sponsors and revised in discussion with Key Informants and Technical Experts. The protocol underwent public review. KQ 6 is included in the analytic framework.
Introduction	Peer Reviewer #4	In KQ7 – you state that studies should account for intervening events of mortality, subsequent surgeries or injuries maintain their ambulation at different time points after amputation. This may be somewhat valuable, although at the time of prosthetic prescription you don't know who will incur one of these events or a key change in medical status. So if one of your goals is to determine if someone should receive a prosthesis, and what type of prosthesis at a given time point you won't be able to account for these effects. So it would seem most appropriate to determine the outcomes at key time periods for the overall population or for specific well definable subpopulations where	We added to the results that studies did not explicitly account for intervening mortality or subsequent surgeries or injuries.

Research review section	Reviewer, Affiliation	Comment	Response
		you could classify a patient in terms of their risk at the time of prescription.	
Introduction	Public Reviewer #19 David Boone	"Over the past few decades of my career as a prosthetist and academic researcher, I have seen the steady improvement of function of persons with limb loss. Thirty years ago it would have been accepted to think that use of a lower limb prosthesis would have a greatly limiting impact on the user's life. The user would commonly walk with a severe limp, would struggle with the weight of the prosthesis, might endure regular skin ulceration from the socket interface and face significant tripping hazards. Understanding the magnitude of these kinds of improvements is vital to continuing the relief of disability that improving technology can provide. I appreciate the effort and scholarship put into the draft AHRQ systematic review entitled "Lower Limb Prosthesis". The continued improvement in amputee function with modern prostheses is so dramatic to those who have witnessed the change over time, it is clear that evaluation assessments must be reviewed for how well they prove out the apparent value. In the hope that focus and brevity will help make a few key comments more impactful, I will provide only a few brief and specific comments on how the draft may also be improved, just like the generation of amputees' lives that have benefited from such improvement."	Thank you.
Introduction	Public Reviewer #2 Andreas Kannenberg, Otto Bock HealthCare LP	Quote: "LLPs replace the functionality of a missing limb to as great a degree as possible." Comment: That sounds like the German Social Security Act but is not really true for the U.S.. The Klevel coverage restrictions may have been appropriate 30 years ago with the prosthetic technology available back then, but no longer today.	We were making a more general statement than a reference to insurer coverage goals. We have changed the sentence to "LLPs replace the functionality of a missing limb, ideally, to as great a degree as possible."
Introduction	Public Reviewer #2 Andreas Kannenberg, Otto Bock HealthCare LP	Quote: "In practice it is difficult for clinicians to assess medical necessity for a patient to receive the most appropriate component (whether of higher or lower level or sophistication). Determination of a patient's potential functional abilities requires an assessment of current condition and ability and	We agree with all the comments and have incorporated them into our discussion. We have removed the concept of "medical necessity" since this has specific legal meaning under CMS coverage.

Research review section	Reviewer, Affiliation	Comment	Response
		potential to ambulate. In practice, therefore, OMTs must both assess and predict function to help guide prosthetists, treating physicians, and beneficiaries. However, it is unclear to what extent measures of current function and status are able to predict future function. A major methodological challenge in addressing selection of OMTs for routine use pertains to the assessment of predictive validity. Predictive tests should be valued with respect to their ability topredict future important outcomes. However, outcomes are determined by the whole patient management strategy which involves the baseline assessment, the LLP that a patient is given based on this assessment, patient health and changes in patient health, and any additional care (e.g., physical therapy, rehabilitation) that the patient receives. Thus, it is inherently challenging to assess the value of a baseline OMT assessment by itself, particularly if the choice of LLP is influenced by the initial OMT assessment." Comment: In fact, neither Medicare nor private insurances have given prosthetists any indication what validated measures they think may be suitable to support the K-level determination.	
Introduction	Public Reviewer #2 Andreas Kannenberg, Otto Bock HealthCare LP	Quote: "A major question is how to match patients with LLPs (both by K levels as well as by other characteristics) to optimize functional and other patient-centered outcomes." Comment: The prosthetic research community would greatly benefit from understanding what the patient characteristics are that should be addressed in future research.	This remains a question. Unfortunately, the evidence does not yet provide an answer. We have stated this more explicitly in the Discussion.
Introduction	Public Reviewer #2 Andreas Kannenberg, Otto Bock HealthCare LP	Quote: "The major contextual challenges in providing data to inform matching of LLPs to patients pertain to the large heterogeneity in patient characteristics and attributes of the LLPs; the fact that it is unclear which patient characteristics and LLP attributes are important to best match a patient to a specific LLP; disagreements about what constitutes an optimal matching of patients with LLPs; and poor clinical outcomes and wasted resources associated with suboptimal LLP allocations. Specifically, patients who are in need of LLPs are	This statement is from the Introduction. While it may be difficult to address all these issues, it remains a pertinent question to try to address, even if one determines that the failure to fully address the question in a single study is perfectly reasonable for the reasons you state. However, our results and discussion highlight that the current evidence almost completely fails to adequately address

Research review section	Reviewer, Affiliation	Comment	Response
		heterogeneous in terms of etiology of limb loss, amputation type (level of amputation, uni-or bilateral), age, comorbidities, frailty, general health status factors, expected life span, mental health status (e.g., depression, posttraumatic stress syndrome), family and social support, and many other factors, including whether they have fragile skin or allergies towards socket liners or other materials. These factors may affect their actual and perceived current and maximum attainable functional ability, and the likelihood that they will receive and use an LLP. 8,16" Comment: That is a theoretical consideration. It would require a study with thousands of patients to address all these different characteristics and the multitude of components available to answer that question. Who would be willing to fund such study?	any aspect of the question of who should receive which component.
Introduction	Public Reviewer #2 Andreas Kannenberg, Otto Bock HealthCare LP	On Key Question 4 Quote: "Study participants characterists of interest include: K-Level" Comment: I refer to earlier comments above that K-levels are non-validated, very vague administrative definitions that are hard to prove and validate. That would require that validated measures be accepted to support the determination of K-levels.	We have added the comment that K levels have not been validated. Yet they are still of interest.
Introduction	Public Reviewer #22 Claire Kilpatrick	"In the Background/Evidence Summary section of the introduction, there is a discussion of "matching LLPs to patients," including: what patient characteristics and LLP attributes are important in determining a match, disagreement in what constitutes an optimal match, and wasted resources associated with suboptimal LLP allocations. A change in language should be made here – suboptimal lower limb prostheses allocations should be changed to suboptimal component allocations.	Thank you. We agree with this modification.
Introduction	Public Reviewer #22 Claire Kilpatrick	In the Objectives of the Systemic Review section, it is stated that the review excludes from evaluation biomechanical and other nonpatient-centered immediate outcomes. If the reviewers are attempting to make an argument for optimal matching of componentry with a specific patient presentation (as Question 4's results attempt to do), biomechanical	At the start of Key Question 4, we have added a statement that the review addresses only clinical and patient-centered outcomes (including quality of life) and why. This is also noted in the abstract, introduction (objectives), and discussion.

Research review section	Reviewer, Affiliation	Comment	Response
		analysis cannot be disregarded. Prosthetic componentry is designed in order to improve function, and function is multifaceted. To limit reviewed research to that which includes clinical outcome measures disregards the continuously established link between motion quality (established via kinematics and kinetics) and patient care. Biomechanical analysis can offer objective analysis on component quality and patient matching. The reviewers seem to minimize findings of many of the studies included in Question 4 due to a determined high risk of bias. Including biomechanical outcomes will open the door to lower risk of bias study inclusion. More importantly, biomechanical analysis can offer further commentary on how to objectively match a patient with componentry; for example, examining extension and flexion moments during the gait cycle often is indicative of user stability. Without included biomechanical outcomes, this review cannot be viewed as comprehensive nor argue that there is insufficient data supporting that prosthetic components have an impact on patient outcomes. Additionally in this section, the reviewers state that this paper does not attempt to review all evidence about comparisons between specific components. Therefore, the introduction to content in Question 4 should be prefaced with the statement that this review is not comprehensive and the results of the Question 4 review cannot be generalized to all prosthetic components."	
Introduction	Public Reviewer #23 Kimberly Lebl	It is stated that studies focusing on biomechanical outcomes are excluded from this review. However, little rationale is given as to why that decision was made. Biomechanical analysis offers an excellent source of objective data that gives insight into the potential benefits (or deficits) of certain lower limb prosthetic components. Also, given that billing codes are often based on the biomechanical function of the components, it seems illogical to exclude these studies.	We have expanded the rationale for this decision and state explicitly that biomechanical outcomes, though important, are not covered. This is noted in the abstract, introduction (objectives), relevant results sections, and discussion.
Introduction	Public Reviewer #4	The draft report begins by discussing the fact that current standards for selecting an appropriate lower limb prosthesis	While these issues are certainly important, they are not topics covered by the Key

Research review section	Reviewer, Affiliation	Comment	Response
	American Orthotic Prosthetic Association	rely on a combination of performance based assessments, self-assessments, and wearable monitoring technologies that record patient activity, although prosthetists and other clinicians often rely on clinical judgement to best match the patient with an appropriate prosthesis. The report then discusses the current functional level classification system used by Medicare to determine coverage guidelines for prostheses provided to Medicare beneficiaries. The current functional level classifications date back over twenty years—is there any evidence that they either are, or are not, still valid and accurate in a world of fast-changing technologies? Data from Dobson-DaVanzo from the 2007-10 data (published in Military Medicine as noted above), highlights that the total health costs of patients who were functionally assigned as K2 patients, but who had received K3 technology devices, were less than K2 [patients who had received K2 devices. In fact, the total health costs of the K2 patients who received K3 devices looks a lot more like those of K3 patients. Is it possible the functional level classification have things upside down or confused? Unfortunately, this vital factor is not addressed in the AHRQ report, and seemingly will need to await another, more visionary approach.	Questions to this review. However, we have added explicit language that the K levels, themselves, have not been validated.
Introduction	Public Reviewer #4 American Orthotic Prosthetic Association	AOPA believes that it is important to note that while Medicare has relied on functional level assessment and classification for many years in order to make coverage decisions, neither Medicare nor other payers have provided an indication as to what validated outcome measures are acceptable when determining a patient's appropriate functional level classification. This creates significant issues for prosthetists when considering what measures should be used to document a patient's functional abilities. The identification of outcome measurement tools that are validated, consistent, and reliable should be a priority of the systematic review and should be addressed in greater detail in the final report.	It is our understanding that CMS asked questions pertaining to test validity and reliability, about heterogeneity of treatment effect, and long term outcomes (and the other questions) to help inform future policy about just these issues, which validated outcome measures are acceptable.

Research review section	Reviewer, Affiliation	Comment	Response
Introduction	Public Reviewer #4 American Orthotic Prosthetic Association	The draft report reviews the seven key questions that were addressed by the systematic review. The key questions were made available for public comment in December of 2016 and AOPA provided extensive comments regarding suggested modifications to the key questions that it believed were relevant to the systematic review and provided references to existing clinical literature that might be helpful to AHRQ in refining the key questions. While AOPA was pleased to see that based on its and other's feedback, AHRQ refined the key questions, we believe that further refinement to the key questions will help AHRQ to perform a more focused review of the clinical literature addressing lower limb prostheses.	Hopefully, a future AHRQ review will ask additional questions that will also be helpful to the research and clinical community.
Introduction	Public Reviewer #4 American Orthotic Prosthetic Association	Key questions 1-3 address assessment techniques, prediction tools, and functional outcome measurement tools respectively. 92 studies were identified in the systematic review that focused on one or more of these three criteria. The report indicates that three tools were evaluated as assessment techniques; the Prosthetist's Perception of Clients Ambulatory Abilities (PROS), Short Form Health Surveys, and the Transfemoral Fitting Predictor (TFP). AOPA is concerned that the identification of only three assessment techniques is short sighted and does not consider other techniques that have been validated through research and are generally accepted within the prosthetic community. As far as AOPA is aware, the PROS assessment tool has not been used in clinical studies and the TFP is not commonly used as an assessment tool within the prosthetic community.	The review of instrument psychometric properties has been completely redone. This resulted in a broader, more inclusive list of instruments assessed as either valid or reliable among either Medicare-like populations or other populations. However, we did not address the extent to which, or whether, these instruments are used in practice.
Introduction	Public Reviewer #4 American Orthotic Prosthetic Association	Key Questions 4-7 address the effect of different prosthetic components on outcomes, expectation of ambulation, patient satisfaction, and long term prosthetic use. The draft systematic review identified a limited number of studies for these key questions and did not identify specific studies that AOPA believes are extremely relevant to these questions. One omission of significance is the Dobson DaVanzo study on the cost effectiveness of lower extremity prostheses that was	This review does not address cost effectiveness or comparisons of components, per se.

Research review section	Reviewer, Affiliation	Comment	Response
		published in Military Medicine in February, 2016 and was appended to AOPA's comments on the key questions under review. The failure of the AHRQ contractor to simply not acknowledge this valuable study is perplexing as it is directly related to key questions 4-7.	
Methods	TEP Reviewer #1	Please provide greater detail to when and how the authors applied the exclusion criteria of "battle-related trauma". To assume that all Veterans who lost one or both lower limbs in a traumatic event did so in "battle", would be false with many of them losing their limb(s) in motorcycle and motor vehicle accidents, cancer, and/or infections not related to DM. By applying this assumption to help guide inclusion criteria would lead to the exclusion of research that could benefit this review	We have added language to the eligibility criteria that only studies that explicitly and clearly included only battle-trauma were excluded and that this does not apply to studies of veterans with multiple amputation etiologies.
Methods	TEP Reviewer #1	Please consider including the following article as part of your review (Gailey et al (2012) study titled, "Application of self-report and performance-based outcome measures to determine functional differences between four-categories of prosthetic feet". The study examined the contribution of different MFCL categorized prosthetic feet on prosthetic mobility as defined by the following outcome measures (LCI, PEQ, AMP, Six-minute Walk Test). It also allows for subgroup analysis between those who did or did not lose their limb due to PVD.	This study has been included. Thank you.
Methods	TEP Reviewer #2	I have limited experience in performing systematic reviews of the published scientific literature, but the study appears to have used established methodologies.	Thank you
Methods	TEP Reviewer #3	There is considerable concern regarding the exclusion of studies that only focus on individuals who lost limbs from trauma, congenital anomalies, and are younger than 65. These individuals are likely to receive insurance through Medicare either through the aging process or through receiving Social Security Disability. Additionally, individuals with limb loss under the age of 18 may receive health insurance through a state Medicaid program or Children's	We did exclude studies of people with congenital anomalies, of people with battle trauma as a cause of their amputation, and of children. However, we did not exclude studies of people with traumatic amputations or who were younger than 65.

Research review section	Reviewer, Affiliation	Comment	Response
		Health Insurance Program (CHIP), which may closely follow CMS policy and regulatory framework.	
Methods	TEP Reviewer #3	The methods for the systematic review do not seem internally consistent. For example, one study highlighted for KQ7 (e.g., Davies 2003) utilizes an assessment tool (e.g. Harold Wood/Stanmore Mobility Grade) that is previously noted in the report as deficient.	We did not exclude unvalidated instruments for Key Questions 4 to 7.
Methods	TEP Reviewer #3	The concept of prosthesis abandonment is not defined in the report nor in the studies included in the discussion of KQ7, making ti difficult to assess this aspect of KQ7. This is very troubling given the potential for any discussion around abandonment of prostheses to justify restricting access to these devices.	The definitions used in the studies are better stated in the KQ 7 summary table.
Methods	Peer Reviewer #1	All search criteria have been explicitly stated. The definitions for outcome measures are appropriate and appropriate statistical methods have been used	Thank you
Methods	Peer Reviewer #1	The inclusion and exclusion criteria are NOT justifiable. A major concern is the exclusion of biomechanical studies from this review. As prosthetic components are designed to normalize lower limb biomechanics, exclusion of biomechanical variables is a major limitation of the study. There is extensive biomechanics literature on lower limb amputees and prostheses. If the authors consistently found low strength of evidence with patient-reported outcomes and performance-based outcomes, it is unclear why they did not choose to change the inclusion criteria to include biomechanical studies	The review was explicitly designed to evaluate patient-centered outcomes, function, and the like. These were the questions posed to us by policymakers. The review was not designed to evaluate the biomechanical properties of LLP components. This is not to diminish the importance of these outcomes, but it was not the purpose of the review. This is noted in the abstract, introduction (objectives), relevant results sections, and discussion.
Methods	Peer Reviewer #1	A second major concern is with the selection of studies for answering key question 7. Out of the eight studies which reported long-term follow-up after prosthesis prescription, only one study was performed in the U.S. As the focus of this review is on the Medicare population, inclusion of studies performed in foreign countries is NOT appropriate because different countries have different criteria for identifying the functional level of amputees and prescribing a	This review does not take the approach that only US studies are applicable to the Medicare population. However, we have added noted limitations to the evidence based on studies mostly being conducted outside the US and mostly being old. The restriction based on sample size of 100 was used in large part because of the lack of precision

Research review section	Reviewer, Affiliation	Comment	Response
		prosthesis. The rationale for selecting foreign studies with more than 100 subjects, as opposed to U.S. based studies with less than 100 subjects, has not been justified. It appears inappropriate to compare long term prosthesis use between foreign citizens' and Medicare eligible U.S. citizens because a number of social and economic issues can lead to reduced use of prosthesis or abandonment of prosthesis, such as, availability of modern prosthetic components, ADA compliant environment, lifestyle etc. Therefore, to answer this critical question, only US-based studies with eligible Medicare population should have been selected	from smaller studies. This has been added to the Methods section where the restriction by sample size is noted.
Methods	Peer Reviewer #2	In general, I think the methods used were appropriate, particularly for questions 1-3 and 5-7. The search strategies seemed complete (except that there are several important studies published since November 2016 that are relevant to this report). The statistical methods and assessment of quality of studies used are standard for these AHRQ reviews and are generally appropriate	Thank you. The review has been updated since November 2016.
Methods	Peer Reviewer #2	The methodological challenges with key question 4 are raised in the general comments above	Thank you.
Methods	Peer Reviewer #3	Excluding military based research is problematic given the already small number of studies for which to rely for the target population. As previously stated many patients with Medicare as their primary insurance are not over the age of 65 and their amputations are traumatic in nature. Including these studies but providing distinct patient demographics would be more helpful to differential how components are authorized and resources are utilized.	Military based research is included. Patient demographics are clearly described for each study and summarized for each instrument or Key Question. These factors were used only to categorize studies.
Methods	Peer Reviewer #3	The justification for using a sample size of 100 in KQ7 (as opposed to 20 in KQ1-3) stated precision and power as the reason for inconsistent 'n' requirement. Of concern is that the lack of evidence based literature in this area suggests the inclusion criteria should be expanded and strength of evidence determined even at the expense of power.	We agree that sample size cutoffs are arbitrary and we understand that not all readers will agree with each threshold used. The decision was based on a balance between usefulness of study findings and available resources, in discussion with both a Key Informant and a Technical Expert panel.

Research review section	Reviewer, Affiliation	Comment	Response
Methods	Peer Reviewer #4	The overall methodological approach is strong. A few questions / comments? exceptions are included below.	Thank you
Methods	Peer Reviewer #4	Search strategy and eligibility criteria. – generally look good, although perhaps it would have been best to include only dysvascular as the trauma population is so fundamentally different and the research design and outcome measures may be completely different in this population. One of the criteria that you use to evaluate different measures is validity. Yet you include studies where subjects have different etiologies of amputation. A measure needs to be validated in a specific population. In this case the dysvascular amputee. So this should be a criteria for inclusion in the report not whether they have been broadly validated.	A literature search that included terms for dysvascular would likely have been too limiting, excluding studies of interest. More importantly, non-dysvascular etiologies were also of interest, even though these populations are less representative of the Medicare-eligible population. The review was not restricted to studies eligible to the Medicare population. Each validation study is, in fact, generalizable to only the subset of people with LLPs in whom the study was conducted. For this reason, we added the statement that researchers and other using the instruments should look at the specific studies that have evaluated specific instruments of interest to see if they match their needs.
Methods	Peer Reviewer #4	It appears with a few exceptions – that this review is primarily of mobility rather than function. There are few other functional domains that were assessed. Was this because there is inadequate numbers and quality of publications. This would be worthy of comment., either to state that it is a review of mobility, or that other domains of function did not have adequate quality to be included.	Functional instruments and domains were reviewed. There are numerous examples of these including SF-12/36, NQ-ACGC, PROMIS-29, and others.
Methods	Peer Reviewer #4	Outcomes of interest - I would have not included gait speed, step count, or walk distance or other gait characteristics. Typically these are collected in a laboratory environment and are irrelevant to actual functional mobility. There are numerous studies that show that prosthetic interventions do not alter step count in community mobility (Orendurff, Klute). Also these data do not follow the ICF classification of mobility and are not necessarily an indication of functional mobility. Functional mobility is really the outcome of interest.	Whether to include gait speed etc. as measures of mobility was discussed at length. For the sake of completeness, in discussion with the Technical Expert Panel, we determined it would be best to include them.

Research review section	Reviewer, Affiliation	Comment	Response
Methods	Peer Reviewer #4	Study selection and data extraction – excellent approach	Thank you.
Methods	Peer Reviewer #4	Post hoc Analyses - Appreciate the effort in the post hoc analysis section to calculate a Bonferoni correction but many published studies do not include all of the analyses that were derived from a single data set in a single paper. Rather results are parsed out in different papers reporting on only a specific subset of data. This is a limitation in this approach and should be stated.	We have added to the description that we did not attempt to further correct for analyses conducted but not reported by the study authors
Methods	Public Reviewer #14 Seung Eun Lee	On Eligible Study Designs (pg 46), the report said they excluded the studies of validation that is not listed in English units. Unfortunately, most of the countries uses SI units, which is metrics, so I strongly believe that excluding non-English unit data will eliminate other evidence based articles that are valid.	What was meant was non-English language instruments. We were not talking about imperial units. We have added the word "language" for better clarity.
Methods	Public Reviewer #19 David Boone	It should be noted that the blanket exclusion of biomechanical measures is a decision that should be reflected in a change in the title of the review since it really is not a comprehensive assessment of Lower Limb Prostheses. The end point functional outcomes are vitally important, but alone do not provide a complete understanding of patient function. I think that it was a wasted opportunity not to similarly identify the validity and reliability of biomechanical measures in concert with the end point functional outcomes. The ability to understand the mechanisms of impact on functional outcomes has been essential to make the improvements we have made to date, and are equally essential to identifying how to improve function for amputees as we move forward.	The review was explicitly designed to evaluate patient-centered outcomes, function, and the like. These were the questions posed to us by policymakers. The review was not designed to evaluate the biomechanical properties of LLP components. This is not to diminish the importance of these outcomes, but it was not the purpose of the review. This is also noted in the abstract, introduction (objectives), relevant results sections, and discussion. The title has been changed to Lower Limb Prostheses: Measurement Instruments, Comparison of Component Effects by Subgroups, and Long-Term Outcomes
Methods	Public Reviewer #2 Andreas Kannenberg, Otto Bock HealthCare LP	On page 18, it is stated that post-hoc Bonferroni corrections were used to adjust significance levels for multiple testing. As much as this may be justified from the statistical perspective, it ignores the specifics of prosthetic studies with relatively low subject numbers that make it impossible to reach such low p-values.	This may be true, but it is a deficiency of the evidence base.

Research review section	Reviewer, Affiliation	Comment	Response
Methods	Public Reviewer #20 Anonymous	"• There is a lack explanation of what constitutes a 'low resource country' and a lack of justification for excluding these studies.	We have referenced the World Bank. We have added that "the interventions, management, and characteristics of people in low-income countries are too different to be applicable to the U.S. population". We believe that low income or low resource is sufficiently descriptive.
Methods	Public Reviewer #20 Anonymous	• There is a lack of justification for excluding 'biomechanical measures'."	The review was explicitly designed to evaluate patient-centered outcomes, function, and the like. These were the questions posed to us by policymakers. The review was not designed to evaluate the biomechanical properties of LLP components. This is not to diminish the importance of these outcomes, but it was not the purpose of the review. This is noted in the abstract, introduction (objectives), relevant results sections, and discussion.
Methods	Public Reviewer #21 James Kling, Georgia Tech MSPO Student	I do not understand why most studies on the biomechanical properties of prosthetic components were excluded. These studies address how different components can help return some biomechanical functions to patients who have lost them due to amputation. These studies highlight why different components are necessary and appropriate for different patients. Every person deserves to have appropriate healthcare, and by eliminating the use of many components, most patients will not receive appropriate treatment, and will not be able to fulfill their full potential during recovery. Including studies on biomechanical properties of components is vital in understanding why these components are a necessity for patients and their quality of life after amputation.	The review was explicitly designed to evaluate patient-centered outcomes, function, and the like. These were the questions posed to us by policymakers. The review was not designed to evaluate the biomechanical properties of LLP components. This is not to diminish the importance of these outcomes, but it was not the purpose of the review. This is noted in the abstract, introduction (objectives), relevant results sections, and discussion.
Methods	Public Reviewer #22 Claire Kilpatrick	In the KQ4 Outcomes of Interest, it is stated again that biomechanical measures were excluded. As described above in my comments on the introduction, I view this as an inappropriate exclusion criteria.	The review focuses on clinical and patient centered outcomes, which is not to diminish the importance of other outcomes for the research community. This is noted in the

Research review section	Reviewer, Affiliation	Comment	Response
			abstract, introduction (objectives), relevant results sections, and discussion.
Methods	Public Reviewer #23 Kimberly Lebl	The guidelines used for this review seem similar to those used in literature reviews for pharmaceutical or similar research. Due to the rather unique nature of the field of prosthetics and orthotics, it's not clear that the same guidelines are necessarily appropriate.	We respectfully disagree and believe the systematic review methods are appropriate.
Methods	Public Reviewer #24 Kellie C	This review must take into consideration studies that analyze biomechanical outcomes to truly and fully understand how to best match prosthesis components to patients. Biomechanical outcomes define the biological system and including them in this review is critical. People who use lower limb prostheses have structural asymmetries in their neuromuscular system and studies have shown that amputees exhibit significant asymmetries between their intact and amputated limb, such as stance and swing times, ground reaction forces, and joint kinematics [1-10]. An understanding of the biomechanics of a person with limb loss and the influence of their prosthetic components can provide insight on features that consequently lead to large collision impact on their sound limb and increased energy expenditure. Though the ideal compensation techniques used in asymmetrical gait is not clearly defined, it is important to consider these studies in the development of prescriptions for assigning specific components.	The review was explicitly designed to evaluate patient-centered outcomes, function, and the like. These were the questions posed to us by policymakers. The review was not designed to evaluate the biomechanical properties of LLP components. This is not to diminish the importance of these outcomes, but it was not the purpose of the review. The review addresses focused questions. This is noted in the abstract, introduction (objectives), relevant results sections, and discussion.
Methods	Public Reviewer #4 American Orthotic Prosthetic Association	In its comments on the key questions for inclusion in the AHRQ systematic review, AOPA cites 19 studies that should have been considered by the researchers when reviewing the existing clinical literature on lower limb prostheses. Of the 21 studies identified by AOPA, 14 were not considered in the systematic review. Table 1 below lists the studies that AOPA identified as relevant to any systematic review that were not considered for inclusion by the AHRQ contractor. See PDF for studies	We have conducted full text screening of all citations recommended by this reviewer and others, and have added any eligible studies found. Appendix B includes the list of excluded studies and reasons for exclusion.
Methods	Public Reviewer #4	The omission of these valuable studies by the AHRQ contractor must call in to question the parameters that were	The specific Key Questions addressed by the various reviews differ from this review, which

Research review section	Reviewer, Affiliation	Comment	Response
	American Orthotic Prosthetic Association	developed in order to conduct the literature search. The failure to recognize and acknowledge valid, published, and reviewed studies is concerning to AOPA as it appears that the AHRQ contractor may have conducted its search in an overly narrow way. Three notable omissions from inclusion in the systematic review are the systematic review of prosthetic intervention in trans-tibial amputees conducted by Highsmith et. al; the significant work of the health economics firm Dobson DaVanzo on the overall cost effectiveness of prosthetic services; and a study by Mundell, Kaufman, et. al. entitled The Direct Medical Cost of Accidental Falls for Adults with Trans-Femoral Amputations. The systematic review of prosthetic intervention in trans-tibial amputees performed by Highsmith et.al was published in the Journal of Rehabilitation Research & Development (JRRD) in early 2016, long before the end of the stated review period in November 2016. It is incomprehensible to AOPA that a systematic review of such importance that was published in a well-established, peer reviewed journal such as JRRD would not be recognized or considered by the AHRQ contractor as relevant to its systematic review.	asks a much more focused question about heterogeneity of treatment effects, which as we describe is rarely addressed in the current literature.
Methods	Public Reviewer #4 American Orthotic Prosthetic Association	The systematic review conducted by Highsmith, et. al. included 8,796 articles which yielded 135 references identified for manual screening. Upon manual screening by the researchers, 31 evidence statements were made, 23 of which were supported by level 2 evidence and 8 of which were supported by level 1 evidence. AOPA strongly encourages the AHRQ to consider this important, peer reviewed study in its final report.	We have screened articles included by Highsmith and included those that met our much narrower eligibility criteria.
Methods	Public Reviewer #4 American Orthotic Prosthetic Association	The significant research on the overall cost of lower limb prostheses, the results of which were published in Military Medicine has been updated to include additional data from 2011-2014. The additional data continues to validate the value of the research and the final report from the updated research is also included as an addendum to these comments.	This review does not address costs.

Research review section	Reviewer, Affiliation	Comment	Response
Methods	Public Reviewer #4 American Orthotic Prosthetic Association	The Direct Medical Cost of Accidental Falls for Adults with Trans-Femoral Amputations study was conducted by researchers at the Mayo Clinic in Rochester, Minnesota and used data from the Rochester Epidemiology Project to determine the direct medical costs associated with falls by trans-femoral amputees. The study results were published in Prosthetics and Orthotics International in June, 2017. As the study was published after the review date deadline of November 30, 2016, AOPA encourages the AHRQ to consider this extremely valuable study prior to issuing its final report.	This review does not address costs.
Methods	Public Reviewer #4 American Orthotic Prosthetic Association	Included with AOPA's comments on the key questions for inclusion in the systematic review was a preliminary report on the study by the RAND Corporation entitled, Economic Value of Advanced Transfemoral Prosthetics. The preliminary report by the RAND Corporation identified 106 relevant studies in its bibliography. Of the 106 studies identified by the RAND Corporation, the AHRQ contractor only identified 9 as relevant to its systematic review. An additional 10 studies were considered but rejected by the AHRQ contractor meaning that 87 studies that were identified as valid and valuable by the RAND Corporation were not identified or considered as part of the AHRQ systematic review. Table 2 identifies the studies that were identified in the RAND Corporation bibliography that were not considered by the AHRQ as part of its systematic reviewTable 2. RAND Corporation Identified Studies That Were Not Included in theA HRQ Systematic Review. See PDF for study citations.	We have conducted full text screening of all citations recommended by this reviewer and others, and have added any eligible studies found. We have included those that met our much narrower eligibility criteria.
Methods	Public Reviewer #5 Orthotic and Prosthetic Alliance	The current review was limited to studies with participants or patients "deemed to be generally applicable to the Medicare population." We strongly disagree with this decision, as we are confident that much of the evidence discounted as a result of this decision would address the Key Questions. Although we acknowledge it is the authors' right to conduct the review in the manner they deem most appropriate, we also believe it is the authors' responsibility to describe the explicit criteria by	Nowhere does it state implicitly or explicitly that only manuscripts examining the Medicare eligible population are included. Nothing in the eligiblity criteria or Key Questions says anything about Medicare. For KQ 1-3 and to some extent KQ 4-7 we do categorize based on generalizabilty to the Medicare population but this is not an eligibility criterion.

Research review section	Reviewer, Affiliation	Comment	Response
		which they determined whether or not study populations were applicable to the Medicare population. We therefore request that the authors provide additional details about how studies were determined (or not determined) to be eligible to be included in this review.	
Methods	Public Reviewer #5 Orthotic and Prosthetic Alliance	For example, we request the authors to specify in the final report the age ranges, mean ages, and/or percent (or number) of people who may have experienced amputation due to dysvascular disease who were required for a study to be deemed relevant to the Medicare population. Further, please specify whether this determination was made by a single author, or by consensus of several authors. Considering the variety of ways in which participants (or patients) in a study are reported, we recognize that this assessment may have been difficult, but additional details are required. As written, it appears that the study authors(s) made this assessment subjectively on a case-by-case basis, which would not be sound.	The study demographics are presented in the relevant summary tables and described for each instrument and/or key question. The criteria for categorizing a study as generalizable to Medicare are listed in the Methods: Mean age >=65 years or dysvacular conditions reported to be >=50% of sample. We acknowledge these are arbitrary and have acknowledged that in the Methods and Limitations section of the Discussion. These criteria were determined in discussion with our Technical Expert panel.
Methods	Public Reviewer #5 Orthotic and Prosthetic Alliance	The metrics for evaluating Key Questions 1-3 are not well-explained or justified. Although the authors provide a table of criteria by which identified outcome measures were evaluated (Table 2), it is unclear how these criteria were determined or how they were applied to determine whether an instrument was “validated” or “found reliable.” (See comments below regarding this terminology.) For example, why do the authors require a N≥30 for testing reliability, but not any other property? How were the thresholds for “excellent,” “adequate/good,” and “poor” for each property determined? Without explanation and citation, these criteria appear to be determined ad hoc for the purposes of this review. This is especially concerning in light of the authors’ criticism of outcome studies relying on ad hoc instruments. We believe the review authors should use (or at least cite) the variety of tools or checklists10-12 that have been developed to assess the quality of outcome measures for Key Questions 1-3, as	We have added in the inadvertently omitted citations to prior work upon which we determined our methods, which included the Terwee 2007 article you cite. We also removed N>=30 requirement.

Research review section	Reviewer, Affiliation	Comment	Response
		they used a standardized appraisal tool (i.e., the Cochrane Risk of Bias Tool) to evaluate studies for Key Questions 4-7.	
Methods	Public Reviewer #5 Orthotic and Prosthetic Alliance	The authors indicated (on p. ES-12 and p. 74) that they "consider variations and modifications of measures to be separate measures that would each need to be validated." We applaud the authors for this distinction because it reflects substantive changes to the underlying instrument, administration method, and resultant metric or score. However, we find it perplexing that the authors did not apply this method of outcome measure assessment consistently throughout the review. For example, the authors list the ABIS and ABIS-R as two separate instruments (Table A). However, they fail to differentiate other instrument revisions, such as the LCI13 and LCI-514, the PEQ-MS15 and PEQ-MS 12/516, or the ABC17 and the Rasch-modified ABC18. These distinctions are important, as the revised versions of these instruments often include different questions, response options, and scoring methods. We encourage the authors to explain why they have selectively applied this rule regarding "variations and modifications" to review of instruments, or indicate why they believe evidence of reliability and validity in one version of a patient-report instrument can be applied to a revised or alternate version.	We have completely reassessed, reanalyzed, and rewritten the sections pertinent to test validity (KQ 1-3). All instruments mentioned here have been included and reevaluated. We have more consistently distinguished different subscales etc.
Methods	Public Reviewer #5 Orthotic and Prosthetic Alliance	Similarly, the method for determining the overall quality of studies considered for Key Questions 4-7 should be explained with greater detail. For example, was the determination of "overall quality" based on an appraisal of a single author (i.e., a reviewer) or multiple authors, or by consensus of multiple authors? Contemporary systematic review standards19, 20 typically require data extraction and methodological appraisal to be performed by two or more reviewers (and consultation of a third, when disagreements exist or consensus cannot be obtained). Based on a number of errors noted in the data extraction (see below), we question whether multiple authors	We have added in the omitted sentences about double assessments of both risk of bias and strength of evidence.

Research review section	Reviewer, Affiliation	Comment	Response
		assessed the included articles. If not, this should be noted as a significant limitation of this review.	
Methods	Public Reviewer #5 Orthotic and Prosthetic Alliance	The authors indicate that "Prospero registration is pending." (See; p. 10.) Post hoc registration of a systematic protocol (similar to post hoc registration of a clinical trial protocol) is generally frowned upon, as it challenges the goal of transparency in the review process. Presumably, the review will be updated to include the appropriate PROSPERO registration (e.g., CRD42017058488).	Thank you and apologies. We didn't enter the existing PROSPERO registration number in the draft. It is now included.
Methods	Public Reviewer #5 Orthotic and Prosthetic Alliance	The authors indicate that "searches will be updated in all databases upon submission of the draft report for peer and public review." (See; p. 10.) We ask the authors to include these updated search dates in the final report.	As per protocol, we have updated the search.
Methods	Public Reviewer #5 Orthotic and Prosthetic Alliance	The authors present a summary of those psychometric properties (e.g., face validity, content validity, criterion validity) relevant to their review (p. 15). Noticeably absent is a description of known groups construct validity. Given that one of the implied goals of this review is identification of instruments suited to distinguish among different types of prosthetic patients, this is a relevant characteristic. We encourage the authors to add this property to their instrument review (and a description of the property on p. 15).	The reviewer must have missed the sections in Table 2 and the end of the paragraphs describing all properties, which has the description of construct validity.
Methods	Public Reviewer #5 Orthotic and Prosthetic Alliance	The term "reliability" is used often throughout the instrument summaries as a means to characterize the psychometric testing performed. The authors rarely qualify what form(s) of reliability have been assessed, as they do with types of validity (e.g., construct, content). We recommend the authors indicate the types of reliability, as this information has important implications relating to how the instruments can and should be used (e.g., evidence of test-retest reliability would be important if the instrument is intended to be used to monitor participants or patients over time).	Following the approaches taken in prior similar work summarizing psychometric properties, cited in the Methods section, and with the goal of comprehensibility, we took the more straightforward approach of dichotomizing reliability into yes/no. Further details are presented in the Appendix table.
Methods	Public Reviewer #5 Orthotic and Prosthetic Alliance	With regard to the property of criterion validity, the authors note that "since 'gold standards' do not exist for the functional outcomes of interest, this specific metric [criterion validity] is	As part of our reanalysis and rewrite, we have removed presentation of criterion validity. For the reasons noted, it was moot point since for

Research review section	Reviewer, Affiliation	Comment	Response
		largely theoretical for our purposes." (See; p.15.) Given this statement, criterion validity should be reported as "not applicable" (n/a), rather than "not reported" (nr) in the instrument review tables. To a casual reader, "nr" may be viewed as an admonishment of the instrument when, as the authors note, this property simply may not apply. We encourage the authors to revise their tables accordingly.	all instruments and studies the property was not evaluated (or "not applicable"). Certainly, there is no implication that individual researcher's failure to evaluate a given instrument property is an admonition of the instrument or the authors. Clearly, all studies evaluated only certain properties for whichever instruments were evaluated.
Methods	Public Reviewer #5 Orthotic and Prosthetic Alliance	The evidence tables include many lines of duplicate information. For example, the same line items appear multiple times (consecutively, often more than once) in the tables on pp. 185-483.	All instruments have been reevaluated and all summary tables updated.
Methods	Public Reviewer #5 Orthotic and Prosthetic Alliance	Data extracted from a number of articles, as they pertain to Key Questions 1-3, also appear to contain errors. Below, we highlight several examples of errors identified in our review.	Thank you.
Methods	Public Reviewer #5 Orthotic and Prosthetic Alliance	The narrative review of the Transfemoral Fitting Predictor (TFP) noted that the instrument was "evaluated in…92 adults (age not reported)." (See; p. 23.) We find it perplexing that the authors determined that age was not reported, but yet this instrument met the criteria for being applicable to Medicare-aged individuals. As the article describes ages for 87 of the 92 individuals (i.e., consenters) in Table 2 (mean age = 68.8±10.6), it may be that the authors deemed it reasonable to conclude that this population was of Medicare age. We encourage the authors to update the review with the ages noted above rather than "age not reported," particularly as age appears to be a criterion for inclusion in this review.	We have corrected this to include the age, as the reviewer noted. All studies of all instruments have been reextracted and reanalyzed.
Methods	Public Reviewer #5 Orthotic and Prosthetic Alliance	Details regarding the Locomotor Capabilities Index (LCI) as cited in Franchignoni, et al.,14 lack context and do not appear to be entirely correct. The authors note that "the LCI was found to have known group validity by differentiating participants by age (r = -0.554)." (See; p. 28.) The authors' conclusion that the LCI differentiated participants by age is not correct, as the variable presented (Pearson's r) indicates that age significantly correlated with LCI. This is evidence of	Thank you. This instrument and study, and all of them, have been reanalyzed and corrected.

Research review section	Reviewer, Affiliation	Comment	Response
		convergent validity, not known groups validity. Later in this same section, the authors note that "the LCI was found to have predictive validity for the RMI (r = 0.752), the TWT (r = -0.667), the FIM instrument (0.617), LCI (0.765), and LCI-5 (0.622)" (p. 28). As written, the data presented are confusing (e.g., it makes no sense that the LCI is correlated with itself with a correlation coefficient of 0.765). Data in the source reference describe correlations between LCI (as measured within the first 72 hours of admission to a rehabilitation center, T0) and other instruments (as measured at the end of the rehabilitation program, T2). Correlations between the LCI (at T0) and other instruments at the conclusion of the rehabilitation program (T2) were as follows: RMI (0.752), TWT (-0.667), FIM (-0.617), LCI (0.765), and LCI-5 (0.788). The 0.622 correlation noted by the authors refers to the correlation between the LCI-5 (at T0) and the FIM (at T2). The authors are encouraged to verify their data extractions and provide proper context to the data (as needed) so that the instrument reviews are accurately presented.	
Methods	Public Reviewer #5 Orthotic and Prosthetic Alliance	The narrative review of the Activities-specific Balance Confidence scale (ABC) notes data from "nine studies with over 2000 participants." (See p. 34.) The authors fail to indicate that these data are drawn from different versions of the instrument and may not be comparable. For example, data presented in Reference 3831 and 3824 were obtained with the ABC that uses a 5-level ordinal scale that is scored 0-4.18 Other references (such as Reference 5832 and 3033) use the original scale, which is scored 0-100, that was developed by Powell and colleagues.17	As part of our reanalysis, we have separated out the 5-level ordinal scale version. This same thing was done for other instruments in our revisions.
	Public Reviewer #5 Orthotic and Prosthetic Alliance		
Methods	Public Reviewer #5 Orthotic and Prosthetic Alliance	In the data extraction table (p. 59), the authors indicate that the PEQ has both a "problems" and "satisfaction" scale, yet no such scales are described in the narrative summary (p. 41).	Thank you. You are correct. More accurate terms are now used.

Research review section	Reviewer, Affiliation	Comment	Response
		We believe the narrative review to be correct, as the PEQ has no subscale for “problems” or “satisfaction” per the instrument’s development paper.34	
Results	TEP Reviewer #1	Please describe in greater detail your process for determining the reliability and validity of outcome measures of your study population. It seems you have excluded the Amputee Mobility Predictor (with and without prosthesis), Prosthetic Limb Users Survey of Mobility (PLUS-M), and the Six Minute Walk Test (pages 16 – 19) which should meet your inclusion criteria. The manuscript addressing PLUS-M validity was published in 2017 (Hafner et al. Construct validity of the Prosthetic Limb Users Survey of Mobility (PLUS-M) in Adults with Lower Limb Amputation. Arch Phys Med Rehabil. 2017 Feb; 98(2): 277-285).	We have completely reassessed, reanalyzed, and rewritten the sections pertinent to test validity (KQ 1-3). All instruments mentioned here have been included and reevaluated. The new article by Hafner was included in our updated literature search.
Results	TEP Reviewer #1	It is unclear why the AMPnoPRO was excluded from review since it meets review criteria and is and is the most widely used performance-based instrument in the United States for determining function as defined by Medicare. (Borrenpohl D, Kaluf B and Major MJ. Survey of U.S. practitioners on the validity of the Medicare Functional Classification Level system and utility of clinical outcome measures for aiding K-level assignment. Arch Phys Med Rehabil 2016; 97(7): 1053–1063.) The AMPnoPRO has been found to be reliable and a valid measure of functional mobility, it is designed to be used without a prosthesis and has the ability to distinguish between known groups as defined by the MFCL. In addition, the subjects’ ages were within the range of Medicare beneficiaries (i.e., K0=77.1 years, K1=74.5 years and K2=65.4) and distributed appropriate across functional levels as expected by Medicare. The AMPnoPRO, also has that has the ability to discriminate across all MFCL levels. (Seker A, Kara A, Camur S, Malkoc M, Sonmez MM, Mahirogullari. Comparison of mortality rates and functional results after transtibial and transfemoral amputations due to diabetes in elderly patients-a retrospective study. Int J Surg. 2018;33:78-82). More	AMPnoPRO has been included. We have updated our literature search and screened in full text all references suggested by reviewers. Thank you.

Research review section	Reviewer, Affiliation	Comment	Response
		recently, the AMPnoPro was compared to the TUG, 2 MWT and LEMOCOT (Lower-Extremity Motor Coordination Test) and was found to be best statistical predictor of mobility outcomes (Spann MH, Vrieling AH, van de Berg P, Dijkstra PU, van Keeken HG. Predicting mobility outcome in lower limb amputees with motor ability tests used in early rehabilitation. Prosthet Orthot Int; 2017; 41(2):171-177.).	
Results	TEP Reviewer #2	I believe the key message 3 in the summary is very misleading without clear description of the included study limitations: These studies "mostly had methodological limitations"	We have added language about methodological limitations to the Key Messages.
Results	TEP Reviewer #2	Table 7.4 summarizes the strength of evidence for each outcome and subgroup analysis with data. For most outcomes of interest, there is low strength of evidence because studies applicable to the Medicare population, some studies were inconsistent with each other, and few studies reported the outcomes of interest. Subgroup analyses in single studies tended to be underpowered to detect differences, mostly leading to determinations that the evidence was insufficient. It then hard to understand why based on six studies, with high degrees of limitations,That the reported conclusions are made: 11 to 22 percent of lower limb amputees who receive a LLP prescription abandon the prosthesis (stop using it) at about 1 year; these studies are generally representative of people with LLP, in particular older adults and those with dysvascular etiologies. Three of these studies provide low strength of evidence that people with unilateral transfemoral amputations are about twice as likely to abandon their LLP than those with unilateral transtibial amputations. Potentialdifferences among other subgroups had insufficient evidence due to conflicting results among three studies or only a single, imprecise study with data. Also based on four, generally representative studies, there is low strength of evidence that 11 to 37 percent of LLP recipients use their prostheses only indoors; however, these studies are somewhat inconsistent and imprecise. There	Although few studies, with low strength of evidence, the studies do have findings, which are summarized. The low strength of evidence does not particularly related to the applicability to Medicare. These are low quality studies, with imprecise estimates, that are not all that consistent. There are also few studies. The listed conclusions are the findings of the studies, regardless of their quality. The strength of evidence addresses the quality and applicability issues.

Research review section	Reviewer, Affiliation	Comment	Response
		is low strength of evidence about how likely different subgroups of people use their prostheses only indoors, suggesting that people with transfemoral amputations, or who are older, or with bilateral amputations are more likely to be limited to indoor use. There is insufficient evidence about why people stop using their prostheses.	
Results	TEP Reviewer #2	Abandonment, although this is used to describe this phenomena, is pejorative , and has the negative connotations. I recommend that the more neutral description of limited use stop using be used, unless additional information, eg. Long-term acceptance of prostheses depends on a number of factors, including cosmesis, ability to perform ADLs without a prosthesis, chronic pain, and weight of the prosthesis. Including the results of this study alone, is very limited and not generalizable without confirmation with repeated study. The abandonment or change rate of lower limb prostheses is approximately 15% within 1 to 5 year of discharge from a rehabilitation program (Gauthier-Gagnon et al., 1999). No rejection statistics by type of lower limb prostheses are available. Of those who did use their lower limb prostheses, approximately 64% used their prostheses for outdoor mobility and 53% used them for ADLs in their homes (Gauthier-Gagnon et al., 1999).	Abandonment is the commonly used term. Although there are few studies, with low strength of evidence, the studies do have findings, which are summarized.
Results	TEP Reviewer #3	The results discussed for KQ7 seem disproportionate to the strength of evidence and the quality of studies underlying these findings. In particular, the issue of prosthesis abandonment is supported by two studies - one with moderate strength of evidence and the other with low strength of evidence. The results from this sub-question will likely have a significant impact on the ability of individuals who lose a limb to access medically necessary prosthetic devices. It is also surprising to see the results of KQ7 reported while primarily relying on studies conducted more than 10 years ago and studies conducted outside of the U.S. A more prudent report would conclude that there is insufficient evidence to report	All outcomes for KQ 7 have insufficient or low strength of evidence. We have added details about study countries and years, noting that these are important limitations.

Research review section	Reviewer, Affiliation	Comment	Response
		results regarding KQ7 and its sub-questions. This was the case for KQ4 which has a substantially larger number of potential studies to draw supporting evidence, most of which were conducted in the US in the last 10 years.	
Results	Peer Reviewer #1	The amount of detail presented in the results section is mostly appropriate. However, a major piece of detail has been omitted in Key Question 7.iii – Use of prostheses only indoors. The functional level (K-Level) of subjects in the studies has not been reported. The K-level classification (table 1) is an important criteria in the prescription of lower limb prosthesis. Amputees functioning at K-Level-0 do not have the ability or potential to ambulate with a prosthesis whereas K-Level-1 amputees have the potential to use a prosthesis only indoors. As K-level information has not been provided for the eight articles that are used to answer key question 7, it is not clear if the indoor use of prosthesis and abandonment of prosthesis is related to the functional level of the amputee or to the prosthesis itself. E.g. A high functioning transfemoral amputee is less likely to use a prosthesis only indoors compared to a low functioning transtibial amputee. So, instead of creating subgroups based on amputation level, it is more relevant to have the subgroups based on functional level.	This is an excellent point. We have added in the information/limitation that none of the studies reported on K level. Because of study limitations, for our overall summary, we focus on the two more generalizable studies.
Results	Peer Reviewer #1	The Dudkiewicz (2011) study which reported indoor prosthesis use of 37% had only 20 subjects who were classified as group 2 amputees, (i.e. Limited community walker) and more than 500 subjects were classified as group one amputees (i.e. Home ambulation only). The low functional level explains the high percentage of indoor use in this study which should be discussed. I would strongly suggest modifying the subgroups to include an analysis of the subjects' functional level. I strongly recommend including an analysis of the functional level of subjects in the studies for key question 7. It is a significant finding that 11-22% of amputees stop using their prosthesis at one year. As the reason for abandonment	Dudkiewicz 2011 is very hard to interpret regarding K level classification, but we have added a description of this lack of clarity and the caveat that almost all people apparently were home ambulators only at the time of prescription. They did not report subgroup analysis based on K level classification. Language was also added that other studies did not report K levels. The studies of abandonment did not report on K levels. This has been added.

Research review section	Reviewer, Affiliation	Comment	Response
		of prosthesis is not provided, an analysis of the functional level will indicate whether the prosthesis were abandoned as the amputees did not have the physical capability to use the prosthesis, or the abandonment was due to the prosthesis itself.	
Results	Peer Reviewer #1	A minor concern is with the percentages reported on Page 115, line 6.: Out of the 3 cited studies, Pohjolainen et al.(1990) have reported that 18.4% of amputees used their prostheses only indoors. This sentence should be modified and "24%" should be replaced by 18%.	18.4% is correct. We used only "Class IV" and "Class V" only. Class VI and VII are people not using prostheses. We used data for Total, not just below knee.
Results	Peer Reviewer #1	Page 97, Line 24: "Studies" should be changed to "study"	Thank you, corrected.
Results	Peer Reviewer #1	Page 98, Line 16: Modify the sentence: "The participants had with K2 or greater function…."	Thank you, corrected.
Results	Peer Reviewer #1	Page 100, Lines 41-42: The total responsiveness values are different in the two sentences.	One set of values related to subject perception, one set related to prosthetist evaluation. These descriptors have been added to clarify.
Results	Peer Reviewer #1	Page 101, Lines 13-32: This paragraph is confusing. The opening sentence states that the paper did not perform analysis of predictive performance and the concluding sentence states that "…the study does not to provide compelling evidence that their model has no predictive performance." The second to last sentence should also be modified for clarity.	We have corrected the typo. The paragraph describes why they did not actually perform an analysis of predictive performance.
Results	Peer Reviewer #2	Key Questions 1-3, 5-7: The literature search appears complete and the evaluation of the studies appropriate. The data is presented clearly and transparently	Thank you
Results	Peer Reviewer #2	Key Question 4: This review importantly states that only one study analyzed the most important aspect of the KQ, whether any patient characteristics can accurately and effectively predict who will benefit most from a component (Hahn, 2016). This large study (n=899) compared the genium knee to other microprocessor knees and had significant methodologic limitations including substantial missing data, use of unvalidated outcome measures and insufficient details about	We agree.

Research review section	Reviewer, Affiliation	Comment	Response
		the logistic regression analysis to be able to accurately determine if the model predicts outcomes. Most importantly, this comparison between the genium versus another microprocessor knee is for the most part clinically irrelevant to the population of interest (Medicare population). Not only were 69% of the patients were traumatic amputees and the vast majority young men, but these were also patients deemed appropriate for a genium who were already using a MPK (for the most part). Clinically the comparison of interest is between microprocessor knees and non-microprocessor knees rather than the genium versus MPK	
Results	Peer Reviewer #2	There are several small studies published since November, 2016 that may be helpful to consider (although they all have methodologic issues and variable relevance to key question 4). Fuenzalida Squella, (July 2017 Prosthet Orthot Int.) which compared a MPK knee (3E80) to non MPK in 13 young, high functioning TF amputees and demonstrated improved balance, satisfaction and falls with the 3E80. Hasenoerhl, April 2017 Disabil Rehabil Assist Technol. was a pilot study of 5 older low functioning adults and compared the GCL-MPK to a non-MPK. This pilot showed promising results favoring the MPK in terms of perception of safety and some biomechanics of gait measures (not validated measures). Another cross-sectional survey study by Moller et al, April 2017 Disabil Rehabil Assist Technol. compared people who currently use MPK versus those who use non-MPK (non-dysvascular transfemoral amputees in Sweden n=42 total). This study focused on self-efficacy and found that higher self-efficacy scores were associated with more prosthesis use, but there were no differences between the groups. Cutti et al (Jun 2017 Prosthet Orthot Int.) was a retrospective cohort that examined the cost utility of the C leg versus non-MPK (n=127 total) and found that the C leg group had better mobility overall but significantly higher costs.	Thank you. We have conducted full text screening of all citations recommended by this reviewer and others, and have added any eligible studies found.

Research review section	Reviewer, Affiliation	Comment	Response
Results	Peer Reviewer #3	The characteristics of the studies were clearly described and key messages explicit and applicable.	Thank you
Results	Peer Reviewer #3	The summary of tables and data were a helpful and a useful tool to summarize the findings of the systematic review.	Thank you
Results	Peer Reviewer #3	Rehabilitation (Stuttg). 1994 May;33(2):69-75. [The geriatric amputee after discharge from successful rehabilitation--a study of former patients of the "Eifelhöhenklinik" rehabilitation center in Nettersheim-Marmagen]. [Article in German] Marmann C1. This article is only available in German and unavailable to review the strength of the evidence; it is outdated and not applicable to the US Medicare population. The prescribing rates in Germany are higher than in the US which will directly affect the abandonment rate in this population and therefore should be excluded.	We did not exclude old or non-US studies. This review is meant to be applicable generally, not just to the Medicare population.
Results	Peer Reviewer #4	Your use of the term assessment techniques and comparison with outcome measures is vague and has little clinical relevance. You state that you called it an assessment technique if it was performed at the time of prosthetic prescription or time of evaluation for a new prosthesis. To me the differentiation is whether or not the measure that was used, influenced the ultimate decision of whether to provide or what type of prosthesis to provide a given patient. There are many standardized measures that are collected at the time of prescription that do not influence the prescription decision.	We agree that the terminology is vague and is not standard or commonly used. We were attempting to get at the concepts of interest to the sponsors and our panel of Key Informants (assessment pre-prescription, predictive tool, and outcome measure). We have added caveats in several places how we categorized instruments. All included instruments are summarized together for KQ 3, including those also assessed for KQ 1 and 2.
Results	Peer Reviewer #4	Page 22 - Two questions: 1. the title of KQ1 includes the phrase “functional ability” This should be more precise. There are many different “ functions” (eg toileting, self care, IADLs, driving etc) that could be considered. Without being semantic/for clarity/do you mean mobility? Should try to be precise.	We did not mean only mobility, but did in fact mean all variations of "function."
Results	Peer Reviewer #4	Page 22 - Two questions: 2. You state that you incorporated only studies that measured function prior to prosthesis use or at the time of evaluation for a new prosthesis. From my understanding of the literature, many other measures have been used to assess patients at time of evaluation for a new	All instruments were included for KQ 1-3 overall (and KQ 3, specifically). For KQ 1, we included only those that used the measure pre-prescription in the study. This is not meant to be a complete list of instruments that could

Research review section	Reviewer, Affiliation	Comment	Response
		prosthesis. Not sure how you arrived at the few measures that you included.	(or should) be used for assessment, only of which instruments that have been evaluated at that time point.
Results	Peer Reviewer #4	One of my primary reservations of this work is your use of the term "predictive tools". Prediction tools take information at one time point and attempt to predict a future event or outcome. None of the studies you refer to appear to actually predict. These prediction tools need to be geographically and temporally validated. The studies that you include do not do this. On page 16, above Table 2 you state that you evaluated predictive validity. None of these studies are true prediction tools and did not validate their prediction tools either geographically or temporally. These are prerequisites to validating a prediction tool.	We agree. We have added a caveat in several places that only one study evaluated test accuracy (sensitivity and specificity) and that the rest evaluated only correlations with future events.
Results	Peer Reviewer #4	What seems to be reported here are "associations" between a given variable and some future outcome. For example in the Dite 2007 reference there was an association between scores at one time point and mobility and a second time point. But knowing the initial scores does not predict what the outcome will be. This section needs to remove the reference to prediction tools.	We agree. We have added a caveat in several places that only one study evaluated test accuracy (sensitivity and specificity) and that the rest evaluated only correlations with future events.
Results	Peer Reviewer #4	I believe the only prediction model that has been developed for mobility outcome is by Czerniecki JM Ann Vasc Surg 2017. This study, however, did not evaluate the effect of prosthetic componentry on outcome.	As you suggest, unfortunately your study did not meet eligibility criteria for our review.
Results	Peer Reviewer #4	KQ4 – As mentioned above this section includes an extensive discussion about what is considered a validated measure. The authors do not include the importance of validation in the population of interest. This review includes traumatic and dysvascular amputations. At a number of points in the document, including the Discussion, reference is made to study subjects and whether they are typical subjects that would be covered by CMS. This raises the question that perhaps there should be a refocus on what is the purpose of this review, and what population should it cover, and from that	The review was not meant to be exclusive to those with dysvascular amputations; although it is an important aspect of the review to highlight the applicability to the Medicare-eligible population.

Research review section	Reviewer, Affiliation	Comment	Response
		studies should be included that specifically are validated in that population. That is, the dysvascular population?	
Results	Peer Reviewer #4	The majority of the studies referenced in KQ4 have a large predominance of traumatic amputees. Is that the focus?	These studies were eligible.
Results	Peer Reviewer #4	Other than this the analysis and interpretation of the studies done to analyse the effects of prosthetic componentry is extremely well done.	Thank you.
Results	Peer Reviewer #4	KQ6 – as noted above unclear why this is included in this review. it appears that it is evaluating the extent to which amputees are satisfied with the process of accessing prosthetists. This KQ does not seem relevant to the overall thrust of the document. It does not seem to fit in the described analytical framework.	This was a key question of interest per the request for the review to AHRQ and the protocol.
Results	Peer Reviewer #4	KQ7 - it is unclear what is being reported. There are a series of paragraphs that have as headers; failure to maintain bipedal ambulation, use of prosthesis only for transfers. Use of prosthesis only indoors, abandonment of prosthesis. It is unclear whether these were the mobility outcomes, or whether these were changes in mobility from some prior level. If the latter is the case this information is only relevant if the prior level of mobility is stated. Ie change of mobility is the outcome. Also if it is change in mobility that is reported, it is unclear from the text what the timeline for the change is. Do prosthetic component types influence these changes? This section should be rewritten for greater clarity.	The outcomes were the simple interpretation of the outcome names. We have included the evidence as it was reported. We included primarily event rates, as reported. We added a comment about this at the end of the introductory summary for the KQ 7 results (just prior to the summary tables and the outcome-specific results).
Results	Peer Reviewer #4	The KQ7 also seems to attempt to quantify the extent to which LL amputees use their prostheses for transfers indoor mobility or ambulation. The alternative to the use of a prosthesis for mobility (transfers, standing, ambulation with device/without device) is the use of w/c or powered mobility aid. The optimum outcome measure would quantify the balance of and the extent to which patients use w/c or other types of wheeled mobility or ambulation and in what environments. It is the balance between these two types of mobility that really define prosthetic use.	We have addressed the key questions as written. This structure was also in line with what outcome data were reported in studies.

Research review section	Reviewer, Affiliation	Comment	Response
Results	Public Reviewer #1 Andreas Hahn, Otto Bock Healthcare Products GmbH	The report comments on our studies of Hahn et al. 2015 and 2016. We respond to clarify questions that were raised in the discussion of our papers. We acknowledge that our work was rated “Moderate” in Study Risk of Bias / Study Quality (Table 4.3 p78). We understand that there may be concerns with the study as stated in table 4.8, p85. We accept that there are clearly limitations that we openly addressed in the publication, both with respect to the used outcomes as well as to the nature of the population we investigated. We also accept that while having the practical side of our investigation in focus, more details on methodology and analysis could have been helpful, specifically with respect to the analysis conducted in this systematic review. However, we are most surprised that the report repeatedly comments on our study as “methodologically and analytically flawed” (ES-13,ES-14, ES-21,p. 102, p. 122). We feel that the point we attempted to make and to which we directed our analysis has been missed and hence the wording used is in this form inappropriate. Such statements disrespect the overall clinical arguments we provide. We will comment in detail on missing information and possible misinterpretations of our work. It is not for us to judge the level of evidence our work provides in answering the questions raised by AHRQ. However, we do request you to consider some very practical implications derived from our work. We believe the value of these conclusions is obscured by a methodological discussion that is in some cases based on misunderstandings – in part facilitated by lack of clarity on our part – and is in general only tangential to the derivation of the key points of our study. Neglecting them may lead to severe practical implications such as the continued inappropriate treatment of patients. PLEASE SEE DOCUMENT ATTACHED PROVIDING THE FULL ARGUMENT	We did not in any way mean to be disrespectful and agree that our phrasing was overly strong. We have toned down the language, removing the term flawed. We appreciate the author's comments. We have removed some of the details of the critique of the review. We have made some other revisions, but in the review we must summarize what is in the published literature that is available to all.

Research review section	Reviewer, Affiliation	Comment	Response
Results	Public Reviewer #1 Andreas Hahn, Otto Bock Healthcare Products GmbH	1. Re Hahn et al. 2015 C-Leg The report states that Hahn 2015 was excluded as there was no description or clear comparison with the previous knee(s). This information is indeed missing; an oversight we became only aware of through the AHRQ report. Previous knees had been: prosthetic knee components total 21-40 41-60 > 60 mob. 2 mob. 3 mob. 4 stance control locked knee 2% 2% 1% 2% 2% 1% 0% friction brake knee 25% 19% 21% 31% 36% 16% 9% mono centric 1% 2% 2% 1% 1% 2% 3% four-bar linkage 50% 53% 50% 50% 49% 54% 44% multiaxial knee 3% 2% 3% 4% 5% 1% 0% hydraulic (NMPK) 18% 21% 22% 11% 7% 26% 43% hydraulic (MPK) 1% 2% 0% 1% 0% 0% 0% magnetorheological (MPK) 0% 0% 0% 0% 0% 0% 1% In essence, Hahn et al. 2015 investigated the effects of switching from mechanical lower limb prosthetic knee components to microprocessor controlled knee components (only 10 out of the 1,223 subjects had previously used a microprocessor knee). We have contacted the editors of the journal and they have agreed to provide this information as a corrigendum to the existing publication.	We have reevaluated the study, particularly in light of the recently published additional data. We now include the study for Key Question 4.
Results	Public Reviewer #1 Andreas Hahn, Otto Bock Healthcare Products GmbH	AHRQ suggests reanalysis of the above work. We are generally open to suggestions and the provision of any further information that may be deemed helpful. We shall reconsider this based on the further evaluation of AHRQ and whether we may be able to contribute to the questions raised.	Thank you
Results	Public Reviewer #1 Andreas Hahn, Otto Bock Healthcare Products GmbH	2. Re Hahn et al. 2016 MODELS AND SUBJECT NUMBERS INCLUDED The report states, e.g., that at most 425 people (p100) could have been included in the regression models and that the final number of cases used in the model was not reported. This interpretation seems to be based on a fundamental misunderstanding of the analysis performed. Multivariate regression was only cited in a few sentences to provide an estimate for the upper bound of an explanatory model. All estimates shown in tables 2 and 3 are estimates from bivariate regression models. Our goal was not to	We have revised to say that no more than 425 people could have been included in a multivariate model that included all the listed variables. The article did not report how missing data were handled, such as imputation. We have fundamental disagreements about the interpretation of the reported analyses and the interpretation conveyed in this comment. However, we do not discuss this disagreement in the review.

Research review section	Reviewer, Affiliation	Comment	Response
		construct the overall optimal model – among other problems it would have been hard to validate such a model against a sample with the selection bias acknowledged in the study and discussed later on. Instead we wanted to provide upper limits on the degree of sensitivity of the outcomes to the explanatory variables. With more explanatory variables in a model the sensitivity of the model to changes in any given variable will tend to decrease. This means that if, e.g., the bivariate model for variable gait speed (table 3) finds an influence of -0.0136 for each year of age then this will tend to overstate the real sensitivity of variable gait speed to changes in age in the dataset. The coefficients of age for all outcomes (table 2) that are significant in all cases demonstrate at the same time that the outcomes are overall consistent with expectations. This can obviously not equal an independently validated measure but together with the responses to the other explanatory variables it shows that the outcomes are fundamentally sound.	We have edited our summary of the study to tone down our comments and have removed some of the details.
Results	Public Reviewer #1 Andreas Hahn, Otto Bock Healthcare Products GmbH	RESPONSE VARIABLES / OUTCOMES Outcomes used for the groups in table 3 were correctly cited by the AHRQ draft report and reported in Hahn et al. 2016: “variable gait speed (22), divided attention (18), safety (14), and change of mobility grade (14)”. We refer to those as “functional benefits”. This categorization may have been imprecise in the publication. “toileting (18), dual tasking (14), alternating stair ambulation (up, down) (13, 12), standing on ramps (11), variable gait speed (11), stepping on small obstacles (10), and carrying objects with visual obstruction (9)” were referred to as “subjects' perception”; and “ascending stairs (25, 29), stance phase resistance adjustment (21), ramps (20), walking backwards (16), small steps (15), obstacles (15), heavy loads (14), and the door-test (13)”, were referred to as “advanced maneuvers”. We refer to all three categories together as “performance indicators”.	Thank you. We have maintained our descriptions of the outcomes as they are reported in the manuscript.
Results	Public Reviewer #1	For table 2, regression models were performed on each single outcome. As mentioned in the paper, linear regressions were	Our summary is of what was reported in the article. Thank you for helping us to

Research review section	Reviewer, Affiliation	Comment	Response
	Andreas Hahn, Otto Bock Healthcare Products GmbH	performed with respect to each outcome in the categories "functional benefits" and "subjects' perception". Logistic regressions were performed with respect to each cited outcome in the category "advanced maneuvers." The most responsive examples thereof are depicted in Table 3 with N reported for each analysis ranging from 393 to 899. Most models exceed 700 subjects. Table 2 reports the ranges of estimates per outcome category. These analyses form the basis of our interpretation. The publication required an appropriate conciseness of the results. Therefore, the critique regarding underreporting certainly depends on the perspective under which the publication is read. Results on each single analysis can be provided. Here, too, we are open to further analyzing our data and reporting further results if this is deemed helpful to answer the AHRQ's questions.	understand some of the nuances of what was reported.
Results	Public Reviewer #1 Andreas Hahn, Otto Bock Healthcare Products GmbH	MULTIVARIATE MODELS Multivariate regression models were as stated in the publication used as an additional check on achievable R^2 values. However, the models, while increasing R^2, did not yield further clinically meaningful information. No consistent set of clinical variables could be identified through stepwise variable selection that would allow a characterization of the specific impact on the outcomes evaluated. The increase in R^2 was just as likely to result from spurious correlations with the specific outcome of the particular model as on meaningful explanatory power. We therefore refrained from a detailed discussion. The AHRQ report rightfully criticizes the lack of explanation of the pseudo-R^2 used for logistic regressions. This was in all cases Nagelkerke's R^2. We apologize for the oversight.	Thank you for this information. We have retained our overall comments about R^2 interpretations but we have removed much of the commentary. We added in that a Nagelkerke's R^2 was used.
Results	Public Reviewer #1 Andreas Hahn, Otto Bock Healthcare Products GmbH	CLASS IMBALANCE / PREDICTIVE POWER / R^2 The AHRQ report discusses the problem of "class imbalance", referring to the high responsiveness in the study population. We report on the upper part of the Likert Scale sensitivities in Figures 1 and 2. This class balance in itself is a positive result as it shows the ability of the prosthetist to select promising patients for trial	We maintain that we are accurate that R^2 values have limitations when it comes to assessing the discriminatory and calibration performance of a predictive model. We believe it is true that the R^2 value is not generally very informative. We have

Research review section	Reviewer, Affiliation	Comment	Response
		fittings. A lack of explanatory power of models for a given explanatory variable therefore simply shows that the prosthetist successfully used all the information contained in the variable regarding the prospective likelihood of the respective benefit. If a significant population of e.g. Mob.2 patients remains, then this illustrates the limits of the variable for selection. The models check for additional explanatory power and confirm whether the relative distribution of the different explanatory variables compared to their distribution in the overall amputee population really is the limit of their selection potential. We reject the blanket criticism of the R^2 value as a measure of explanatory power. Especially the statement that the R^2 value is "not generally a very informative" metric of predictive performance is in our opinion at least a controversial opinion that should not be stated as an undisputed fact. The difference in assessments might be the result of the misunderstanding regarding the nature of the models (bivariate vs. multivariate) and outcomes (binary vs. metric) mentioned above. Fitting outcomes occur on an underlying continuous scale and are then classified on a five point scale. The concentration of results in parts of the Likert Scale is a challenge. But as the R^2 of a model compares the residual variance to the null model the shift in mean is accounted for. Additionally, in most outcomes no category has more than 60% of the data (see figures 1 and 2) and the class imbalance is not strong enough to set a relevant ceiling for the R^2 measure. We do not dispute the shortcomings of the R^2; but any performance metric does have shortcomings as well as advantages. It is mostly a question of choosing the right tool for the right job with us standing by our choice for the research question investigated in the study. This does not mean, however, that we are not ready to supply other performance measures to address the questions of the AHRQ. As an alternative we would suggest the following approach – at least for categorical explanatory variables: For every level	referenced this statement. The author here notes the lack of variability which makes any modeling effort challenging. We highlighted these concerns. We encourage the authors to publish a revised analysis that deals with many of the issues that the author describe.

Research review section	Reviewer, Affiliation	Comment	Response
		of, e.g., mobility grade we fit a normal distribution to the results for, e.g., safety – i.e. we assume underlying normally distributed data that has then been binned in the classes 1 to 5. We then search for the mean and variance that minimize the deviation of the class counts to the expected counts. Doing so allows us to extract more meaningful comparisons between concentrated data while offering more stable estimates of the true parameters than taking the group means.	
Results	Public Reviewer #1 Andreas Hahn, Otto Bock Healthcare Products GmbH	MODEL INTERPRETATION Some variables (like e.g. mobility grade) are in some geographies including the USA used in clinical practice as if they had classifying power and access to MPK technology is often denied on this parameter only. We can, however, see that our population shows clear improvements in a high number of performance indicators independent of the value of the independent clinical variable investigated. We therefore conclude that models that would rely on such a variable as a predictor (or rather, as classifiers for denial) would lead to objectively wrong clinical decisions. We deny that such a variable has the predictive power to make it the sole or predominant selection criterion. The low R^2 values that are associated with these models help us characterize the lack of explanatory potential of this variable. They were used as a means to characterize the potential of a variable to guide a clinical decision. In the regression models we also report on the effect estimates. We observed that in the linear regression models estimates were high and a large portion of the regression models showed statistical significances. The participants span the entire range of the clinical variable in question and were (with one exemption) sufficiently distributed. We can therefore conclude that the variable showed effects on the respective Likert scales. Most of such effects were highly plausible and – as mentioned – can be seen as an affirmation of the validity of the assessments used as outcome variables. Most striking is the (lack of) influence of parameters like mobility grade rating, age	Thank you for this comment. We understand that your interpretation is that the model fails to reliably predict who will benefit. However, we have pointed out a range of technical issues in the analysis that call into question this interpretation.

Research review section	Reviewer, Affiliation	Comment	Response
		or etiology of vascular disease. While we accept that a class imbalance is likely, we still believe that in the way we analyzed and interpreted the data, our conclusions are valid. We accept that, if we had claimed predictiveness, more caution and methodological finesse would have been required to characterize such findings. We believe that as we only claim the failure of the variables in question to sufficiently characterize all likely responders, our argument is sound and fair.	
Results	Public Reviewer #1 Andreas Hahn, Otto Bock Healthcare Products GmbH	SELECTION BIAS The question of class imbalance also refers to the generalizability of our findings. How would a more general population be characterized? And who of those would not have responded as positively to Genium (or C-Leg)? If we look at the demographic description, our populations do not seem to be too different from LLP populations described elsewhere and in Hahn 2016 include 131 individuals with an age > 60 years. This portion is significantly higher in Hahn et al. 2015. It is striking that we did find that a subject's capability to utilize the functional benefits of Genium was almost independent of age. The finding was similar to that with C-Leg in Hahn et al. 2015. While we stated that a pre-selection has occurred, we observe that this shows a limited shift in the demographic characteristics of our population. The criteria of “pre-selection” used by the prosthetist remain unclear. Generally, selection consciously or unconsciously based on the variables discussed in the paper seems likely; in particular due to the lack of remaining explanatory power left in the variables. However the demographics of the selected population and their successful response to the test fittings clearly illustrate the limits of the explanatory power to classify the population at large. The selection bias is therefore not a flaw of the paper as the demographics of the selection are part of the result.	Our categorization of this study as not generalizable to the Medicare population is based on a standard system we used. The people receiving the Genium knee were relatively young on average and few had dysvascular etiologies. Regarding our comment about selection bias, it appears that the author agrees with our interpretation; namely, that the population was selected based on the prostheticists' assessment that they were most likely to benefit from the Genium knee.
Results	Public Reviewer #1	CONCLUSIONS Our primary clinical conclusions reach beyond the statistical analysis: If, in a more generalized	We understand the author's interpretation but we note that these conclusions do not readily

Research review section	Reviewer, Affiliation	Comment	Response
	Andreas Hahn, Otto Bock Healthcare Products GmbH	population, a parameter such as mobility grade or K-level were used to determine access to a specific prosthetic component, a clinically significant portion of likely responders would be withheld their eligibility for certain components and hence deprived of their chance to benefit from safety and functional gains. Another parameter worthwhile to be highlighted in this discussion is BMI that in our analysis and within the range investigated in our study did not even show sufficient sensitivity. Thus, we conclude that the parameters we investigated lack the power to predict individual outcomes and to justify individual clinical decision making. The multivariate analysis did not help to identify meaningful combinations of such variables either. The exception we make is related to daily walking distance. We reported that those in our sample having daily walking distances below 400m were clearly underrepresented. Hence, the “pre-selection” of likely responders may find that its representation somewhat correlated to this variable. Due to the power of our sample we found this variable may indeed show threshold-like behavior and seems to be interesting enough to be studied further. We concur with the conclusion of AHRQ that there is no evidence which particular patient may benefit from different specific components. We also concur that the lack of such evidence may not imply the general absence of such characteristics. We do, however, feel that we contributed to the discussion that there may be some characteristics that we may already assume to be less meaningful.	extend to all patients. They apply only to patients chosen by their prostheticist based on their likelihood of benefiting.
Results	Public Reviewer #1 Andreas Hahn, Otto Bock Healthcare Products GmbH	Our personal opinion is that, until this has been resolved, individual assessments may be the best choice to identify whether or not a patient benefits from specific prosthetic components at least for lower mobility grades and with other parameters in doubt. The use of outcome measures with high validity and sensitivity in LLP to determine the effects of a trial fitting is undoubtedly recommended.	We have summarized our understanding of the analysis. We do not make clinical or policy recommendation.

Research review section	Reviewer, Affiliation	Comment	Response
Results	Public Reviewer #1 Andreas Hahn, Otto Bock Healthcare Products GmbH	We would also like to indicate that some of AHRQ's conclusions, if not being read carefully, may be misinterpreted as if there was no difference between the benefits of different categories of prosthetic components. This is highly relevant for the current discussion of differential benefits of various prosthetic component categories, be it energy restoring and return feet (ESAR) or microprocessor controlled exo-prosthetic knee components. We recommend to more clearly state in the final report that AHRQ's research focus had not been to discuss the differentiation of prosthetic component categories to avoid such possible misinterpretations	Thank you. This has already been a frequent misinterpretation. We have added several new explicit statements stating what the review does and does not address.
Results	Public Reviewer #1 Andreas Hahn, Otto Bock Healthcare Products GmbH	The AHRQ report repeatedly refers to our work as being analytically and methodologically flawed. With all due respect, we request to abstain from such characterization. We hope that we were able to provide more insight and could clarify some misinterpretations of our work. We acknowledged and disclosed all the limitations we had been aware of when analyzing and discussing real-world data. We appreciate the limitations that are associated with the choice of outcomes in this dataset. We acknowledge that we do not present data on a controlled trial designed to possibly answer the question set out by AHRQ. It is of course up to AHRQ to decide to which extend they find our work qualifies to contribute to the question that was set out in this discussion. In our opinion, our conclusions are even in the light of all limitations sound and valid. Hence the contribution to the debate may likely be more than "insufficient". The wording "analytically and methodologically flawed", however, fails to capture the main point of our work, i.e. the characterization of a responder population and the identification of the inappropriateness of a number of clinical variables to decide on an individual's fate within such or similar populations. Rather, the wording chosen in an AHRQ report has the potential to lead to a general devaluation of our contribution to the field which we feel is	We agree that the language was too strong and have removed the offending phrase. We thank you very much for your comments and encourage you to publish any appropriate re-analyses or clarifications about the study.

Research review section	Reviewer, Affiliation	Comment	Response
		completely out of perspective and we assume possibly beyond the intention of AHRQ.	
Results	Public Reviewer #12 Elizabeth Allen	The answer of KQ 5, "How do study participants' preprescription expectations of ambulation align with their functional outcomes?" claims that no studies are available to answer this question. I submit that studies which compare an amputee's pre-amputation activity level and motivation level regarding rehabilitation to their functional outcomes would help answer this question. Some examples include: George, Jaiben, et al. "Predicting functional outcomes after above knee amputation for infected total knee arthroplasty." The Journal of arthroplasty 32.2 (2017): 532-536. Schnall, Barri L., et al. "Functional Outcomes of Service Members With Bilateral Transfemoral and Knee Disarticulation Amputations Resulting From Trauma." Military medicine 181 (2016). Greive, A. C., and G. J. Lankhorst. "Functional outcome of lower-limb amputees: a prospective descriptive study in a general hospital." Prosthetics and orthotics international 20.2 (1996): 79-87. Dunne, Simon, et al. ""If I can do it I will do it, if I can't, I can't": a study of adaptive self-regulatory strategies following lower limb amputation." Disability and rehabilitation 36.23 (2014): 1990-1997.	We have screened these articles in full text. We required that studies correlate (or otherwise analyze) measures of pre-prescription expectations with post-prescription functional outcomes.
Results	Public Reviewer #13 Tim Bump	Regarding question 6. I appreciate that "satisfaction of the patient" is considered within this literature review, however, while the SAT-PRO was considered a validated measure, it wasn't used when pulling studies for Question 6. I understand that they stated they are looking at the process (including payer and prosthetist), but why are those placed under "satisfaction" with their payer and prosthetist, when it is really the prosthesis that is provided that often makes the largest difference in their activities of daily living? I understand questionnaires do not often provide high evidence, but if they could be at least standardized it should provide a higher level of validity.	KQ 6 refers to satisfaction with the process, not satisfaction with prostheses (as per SAT-PRO). We have added a statement to this effect.

Research review section	Reviewer, Affiliation	Comment	Response
Results	Public Reviewer #14 Seung Eun Lee	1) Overall, I agree that 2MWT, ABC, PEQ, TAPES, and TUG are valid and reliable. 2) On page 101, the report didn't include AMP test as Medicare applicable outcome measurements even though Table 1-3.1 shows AMP have suitable reliability, validity and MDC. Please add AMP. 3) Plus-M: I understand it is new and there are not many papers to evaluate, but it has higher potential since Plus-M has no floor effect nor ceiling effect. Please re-consider article [Hafner 2017 27590443] for Construct Validity for Plus-M. The article compared correlation of Plus-M and other outcome measurements that Medicare approved, such as ABC, PEQ, and TUG (pg 101). The Plus-M was compared with various approved measurements, this article should be strongly considered.	All studies of instruments have been reevaluated and reanalyzed. Hafner 2017 has been included.
Results	Public Reviewer #18 William Hendrix	"In regards to KQ 6: I would be interested in seeing the comparison of satisfaction with the process of accessing a LLP versus satisfaction with the process of accessing other medical devices and/or medical services outside of the prosthetic field. Findings may suggest that the satisfaction with prosthetists is higher than the average satisfaction among medical professionals and that the technical skills, information giving, and interpersonal manner qualities of prosthetists should be sought after in other fields of medicine.	This is a very interesting question. We did not review other topics and did not find a review of a sufficiently similar topic for comparison. However, satisfaction with care was generally high.
Results	Public Reviewer #18 William Hendrix	In regards to KQ 7: This section seems fairly biased giving the outlook on receipt of a LLP very dim. This section asks questions that lead to negative connotations surrounding future use of prostheses. Of course the main group of people with amputations that would abandon their prostheses has dysvascular conditions. This is in part due to the nature of the pathology. For 1.9 million people living in the US with limb loss, these studies do not accurately predict the amputee population. This section does not take into account all of those who did not participate in the study because they have been able to live beneficial lives with a high quality of life. This section can be applied to various medical devices or medical	We do not agree that posing the questions regarding long-term outcomes in a population that may have "dim outlooks" is inherently biased. We present the evidence to best support our knowledge about the long-term outcomes posed. It is for policymakers and others to determine if these rates are "appropriate" or not, if anything needs to be done to improve them, and if so, what.

Research review section	Reviewer, Affiliation	Comment	Response
		services (abandonment of dentures, abandonment of mobility aids, etc.)."	
Results	Public Reviewer #2 Andreas Kannenberg, Otto Bock HealthCare LP	Literature flow (page 20) Here it would be helpful to understand how many total patients the studies reviewed for each key question had enrolled.	The numbers of included patients are in the results section for each Key Question (or for each instrument).
Results	Public Reviewer #2 Andreas Kannenberg, Otto Bock HealthCare LP	Key Questions 1-3 Noteworthy again, is that the report does not mention that many prosthetic studies had used Klevel as an inclusion or exclusion criterion or to characterize patients or even as an outcome measure. Therefore, the report should list K-level as a non-validated assessment/measure. Also, it should be stated that Medicare does not accept any of the listed measures to support the Klevel determination of a patient.	In the summary for KQ 1-3 we have added that K levels have not been validated. We make no comment on Medicare policy.
Results	Public Reviewer #2 Andreas Kannenberg, Otto Bock HealthCare LP	The study of Theeven et al, 2011, was criticized for not having validated their definitions of their 3 subgroups of K2 patients. That criticism raises the question of how subgroups of a nonvalidated classification/measure could ever be validated?	The main point is that it was an ad hoc subgroup classification. We have added that K levels overall have not been validated. Above, under KQ 4 validated outcomes, we note that we assessed K levels together with validated predictors (subgroups) but in fact they have not been validated.
Results	Public Reviewer #2 Andreas Kannenberg, Otto Bock HealthCare LP	Quote: “Overall studies that investigated subgroup effects did not identify participant characteristics that predict which lower limb amputees would most benefit from a given component. Based on the methodology used to assess strength of evidence, the studies warrant a low strength of evidence that evaluated patient characteristics do not predict which patients would most benefit from a given LLP component. However, it may be more accurate to conclude that the evidence is currently sparse and fails to adequately address whether different subgroups of amputees are more or less likely to benefit from given specific components.” Comment: In order to prevent misinterpretation of this conclusion, it should be accompanied by a statement saying that this systematic	This caveat has been added in several prominent locations.

Research review section	Reviewer, Affiliation	Comment	Response
		review did not evaluate all the evidence for differential overall effects of different types of prosthetic components.	
Results	Public Reviewer #2 Andreas Kannenberg, Otto Bock HealthCare LP	Key Question 7 Quote: "Most studies of amputees with outcomes of interest were rejected because the analyses were not restricted to people with prescribed prostheses and were thus mostly analyses of predictors for not receiving a prescription for LLP." Comment: This is an important point that needs to be addressed in future updates because the prosthesis fitting rate of new amputees in the U.S. is considerably lower than in comparable countries in Europe.	We have added a sentence to the Future Research section for KQ 7 that studies should include unbiased participant samples.
Results	Public Reviewer #2 Andreas Kannenberg, Otto Bock HealthCare LP	Quote: "However, we found a moderate strength of evidence, based on six studies, that about 11 to 22 percent of lower limb amputees who receive a LLP prescription abandon the prosthesis (stop using it) at about 1 year; these studies are generally representative of people with LLP, in particular older adults and those with dysvascular etiologies. … Also based on four, generally representative studies, there is low strength of evidence that 11 to 37 percent of LLP recipients use their prostheses only indoors; however, these studies are somewhat inconsistent and imprecise. There is low strength of evidence about how likely different subgroups of people use their prostheses only indoors, suggesting that people with transfemoral amputations, or who are older, or with bilateral amputations are more likely to be limited to indoor use." Comment: Here it should be made clear that there is no evidence for the reasons for abandonment or indoor use only, so that this finding may not be used to deny coverage or fitting of prostheses to certain subgroups of patients, especially those with higher amputation levels based on the higher risk of prosthesis abandonment. Also, I would like to re-emphasize that studies conducted outside the U.S. may not yield results that are transferable to the situation here. This is due to the fact that primary fitting rates of new amputees are usually higher in European countries, likely resulting in higher	We already clearly state there is no evidence among eligible studies about reasons for abandonment. This was a specific outcome of interest.

Research review section	Reviewer, Affiliation	Comment	Response
		abandonment rates as many of these patients would probably never receive a prosthesis in the U.S.	
Results	Public Reviewer #20 Anonymous	"• The report gives the impression that “no significant difference” can be interpreted as a “lack of evidence” for the benefits of technologies for different sub-populations. I don’t think this is the authors’ intent but this point could be made clearer. One might interpret the findings differently as there being evidence that prosthetic intervention is “equally beneficial” to each sub-population group, regardless of amputation aetiology.	Given the overall state of the evidence, we do mostly conclude that here is a lack of (or at least insufficient or low strength) of evidence. For the most part, for this evidence base it would be incorrect to say that the evidence supports equivalence.
Results	Public Reviewer #20 Anonymous	• The statistical analysis on low sample size is due to the very large variability in the amputee population and this is confounded by the fact that the technology is, in reality, adjusted to suit the individual wearer. For example, micro-processor devices are programmed, all devices require alignment. Hence, the clinical reality is there is no “standard” prosthetics intervention. If anything, a case series study is the more representative of the clinical realities. Significance in a single subject is still a result for one subject and should not be dismissed (I appreciate that a case study does not permit comparisons of effects for different demographics - the goal of KQ4 - but I felt this may make an interesting discussion point).	Thank you for this comment. We have added the concept into the Limitations of the Evidence section of the discussion.
Results	Public Reviewer #20 Anonymous	• The standard for judgement of potential bias in studies may be set too high. The ideal situation is a double-blind, randomised control trial but often, due to clinical realities, this just isn’t feasible in prosthetics. Very often blinding is impossible, purely on safety grounds i.e. not knowing the limitations of a device could lead to falls. P&O studies are broadly representative of what actually happens clinically and therefore should not be dismissed due to the “perceived weakness” of not being a DBRCT. The models and standards used for areas such as pharmaceutical medicine are not directly translatable to the prosthetics field."	We included randomization, blinding, etc, in the evaluation of risk of bias, but acknowledge the inherent difficulties in achieving these goals.

Research review section	Reviewer, Affiliation	Comment	Response
Results	Public Reviewer #22 Claire Kilpatrick	The comments here relate directly to the results of Key Question 4. Reviewers recognize that patients and providers were virtually impossible to blind, as components are visible and the impacts of those components are tangible. They also recognize that it would have been difficult to blind outcome assessors. Despite these acknowledgements, in the tables detailing the results of the review the authors continued focus on risk of bias is misleading. An amendment to the results presentation in Table 4.3 which refers to risk of bias should be made where it is highlighted that the circumstances of prosthetic component research make it very difficult to blind the subjects/reviewers, and therefore risk of bias should not be interpreted as indicative of the overall quality of the study nor a reason to discredit the results of a study.	The summary of the risk of bias of the studies does acknowledge that blinding is difficult and may be impossible.
Results	Public Reviewer #5 Orthotic and Prosthetic Alliance	The brief public review period did not offer sufficient time to examine every instrument review in detail, but we examined several measures to determine whether the authors' overall assessments were accurate. Unfortunately, we have identified a number of concerning issues related to the characterization of specific outcome measures included in this review.	We have completely reassessed, reanalyzed, and rewritten the sections pertinent to test validity (KQ 1-3).
Results	Public Reviewer #5 Orthotic and Prosthetic Alliance	The authors indicate that several subscales of the Patient Reported Outcome Measurement Information System profile (PROMIS-29) have been assessed for validity in people with lower limb amputation (p. 39-40). One cited article22 is an abstract; the other23 reports normative-type data on seven PROMIS-29 subscales in a large sample of LLP users. We see no evidence in this paper to explain the authors' assessment of validity of the Depression, Physical Function, and Social Role Satisfaction subscales. The article reports that prosthesis users' scores on these scales are significantly different from a T-score of 50 (an average score based on the U.S. General Population norms) and that scores on other scales are not significantly different than 50. However, such data does not, in itself, constitute evidence of validity.	We have completely reassessed, reanalyzed, and rewritten the sections pertinent to test validity (KQ 1-3), including this instrument. The article cited here has reported reliability and MDC(90). Test validity has been removed.

Research review section	Reviewer, Affiliation	Comment	Response
Results	Public Reviewer #5 Orthotic and Prosthetic Alliance	The authors report that the Quality of Life in Neurological Conditions Applied -Cognition/General Concerns (NQ-ACGC) instrument has been tested for reliability and construct validity in people with lower limb amputation. While one cited reference24 indicates it has been assessed for reliability, neither of the cited references24, 25 indicates the instrument has been assessed for construct validity.	We have corrected the inclusion of test validity for this instrument.
Results	Public Reviewer #5 Orthotic and Prosthetic Alliance	The authors also indicate that the Prosthetic Limb Users Survey of Mobility (PLUS-M, p. 42) lacks evidence of validity. However, one of the cited references (Reference 74) "Construct Validity of the Prosthetic Limb Users Survey of Mobility," indicates that this instrument has been assessed for content validity (i.e., floor and ceiling effects), convergent construct validity, and known groups construct validity in a large sample (n=199) of LLP users seemingly reflective of the Medicare population (mean age 55, 43% of the sample experienced amputation due to dysvascular disease, participants classified as K-levels 2-4).26	We have completely reassessed, reanalyzed, and rewritten the sections pertinent to test validity (KQ 1-3), including this instrument. PLUS-M was found to have evidence of both test validity and reliability. We did not use floor and ceiling effects to assess content validity. They are evaluated separately.
Results	Public Reviewer #5 Orthotic and Prosthetic Alliance	One of the more concerning sections of this review relates to the Trinity Amputation and Prosthesis Experience Scales (TAPES). The authors note that the TAPES instrument has 26 subscales, including such scales as "gender," "age," and "level of amputation." It would appear that the authors have inappropriately elected to characterize the demographic form attached to the TAPES instrument as a series of measurement scales. Further, they have determined that these "scales" all exhibit evidence of validity. The TAPES instrument, in fact, has only nine scales, not 26,27, 28 including three Psychosocial scales (General Adjustment, Social Adjustment, Adjustment to Limitation), three Activity Restriction scales (Functional, Social, and Athletic Restriction), and three Satisfaction with the Prosthesis subscales (Functional, Aesthetic and Weight Satisfaction). The authors' review of this instrument is particularly concerning because, if they are unable to distinguish a	We have reevaluated TAPES (and all other instruments). The concerns noted have been corrected.

Research review section	Reviewer, Affiliation	Comment	Response
		demographics form from a health status instrument, readers may legitimately question the authors' ability to accurately extract detailed information from other instruments or articles.	
Results	Public Reviewer #5 Orthotic and Prosthetic Alliance	It is unclear how instruments classified as "assessment techniques" were determined. The authors stipulate "here, we limit the list of assessment techniques to those measures either described by studies as assessment techniques or studies that explicitly included lower limb amputees prior to prosthesis use or at the time of evaluation for a new or replacement LLP." (See; p. 22.)	We believe this description is accurate, complete, and clear.
Results	Public Reviewer #5 Orthotic and Prosthetic Alliance	One of the references26 in the draft report indicated that the study sample included "current lower limb prosthesis users (N=199; mean age±SD, 55.4±14.3y; 71.4% men) that were assessed before receiving a replacement prosthesis, prosthetic socket, and/or prosthetic knee," (p. 277) and "mobility outcomes were collected as part of a longitudinal study to evaluate the effects of prosthetic interventions; data presented here were collected pre-intervention." (See; p. 278.) Thus, it would appear that the measures included in this study (AMP, TUG, PLUS-M, ABC, PEQ-MS, and PROMIS-PF) may meet the authors' stated criteria for being considered an "assessment technique." The study had a large sample of individuals, many of whom were K2 or K3 amputees, and had amputation due to dysvascular disease (79.2% and 45.0% among the K2 and K3 groups, respectively).	We respectfully disagree with this interpretation. The intervention in question in Hafner 2017 (PMID 27590443) was the rehabilitation, not the prosthesis. Eligibility criteria for the study included use of a prosthesis for at least 4 months.
Results	Public Reviewer #5 Orthotic and Prosthetic Alliance	A second study29 described development of the Amputee Mobility Predictor (AMP), an instrument "that would enable physicians, prosthetists, and physical therapists to assess objectively an amputee patient's ability to ambulate with a prosthesis." (See; p.614.) This study also included a large sample (n=191) of people who were classified as K0 to K4 (mean ages of 77.1, 74.5, 65.4, 53.1, and 36.7 for K0, K1, K2, K3, and K4, respectively), and a large number of people with amputation due to dysvascular disease (45.5% of the sample). Thus, instruments used in this study (AMP, 6MWT, AAS)	AMPnoPRO has been included as an assessment technique.

Research review section	Reviewer, Affiliation	Comment	Response
		would also seemingly qualify as “assessment instruments.” The AMP, in particular, would seem to be an ideal “assessment technique,” as it was designed to be administered to people with and without a prosthesis29 and has the ability to distinguish between groups of individuals classified by different K-levels.26, 29 It is also the instrument most commonly used by U.S. prosthetists for determining prosthetic patients’ K-level.30	
Results	Public Reviewer #5 Orthotic and Prosthetic Alliance	For Key Questions 1-3, the authors elected to characterize instruments and make recommendations for their use (as assessment techniques, prediction tools, or functional outcome measurement tools) according to psychometric properties, such as reliability and validity. While these properties are critically important, we submit they are not sufficient for conclusive recommendations regarding use of these instruments in research and routine clinical practice. Below, we highlight several reasons why results of this initial psychometric review may be limited:	Thank you. Please note that we are not making recommendations about the use of the instruments.
Results	Public Reviewer #5 Orthotic and Prosthetic Alliance	Many of the instruments identified in the review were developed, validated, and/or tested in an era when psychometric standards were lower than they are at present. Therefore, the authors also applied psychometric criteria that are generally lower than current standards require. For example, while contemporary standards call for the inclusion of a minimum of 50 participants in reliability testing,11 the authors used a minimum of 30 participants. Thus, measures may have been included in the review (and by inference recommended for future research), when they do not meet contemporary standards. We encourage the authors to discuss in the final report changes in psychometric standards over time to provide context to the measures identified in the review.	Thank you. This is a good point to make more explicitly. We have added to the Discussion: Furthermore, the reader is reminded that lack of evidence regarding the psychometric properties of instruments does not imply that they are not valid or reliable, only that they have not been (adequately) evaluated. Standards for psychometric testing have changed over the years, so older instruments, evaluated by older studies, may not have psychometric property evaluations that meet modern criteria.
Results	Public Reviewer #5 Orthotic and Prosthetic Alliance	The details of an instrument’s development and testing can provide information that challenges if that instrument can (or should) be recommended for use in research or clinical care.	We did our best to avoid using study authors' conclusions about the psychometric properties, including specific types of validity.

Research review section	Reviewer, Affiliation	Comment	Response
		For example, the Transfemoral Fitting Predictor (TFP)35 was recommended by the review authors as an "assessment technique." However, available evidence does not adequately support this determination. Although the TFP developers indicated they assessed "construct validity," they did so using principal component analysis, a method not considered to provide sufficient evidence of construct validity.36 The developers' assessment of discriminant validity showed that only the two highest tasks (i.e., stand with an external walking aid for 30 seconds, and walk to the end of the parallel bars with a walking aid and turn around) were significant predictors of who would be fit for a prosthesis. While the developers did not report scores for individual participants, we suspect many individuals with amputation would be able to walk the parallel bars with a walking aid without difficulty (i.e., exhibit a ceiling effect on the TFP). Lastly, the developers used questionable methods to assess inter-rater reliability (i.e., having a group of therapists watch a video of another therapist administering the test—a technique used more appropriately to assess intra-rater reliability) and performed no assessment of test-retest reliability.	Instead, we relied as much as possible on the reported data and analyses.
Results	Public Reviewer #5 Orthotic and Prosthetic Alliance	Assessment of validity that does not include comparisons to other standardized instruments also may be questionable. For example, the systematic review authors characterized the SF-36 as a valid and reliable "assessment technique" based on a study by Hart and colleagues.37 The only efforts to establish convergent construct validity in this article were performed by comparing the SF-36 domain scores to the SF-36 component scores (which are, in turn, derived from the domain scores38) or by an ad hoc measure of the prosthetists' perception of the prosthesis user. The degree to which component scores can be used to establish the validity of the domain scores (and vice-versa) is debatable at best. Similarly, it is concerning that the revised PF-15 (an ad hoc modified version of the physical functioning domain score) could not differentiate people by	The study provided evidence of divergent validity by K levels for the SF-12 subscales. The study reported only Rasch analysis for PF-15. This study, and all, were reextracted and summarized.

Research review section	Reviewer, Affiliation	Comment	Response
		Medicare Functional Classification Level (MFCL) and exhibited a ceiling effect.	
Results	Public Reviewer #5 Orthotic and Prosthetic Alliance	While we acknowledge that details of these instruments may have been overlooked in a top-level assessment of reliability and validity performed in this review, they seem particularly pertinent when considering if a measure can be recommended for a given purpose. While the above examples illustrate practical and/or psychometric issues with select instruments, we expect that other measures included in this review may exhibit similar issues. We therefore urge caution in characterizing measures as “valid” or “not valid” based on the preliminary analysis conducted by the authors of this review. We instead encourage the authors to suggest additional efforts that may be required to further vet the instruments identified in this review. For example, the authors may wish to reference consensus efforts by other professions (e.g., physical therapists) to review and recommend outcome measures.39 Such efforts could serve as a model for instrument reviews applicable to measuring outcomes in people with lower limb amputation.	It is important to note that this report reviews and summarizes the evidence. We do not make any recommendations for policy or clinical decisionmaking, including recommendations whether instruments should be recommended. We defer to other experts to recommend outcome measures. Nevertheless, we agree that we should not be calling instruments valid or not valid, but instead more clearly talk about "evidence of validity", which we now do. We have removed listings or "not valid" instruments since this is likely an unfair interpretation of the evidence (since instruments were not evaluated for all types of validity and for validity against all other potential measures). We have explained this in the Methods.
Results	Public Reviewer #5 Orthotic and Prosthetic Alliance	We noted a number of minor issues with the narrative instrument reviews.	Thank you.
Results	Public Reviewer #5 Orthotic and Prosthetic Alliance	The “summary of studies and participant characteristics” reference data in Appendix C, but the corresponding evidence tables are not labeled as Appendices.	Thank you. This has been corrected.
Results	Public Reviewer #5 Orthotic and Prosthetic Alliance	Details for the Activities-specific balance confidence scale (ABC) are provided without reference (p. 26).	This has been corrected for ABC and all instruments.
Results	Public Reviewer #5 Orthotic and Prosthetic Alliance	In several instances, the authors indicate that instruments “displayed reliability” (pp. 28, 33, 34, etc.). Given the implied uses for these instruments (e.g., monitoring patients over time, comparing between groups of patients), the form of reliability seems important. For example, if measures are to be used for monitoring patients over time, evidence of test-retest reliability	Following the approaches taken in prior similar work summarizing psychometric properties, cited in the Methods section, and with the goal of comprehensibility, we took the more straightforward approach of

Research review section	Reviewer, Affiliation	Comment	Response
		would be important. The authors are encouraged to clarify the types of reliability assessed for each measure in the narrative summaries.	dichotomizing reliability into yes/no. Further details are presented in the Appendix table.
Results	Public Reviewer #5 Orthotic and Prosthetic Alliance	The authors characterize the PEQ scale as a "Likert" scale (p. 41). A Likert scale is a scale of agreement. Here, the PEQ scale would be better classified as an "ordinal" scale.	We have changed the description to the terminology used in the studies.
Results	Public Reviewer #5 Orthotic and Prosthetic Alliance	The narrative summaries selectively refer to "weak" validity (e.g., Assessment of Quality of Life on p. 35, Patient Generated Index on p. 39, Russek's Code on p. 43, TAPES on p. 46-47, WHODAS on p. 50), but in general, the strength of evidence related to validity or reliability is not discussed for all other measures. We recommend the authors consistently apply terms related to the strength of evidence	We have removed the concept of weak validity, etc.
Results	Public Reviewer #5 Orthotic and Prosthetic Alliance	A number of minor issues were noted with the narrative summaries of the remaining Key Questions.	Thank you.
Results	Public Reviewer #5 Orthotic and Prosthetic Alliance	The authors note that "Ten studies included between 5 and 168 users of LLP; one included 899 amputees." (See; p. 73.) However, the data table (p. 75) indicates that the study by Alaranta and colleagues46 includes 208 people. That data is incorrect, as the study included 168 participants.	The 208 referred to the number of people enrolled, but we have changed the table column to number analyzed. 168 is correct.
Results	Public Reviewer #5 Orthotic and Prosthetic Alliance	We suggest replacement of the term "nonsignificant," which is used in Table 4.4, with the phrase "not statistically significant."	Nonsignificant is standard terminology, but in most instances, we replaced with "NS" which is now defined as not statistically significant.
Results	Public Reviewer #5 Orthotic and Prosthetic Alliance	We suggest citing the "similar study" noted in the review of the study by Hahn and colleagues (p. 100, first paragraph).	Thank you. This accidental omission has been corrected.
Results	Public Reviewer #5 Orthotic and Prosthetic Alliance	We suggest revising "…suggesting that few people failed to have some improvement…" (p. 100) to "…suggesting that most people had some improvement…"	Done.
Results	Public Reviewer #5 Orthotic and Prosthetic Alliance	We question why the study by Webster et al.,47 was not included in the results pertaining to Key Question 6. This was a prospective, multi-center study that was designed to assess prosthetic fitting, use, and satisfaction in LLP users. While the study population, which included people with transmetatarsal,	Webster evaluated satisfaction with the prosthesis, not satisfaction with the process of accessing a prosthesis

Research review section	Reviewer, Affiliation	Comment	Response
		transtibial, or transfemoral amputation, consisted of Veterans, their mean ages were 61-63 years and included a high percentage of people with diabetes (50-100%).	
Results	Public Reviewer #5 Orthotic and Prosthetic Alliance	The authors may wish to examine the document for text that is pulled directly from the source publications. For example, the sentence, “Less favorable ratings related to being able to depend on the prosthetist for the individual’s wellbeing (26% disagreed or strongly disagreed)” (p. 105, first paragraph) is a direct quotation from the cited reference,48 but is not presented as such.	Thank you. We have made it clear this was a quote.
Results	Public Reviewer #5 Orthotic and Prosthetic Alliance	We also suggest revising the phrase “…estimated exact…” (p. 114, first paragraph) to include one term or the other, as they seem to be mutually exclusive.	We removed exact. Exact confidence interval is correct; it is a specific method to estimate the CI that is not based on the standard error. But we agree the term is potentially confusing and is not necessary.
Discussion / Conclusion	TEP Reviewer #1	The authors have done a good job summarizing the major findings of the review. The authors made a fantastic point on page 22, line 21-26 stating that the evidence is currently sparce and fails to adequately address the sub-groups. This is one of the main points of this review. How will the reviewers use this point to suggest recommendations for future research. It is felt that once the above issues are addressed, the future research section could be finely tuned.	We have a large section on future research suggestions for studies of heterogeneity of treatment effects. We have added in the point that the evidence is currently sparse.
Discussion / Conclusion	TEP Reviewer #2	The major finding that current validated measures and a requirement for new, "improved" to be validated measures, should form the basis of needed future research is clear.	Thank you
Discussion / Conclusion	TEP Reviewer #2	The major limitations of the "few studies" used to characterize abandonment and limited uses outdoors, should be boldly included in any statements alleging the "relative effectiveness" and "long-term use" of LLP's.	The methodological limitations (and low strength of evidence) are stated in overall summaries.
Discussion / Conclusion	TEP Reviewer #2	I can only imagine a "downward spiral" of limited performance componentry in the future looking to cut costs, carrying warning labels, "indoor use only".	We hope this is not the case. We do not believe this review supports such an interpretation.
Discussion / Conclusion	TEP Reviewer #3	I do not think that the implications of this report are clearly stated. Although the authors feel that the report will primarily	The discussion does not suggest that the report will be used primarily to improve LLP

Research review section	Reviewer, Affiliation	Comment	Response
		be used to improve lower limb prostheses prescription for Medicare, its implications will likely extend throughout the U.S. health care system.	prescription for Medicare. The mentions of Medicare in the Discussion mostly pertain to KQ 1-3, for which we subdivided instruments based on Medicare generalizability. In the final sentence of the Conclusions about future research on instruments we removed the phrase "particularly for the Medicare population".
Discussion / Conclusion	TEP Reviewer #3	The report rightful concludes that more 'research is needed to adequately address most of the questions in this review.' This conclusion needs to be more clearly stated throughout the report. Further, the language and results included in the report should reflect the lack of evidence. Currently, the report appears to diminish the paucity of evidence for certain key questions (e.g., KQ7) in order to report findings that are supported by a questionable level of evidence.	We added a sentence about sparseness of evidence for KQ 5-7 to the opening paragraph of the discussion.
Discussion / Conclusion	Peer Reviewer #1	Yes. The implications and are limitations are clearly stated and adequately described. The future research section is also clear.	Thank you
Discussion / Conclusion	Peer Reviewer #2	Overall the discussion is appropriate and highlights the challenges with the available data to answer the key questions and provides suggestions about future research efforts that are needed. One particular challenge in the field of prosthetics research is that the major funding source for prosthetic research is the DOD – more funding and specific RFAs addressing older adults need to be made available in order to answer these questions.	Studies mostly did not report on funding source. The reviewers' comment is likely correct, but have not added this comment as we are not confident the degree to which it pertains to the studies included in the review.
Discussion / Conclusion	Peer Reviewer #2	Key question 4 is the one most difficult to address given the current literature. The discussion of this question in particular is very appropriate and highlights the limitations in the existing data. “Although not a limitation, per se, it should be noted that this review makes no attempt to make conclusions about the overall effects of different LLP components.” I think this sentence is very important to highlight, particularly as the conclusions drawn about key question 4 may lead some to	Thank you. We have added further text to clarify what the question addresses (and what it does not).

Research review section	Reviewer, Affiliation	Comment	Response
		argue that MPKs and other components should not be covered as there is no clear difference in outcomes between people who have these and those who don't. Clearly, this is not the appropriate conclusion from the data and further research is needed to better understand who should be prescribed MPKs and who should not. My sense from reading this report is that the authors fully understand the limitations of the literature and the challenges associated with answering these key questions, but I worry that these conclusions may be taken out of context and applied inappropriately. The discussion states that although there is low evidence to suggest that patient characteristics predict outcomes with specific components, it may be more accurate to state that the evidence is sparse and fails to adequately address the key question. Only 5 of the 11 studies used validated outcome measures, all were small and underpowered and none of the studies truly represent the population of interest as they were for the most part young and had traumatic amputations.	
Discussion / Conclusion	Peer Reviewer #2	In the discussion, it is also stated "A small evidence base does not support which components should be selected for which patient to maximize their ambulation, function, and quality of life or to minimize abandonment or limited use. However, this does not imply that there is evidence that no patient characteristics could effectively predict which patients would most benefit from one or another specific component." This also is very appropriate from the data at hand and recognizes the challenges in the existing literature	Thank you
Discussion / Conclusion	Peer Reviewer #3	As above the implications for this SR are critical and far-reaching for the limb loss community. Opportunity exists to clearly indicate upon which future research should focus. Accurately determining specific Outcome Measures upon which pending, imminent and future research should use is crucial for obtaining the evidence based data that is needed to answer the SR key questions. In order for clinicians to advocate for patient access and work with policy makers to	Thank you for your comment.

Research review section	Reviewer, Affiliation	Comment	Response
		effectively use resources, the field cannot afford wasting time and effort that is inevitable without more specific direction.	
Discussion / Conclusion	Peer Reviewer #4	Evidence and Analysis limitations - Nicely written, and comprehensive with the exceptions I have mentioned above.	Thank you
Discussion / Conclusion	Peer Reviewer #4	Future Research Recommendations – once again nicely written and a nice addition to assist potential researchers in formulating their research strategies to overcome the limitations in prior investigations.	Thank you
Discussion / Conclusion	Public Reviewer #2 Andreas Kannenberg, Otto Bock HealthCare LP	Quote: "Suboptimal matching of patients to LLPs may unnecessarily increase health care utilization, prevent attainment of maximal patient function, and defer realization of improved quality of life attainable with an appropriate prosthetic." Comment: Again, this statement only emphasizes the threat of over-utilization. I would recommend the more balanced statement that suboptimal matching of patients to LLPs may result in over- and under-utilization that both may have undesirable outcomes.	We have moved the phrase about increased utilization to the end of the sentence and changed utilization to expenditures. The majority of the sentence is already about underutilization.
Discussion / Conclusion	Public Reviewer #2 Andreas Kannenberg, Otto Bock HealthCare LP	In the analysis of the individual publications, I noticed that papers published in Prosthetics and Orthotics International, Archives of Physical Medicine and Rehabilitation, and Annals of Physical and Rehabilitation Medicine have been excluded from the systematic review for not being peer reviewed. The authors are advised that all three journals have a rather strict peer review process. Exclusion of papers published in these journals as "not peer-reviewed" is therefore incorrect and inadequate.	These would have been conference abstracts/posters that were published in these journals' supplements. No study was excluded based on publication journal. No journals were excluded.
Discussion / Conclusion	Public Reviewer #22 Claire Kilpatrick	In order to improve this review, I believe the scope must be expanded to include studies with biomechanical outcome measures. Much of the reviewer's critique, especially on Question 4, lies with a lack of studies establishing a heterogeneity of treatment effect. I believe this standard of research is inappropriate for lower limb prosthetic components due to the incredibly complex set of presenting characteristics which include and are not limited to etiology of amputation, residual limb characteristics, age, weight, functional level,	The scope of the review has not been expanded beyond what was included in the protocol. This review did not aim to address a broad scope.

Research review section	Reviewer, Affiliation	Comment	Response
		comorbidities, activities of daily living, date of amputation, access to care, and even patient hygiene. In order to appropriately establish the heterogeneity of treatment effect for a given prosthetic component, a study would have to find multiple amputees which have all of the same presenting characteristics. This is an impossible standard."	
Discussion / Conclusion	Public Reviewer #22 Claire Kilpatrick	"The reviewers indicate that they are seeking literature which will predict which lower limb amputees would most benefit from a given component. In this document, what the reviewers have done is to limit the pool of available research by excluding biomechanical data, then filter through the already reduced pool to studies which offer patient characteristics so that a heterogeneity of treatment effect can be established, and label the study as not-validated due to a lack of established heterogeneity of treatment effect, outcome measure selection or a statistical significance. This has the potential to be misleading for CMS and insurance providers. The scope of this review is too narrow; in a worst-case situation, I believe the information presented in Question 4 could be used to argue that as there is insufficient literature (as deemed by these reviewers) to indicate that a specific amputee will benefit from a specific component, then the component will labeled as not medically necessary. I do not believe this was the intent of the reviewers. However, the potential impact of this review must be thoroughly vetted as it will directly influence the availability of advanced components for all lower limb amputees in the United States.	The review should only be interpreted in regards to the Key Questions asked. The review is narrow in focus, which is explicated repeatedly. We have added a list of topics that the review addresses (and does not address) to the bullet list of the Evidence Summary in the Discussion.
Discussion / Conclusion	Public Reviewer #4 American Orthotic Prosthetic Association	As discussed earlier in AOPA's comments, the final report of the RAND Corporation regarding the Economic Value of Advanced Transfemoral Prosthetics is included as an addendum to this document. AOPA strongly believes that the AHRQ systematic review cannot be considered valid and complete without inclusion and consideration of the RAND Corporation study that focuses on trans-femoral prostheses, the Highsmith study that focuses on trans-tibial prostheses,	The two reviews address different topics.

Research review section	Reviewer, Affiliation	Comment	Response
		and the updated Dobson DaVanzo research that reviews the overall cost effectiveness of lower limb prosthetic intervention.	
Discussion / Conclusion	Public Reviewer #4 American Orthotic Prosthetic Association	In addition, AOPA strongly encourages the AHRQ to consider the additional studies highlighted in AOPA's comments for consideration in AHRQs systematic review.	We have conducted full text screening of all citations recommended by this reviewer and others, and have added any eligible studies found.
Discussion / Conclusion	Public Reviewer #4 American Orthotic Prosthetic Association	AOPA appreciates the opportunity to provide comments regarding the draft report on the AHRQ systematic review of the literature regarding lower limb prostheses. We sincerely hope that our comments are helpful in ensuring that the final review conducted by the AHRQ and its contractors is thorough, complete, and comprehensive.	Thank you for your comments.
Discussion / Conclusion	Public Reviewer #5 Orthotic and Prosthetic Alliance	The decision to integrate limitations of the evidence, which are raised throughout the review, with limitations of the review methodology and analysis in the "Evidence and Analysis Limitations" section (p. 122-124) is highly questionable. This approach would seem to obfuscate decisions made by the authors with limitations of the available evidence. Further, the choice to label this section in this manner is inconsistent with other contemporary AHRQ reviews, which more appropriately discuss limitations of the review,42 or differentiate limitations associated with the review process and limitations associated with the evidence.43, 44 We encourage the authors to differentiate the limitations so readers may objectively assess the merits of the evidence and review. Below, we highlight several examples of process-related limitations that should be distinguished from limitations of the evidence.	We are happy to split the section into Evidence Limitations and Analysis Limitations
Discussion / Conclusion	Public Reviewer #5 Orthotic and Prosthetic Alliance	A key limitation to the review process is that evidence was limited to studies considered by the authors to be applicable to the Medicare population. While we understand the motivation for the decision (i.e., that this study was sponsored by the Centers for Medicare and Medicaid Services), this decision results in the exclusion from consideration of a large body of evidence. Numerous studies of people with primarily traumatic amputation and with Veterans who are Medicare-aged appear	It is not the case that the evidence was limited to studies generalizable to the Medicare population. This designation was used only for categorization. We discuss the limitation of the arbitrary decisions made for this categorization. The methodology used to make the categorization is now more clearly reported.

Research review section	Reviewer, Affiliation	Comment	Response
		to have been excluded. The determination of applicability seems to have been made subjectively rather than using explicit criteria, which is also concerning. We request the authors explicitly state in the final report this methodological decision as a limitation of the review.	
Discussion / Conclusion	Public Reviewer #5 Orthotic and Prosthetic Alliance	Other limitations include the methods used by the authors to assess the validity and reliability of outcome measures. Although the authors indicate that "assessment of reliability, validity, and other measure properties is open to interpretation," there is general consensus on the level of evidence needed for instruments to be recommended for specific applications.10-12 The liberal approach used by the authors to assess the available instruments may have resulted in the consideration and/or recommendation of suboptimal measures for future studies. We feel it appropriate for the authors to acknowledge that the methods used to evaluate instruments in this review are not consistent with contemporary standards.	We disagree that we have used an approach not consistent with contemporary standards. There is no universally agreed upon method to summarize psychometric properties. We have used an approach based on a body of prior work. We have added in the citations to these studies that were inadvertently omitted in the draft report.
Discussion / Conclusion	Public Reviewer #5 Orthotic and Prosthetic Alliance	As noted above, the authors' decision to develop ad hoc definitions for "assessment techniques," "prediction tools," and "outcome measures," rather than adhering to conventional definitions of "evaluation," "prediction," and "discrimination"21 may be considered a limitation of the review. We disagree strongly with the authors' suggestion that "most, if not all, measures can be used for any of these contexts." (See; p. 123.) An instrument developed for discriminative purposes cannot be used for predictive purposes (and vice versa) without justification and supporting evidence. That the authors believe this to be true is concerning and suggests a fundamental misunderstanding of instrument development and application.	We have maintained the original terminology. We removed the sentence about using the measures for any context. We were considering how they would be used by clinicians and researchers, but we agree that this is an overstatement.
Discussion / Conclusion	Public Reviewer #5 Orthotic and Prosthetic Alliance	Other methodological issues, such as the review of each article by a single author, should also be noted as a limitation. (Although the review does not explicitly state articles were	It is not the case that each article was reviewed by a single author.

Research review section	Reviewer, Affiliation	Comment	Response
		reviewed by a single author, we assume as much based on the number of errors noted in the draft.)	
Discussion / Conclusion	Public Reviewer #5 Orthotic and Prosthetic Alliance	The authors indicated that "extremely few studies reported the type of medical insurance study participants had (although, many of the studies were conducted in Europe and other countries other than U.S.)." (See; p. 123.) We encourage the authors to add a recommendation that researcherscollect this type of information in the future so that the generalizability of study results to Medicare-eligible individuals can be assessed directly, rather than assessed subjectively as was done for this review.	This is a good suggestion. We have added this to the general future research recommendations.
Discussion / Conclusion	Public Reviewer #5 Orthotic and Prosthetic Alliance	We recommend the authors generalize the term "function" by replacing it with the phrase "health, function, and quality of life" in the following statement: "From the amputee's and the clinician's perspective, among the most important questions is which prosthesis (comprised of which prosthetic components) would best enable maximal function for a given individual." (See; p. 120, first paragraph.) Prosthetists, like other health care providers, are interested in maximizing patient outcomes including, but not limited to, function. Further, many of the outcomes included in this review assess constructs other than function (which may be perceived by readers to mean "physical function").	We agree. We did not mean for this sentence to be limiting. We have added health and QoL.
Discussion / Conclusion	Public Reviewer #5 Orthotic and Prosthetic Alliance	While we appreciate the authors' suggestion that "those [instruments] that have been validated should be used to form a core set of measures for use in future research studies," (p. 126) and we agree that greater standardization of outcomes assessment is needed in prosthetics research, we submit that more work is needed to assess the reliability, validity, and applicability of the studied instruments before they should be advocated as part of a core set.	We agree. We removed the sentence.
Discussion / Conclusion	Public Reviewer #5 Orthotic and Prosthetic Alliance	The preliminary review of instruments conducted for this report may, at best, serve as a starting point for a review of measures that can be used in research and/or clinical practice. Given the numerous issues and concerns we have	We hope and expect that the revision has addressed the reviewer's concerns. We expect that this review will form the basis to continue the discussion about choice of

Research review section	Reviewer, Affiliation	Comment	Response
		identified with the draft report, we believe this review is not sufficient to serve as the basis for recommending instruments for a core set. We encourage the authors to revise or more thoroughly justify their recommendation, so such a statement is not perceived as an endorsement of the measures included in this review.	measures to be used in research and/or clinical practice.
Discussion / Conclusion	Public Reviewer #5 Orthotic and Prosthetic Alliance	We also question the authors' conclusion that "the majority of the evidence addresses the question of which components maximize ambulation and function in the average patient, as opposed to which component would best suit the needs of a given individual." (See; p. 126.) We submit that studies are (generally) designed to assess a range of individuals, not a specific one. It is unlikely that one would find, in this body of literature or any other, evidence to indicate which intervention is suited to a particular individual. Rather, it is up to the reader (i.e., a clinician) to assess the studied population relative to a specific individual and make the determination as to whether the evidence from the study would be generalizable and applicable to the individual patient. We encourage the authors to revise or clarify this statement so that their conclusion is more objective.	We have added the sentence "In other words, few studies address the issue of heterogeneity of treatment effect." The Key Question pertained to this concept, in contrast to an overall assessment of comparative effectiveness.
Discussion / Conclusion	Public Reviewer #5 Orthotic and Prosthetic Alliance	We also recommend the authors avoid the double negative in the statement "however, this does not imply that there is evidence that no patient characteristics could effectively predict which patients would most benefit from one or another specific component." (See; p. 126.) Instead, we suggest the authors consider the following statement: "[h]owever, this does not imply that the evidence suggests patient characteristics cannot predict which patients would benefit from one specific component or another."	Thank you. We agree with this modification.
Discussion / Conclusion	Public Reviewer #5 Orthotic and Prosthetic Alliance	The authors seem unusually critical of the study by Hahn and colleagues.49 We recommend the authors reconsider and revise characterizations such as "the study was methodologically and analytically flawed." (See; p. 102.) While we agree the study is limited, we encourage the authors to	We have toned down our description of this study.

Research review section	Reviewer, Affiliation	Comment	Response
		use more objective language to avoid the perception of bias toward this study. Conversely, if the authors elect to characterize the study in this manner, they should verify that they have applied equitable scrutiny and descriptions when discussing other studies included in this review.	
Discussion / Conclusion	Public Reviewer #5 Orthotic and Prosthetic Alliance	As members of the Orthotic and Prosthetic Alliance, we invite the Agency for Healthcare Research and Quality and the Centers for Medicare and Medicaid Services to partner with the prosthetics community in examining and revising any public policies based on the results of this review.	Thank you.
Discussion / Conclusion	Public Reviewer #5 Orthotic and Prosthetic Alliance	We request that before this review is finalized, stakeholders have the opportunity to review, provide contextual history, propose areas of scientific uncertainty or debate, and provide consumers' perspective on the topic of LLPs.	We convened a Key Informant panel that did include consumer perspectives. We welcome publication or dissemination of consumers' and other stakeholders' perspectives regarding this report.
Discussion / Conclusion	Public Reviewer #5 Orthotic and Prosthetic Alliance	We request that AHRQ create a forum to discuss and develop consensus among stakeholders on the terminology presented in the systematic review. The review authors have proposed unique terminology that has not been often used in the body of scientific literature. We are concerned that use of these terms may cause confusion among public readers or have other unforeseen consequences.	We have maintained the original terminology from the draft review. This was the terminology used by the review sponsors and in the protocol. In addition, our Key Informant and Technical Expert panel members agreed with the terminology.
Discussion / Conclusion	Public Reviewer #5 Orthotic and Prosthetic Alliance	We request AHRQ form an Advisory Panel of researchers, methodologists, clinicians and stakeholders to examine the draft systematic review, comments submitted during the open comment period, and final report to ensure important issues have been adequately addressed.	Thank you for your comment. We will strive to adequately address all reviewer and public comments.
Discussion / Conclusion	Public Reviewer #5 Orthotic and Prosthetic Alliance	We request that CMS partner with the prosthetics community and other relevant federal agencies to develop a strategic plan, based on results of this review, consensus of stakeholders, and best clinical practices, to acquire the data necessary to answer the Key Questions posed in this review. We are confident that implementation of such a plan would both reduce healthcare costs and optimize health, function,	Thank you for this suggestion.

Research review section	Reviewer, Affiliation	Comment	Response
		and quality of life for the many Americans with limb loss who require prosthetic services.	
Discussion / Conclusion	Public Reviewer #5 Orthotic and Prosthetic Alliance	Finally, we request that CMS collaborate with the prosthetics community to acquire and analyze data collected as a result of the strategic plan, and develop revised policies based on sound outcomes data and best clinical practices.	Thank you for this suggestion.
Discussion / Conclusion	Public Reviewer #5 Orthotic and Prosthetic Alliance	The importance of CMS, AHRQ, and other relevant federal agencies collaborating in the future cannot be overstated. As new evidence becomes available, coverage policies should be revised to meet the needs of users of prostheses. For instance, during the development of this draft systematic review, the RAND Corporation issued a seminal report, published on September 3, 2017, that is not referenced in the AHRQ literature review. RAND's analysis underscores substantially increased risks of falls and osteoarthritis in the contralateral limb for patients with non-micro-processor knee (non-MPK) technology, as well as economic costs—demonstrating that MPKs are safer for patients.	We agree that the RAND report and other recent reviews should be considered together with this review.
Discussion / Conclusion	Public Reviewer #5 Orthotic and Prosthetic Alliance	These RAND findings, together with the conclusions from a Dobson-DaVanzo September 2017 study report (Medicare data on years 2011-14, now pending journal publication) showed total 15-month health costs for K2 patients at over $16,000 higher than for K3 patients (roughly 5 times higher than the spread in parallel 2007-10 Medicare data). In addition, a Mayo Clinic paper by Kaufman on amputee fall frequency and costs collectively highlights the deficiencies of the current K-levels for accurately guiding prosthetic component prescription.	Thank you. We did not review costs or the K levels, per se.
Discussion / Conclusion	Public Reviewer #5 Orthotic and Prosthetic Alliance	While an early iteration, we submit for AHRQ consideration in these comments the following preliminary outline of an improved, alternative approach to the existing K-levels, subject to further discussion, that would assure greater safety and potentially improved value in amputee prosthetic treatment. The following serves as a proposal to revise the existing functional K-Level system to response to new	We have thoroughly revised the report.

Research review section	Reviewer, Affiliation	Comment	Response
		developments in the prosthetic care evidence base, referenced immediately above.	
Discussion / Conclusion	Public Reviewer #5 Orthotic and Prosthetic Alliance	See PDF for porposed alterations to existing K levels	Thank you
Tables	Public Reviewer #22 Claire Kilpatrick	"Table 4.5 In this table, in the column ""Findings"" the reviewers highlight ""younger age weakly correlated with favoring ES"" and ""lighter body weight weakly correlated with favoring ES."" Highlighting these two correlations is misleading; this article did not indicated that old age and increased body weight were not correlated with ES. Reporting these two pieces of information implies that there is less of a benefit to providing energy storage feet to the elderly and higher body weight population which is simply not the case nor supported by Alaranta's work."	Although there are caveats to the simpler findings the study reports "The beneficial trend for the ES prosthesis was weakly correlated with the age at the phase of the interview ($r = -0.30$, $p < 0.01$). The younger amputees gained more benefit than the older ones." and "more benefit from the ES prosthesis compared to heavier ones. The benefit of the ES prosthesis was inversely correlated with the body-weight ($r = 0.29$, $p < 0.01$)." Regardless, our overall conclusion is that the study does not support a difference in relative outcomes based on age or weight (or other factors).
References	Public Reviewer #24 Kellie C	"[1]. Cheung C, Wall JC, Zelin S. A microcomputer-based system for measuring temporal asymmetry in amputee gait. Prosthet. Orthot. Int. 1983; 7(1):131–140. [PubMed: 6647009] [2]. Dingwell JB, Davis BL, Frazder DM. Use of an instrumented treadmill for real-time gait symmetry evaluation and feedback in normal and trans-tibial amputee subjects. Prosthet. Orthot. Int. 1996; 20(2):101–110. [PubMed: 8876003] [3]. Isakov E, Keren O, Benjuya N. Trans–tibial amputee gait: Time–distance parameters and EMG activity. Prosthet. Orthot. Int. 2000; 24(3):216–220. [PubMed: 11195356] [4]. Nolan L, Lees A. The functional demands on the intact limb during walking for active trans- femoral and trans-tibial amputees. Prosthet. Orthot. Int. 2000; 24(2):117–125. [PubMed: 11061198] [5]. Nolan L, Wit A, Dudziñski K, Lees A, Lake M, Wychowañski M. Adjustments in gait symmetry with walking speed in transfemoral and trans-tibial	Thank you. We have conducted full text screening of all citations recommended by this reviewer and others, and have added any eligible studies found.

Research review section	Reviewer, Affiliation	Comment	Response
		amputees. Gait Posture. 2003; 17(2):142–151. [PubMed: 12633775] [6]. Seliktar R, Mizrahi J. Some gait characteristics of below-knee amputees and their reflection on the ground reaction forces. Eng. Med. 1986; 15(1):27–34. [PubMed: 3699235] [7]. Gard SA. Use of quantitative gait analysis for the evaluation of prosthetic walking performance. JPO J. Prosthet. Orthot. 2006; 18(6):P93–P104. [8]. Powers CM, Torburn L, Perry J, Ayyappa E. Influence of prosthetic foot design on sound limb loading in adults with unilateral below-knee amputations. Arch. Phys. Med. Rehabil. Jul; 1994 75(7):825–829. [PubMed: 8024435] [9]. Snyder RD, Powers CM, Fountain C, Perry J. The effect of five prosthetic feet on the gait and loading of the sound limb in dysvascular below-knee amputees. J. Rehabil. Res. Dev. 1995; 32:309–315. [PubMed: 8770795] [10]. Winter DA, Sienko SE. Biomechanics of below-knee amputee gait. J. Biomech. 1988; 21(5): 361–367. [PubMed: 3417688]"	
Appendixes	Public Reviewer #20 Anonymous	"• There are a number of studies (relating to KQ4) that appear to have not been considered (see attached spreadsheet) despite comparing prosthetic components. (N.B. They may be excluded on other criteria but they are not listed in Appendix B, which names all the excluded studies and the reasoning).	Thank you. We have conducted full text screening of all citations recommended by this reviewer and others, and have added any eligible studies found.
Appendixes	Public Reviewer #20 Anonymous	• The table in Appendix B is poorly constructed. Journal names and article titles are often the wrong way around.	We have updated and reformatted the appendix.
Appendixes	Public Reviewer #20 Anonymous	• In Appendix B, some studies are cited as being from 'Prosthetics & Orthotics International', but seem to refer to ISPO conference presentations.	These would have been conference abstracts/posters that were published in these journals' supplements. No study was excluded based on publication journal. No journals were excluded.
Appendixes	Public Reviewer #20 Anonymous	• Do PhD theses not count as being peer-reviewed? (e.g. MJ Highsmith's)"	We did include PhD theses that were available to us. We excluded Dr. Highsmith's thesis as it did not provide the heterogeneity of treatment effect analyses required.
General	TEP Reviewer #1	The title of the systematic review is misleading and does not reflect what the systematic review is actually presenting and	We have changed the title to Lower Limb Prostheses: Measurement Instruments,

Research review section	Reviewer, Affiliation	Comment	Response
		assessing. It does not reflect the review of the psychometric properties of existing outcome measures used clinically for the lower limb amputee population. It does not reflect the population which this reviews is intended for, which is the Medicare eligible population, impact of prosthetic components on functional and the long term use of prosthetic lower limbs. Therefore, there is a need to re-do the title of the review	Comparison of Component Effects by Subgroups, and Long-Term Outcomes
General	TEP Reviewer #1	It is felt that the target population needs to be defined in a clear and concise fashion. The authors state that published manuscripts only examining the Medicare eligible population would be included in this review. As you know, the Medicare eligible population not only includes those individuals between the ages of 65 – 74 (> 50%) and greater than 85 years of age (> 10%), but also those who receive disability benefits who are younger than 45 years and between the ages of 45-55, which makes up the remaining percentage. In addition, due to the rise in obesity and cardiovascular disease, clinically, individuals are being diagnosed with Type II Diabetes Mellitus (DM) at a younger age and there is a growing population of traumatic lower limb amputees that are 10-15 years younger than the standard age of 65. Therefore, it is critical that the ages of those study samples are included in the review which would allow for the inclusion of research that may have been excluded based on the age of the study population.	Nowhere does it state implicitly or explicitly that only manuscripts examining the Medicare eligible population are included. Nothing in the eligibility criteria or Key Questions says anything about Medicare. For KQ 1-3 and to some extent KQ 4-7 we do categorize studies based on their generalizability to Medicare recipients, but we do not exclude any studies based on this criterion. The ages of all study samples are described.
General	TEP Reviewer #1	Key questions 1-3 need to be further defined or need to be compressed into 1 question. Those measures may have been developed for one purpose which would be to assess current prosthetic mobility but have also been examined to establish its predictive capabilities. It is felt that these questions could be compressed into one question.	We have re-organized this section to first describe all eligible instruments and only then address the specific KQs. We believe this addresses the concern, which was one we shared also.
General	TEP Reviewer #1	The authors have done a great job structuring and organizing this review. But they need to address the issues in the above sections before clearly addressing the policy and practice decisions.	Thank you

Research review section	Reviewer, Affiliation	Comment	Response
General	TEP Reviewer #2	Feel that questions 1-3 combined, do addressing a primary purpose , the assessment of assessments / the validity of the measures used in patients with lower limb amputations and are clearly stated	We have re-organized this section to first describe all eligible instruments and only then address the specific KQs. We believe this addresses the concern, which was one we shared also.
General	TEP Reviewer #2	Followed by question 7, are observed pt outcomes at 6 months, 12 months and 5 years appropriate and clearly stated, yes	Thank you
General	TEP Reviewer #2	question 4, would be better stated “Are there definable pt characteristics that can predict pt outcome / “relative effectiveness” of different LLP componentry and /or predict pt’s short and long term use of LLP’s	We like the language proposed by the reviewer, but it does not lend itself well to a "PICO" research question. We have simplified the question to how do ambulatory, functional, and patient-centered outcomes with different prosthetic components vary based on study participant characteristics?
General	TEP Reviewer #2	Questions 5,6 :expectation management and pt satisfaction are appropriate concerns, but don’ carry the weight / not really studied at the level of the other key questions in this review	True. Thank you.
General	TEP Reviewer #2	yes, there is practical importance in assessing the assessments, validity of measures used in pts with LLP and No, unable to predict the effectiveness of various component/ try how long pts will use them effectively	True. Thank you.
General	TEP Reviewer #2	“The key to improving outcomes for those who have lost limbs is to ensure that they receive appropriate and comprehensive interdisciplinary care to address both their physical and psychosocial needs. Fundamental to the rehabilitative care and recovery of many people who have lost limbs is their fitting for and training on the use of prostheses. 1. Pasquina, PF, Carvalho, AJ,Sheehan, TP. Ethics in Rehabilitation: Access to Prosthetics and Quality Care Following Amputation. AMA Journal of Ethics. June 2015, Volume 17, Number 6: 535-546.	Thank you for this comment. We have not, though, included this policy-related opinion, although it is sensible.
General	TEP Reviewer #2	My bias, that in the absence of clear evidence, my recommendation is that patients involve a well trained physiatrist working in an interdisciplinary limb restoration	Thank you for your opinion.

Research review section	Reviewer, Affiliation	Comment	Response
		program with long term follow-up and direct, clinical management responsibilities.	
General	TEP Reviewer #2	The lightly evidenced conclusions with regard to limited use of LLP may have the potential to create new barriers to access, even in highly qualified and motivated patients.	We believe that the proper interpretation is more that the evidence is sparse and of low strength of evidence.
General	TEP Reviewer #2	Using the results with regard to assessment should galvanize efforts to improve functional outcome research in this challenging populations, if funding can be secured and maintained.	We agree
General	TEP Reviewer #3	The report addresses a very important topic and will likely have impacts beyond the stated goal of the report to 'assist CMS to better understand the state of evidence regarding how best to match patients with LLPs that would yield the best outcomes for them'. As is evident from payers regarding CMS' draft LCD on Lower Limb Prostheses in 2015, private payers frequently establish policies and/or regulations that mirror those that apply to the Medicare population. For example, after the draft LCD for Lower Limb Prostheses was issued, private insurers started incorporating language similar to this draft proposal in their letters notifying beneficiaries of the decision to deny them a prostheses. As a result, this report, its contents, and especially, its conclusions should be very carefully considered.	Thank you
General	TEP Reviewer #3	While the target population is explicitly defined (e.g., persons over the age of 65 who lost a limb from dysvascular disease), the findings from this report will likely impact patients outside this population. As described above, private payers frequently enact policies that reflect Medicare/Medicaid policies. Additionally, many individuals with limb loss who receive Medicare benefits may be less than 65 years old and lost a limb due some other etiology. Individuals who lose a limb become eligible for Medicare two years after their amputation. For many, this represents their best option to receive consistent and reliable health insurance after their amputation.	The review covers essentially all adult lower limb amputees in high-resource countries (except those with battle injuries). While we categorize some Key Questions based on generalizability to Medicare, the review summarizes all evidence, regardless of population.

Research review section	Reviewer, Affiliation	Comment	Response
General	TEP Reviewer #3	As has been stated by Key Informants through-out the process of conducting this systematic review, some of the key questions guiding this project are somewhat problematic. For example, KQ 1-3 largely focus on tools used to assess a patient's k-level, which is then used to inform the prescription of prosthetic componentry. Yet, the K-level system itself is never explicitly called into question or placed under scrutiny. If the overarching goal of the review is to better understand how to match patients with a lower limb prostheses, it seems reasonable to examine the mechanism that most explicitly guides this process - the Medicare Functional Classification Level system.	We have added to the discussion paragraph about test validity that no study has evaluated the psychometric properties of the K level system.
General	TEP Reviewer #3	KQ7 also contains some problematic elements, especially regarding the topic of prostheses abandonment. Despite being a sub-topic under KQ7, the concept of prostheses abandonment is not clearly defined, nor is it clear that all of the studies cited in KQ7 as evidence of prostheses abandonment define and measure the concept consistently.	We do not believe the definitions are, in general, not clearly defined. We have added more details into the 7 KQ summary table about the definitions.
General	TEP Reviewer #3	This report does not contribute new information or understanding to the topic at hand. Anyone familiar with the evidence surrounding the topic of lower limb prosthesis prescription would reach similar conclusions that there is largely a lack of evidence to support creating evidence-based practice guidelines. It is reassuring to hear the recommendation for future funding for research on this topic - especially the need for long-term follow-up studies to understand problems and limitations people have with their prosthesis, rates of abandonment or limited use, and reasons for limited use or abandonment. Perhaps this may lead to creating targeted funding opportunities to create a research infrastructure to support this type of research. One possibility is the creation of a Limb Loss Model System, as the current Model Systems programs have done a very good job of supporting longitudinal research on patients that has an impact on clinical care.	Thank you. We agree that more research is needed for this important healthcare topic.

Research review section	Reviewer, Affiliation	Comment	Response
General	Peer Reviewer #1	The authors should be commended for performing a thorough review of the prosthetics literature. This report is very well organized and is easy to read. The target population, the search methodology and a summary of the articles have been explicitly defined and clearly presented	Thank you
General	Peer Reviewer #1	All key questions are appropriate and explicitly stated. However, the wording for Key Question 4 is confusing and vague. I would recommend rewording KQ4. The meaning of "relative effects" and "levels of components/prostheses" is not clear	We agree and have simplified the question.
General	Peer Reviewer #1	Secondly, KQ 4a.ii and 4b.ii are related to association between prescription assessment techniques and validated outcomes. The discussion or results do not discuss anything about the proscription techniques. The results / discussion should be modified, or these KQs should be removed	We have added to the Discussion limitations that none of the studies for KQ 4 evaluated heterogeneity of treatment effect based on any instruments. All evaluated only patient characteristics.
General	Peer Reviewer #1	This report is very well structured and organized. The study findings have been clearly presented and significantly contributes to our understanding of the lower limb the prosthesis literature. However, the relevance of this review for policy or practice decisions is questionable for two reasons.	Thank you
General	Peer Reviewer #1	1. This review excluded all the biomechanical studies. As the vast majority of lower limb prosthesis literature consists of biomechanical evaluation of prosthetic devices, results of this review are based only on a small subset of amputee research, and thereofer cannot be used for policy/practice decisions.	The review was explicitly designed to evaluate patient-centered outcomes, function, and the like. These were the questions posed to us by policymakers. The review was not designed to evaluate the biomechanical properties of LLP components. This is not to diminish the importance of these outcomes, but it was not the purpose of the review. This is noted in the abstract, introduction (objectives), relevant results sections, and discussion.
General	Peer Reviewer #1	2. The review did not categorize amputees based on their functional level. The functional level of an amputee is a critical criterion recommended by Medicare which indicates the likelihood of success with a prosthesis. The data on "indoor prosthesis use" and "abandonment of prosthesis" has	The little data about K levels that were reported by studies has been added. We also added a limitation about this issue. Because of study limitations, for our overall summary,

Research review section	Reviewer, Affiliation	Comment	Response
		significant limitations as it does not address the functional level of study subjects	we focus on the two more generalizable studies.
General	Peer Reviewer #2	This AHRQ systematic review attempts to answer key questions related to the metrics used to evaluate performance and function in people with lower extremity amputations and prosthetic users and to identify how different prosthetic components affect outcomes. The overarching goal of the review is to determine if we can match patients to appropriate prosthetic components using existing outcome tools/metrics and whether use of specific prosthetic components are associated with outcome. The focus of this review is for CMS patients and examined published literature through November, 2016. These findings can be used to determine coverage for specific prosthetic devices. This is an extremely important topic and clinically very relevant.	Thank you
General	Peer Reviewer #2	There are however several important challenges to this endeavor and these are acknowledged in the discussion section to some extent. First, if there is insufficient evidence in key questions 1-3 related to the availability of prediction tools and outcome measures for people who use a LLP, it is impossible to answer key question 4 accurately. The second is that if the studies that attempt to answer key question 4 do not use appropriate prediction tools and outcome measures identified in key questions 1-3 (if there are such tools), it is impossible to determine if the results are due to the use of inappropriate/inadequate prediction and outcome tools or because there is no relationship between prosthetic components and outcome. The third issue is that because the populations are so diverse with many different factors that influence a person's success with a given prosthetic device and so many different ways to measure outcome (diverse patients, predictive factors, prosthetic components (sockets, knees, feet), timing of fitting with a specific component, methods for fitting and training and outcome measures), it makes it next to impossible to synthesize these small studies	Thank you. We believe we have covered the described points.

Research review section	Reviewer, Affiliation	Comment	Response
		examining specific prosthetic components to come to any meaningful conclusions. The clinically most relevant questions are related to comparison of different categories of devices (MPK vs non MPK as the most important and then comparison of different categories of feet) – these are important because of the cost differential and the need for payers to be able to rationally develop coverage policies for the different categories of components.	
General	Peer Reviewer #2	In an effort to be thorough, this report examines studies with many different types of comparisons and includes socket designs as well – this broad focus obscures the clinically relevant questions. What is needed is the development of a clinical prediction model that can take into consideration this myriad of different factors to help match individual patients to specific categories of prosthetic components. But this doesn't currently exist and can't be developed with the existing literature	We agree.
General	Peer Reviewer #2	As a clinician, I frequently see firsthand the improvement in outcomes and experience of patients who switch from non-MPK to MPKs – the influence of MPK on falls for example is striking and is one of the primary reasons that I prescribe MPKs particularly for older adults, even at (and particularly at) lower activity levels. A recently published study by Mundell et al (June 2017) highlights the cost of falls in transfemoral amputees. They report that the mean 6-month direct cost of falls requiring hospitalization is $25,652 and of falls requiring ED visits is $18,091. These are extraordinary costs and represent significant clinical outcomes – future prosthetic studies need to better assess how the cost of MPKs offset the cost of fall treatments and impact. This is something that would be very difficult to do with a small trial and could probably only be accomplished with a large, well-designed cohort or registry	Thank you for your comment. We did not directly address cost, which is certainly an important issue.
General	Peer Reviewer #2	Overall this is a well structured and thorough report	Thank you

Research review section	Reviewer, Affiliation	Comment	Response
General	Peer Reviewer #2	It seems that to avoid concern about misapplying the results of this review, the italicized statement above "the evidence is sparse and fails to adequately address the key question" should be one of the key conclusions rather than the conclusion as stated in the abstract. "Currently, there is not evidence to support the selection of specific components for patient subgroups to maximize ambulation, function, and quality of life or to minimize abandonment or limited use	We have amended the sentence to say the evidence is too sparse.
General	Peer Reviewer #2	This report helps to identify important gaps in the research evidence, but is not particularly helpful in terms of policy or practice decisions. I don't see how this report can help an individual physician or prosthetist determine which prosthetic device is most appropriate and I do not think coverage decisions should be made about specific devices based on this report. I do think it highlights the need for collection of standardized outcome measures clinically and in the research setting - we should develop a national registry to allow standardized collection of data on these patients.	Thank you for this comment. We are hopeful that it will be of value to at least policy makers, at a minimum to guide future research.
General	Peer Reviewer #3	The importance of this systematic review on lower limb prosthetics impacting the field of limb loss cannot be overstated. Not only will this report be clinically meaningful, this document will have far reaching implications in Amputation Medicine for many years to come. Care must be given in regards to how this will affect not only amputees covered by Medicare, but all demographics of people with limb loss. The significance of such a document and how it will undoubtedly it will be used by all stakeholders in the process is of paramount distinction. Any information obtained and surmised from this document will be cited by all insurance companies as the gold standard for what is covered from a prosthetic standpoint. The potential for misuse of a document of this stature to attempt to apply the conclusions across all ages and demographics regardless of the intended application is real and historic. Clinicians caring for amputees look to institutions like CMS to provide guidelines for appropriateness	The review is not specific to Medicare, but it is the important factor regarding generalizability. We have changed the title to Lower Limb Prostheses: Measurement Instruments, Comparison of Component Effects by Subgroups, and Long-Term Outcomes. We maintained the term prostheses, as this seems to be more commonly used, including by CMS.

Research review section	Reviewer, Affiliation	Comment	Response
		of prosthetic candidacy to ensure not only effective utilization of resources but patient access. This SR will serve as a guideline for all amputees regardless of the age and demographics of the patient population and therefore must be more inclusive rather than exclusive if it is to provide useful information. The title, Lower Limb Prosthesis (should be "Prosthetics" or "Prostheses") implies inclusiveness rather than specifically stating that the SR is only intended to review Medicare patients. This is nebulous and must be more reflective of what was reviewed or ideally must be more inclusive to reflect all people with limb loss. By being more inclusive, a more thorough analysis of the data according to demographic subsets.	
General	Peer Reviewer #3	While this draft report highlights the deficiencies in evidence for the stated target patient population for Medicare amputees, it fails to really outline how that population is defined specifically. There needs to be more explicit detail in how this target population was chosen and the justification for why (i.e. what percentage of amputees that are prosthetic candidates are from vascular causes versus traumatic amputees that are much younger than 65 but qualify for Medicare after disability). Of particular concern is that a number of trauma related amputations of younger age and fewer comorbidities are more likely to benefit and qualify for a prosthesis when compared with the typical Medicare vascular amputee greater than 65 years old; both of these populations are covered by Medicare.	There is no clear definition of a population that is Medicare eligible, at least in terms of comparing study participants to those covered by Medicare. We came up with a somewhat arbitrary threshold in discussion with representatives from CMS.
General	Peer Reviewer #3	Limiting the systematic review to include only "Medicare" patients and excluding military is fundamentally flawed reasoning in this reviewer's opinion. A valid argument for including military based articles upfront but then categorizing according to demographics such as age and etiology would provide much better information for all people with limb loss, not ones arbitrarily excluded, particularly when these patients may actually be more appropriately matched to prosthetic	The review is not restricted to include only Medicare patients. This was not part of study eligibility criteria. Instead, particularly for KQ 1-3, we highlight instruments tested in studies generalizable to Medicare patients. Military based articles were also included, except for those including only people with battle injuries.

Research review section	Reviewer, Affiliation	Comment	Response
		candidates who have Medicare insurance. This is far too important document with far and long reaching implications for it not to be inclusive.	
General	Peer Reviewer #3	An additional critique is that military funded research often has better strength of evidence particular in recent years with Department of Defense funding and frequently involves the technology that will help identify resource management of component, one of the key areas needing to be answered by this work. Military amputees are probably more closely matched in age and comorbidities with traumatic amputees who qualify for Medicare via disability and are appropriate for prosthetic candidacy. Excluding studies done on the military population limits important data particularly given the paucity of available studies in people with limb loss of all ages and etiologies.	The review excludes only studies of people with battle injuries. Several included studies were of veteran populations.
General	Peer Reviewer #3	In part the key questions are appropriate and explicitly stated however samples size cutoff for KQ 1-3 were 20, whereas KQ7 was 100. While the explanation was provided that 100 would provide more precise and stronger evidence that should be addressed when rating the strength of the publication in the analysis rather than exclude the article upfront.	We agree that sample size cutoffs are abritrary and we understand that not all readers will agree with each threshold used. The decision was based on a balance between usefulness of study findings and available resources.
General	Peer Reviewer #3	Of serious concern is the section on Outcome Measures, which this SR tabulates as reliable and valid. Well-respected and well-funded clinical research projects currently are using tools, including surveys like PLUS-M and the PROMIS database and functional measures such as AMP-Pro, TUG, 6 min Walk and 10 M Walk. These are tools that can be stratified to answer the key questions across the subset groups. But according to this SR some of these tools do not meet the basic requirements for validity and reliability and yet these tools are agreed upon by experts in the field to be much more useful than the antiquated tools that according to this SR analysis are superior. If the SR identifies accepted tools as problematic based on the literature reviewed when these tools are accepted by experts as better tools, this indicates that	The review of instrument psychometric properties has been completely redone. This resulted in a broader, more inclusive list of instruments assessed as either valid or reliable among either Medicare-like populations or other populations. The Medicare applicability criteria were used only for categorization of studies, not as eligibility (exclusion) criteria.

Research review section	Reviewer, Affiliation	Comment	Response
		there is a problem in the analysis, which leads me to question how these Outcome Measures were screened and rated. Were the articles screened and rated by multiple reviewers to ensure the information identified is accurate? If articles were excluded by applying "Medicare" applicable criteria, how was this specifically determined?	
General	Peer Reviewer #3	Publishing a document that appears to promote antiquated tools that are impractical to implement clinically and discredit tools that are meaningful and currently being used in respected studies by experts in the field will only cause more problems in establishing limb loss research. If publications were omitted which would change the review, this should be considered before final draft of this report. Likewise, if analysis of current articles were misinterpreted, this needs to be revisited.	The review of instrument psychometric properties has been completely redone. This resulted in a broader, more inclusive list of instruments assessed as either valid or reliable. We do not believe that relevant studies have been omitted.
General	Peer Reviewer #3	As evident from this SR which highlights the lack of KQ answered in past available research, clear and unequivocal guidelines need to identify which OMs are meaningful and useful for pending and future research. If this SR fails to identify appropriate OMs, this field will continue to lack evidence to answer the key questions in many years to come. Again, the implications of this document in providing information for limb loss is critical and will have far reaching effects.	We agree
General	Peer Reviewer #3	The cut-off date for this review was 11/30/16. Given the importance of this report, any subsequent publications since that date should be added to this SR.	As part of the protocol, the review has been updated since the draft report.
General	Peer Reviewer #3	The report is well constructed and organized in identifying the KQs, however, the approach to the literature review should be re-evaluated. The goal of this SR should be more inclusive to obtain better data resulting in conclusions that are applicable to all people with limb loss based on subset populations since it inevitably will be used as an extrapolating tool despite the intended target. The gravity of this implication is profound and	We agree. The primary population of interest is the Medicare-eligible population, but the review was not restricted to this population as evidenced by the non-U.S. literature and the studies of people who would mostly not be eligible for Medicare. We highlight the Medicare issues but also the findings relevant to the broader population.

Research review section	Reviewer, Affiliation	Comment	Response
		cannot be overlooked by the task force and those reviewing this important work.	
General	Peer Reviewer #3	The potential conclusions are relevant and influential to both policy and practice decisions and will impact all people with limb loss regardless of coverage by Medicare or some other third party payer. The responsibility of this task force by way of this SR is to determine effective resource utilization as well as promote patient access. We all must understand that producing a clinically relevant document to achieve these two goals for all stakeholders involved and ultimately ensure the right components for the right patient. The potential conclusions are relevant and influential to both policy and practice decisions and will impact all people with limb loss regardless of coverage by Medicare or some other third party payer. The responsibility of this task force by way of this SR is to determine effective resource utilization as well as promote patient access. We all must understand that producing a clinically relevant document to achieve these two goals for all stakeholders involved and ultimately ensure the right components for the right patient. The potential conclusions are relevant and influential to both policy and practice decisions and will impact all people with limb loss regardless of coverage by Medicare or some other third party payer. The responsibility of this task force by way of this SR is to determine effective resource utilization as well as promote patient access. We all must understand that producing a clinically relevant document to achieve these two goals for all stakeholders involved and ultimately ensure the right components for the right patient.	Thank you. The review (of course) and the reviewer comments will be available to relevant policymakers and other thought leaders.
General	Peer Reviewer #4	Yes overall the report is well structured and organized. main points are clearly presented. It is generally an excellent summary of the literature with the exceptions noted in my comments. I agree with its conclusions and the policy /practice implications of its conclusions.	Thank you

Research review section	Reviewer, Affiliation	Comment	Response
General	Peer Reviewer #4	I don't know that it contributes new information but it presents information in a format that integrates an extensive body of research, in an organized fashion with a clearly stated purpose and a defined organizational approach.	Thank you
General	Public Reviewer #10 Anonymous	This comment section should have allowed a longer period than the allotted 3 weeks in order to thoroughly assess the research articles that were investigated. This time frame is too short to begin to address responses to the literature review.	The review period was based on AHRQ policy.
General	Public Reviewer #10 Anonymous	Although the authors do demonstrate the limitations in current research and the need for higher quality studies to demonstrate the evidence of prosthetic outcomes, there is little evidence to support the argument against.	We do not argue against the use of prosthetics or any given component. The review (KQ 4-7) mostly highlights the lack of high quality evidence.
General	Public Reviewer #10 Anonymous	With respect to outcomes, the authors that do use validated outcome measures that are reliable such as TUG, Berg, AMP-Pro, FIM, and 10 MWT demonstrate improved outcomes with prosthetic use components over others. Although there are issues with methodology, the majority of the studies demonstrated improved outcomes with respect to increased ambulation, functional activities, and decreased fall risk with prosthesis use, the results do not demonstrate user outcome regression as a result of using prosthetic components.	For KQ 4, we included only studies that addressed heterogeneity of treatment effects; ie, differences in relative effects among subgroups. We did not address the research question addressed by this reviewer, whether people, on average, benefit from a given component.
General	Public Reviewer #10 Anonymous	The authors allude to “moderate evidence” with six articles to support that users (11-22%) abandon use of their limb prosthesis after 1 year. However, the patient’s in these studies are one who fit criteria of poor prognosis with prostheses d/t comorbidities and cognition. This was seen in the study by Remes et al 2009 that older adults with PAD with comorbid conditions influencing their functional capacity hindered their ability to ambulate with a prosthesis. In addition the evidence to support negative functional outcomes (loss of bipedal ambulation, etc.) is insufficient and low, there is not a strong argument to support it. In order to continue treatment, rationale for preventing further comorbidities, decline of condition, regression, or worsening prognosis is the main target for clinicians in justifying providing care to insurance	We have downgraded this to low strength of evidence based on the fact that the studies are all either old or non-US based. The findings of low or insufficient evidence are not to imply that the outcomes are not important or do not commonly occur. They reflect only the weak state of the evidence.

Research review section	Reviewer, Affiliation	Comment	Response
		payers and should have been one of the questions to address in this review.	
General	Public Reviewer #21 James Kling, Georgia Tech MSPO Student	Key question 7 of this systematic review addresses the percentage of individuals who after 6 months, 1 year, and 5 years after the receipt of a lower limb prosthesis, maintain bipedal ambulation, use prostheses for transfers only, use prostheses only indoors, abandon their prostheses, and have major problems with their prostheses. This section looks at longitudinal studies that assess these areas with feedback surveys.	This is correct.
General	Public Reviewer #21 James Kling, Georgia Tech MSPO Student	One criticism I have with the methods of this section is that studies where patients failed to respond to the survey, were criticized and labeled as having a high risk for sample bias. The researchers sent these surveys to patients, and those patients decided not to respond. This was not the fault of the researchers, so I don't believe labeling their studies as having a high risk for sample bias is a fair assessment, as the researchers were not at fault for the lack of responses.	While not the "fault" of the researchers, very low response rate is an important source of bias. People who respond to surveys are inherently different (maybe healthier, maybe angrier) than those who do not respond (maybe more depressed). This is a universal problem with surveys. It is not indicative of poor methodology, but is an important bias concern.
General	Public Reviewer #21 James Kling, Georgia Tech MSPO Student	I also believe that the low numbers of prosthesis abandonment should be labeled as a positive, not a negative outcome. The reasons for prosthesis abandonment in studies that address this, are primarily patient caused, not prosthetist caused. This should be noted, as the prosthetists and the devices they make for patients caused a very low percentage of issues that led to prosthesis abandonment. This highlights the expertise in socket design and fabrication, as well as componentry selection that prosthetists have.	Abandonment is a "negative" outcome, although we do not characterize it as either negative or positive. Nothing in KQ 7 suggests blame or who causes abandonment. The evidence regarding reasons for abandonment was insufficient. The single study reporting on reasons for abandonment also does not assign blame.
General	Public Reviewer #21 James Kling, Georgia Tech MSPO Student	There is an area of research that I believe should be considered when addressing the long-term effects of receiving a prosthesis. Longitudinal studies where the quality of life of patients that receive prostheses should be addressed, more specifically, cost effectiveness of the device when compared to individuals who do not receive a prosthesis. I believe that looking at these studies will show a positive return on	These are important issues that may be able to be addressed in a decision model. As noted, Dobson & DaVanzo report an economic evaluation. The current review neither addresses costs or economic outcomes nor does it include a decision analysis.

Research review section	Reviewer, Affiliation	Comment	Response
		investment for prosthetic care from Medicare. Individuals who receive prostheses, are more likely to return to the workforce after the initial recovery period after amputation. Those who return to work have the chance to get off of disability, and will pay taxes on their income. The Dobson & DaVanzo, 2013, "Retrospective Cohort Study of the Economic Value of Orthotic and Prosthetic Services Among Medicare Beneficiaries" study attempted to look at this, however they were only limited to the first year after amputation. The first year after amputation is inherently the most expensive year of an amputee's life, due to surgery and rehabilitation costs, regardless of whether or not they receive a prosthesis, so no significant difference in Medicare payments was found between amputee's who received prostheses, and those who did not. However this study is inherently flawed since it only looked at the first year after amputation, but if Medicare costs were evaluated after 2 years, 5 years, and possible more, the results would likely see amputees who receive prostheses would have a statistically significantly lower cost burden on Medicare than amputees who do not receive prostheses. Another study that attempts to address this question is the Childers et. al, "Vocational rehabilitation services benefits people with amputation" study. This study was not performed on Medicare patients, however, it highlights a very important result that can be applied to any amputee population. This study found that 77.7% of people who received a prosthesis were able to find employment, and were 90.2% less likely to have public support as their primary income. This is incredibly important to the economical aspect of providing a prosthesis. It shows that amputees who receive prostheses are very likely to have a decreased financial burden on Medicare after 2 years. More of these studies should be looked at, and more of these studies should be done to show the financial importance of an amputee receiving a prosthesis.	

Research review section	Reviewer, Affiliation	Comment	Response
General	Public Reviewer #13 Tim Bump	I appreciate that the authors consider 19 measures to be validated. This seems like a positive thing to helping further research unfold. I think the K-levels, although used currently, should be removed totally in favor of looking at outcome measures. I think it would be appreciated if the tone within the paper was in a more positive manner. I also think the evidence levels should be based off of a rehabilitation model rather than the AHRQ Methods Guide.	We make no recommendations about policy issues surrounding instruments, including K levels. The "evidence levels" are Strength of Evidence based on the study evidence, not on levels of rehabilitation or the like.
General	Public Reviewer #15 Phil Stevens	"The authors are commended for their substantial attempts to aggregate and synthesize the current body literature related to lower limb prostheses. I. The current narrative is largely dismissive of the value of clinical judgment in prosthetic rehabilitation despite the mandate of its use by the current language of the Local Coverage Determination (LCD) for lower limb prostheses and its stated value as a tenant of Evidence-Based Practice. While the stated objective of the authors is to synthesize the available academic evidence, the authors are encouraged to clarify that this should be performed to compliment clinical judgment and individual patient consideration, rather than at their expense. The authors are encouraged to modify the narrative in the background to accurately reflect the current standard in matching patients to prostheses. The current language correctly summarizes that “prosthetists often rely on clinical judgment to match patients to prostheses.” However, this statement dismisses the role of the treating physician and fails to disclose that the current reliance on clinical judgment is mandated by the language of the current LCD, included below: ""A determination of the medical necessity for certain components/additions to the prosthesis is based on the beneficiary’s potential functional abilities. Potential functional ability is based on the reasonable expectations of the prosthetist, and treating physician, considering factors including, but not limited to: 1. The beneficiary’s past history (including prior prosthetic use if applicable); and 2. The	Thank you. The review does not address, dismiss, or impugn clinical judgment. The review only addresses the evidence to support the Key Questions. We make no recommendations regarding policy or clinical implementation. We have added "and treating clinicians" to the sentence about use of clinical judgment. We have also added a sentence that the choice of prosthesis in the US is usually restricted based on Medicare LCD requirements. We have further removed the concept of “medical necessity”.

Research review section	Reviewer, Affiliation	Comment	Response
		beneficiary's current condition including the status of the residual limb and the nature of other medical problems; and 3. The beneficiary's desire to ambulate."" The current standard is defined as the "reasonable expectations of the prosthetist and treating physician." The inadequacy of current standardized assessment tools in lower limb prosthetic prescription practices is due in part to a well-established policy based on the "reasonable expectations" of the treating clinicians. Further, any document that attempts to synthesize the available published evidence related to medical management, especially when the document concludes that such evidence is largely inadequate has a responsibility to place the shortcomings of the evidence within the larger context of evidence-based practice. These standards, as articulated by Sacket, are defined below: ""Evidence based care is the integration of clinical expertise, patient values, and the best evidence into the decision making process for patient care. Clinical expertise refers to the clinician's cumulated experience, education and clinical skills. The patient brings to the encounter his or her own personal and unique concerns, expectations, and values. The best evidence is usually found in clinically relevant research that has been conducted using sound methodology (Sacket, 2002)."" When the "best evidence" is inadequate or requires further development, evidence based care suggests increased reliance on the remaining considerations of clinical expertise and the individual concerns, expectations and values of the patient. This simply reinforces the position that the current standard of clinical judgment based on individual presentation complies with the standards of evidence based care within the limitations of the current academic evidence. The authors are encouraged to treat these important considerations in their interpretations of the findings. References Sackett, D. (2002) Evidence-based Medicine: How to Practise and Teach EBM, 2nd edn. London: Churchill Livingstone. Humensky J et	

Research review section	Reviewer, Affiliation	Comment	Response
		al. Service utilization of veterans dually eligible for VA and Medicare Fee-For-Service: 1999-2004. Medicare & Medicaid Research Review. 2012; 2(3):E1-E22."	
General	Public Reviewer #15 Phil Stevens	II. The general emphasis on elderly individuals with amputations of dysvascular etiology overlooks large segments of Medicare's current constituency and the role of CMS in establishing healthcare standards within the private sector. It also presumes differential response rates to prosthetic components among certain subgroups, when this is not evidenced in the review itself. Notably, according to the Center for Medicare Advocacy, roughly 16%, or 1 in 6 Medicare beneficiaries is under the age of 65 (http://www.medicareadvocacy.org/under-65-project/). Further, Humensky et al suggest in their 2012 publication that there are over 5 million dually enrolled veterans that are also Medicare beneficiaries, many of whom are either under the age of 65 or have amputations of traumatic etiology (Humensky, 2012). The decision to largely exclude studies that report on middle aged adults or individuals with traumatic amputations neglects large segments of Medicare's current beneficiaries. Given that most private sector coverage policies are based on Medicare guidelines, the decision to discount studies on patients that do not meet the current stereotyped standard of a geriatric patient with a dysvascular amputation will eventually undermine access for patients in the private sector that do not meet this stereotype.	The review does emphasize the generalizability to the Medicare population, but thoroughly includes and summarizes evidence for all populations (in high resource countries). The categorization to potential generalizability to the Medicare population included both age and dysvascular conditions.
General	Public Reviewer #15 Phil Stevens	Finally, the authors are encouraged to consider their decision to base their search on the premise that the subgroup of patients with geriatric-onset, dysvascular amputations respond differently to prosthetic interventions. This premise was ultimately used to exclude "79 studies that compared lower limb prosthesis components but did not report subgroup analyses, regression analyses or individual patient data which would allow subgroup analyses," a number that approximates	The 79 articles (now 97) referred to here did not address heterogeneity of treatment effect, which was what the Key Question addressed. We did not address the research question suggested by this reviewer, whether people, on average, benefit from a given component.

Research review section	Reviewer, Affiliation	Comment	Response
		the 92 articles that were included in consideration. Yet, the authors' current review will ultimately declare, "evidence that evaluated patient characteristics do not predict which patients would most benefit from a given LLP component." Until such predictive evidence is established, the fundamental decision to exclude a large body of published evidence that fails to allow sub-group analysis is unfounded by the authors' own work.	
General	Public Reviewer #15 Phil Stevens	III. The authors are encouraged to discuss their findings within the context of the now universally understood "triple-aim" of health care. This includes improving the patient experience of care (including quality and satisfaction), improving the health of populations, and reducing the per capita cost of health care. Within this construct, the pursuit of one aim should not occur at the undue expense of the others. More specifically, until adequate evidence exists that could simultaneously inform improvement in the patient experience and reductions in per capita cost, the relatively high marks related to patient experience should not be compromised. The ultimate intent of this review is found in the statement that, "Suboptimal matching of patients to LLPs may unnecessarily increase health care utilization, prevent attainment of maximal patient function, and defer realization of improved quality of life attainable with an appropriate prosthetic." Unfortunately, the review concludes that: "the studies warrant a low strength of evidence that evaluated patient characteristics do not predict which patients would most benefit from a given LLP component…However, it may be more accurate to conclude that the evidence is currently sparse and fails to adequately address whether different subgroups of amputees are more or less likely to benefit from given specific components." Within these limitations, the authors find that "at least three-quarters of people receiving a LLP were satisfied with the process of accessing their LLP" and that "on average clients were satisfied with their visits to their prosthetists' offices." In consideration of the triple-aim of	These are worthwhile considerations but venture too far toward policy recommendations, particularly about costs to be included in this evidence review.

Research review section	Reviewer, Affiliation	Comment	Response
		health care the authors are encouraged to identify the relatively high marks within the patient care experience and caution against the pursuit of per-capita cost reduction until the evidence exists to support the pursuit of this aim.	
General	Public Reviewer #15 Phil Stevens	IV. The authors are encouraged to place the utilization and abandonment rates within the context of other healthcare services. For example, how does the 78-89% percent success rates observed with lower limb prosthetic interventions compare to the one-year success rates observed with other major health care episodes such as cardiac events or cancer among a geriatric population with overlapping co-morbid health conditions? When viewed in terms of success rather than abandonment and compared against the success rates observed in other medical scenarios involving older adults with dysvascular etiologies, the one year success rates in lower limb prosthetic acceptance can be rightly recognized as being reasonably high.	Cross-clinical comparisons are fraught with risks of inappropriate equivalencies. We leave it to policymakers and others to make these comparisons and stick to presenting the evidence.
General	Public Reviewer #15 Phil Stevens	V. The authors are obligated to address the limitations and pitfalls associated with predictive analytics and sub-group stereotyping that they advocate. The data of Roffman et al (reference 130 in the authors' current review) is indicative of these limitations. For example, while 46% of those who abandoned their prostheses had type 2 diabetes, 37% of the prosthesis users had the same diagnosis. Similarly, peripheral arterial disease was observed in 61% of the non-users, but also in 47% of the users. Traumatic etiology of amputation was observed in 31% of the users, but also in 22% of the non-users. Thus, single variables have failed to define appropriate care pathways. Similarly, the authors' observation of "low strength of evidence that 11 to 37 percent of LLP recipients use their prostheses only indoors; …. suggesting that people with transfemoral amputations, or who are older, or with bilateral amputations are more likely to be limited to indoor use," must acknowledge that there are numerous individuals within each of the subgroups that do not	We agree that the evidence is of low quality and the between-subgroup data particularly so. This, together with the sparseness of data explain the mostly insufficient evidence conclusion. We think the finding that 11 to 37 percent of people use their prostheses only indoors is clear enough that the majority are not thus restricted.

Research review section	Reviewer, Affiliation	Comment	Response
		limit their prosthetic wear to indoor use. Even the aggregated prediction rules of Roffman et al were unable to consistently predict users and non-users. To justify application in healthcare policy, prediction rules would need to be able to determine those at risk for nonuse without falsely classifying users and likely nonusers. Until such standards are established, predictive analytics run the risk of denying care to patients who would have otherwise benefited from that care.	
General	Public Reviewer #15 Phil Stevens	V. The authors are encouraged to honor the spirit of their comment periods. It is concerning that Search Strategy states the literature searches were done on November 30, 2016. This indicates that the literature search was performed prior to the closing of the comment period that AHRQ opened for that review. This implies that the literature searches were performed with a lack of regard for external comments, as these had not yet been fully received.	The search date should have been explained better in the draft review. It represented the date that the actual search was done but not the dates of the screening process, which took place in 2017. Subsequently the search (and screening) was updated through October 30, 2017. In discussions with Key Informants and Technical Experts, it was decided to not exclude older studies.
General	Public Reviewer #15 Phil Stevens	VI. The a priori decision to limit those studies addressing KQ 1-3 to those with more than 20 subject and those addressing KQ 7 to those with more than 100 subjects appears arbitrary. The authors are encouraged to provide some level of rationale for this decision to further limit a small evidence base.	For KQ 1-3, we have added the explanation: an arbitrary threshold chosen to ensure a sufficient number of study participants for statistically meaningful correlation and comparison analyses within each study. For KQ 7 we added the explanation: smaller studies are numerous but lack precision.
General	Public Reviewer #15 Phil Stevens	VII. The authors state study participant characteristics of interest to include K-levels. It should be noted that K-level assignment is a payer driven categorization of patients, which fails to meet the validity and reliability standards established for outcome measures. As such, K-level assignment should not be included the current review as it falls short of the review's stated purpose to examine evidence in prosthetic rehabilitation.	We chose to include K levels as they are universally applied to categorize patients and their prosthesis requirements. We acknowledge and state that they have not been validated. Given the insufficient state of the evidence, inclusion or exclusion of K levels would make no difference to the conclusions. Readers who disagree with their inclusion are free to dismiss the relevant rows within the summary tables.

Research review section	Reviewer, Affiliation	Comment	Response
General	Public Reviewer #15 Phil Stevens	VIII. The authors are encouraged to include within their review studies published in the peer-reviewed Journal of Prosthetics and Orthotics which is not listed on the databases noted.	All eligible studies from this journal have been included.
General	Public Reviewer #15 Phil Stevens	IX. With regard to outcomes measures that meet reliability and validity standards, within the 19 that are concluded to meet all requirements, the PLUS-M (references 70 and 74) was omitted. However, this measure meats all of the definitions of your inclusion criteria and should be included.	We have completely reanalyzed the studies for KQ 1-3. PLUS-M is now included among those instruments with test validity and reliability (from studies not generalizable to the Medicare population).
General	Public Reviewer #16 Tyler Murphey	"Key Question #2 Comments: The results from these outcome measures are being compared with other outcome measures without a gold standard. There is no emphasis on biomechanics to compare these outcome measurements to. Gait analysis (range of motion, ground reaction forces, socket pressures, metabolic costs, muscular power, step length, center of gravity distance, load line, etc) should be analyzed and used to predict validity of outcome measures for each amputation level (transfemoral, transtibial, symes, knee disarticulation, etc). This should be done over a variety of population demographics and k-levels. Various prosthetics components should also be assessed across populations (different socket type, components, feet, microprocessor knees (MPK) / non-microprocessor knees (NMPK's), etc). There were also several important outcome measures missing from the study such as the AMP pro and the 10 meter walk test."	The review was explicitly designed to evaluate patient-centered outcomes, function, and the like. These were the questions posed to us by policymakers. The review was not designed to evaluate the biomechanical properties of LLP components. This is not to diminish the importance of these outcomes, but it was not the purpose of the review. This is noted in the abstract, introduction (objectives), relevant results sections, and discussion. KQ 4 addresses prosthetic components compared by populations (patient characteristics). This is noted in the abstract, introduction (objectives), relevant results sections, and discussion The instruments have been reanalyzed. AMPPRO and 10 meter walk test were and are still included.
General	Public Reviewer #18 William Hendrix	"Clinician experience fails to be mentioned. Clinician experience should have some weight when it comes to deciding which prosthetic components are suitable for the patient. Decades of experience fitting and producing prostheses for thousands of patients can lead to a decision-making strategy that becomes second nature for the clinician. Some aspects of this draft may not have hard evidence to support claims because clinicians cherish their intuition.	We have added in clinicians in the statement about prosthetists selecting prostheses for patients. This review covers the evidence and does not cover expert opinion.

Research review section	Reviewer, Affiliation	Comment	Response
General	Public Reviewer #18 William Hendrix	A three week long comment period for a 440 page document is part of the bias in favor of those who produced this document."	We apologize for the short comment period. We did not have control over this decision.
General	Public Reviewer #19 David Boone	There is general agreement that additional research is needed to more fully understand the functionality and benefits of prostheses and their components. However, any present lack of evidentiary proof for specific patient benefits provided by prosthetic technologies today is not the same as support that there is not benefit. This report should be written to be carefully and explicitly point this out. Even the title is misleading and should be changed to reflect the results: "A Systematic Review of Currently Validated Outcome Measures for the Medicare Population with Lower Limb Loss."	We now more clearly state that lack of evidence is not the same as evidence of lack of effect. This review does not directly address the question about effectiveness of LLP. The suggested title is incorrect as this is not a review of validated outcomes or of the Medicare population.
General	Public Reviewer #2 Andreas Kannenberg, Otto Bock HealthCare LP	First, I would like to commend and congratulate the AHRQ and the commissioned research group on the comprehensive and thorough systematic review. I basically concur with most of the results and conclusions. Nevertheless, I would like to bring a number of points to your attention that may warrant reconsideration and edits or additions to the report. The intention of my comments is to support the systematic review by helping sharpen and preventing misinterpretation of some of the conclusions.	Thank you
General	Public Reviewer #2 Andreas Kannenberg, Otto Bock HealthCare LP	The comments made on the abstract and the evidence summary do also apply to the respective sections and aspects in the main report. In addition, I would like to provide comments on specific aspects in the main report that have not been covered by the structured abstract or the evidence summary.	Thank you
General	Public Reviewer #20 Anonymous	"• Limited participant numbers is a well-known limitation of amputee based studies in the field. To further restricted the acceptance criteria to studies that compare (or allow comparison of) sub-populations is likely to prove overly-specific (KQ4). Indeed, many studies that compare prosthetic components will deliberately try to avoid variation for fear of	We have more explicitly stated that the purpose of KQ 4 is not to address the overall comparative effectives. The question addresses heterogeneity of treatment effect.

Research review section	Reviewer, Affiliation	Comment	Response
		external influence on the results from things such as comorbidities, mobility level etc."	
General	Public Reviewer #21 James Kling, Georgia Tech MSPO Student	The authors of this systematic review do not have experience in the field of prosthetics, so they could never know the importance of the work done by practitioners, and the importance of a variety of quality components that give function back to patients. Every patient is different, and no one set of components will work the same on every patient. The technical experts, key informants, and professional assistants have not been disclosed, and are listed as pending, which is a disservice to the credibility of the authors of this review.	We collaborated with an expert in prosthetics, healthcare of amputees, and psychometric evaluation. As per AHRQ policy, the authors, Key Informants, and other were redacted from the draft. They are listed in the final report.
General	Public Reviewer #3 Anonymous	The systematic review draft for Lower Limb Prosthesis by the Agency for Healthcare Research and Quality brings up some key points about the validity and reliability of outcome measures used to assess the effects of a prosthetic intervention. The systematic reduction of 61 to 19 outcome measures will help simplify how prostheses may be assessed in the clinic and will reduce confusion among clinicians and researchers as to which outcome measures may be the most appropriate. However, there are several key issues I feel should be resolved if this literature review will be used to set Medicare reimbursement policy	Thank you
General	Public Reviewer #3 Anonymous	The L-code reimbursement system for lower limb prostheses is grounded in describing mechanical components based on their mechanical function. For example, L5980 - flex foot system, describes a prosthetic foot designed to store energy during loading response, facilitate a transfer of that energy to be used during late stance for propulsion. These systems are typically made of carbon fiber laminate and include an integrated carbon fiber shank section. These components are mechanical in nature and biomechanical analyses are able to investigate these components to ensure that they are working as intended. Excluding these studies overlooks the core reason why they are used to evaluate prosthetic components	As is more clearly stated in the revision, the review does not assess overall comparisons between components. We were evaluating heterogeneity of treatment effect. We were not interested in outcomes that were not patient-centered. This review is not designed to address biomechanical performance. This is noted in the abstract, introduction (objectives), relevant results sections, and discussion.

Research review section	Reviewer, Affiliation	Comment	Response
		and ignores the connection between mechanical performance of the prosthetic component and the Medicare L-code reimbursement system.	
General	Public Reviewer #3 Anonymous	The rationale to exclude biomechanical studies based on them being "nonpatient centered" is unrealistic. Biomechanical studies evaluate how human beings are able to use these devices. Given that the patients are also human beings, it would seem that these analyses are indeed patient centered. These analyses offer more sensitivity to detect changes in neuromechanical strategies (the combination of measures that involve assessment of neuromuscular performance in the context of the person's biomechanics, and task being performed) that would go otherwise undetected in the clinically based outcome measures listed in the review. This enables a very thorough method to evaluate individual performance of a prosthetic component and meets the very definition of patient centered. For example, biomechanical measures can measure the knee joint moment in the coronal plane and define how a different prosthetic treatment may influence this moment. The knee adduction moment is highly predictive of the development of knee osteoarthritis in the sound limb. A condition people with uni-lateral amputation are already at high risk of developing. The tuning of prosthetic treatments based on these approaches are prevalent in the literature and represent very relevant and patient centered measures to evaluate prosthetic care.	We acknowledge the importance and value of studies evaluating biomechanical and other surrogate outcomes, but this review was designed to focus on the heterogeneity of treatment effects for outcomes of primary interest to patients and clinicians, as opposed to prosthesis designers, engineers, and researchers. This is noted in the abstract, introduction (objectives), relevant results sections, and discussion.
General	Public Reviewer #3 Anonymous	This review was difficult to read in many places and lacked clarity. The way studies were referenced were inconsistent throughout the article and many times a sentence would make a statement about a study but without a reference back to which specific study the author was referring. For example, the first paragraph on page #72 includes articles referenced in number format and in a first author and year format. The second sentence of the paragraph states, "Ten studies included between 5 and 168 users of LLP; one included 899	We have improved referencing. However, it is still necessary for readers to refer to the summary tables for specifics of studies.

Research review section	Reviewer, Affiliation	Comment	Response
		amputees." Yet no indication of which study had 5, which had 899, etc. The literature review is riddled with sentences like this that make it difficult to accurately review.	
General	Public Reviewer #3 Anonymous	. Many of the tables have text that are cropped within a cell so the reader cannot see the entire comment made by the author. Tables are poorly and rarely referenced in the text, again making it difficult to review, and underscore the need for more time to accurately assess the work.	Referencing and table formatting have been improved.
General	Public Reviewer #4 American Orthotic Prosthetic Association	As stated in the draft report, the purpose of the systematic review was to "assess validity of measures used in adults with lower limb amputation, whether characteristics can predict relative effectiveness of different lower limb prosthesis (LLP) components, and long-term use of LLPs." AOPA supports each of these criteria as crucial to determining the clinical effectiveness of lower limb prostheses and believes that the AHRQ appropriately focused on these areas when conducting its systematic review. Unfortunately, AOPA believes that there are significant deficiencies in the draft report that fail to consider much of the existing clinical research that achieves many of the stated purposes of the systematic review, and misinterprets much of the existing clinical literature and its findings, accentuating a few negative points while missing important positive conclusions about benefits of advances in prosthetic treatments, and how newer lower limb prostheses provide improved patient outcomes. AOPA is also disappointed that the AHRQ did not appear to take into account many of the suggestions that AOPA submitted as part of its extensive comments on the key questions that would be used when performing the systematic review. Finally, AOPA believes that it provided valuable resources to AHRQ regarding in-progress studies, both by the RAND Corporation and the health economics firm Dobson-DaVanzo that should have been included in the systematic review by AHRQ. While those studies were not complete when AHRQ requested comments on the key questions, AOPA submitted preliminary	Thank you. We have followed the protocol for Key Questions, study eligibility (including publication history), and other specifics. We have added caveats to clarify that this review is not designed to address all questions but stays focused on the specific Key Questions asked.

Research review section	Reviewer, Affiliation	Comment	Response
		reports from both organizations with its earlier comments and later submitted the completed reports, both of which were transmitted well in advance of the AHRQ's publication, so that these important contributions to the literature on lower limb prosthetics, both could have been, and should have been incorporated in the AHRQ document. The failure of the AHRQ to acknowledge or recognize these studies in its draft report is extremely disappointing as both studies are focused on the cost effectiveness and clinical advantages of certain lower limb prosthetic components. AOPA's full comments on the AHRQ systematic review draft report are presented below	
General	Public Reviewer #5 Orthotic and Prosthetic Alliance	The scope and depth of the draft review, "Lower Limb Prosthesis," is considerable and reflects the broad Key Questions proposed in the protocol. However, the draft review appears to be hastily conducted and authored, as evidenced by spelling and grammatical issues, selective or improper citations, data extraction errors, and inconsistent formatting. Below, we highlight key substantive concerns and provide suggestions to improve the quality and transparency of the review. Given the significance of the topic and the potential impact of this review on the care of hundreds of thousands of people with limb loss living in the United States, we strongly encourage the authors (and AHRQ as the study sponsor) to review the draft report for clarity, consistency, content accuracy, and formatting before final public dissemination. Although we recognize that it may be too late to change the overall course of this review, we believe it is important for the authors to consider and address the issues we have raised in this comment letter. We implore the authors to consider the significant implications of this systematic review as they finalize the report.	We believe we have corrected all typographical and other errors of the draft report.
General	Public Reviewer #5 Orthotic and Prosthetic Alliance	The review was intended to assess the validity of outcome measures used to evaluate patients with lower limb amputation and determine if patient characteristics can predict the relative effectiveness of different LLPs (or specific LLP	These are accurate statements as of the draft report.

Research review section	Reviewer, Affiliation	Comment	Response
		components). Based on this goal, the authors proposed seven "Key Questions," each of which included between one and eight sub-questions (for a total of 30 questions). The corresponding search for evidence identified 10,178 candidate articles, which were screened against stringent selection criteria. Ultimately, the authors identified 95 articles to consider that were responsive to the Key Questions (n=3, 6, 64, 11, 0, 1, and 8 for each Key Question, respectively).	
General	Public Reviewer #5 Orthotic and Prosthetic Alliance	After a thorough review of these articles, the authors concluded that, "relatively few studies address the questions of interest for this review, particularly related to heterogeneity of treatment effect, patient expectations and satisfaction, and long-term use of LLP after prescription." (See; p. 122.) While the authors ultimately suggest that the quantity and quality of the literature in this area is not sufficient to address the Key Questions, we question whether the Key Questions were sufficiently developed and tested (e.g., via a preliminary search of the literature) prior to conducting the review. For example, it seems the only question that concluded with a reasonable amount of evidence was key question 3, "What functional outcome measurement tools used to assess adults who use a LLP have been evaluated in the published literature?" While we acknowledge the importance of identifying outcome measures available to assess Medicare-eligible prosthesis users, we question if the information presented in this review is distinct from systematic reviews conducted in recent years by other experts in this area.8, 9	We believe that this review more thoroughly covers psychometric properties of a wider range of instruments than prior reviews. It is also more current.
General	Public Reviewer #5 Orthotic and Prosthetic Alliance	Title of the Final Report The title "Lower Limb Prosthesis," (or more appropriately, "Prostheses") is not an accurate reflection of the scope and depth of the report. This is evidenced by the authors' own words, "It is important to note that this review does not fully cover the field of evaluation of LLPs [lower limb prostheses]" (p.ES-1). Given the scope of the review, a more specific title is strongly recommended, such as "Measurement and Prediction of Health Outcomes among Medicare-Aged	We have revised the title. We struggled to find a title that covers the disparate Key Questions and hoped a simple title might suffice. We have expanded the title to Lower Limb Prostheses: Measurement Instruments, Comparison of Component Effects by Subgroups, and Long-Term Outcomes. Note, though, that this is not a review of Medicare

Research review section	Reviewer, Affiliation	Comment	Response
		Prosthesis Users with Amputation Due to Dysvascular Disease." Other AHRQ report authors have used a similar approach to specify the content of their review.42, 43, 45	aged prosthesis users with amputation due to dysvascular disease.
General	Public Reviewer #5 Orthotic and Prosthetic Alliance	The review authors acknowledge that "the distinction between assessment techniques (used to assess patient function prior to new or replacement prescription of a LLP), prediction tools (used to assess future outcomes), and outcome measures (used to assess patient function, etc. with their new or replacement LLP) is not as clear-cut as their definitions would imply" (p. 22). One reason these terms are confusing is that the authors' brief definitions appear to confound well-established definitions of the purpose of an outcome measure (i.e., evaluation, discrimination, and/or prediction21) with the times at which the instrument can be used (i.e., prior to or following delivery of a LLP). We encourage the authors to consider revising their definitions to clarify issues of instrument purpose and timing. In lieu of such changes, we request the authors to explicitly define their terminology at the beginning of the final report.	We have rewritten our descriptions of how we categorized instruments as assessment techniques and prediction tools.
General	Public Reviewer #5 Orthotic and Prosthetic Alliance	We respectfully submit that terms such as "validated" and "found reliable" are generally not appropriate for characterizing predictor or outcome measures. Instrument development is an ongoing process, wherein evidence of testing—in different populations, settings, and applications—is continually added to the body of knowledge. The terms "validated" and "found reliable" imply that measures possess a fixed and equivalent characteristic, rather than evidence of desirable psychometric properties. We encourage the authors to use terms such as "evidence of validity" or "evidence of reliability" to characterize the appraised measures.	We agree completely and have made this change throughout.
General	Public Reviewer #5 Orthotic and Prosthetic Alliance	We question if "assessment techniques" would be better-termed as "discrimination measures" (i.e., those intended to classify people into subgroups). Given the topic of this review (LLPs) and corresponding relevance of the Medicare Functional Classification Levels (the functional classifications	It was decided to retain the protocol and terminology. The assessment does include differentiation by MFCL levels.

Research review section	Reviewer, Affiliation	Comment	Response
		that effectively determine how prosthetic components are designed, marketed, prescribed, and reimbursed in the U.S.),40 we question why the systematic review authors do not appear interested in identifying instruments capable of detecting differences in groups of people classified by different MFCLs. For example, the following studies26, 29, 41 examined outcome measure scores across people of different K-levels. These instruments seem far better suited to measuring people across a range of functional ability, rather than most of the measures currently listed, which appear to be focused on measuring low-level functional abilities. We suggest revising Key Question 1 to include identifying instruments such as these as "assessment techniques."	
General	Public Reviewer #6 Amputee Coalition	A prosthesis is an individualized and custom device that is tailored to the goals and ability of each person. We applaud the overarching goal of this work, which is to ensure that an individual who loses a limb receives the appropriate prosthetic devices to optimize their outcomes. Likewise, we concur with the general conclusions that further research is needed "to inform optimal matching of prosthetic components to patients and to assess patient expectations and satisfaction with care." As the nation's leading patient advocacy organization serving the limb loss community, we enthusiastically offer our assistance toward this goal, especially to any study seeking to assess patient expectations and satisfaction with prosthetic care.	Thank you.
General	Public Reviewer #6 Amputee Coalition	Additionally, we agree that 'long term follow-up studies are needed to understand problems and limitations people are having with their prostheses, rates of abandonment or limited use and reasons for these limitations and abandonment.' Our 2015 Limb Loss Task Force Summit of experts carefully considered how to best address this need. That group recommended the creation of a Limb Loss Model System similar to the Spinal Cord Injury (SCI), Traumatic Brian Injury (TBI), and Burn Model Systems. The longitudinal databases	Thank you for this comment.

Research review section	Reviewer, Affiliation	Comment	Response
		supported by these Model Systems have generated an enormous amount of research to inform efforts to optimize the care of these patients and improve their quality of life.	
General	Public Reviewer #6 Amputee Coalition	Despite these agreements, there are elements of the project and subsequent report that we find problematic and concerning. The focus on individuals over 65 who have lost a limb due to dysvascular disease is somewhat puzzling. We are also concerned about the section of the report related to KQ7 – especially the language used in that section and the possibility of unintended consequences from how this section is presented in the report. Finally, the results from KQ1-3 appear at odds with current clinical practice and appear to reflect an overtly academic approach to the topic, rather than a more pragmatic approach that is implementable in clinical settings	Part of the review, particularly related to instrument psychometric properties, emphasizes generalizability to the Medicare population, as per protocol. The review of the instruments (measures), and the rest of the review, primarily addresses the state of the evidence. We do not attempt to address whether use of each instrument is implementable in clinical settings. Clearly, some instruments are designed only for use in research settings.
General	Public Reviewer #6 Amputee Coalition	Although the reports clearly states that the 'review does not fully cover the field of evaluation of LLPs', it is very likely that findings from this report will be used to enact or inform the development of policies or regulations that impact the non-Medicare population. This has already occurred with the draft LCD for Lower Limb Prosthesis issued by the Centers for Medicare/Medicaid Services in 2015. Even though that was a draft proposal for reimbursement for LLPs – and was ultimately rescinded by CMS – private insurers have used language from that proposal to deny coverage for medically necessary prosthetic devices.	This review addresses the state of the evidence only. We make no policy recommendations.
General	Public Reviewer #6 Amputee Coalition	Additionally, Medicare may cover individuals with limb loss who are under the age of 65 or lost their limb from some etiology other than dysvascular disease. Currently, Medicare covers nearly 8 million individuals under the age of 65. Since a person who loses a limb is eligible for insurance through Medicare after two years, it is reasonable to expect that some of the individuals under 65 who receive insurance through Medicare have limb loss. It is also reasonable to assume that at least some of the 185,000 individuals who experience a	Based on discussions with CMS representatives and with the Key Informants, we determined general rules to categorize studies as generalizable to the Medicare population, including age and dysvascular etiologies. The review acknowledges that this system is imperfect and open to debate.

Research review section	Reviewer, Affiliation	Comment	Response
		lower limb amputation already receive insurance through Medicare or may be under the age of 65. According to 2014 hospital discharge data, 56.5% of lower limb amputations were performed on individuals under the age of 64 while Medicare paid for 57.7% of lower limb amputations performed (See HCUP 2014 NIS Accessed 11/9/2017). Furthermore, it is hopeful that many of the younger individuals who have an amputation or congenital limb difference and use a prosthesis will live long enough to age into the Medicare program.	
General	Public Reviewer #6 Amputee Coalition	The results for KQ 1-3 are somewhat surprising and seem slightly at odds with clinical practice. Many O&P practices use some version of the 6MWT and the Amputee Mobility Predictor to assess patient's functional status prior to prosthetic prescription. Yet, these assessments are scored lower than others that are not as widely used in actual clinical practice (e.g., FAI-15). Additionally, some of the tools included in KQ1-3 are not designed to assess function of adults with lower limb prosthesis (e.g. – Amputee Body Image Scale) while others (e.g., TAPES) are more appropriately used to assess quality of life after an amputation rather than function.	The evidence for KQ 1-3 has been completely reanalyzed with the addition of several instruments among the lists of validated and reliable instruments. Also the removal of other instruments, such as ABIS.
General	Public Reviewer #6 Amputee Coalition	The Amputee Coalition has serious concerns regarding the discussion around KQ7. These reservations include a lack of definition around the concept of prostheses abandonment, the language used in the report to discuss KQ7, and the quality of the evidence included. Each of these will be addressed below.	Thank you.
General	Public Reviewer #6 Amputee Coalition	Although a subcomponent of KQ7 is explicitly concerned with the topic of prostheses use and abandonment, these terms are never clearly defined within the report, nor is it made clear in the discussion of the evidence whether each study cited conceptualizes, operationalizes, and measures it consistently. Thus, it is difficult to determine what is actually being discussed in these studies and in the report.	The revision better explicates the definitions of abandonment (and other outcomes) used in the studies.
General	Public Reviewer #6 Amputee Coalition	The term prostheses abandonment implies of a lack of prostheses use on behalf of individuals with limb loss. It is highly likely that payers could use such language to justify	The revision better explicates the definitions of abandonment (and other outcomes) used in

Research review section	Reviewer, Affiliation	Comment	Response
		restricting access to medically necessary prosthetic devices, even for patients who would use them. Yet, the studies cited as evidence for KQ7 do not explicitly focus on the issue of prosthesis abandonment. For most of these studies, abandonment rates are reported anecdotally. Only two studies focus on the issue of 'abandonment' of prosthesis' outcome as an outcome – one with a moderate strength of evidence and the other with low strength of evidence.	the studies. We equated "do not use" with "abandonment".
General	Public Reviewer #6 Amputee Coalition	It is especially troubling that many of the studies cited as evidence for KQ7 possess a relatively low strength of evidence and are from non-U.S. countries and health systems. These studies also appear to contradict findings reported earlier in this report. For example, the Davies (2003) study utilizes an assessment tool (Harold Wood/Stanmore Mobility Grade) that is previously noted in the report as not being validated, reliable, or have MDC/MID identified (see Table 1-3.1). It would seem that this study – and any study using non-validated assessments- would be excluded for being cited as evidence. Doing so would be consistent with the overall spirit of KQ 1-3. If assessments are not valid, reliable, or able to determine clinical relevance, it is questionable to cite studies that utilize them as evidence for this report.	We have more explicitly described that almost all these studies are old or non-US based. This has important implications for the conclusions.
General	Public Reviewer #6 Amputee Coalition	In conclusion, while the Amputee Coalition supports the conclusion that more evidence is needed on this topic, we have serious concerns regarding the language included in the report around the issue of prosthesis abandonment, the focus on individuals over the age of 65 who lost a limb due to dysvascular disease and the divergence between the findings of this report and clinical practice. We wish to urge the study team to consider the potential impact of the findings and language included in this report beyond its intended purpose. As is already evident from the draft LCD on Lower Limb Prosthesis, reports issued by federal agencies often have far reaching consequences beyond their intended effect. We have no doubt that the intention of this report is to further improve	We believe we have presented a neutral, accurate description of the evidence to date.

Research review section	Reviewer, Affiliation	Comment	Response
		the quality of life of those with limb loss while ensuring that tax payer resources are not wasted. The Amputee Coalition supports these goals. However, we are not confident that this report will not be used for other purposes that negatively impact the limb loss community. Our comments submitted in response to this draft report are intended to attempt to prevent this from occurring.	
General	Public Reviewer #7 Anonymous	I would like to start off by commenting on the "validation" process for effective evidence. It is stated throughout the review how crippling it is that there is a lack of research to validate if these prosthesis and prosthetic components are effective or appropriate, but as the review continues it is evident that it is nearly impossible for studies to be validated. The unnecessarily strict validation process is almost as disabling to our field as is the window of merely three weeks that experts on these topics were allowed to comment. What this review is clearly lacking most is the insight of a Prosthetist. That being said, invalid testing or lack of testing does not mean these prostheses and prosthetic components are not what they claim to be, or not what the patient needs.	We have added that a lack of evidence on test validity does not imply that the instruments are not valid. Most evaluated instruments were found to be validated, although many only in studies we did not deem to be generalizable to the Medicare population.
General	Public Reviewer #7 Anonymous	When referring to people living with an amputation being "over-prescribed" with "more complex" components, I think another key understanding is missing. While this review suggests that older, less active, weaker patients may be better suited for "less complex" components, I do not entirely agree. The "complexity" of these components are designed to help users ambulate with less energy cost, less deviations from natural healthy gait, and most importantly to increase stability and safety! These are the same patients that must spend weeks in the hospital and have major surgeries if they use too much energy, adopt unhealthy gait patterns, or lose stability!	We have improved our language on this issue to better include the flip side of "under-prescription".
General	Public Reviewer #7 Anonymous	A major concluding point of the review was the low strength of evidence that patients are satisfied. I agree that a high level of satisfaction should be a constant goal for each prosthetist. However, the patient is most likely not the expert in	The Key Question refers to satisfaction with the process, not satisfaction with the prosthesis. We have added a statement to

Research review section	Reviewer, Affiliation	Comment	Response
		prosthetics. Certified prosthetists are trained to analyze patient health, determine proper care, then design, fabricate, fit, and align a prosthesis with multiple parameters to consider. They must consider alignment of each component, gait, forces on certain areas of the limb, range of motion at joints, velocity of the limb and angular velocities of the joints, and so on. Not all patients fully understand each of these considerations, which could cause satisfaction reports to be invalid tests.	this effect under the results summary for KQ 6.
General	Public Reviewer #8 Anonymous	A major shortcoming in the objectives of this systematic review is a failure to consider biomechanical outcomes of choices of lower limb prosthetic componentry. The primary purpose of a lower limb prosthesis is to be a mechanical interface between a patient and the environment to provide efficient, safe locomotion. Without addressing biomechanical comparisons of componentry, the review does not consider evidence that connects reduced loading, which may increase risk of (costly) repetitive injuries (e.g. low back pain, knee osteoarthritis),1 with the use of more advanced technologies, such as energy storing and returning feet2, elastic prosthetic ankles,3 torsion adapters,4,5 or shock absorbing pylons. 6	The review was explicitly designed to evaluate patient-centered outcomes, function, and the like. These were the questions posed to us by policymakers. The review was not designed to evaluate the biomechanical properties of LLP components. This is not to diminish the importance of these outcomes, but it was not the purpose of the review. This is noted in the abstract, introduction (objectives), relevant results sections, and discussion.
General	Public Reviewer #8 Anonymous	Additionally, exclusion of studies composed of individuals with lower limb amputation as a result of battle-related trauma prevents inclusion of larger scale studies, because centralized reporting of results for large numbers of persons with amputation are more likely in DoD/VA healthcare systems.	Very few studies were excluded for this reason. The review does include several studies of veterans from the DOD/VA healthcare system. We excluded only studies that were restricted to people with battle injuries.
General	Public Reviewer #9 Anonymous	See Excel list of missing papers	Thank you. We have conducted full text screening of all citations recommended by this reviewer and others, and have added any eligible studies found.