Comparative Effectiveness Review
Number 213

Effective Health Care Program

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

Evidence Summary

Introduction

Background

An estimated 1.9 million people in the United States are living with limb loss, a number expected to double by 2050, mostly due to the rising prevalence of diabetes.[1, 2] However, fewer than half of amputees ever receive a prescription for a prosthetic device.[3, 4] The management of lower limb amputees with respect to lower limb prostheses (LLPs) is complicated. LLP candidates are a heterogeneous group with distinct needs dependent upon age, etiology of limb loss, level of amputation, comorbidities and health status, postoperative stage, and rehabilitation status. Many LLP options exist, comprising numerous permutations of components, the anatomy they replace, their sophistication, and other attributes, including those pertaining to cosmesis and comfort.

The current standard approach for matching patients to prostheses relies heavily on performance-based assessments, self-assessments, and, in some instances, wearable monitoring technologies that record patient activity;[5] although prosthetists and treating clinicians often rely on clinical judgment to match patients to prostheses. Insurance coverage policies often dictate which prostheses and components are selected for a given

Purpose of Review

To assess validity of instruments used in adult lower limb amputees, whether patient characteristics can predict relative effectiveness of different lower limb prosthesis (LLP) components, and long-term LLP use.

Key Messages

- Thirty of 50 evaluated instruments (ambulatory/functional outcomes and other measures) have evidence of validity and reliability. Many studies use nonvalidated instruments.
- Based on a small number of studies, patient characteristics do not predict who would most benefit from a given LLP component. Half of studies used nonvalidated instruments and analyses were inadequate.
- Only a few studies assessed long-term LLP use; 11 to 22 percent of patients abandon their LLP after 1 year; people with above-the-knee amputations are more likely to abandon their prostheses than people with below-the-knee amputations; 24 to 29 percent of people with LLPs use them only indoors 1 year after they first receive the prostheses. The studies, though, had important methodological issues.

Agency for Healthcare Research and Quality
Effective Health Care Program

patient. Numerous instruments exist to assess the patient functional status, but no consensus "gold standard" assessment schema exists.

The major contextual challenges in providing data to inform matching of LLP components to patients pertain to the large heterogeneity in patient characteristics and attributes of LLPs; the lack of data on patient characteristics and LLP attributes that 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 matching.

Objectives

This review's Key Questions and study eligibility criteria were designed to assist Centers for Medicare & Medicaid Services to better understand the state of the evidence regarding how best to match patients with LLPs that would yield best outcomes for them, and related issues. It is important to note that this review does cover all aspects of LLP evaluation. Specifically, it excludes from evaluation biomechanical and other nonpatient-centered intermediate outcomes. It also does not attempt to review all evidence comparing specific components. Instead, it largely focuses on those comparisons that provide within-study data to allow assessment of heterogeneity of treatment effects (i.e., whether outcomes with specific devices vary across individuals based on different characteristics such as age or health status). The review also focuses on people who may be eligible for Medicare coverage, whether due to age or disability. Thus, we categorize studies based on their likely generalizability to amputees with Medicare. Based on discussions with our Technical Expert Panel, this includes studies with a mean age at least 65 years and those in which the percentage of participants with dysvascular disease (including diabetes) is broadly similar to the Medicare amputee population (i.e., at least 50%). Furthermore, the review excludes studies of exclusively military amputees with battle-related trauma (who are generally covered by Department of Defense and/or Veterans Health Administration insurance); however, we do include studies of veterans with multiple etiologies of amputation. Furthermore, the review excludes studies from low-income or low-resource settings not applicable to the United States.

Key Questions

The following summarized Key Questions (KQs) are addressed by the review:

KQ 1. What assessment techniques used to measure functional ability of adults with major lower limb amputation have been evaluated in the published literature?

KQ 2. What prediction tools used to predict functional outcomes in adults with major lower limb amputation have been evaluated in the published literature?

KQ 3. What functional outcome measurement tools used to assess adults who use an LLP have been evaluated in the published literature?

KQ 4. In adults who use a lower limb prosthesis, how do ambulatory, functional, and patient-centered outcomes with different prosthetic components vary based on study participant characteristics?

KQ 5. How do study participants' preprescription expectations of ambulation align with their functional outcomes?

KQ 6. What is the level of patient satisfaction with the process of accessing an LLP (including experiences with both providers and payers)?

KQ 7. At 6 months, 1 year, and 5 years after receipt of an LLP, (accounting for intervening mortality, subsequent surgeries, or injuries) what percentage of individuals…?

- **i. Maintain bipedal ambulation**
- **ii. Use their prostheses only for transfers**
- **iii Use prostheses only indoors**
- **iv. Have abandoned their prostheses**
- **v. Have major problems with prosthesis**

Methods

The Brown Evidence-based Practice Center conducted a systematic review of the published scientific literature, using established methodologies as outlined in the Agency for Healthcare Research and Quality (AHRQ) Methods Guide for Effectiveness and Comparative Effectiveness Reviews.[6] The review was registered with PROSPERO (CRD42017058488).

The searches were conducted on October 30, 2017. Study eligibility criteria are described in the full report.

Results

Summary of Studies

The literature searches yielded 10,765 citations and an additional 357 references were screened from review articles, existing systematic reviews, and from reviewers of the draft report. Of these, 425 articles were retrieved in full text. We excluded 348 articles. Of note, 89 studies compared lower limb prosthesis (LLP) components or configurations but did not report either subgroup analyses, regression analyses, or individual patient data which would allow subgroup analyses. Overall, we found 80 eligible studies (in 77 articles), of which 55 studies evaluated psychometric properties addressing Key Questions (KQ) 1 to 3, 14 studies provided data relevant to KQ 4, no studies for KQ 5, two studies for KQ 6, and eight studies for KQ 7.

Key Questions 1 to 3. Assessment Techniques, Prediction Tools, Functional Outcome Measurement Tools

Studies provided evidence regarding psychometric properties of 50 instruments for people with lower limb amputations. In total, 55 studies in 52 articles met criteria to provide evidence regarding instrument psychometrics in people with lower limb amputations. The evidence is summarized for each instrument in the main report.

We categorized instruments (or subscales, etc. of instruments) by whether studies that evaluated them were generalizable to the Medicare population (i.e., study mean age ≥65 years or ≥50% of participants had dysvascular disease) and by whether there is supporting evidence for validity and/or reliability.

The instruments evaluated are:

- 1 Leg Standing Balance
- 180 Degree Turn Test
- 2MWT (2 Minute Walk Test)
- 6MWT (6 Minute Walk Test)
- AAS (Amputee Activity Survey)
- ABC (Activities-specific Balance Confidence)
- ADAPT (Assessment of Daily Activity Performance in Transfemoral Amputees)
- AMP (Amputee Mobility Predictor with, AMPPRO, or without prosthesis, AMPnoPRO)
- AMPSIMM (Amputee Single Item Mobility Measure)
- Barthel Index
- BBS (Berg Balance Scale)
- Climbing Stairs Questionnaire
- Employment Questionnaire
- FAC (Functional Ambulation Categories)
- FAI (Frenchay Activities Index)
- FIM (Functional Independence Measure)
- FSST (Four Square Step Test)
- Functional Reach Test
- Houghton Scale
- L Test (L Test of Functional Mobility)
- LCI (Locomotor Capabilities Index)
- LEMOCOT (Lower-Extremity Motor Coordination Test)
- NQ-ACGC (Quality of Life in Neurological Conditions – Applied Cognition/General Concerns)
- OPCS (Office of Population Censuses and Surveys Scale)
- OPUS (Orthotics Prosthetics Users Survey)
- Patient Activity Monitor
- PEQ, PEQ-MS (Prosthetic Evaluation Questionnaire, Mobility Subscale)
- PFI (Physical Function Index)
- PGI (Patient Generated Index)
- PLUS-M (Prosthetic Limb Users Survey of Mobility)
- PPA (Prosthetic Profile of the Amputee)
- PROMIS-29 (Patient-Reported Outcomes Measurement Information System 29-Item Profile)
- PROS (Prosthetist's Perception of Client's Ambulatory Abilities)
- PSFS (Patient-Specific Functional Scale)
- Q-TFA (Questionnaire for Persons with a Transfemoral Amputation)
- Rising and Sitting Down Questionnaire
- RMI (Rivermead Mobility Index)
- SAT-PRO (Satisfaction with Prosthesis Questionnaire)
- SCS (Socket Comfort Score)
- SF-12/SF-36/SF-36V (Short Form Health Surveys 12, 36, and 36V)
- SIGAM (Special Interest Group of Amputation Medicine)
- Single beam test
- SIP-PD (Sickness Impact Profile-Physical Dimension)
- Tandem Test
- TAPES (Trinity Amputation and Prosthesis Experience Scales)
- TFP (Transfemoral Fitting Predictor)
- TUG (Timed Up and Go)
- TWT (Timed Walking Test)

- Walking Questionnaire
- WHOQOL-BREF (World Health Organization Quality-of-Life Scale – Brief Version)

Key Question 1. Assessment Techniques

Based on explicit reporting within articles that instruments were evaluated at the time of initial assessment or prosthesis fitting, 10 studies evaluated 12 instruments as initial assessment tools.

Eleven of the instruments have evidence of test validity from studies generalizable to the Medicare population. These include 1 Leg Standing Balance, 2MWT, AMPnoPRO, FAC, FAI, FIM, LEMOCOT, OPCS, PROS, SF, and TFP. For SF, more specifically, test validity has been found for SF-12 Physical Component Score, SF-12 Role Physical, SF-12 Bodily Pain, SF-36 Physical Functioning (where a modified 15-item version performed better than the original 10-item version). Three of the 11 instruments were also reported to have evidence of test reliability when evaluated at initial assessment: AMPnoPRO, TMP, and for SF-12 the subscales for Role Emotional, Role Physical, Bodily Pain, and Mental Health.

One instrument, LCI, was evaluated at initial assessment only in a study that is not generalizable to the Medicare population. Both the LCI-4 and LCI-5 versions of the instrument were found to have evidence of test validity, reliability, and responsiveness. Floor and ceiling percentages were reported for LCI-4, and no such effects were found.

Key Question 2. Prediction Tools

Based on reporting of metrics relevant to predictive validity, eight studies evaluated 13 instruments as prediction tools. However, all but one study reported only correlations of the instrument results with occurrence or test scores at a future time point. Thus, these are not true evaluations of the predictive accuracy of these instruments. Only one study reported on diagnostic test accuracy (sensitivity and specificity) for several instruments.

Twelve instruments have been reported to have predictive validity in whole or in part in studies that are generalizable to the Medicare population. These include the 1 Leg Standing Balance, 180 Degree Turn Test, 2MWT, AMPnoPRO, FAC, FAI, FIM, FSST, LCI-4 Advanced,

LEMOCOT, OPCS, and TUG. Two instruments were evaluated for predictive validity only in studies that were not generalizable to the Medicare population. Both AMPSIMM and LCI-5 were reported to be correlated with future functional status.

One study evaluated four of these instruments in a study deemed generalizable to the Medicare population for test accuracy to predict two or more falls during a 6-month followup period.[7] The Turn Time and Turn Test components of the 180 Degree Turn Test, FSST and TUG all had high sensitivity (85% to 100%) and specificity (74% to 93%) to predict falls. The Advanced components portion of LCI-4 had high specificity (91%) but low sensitivity (43%) to predict falls, which overall was reported to be statistically significant ($P<0.01$). The Turn Steadiness component of the 180 Degree Turn Test also had high sensitivity (85%) but low sensitivity (31%) to predict falls, but this test overall was not statistically significant ($P=0.22$).

Key Question 3. Functional Outcome Measurement Tools

All 50 evaluated instruments were deemed to be relevant functional outcome measurement tools. The findings are summarized in Tables A to D. In brief, 34 instruments (in whole or in part) had supporting evidence generalizable to the Medicare population, of which, in Table A, 17 instruments (or parts thereof) have evidence to support validity and reliability, and in Table B, 13 instruments have evidence of validity alone and 7 instruments have evidence of reliability alone. As noted in the tables, two of the instruments (PEQ and SF-12/36/36V), specific instrument items have supporting evidence for both validity and reliability, or for either validity or reliability alone. There are also 19 instruments (in whole or in part) that have supporting evidence only from studies not generalizable to the Medicare population. Of these, in Table C, 13 instruments (or parts thereof) have evidence to support validity and reliability, and in Table D, four instruments have evidence of validity only, three more have evidence of validity but explicitly not reliability, and 4 have evidence of reliability only. As noted in the tables, five of these instruments with evidence not generalizable to the Medicare population also have evidence for specific items that was generalizable to the Medicare population. Also, as noted in the tables, five other instruments have evidence for both validity and reliability for some subscales not only validity or reliability for others.

Table A. Instruments with evidence of both validity and reliability generalizable to the Medicare population

Rep*	Instrument	MCare	Validity	Reliability
	2MWT (2 Minute Walk Test)	Gen	Valid	Reliable
	6MWT (6 Minute Walk Test)	Gen	Valid	Reliable
	ABC (Activities-specific Balance Confidence)	Gen	Valid	Reliable
	AMP (Amputee Mobility Predictor)	Gen	Valid	Reliable
	Both AMPnoPRO (without prosthesis) and AMPPRO (with prosthesis)			
	Climbing Stairs Questionnaire	Gen	Valid	Reliable
	Functional Reach Test	Gen	Valid	Reliable
	Houghton Scale Both total Scale score and a subscale of items 1 to 3 (on prosthesis wear and use)	Gen	Valid	Reliable
1	LCI (Locomotor Capabilities Index)† Specifically: LCI-4 (LCI with a 4-point ordinal scale)	Gen	Valid	Reliable
2	PEQ (Prosthetic Evaluation Questionnaire)‡ Specifically, the PEQ-MS 13/11 (the Mobility Subscale with 13 items and 11 categories)	Gen	Valid	Reliable
	PPA (Prosthetic Profile of the Amputee)§ Specifically: Prosthesis use (outdoors), and Acceptance / Adaptation	Gen	Valid	Reliable
	Rising and Sitting Down Questionnaire	Gen	Valid	Reliable
	RMI (Rivermead Mobility Index)	Gen	Valid	Reliable
	SCS (Socket Comfort Score)	Gen	Valid	Reliable
3	SF-12 (Short Form Health Survey 12) Specifically: SF-12 PCS (Physical Component Score), SF 12 RP 2 (Role Physical), and SF 12 BP 2 (Bodily Pain)	Gen	Valid	Reliable
	TFP (Transfemoral Fitting Predictor)	Gen	Valid	Reliable
	TUG (Timed Up and Go)	Gen	Valid	Reliable
	Walking Questionnaire	Gen	Valid	Reliable

Abbreviations: Gen = generalizable to the Medicare population, MCare = Medicare (generalizability), Rep = repeated (see note for explanation).

Note: Instruments are organized across Tables A to D by whether the studies that evaluated them were generalizable to the Medicare (MCare) population (indicated by "Gen" in the MCare column) and by whether there was evidence of both validity and reliability, validity only, reliability only, and evidence of lack of reliability (indicated by "No" in the Reliability column of the relevant tables). Across tables, blank cells in the Validity or Reliability columns indicate that there was not evidence to support validity or reliability (not that the studies were found to be not valid or not reliable). In the Rep column, instruments that had parts that were variably generalizable, validated, or found to be reliable, and are thus repeated in different tables (or sections of tables), are noted with unique indicators (unique numbers) for each instrument. Instruments with blank cells in the Rep column are presented only in this table.

*Instruments that are included in multiple sections (repeated [Rep]) are indicated by unique numbers.

†LCI-4 (the total instrument) has been reported to be both valid and reliable in studies generalizable to the Medicare population. LCI-4 Basic, LCI-5, and LCI1-4 were not evaluated among studies generalizable to the Medicare population.

‡PEQ MS 13/11 has been reported to be both valid and reliable in studies generalizable to the Medicare population. The overall PEQ scale and each of the items, except shower and bathe safely (version with seven categories, 1 to 7) were reported to have reliability but were not evaluated for test validity in a study generalizable to the Medicare population.

§Also see listings for LCI, which is included in the PPA, but is evaluated separately in this table.

Table B. Instruments with evidence of either validity or reliability generalizable to the Medicare population

	Rep*	Instrument	MCare	Validity	Reliability
Instruments With Evidence of Validity (Only) Generalizable to the Medicare Population		1 Leg Standing Balance	Gen	Valid	
		180 Degree Turn Test Specifically: Turn Time and Turn Steps components	Gen	Valid	
		AAS (Amputee Activity Survey)	Gen	Valid	
	4	BBS (Berg Balance Scale)†	Gen	Valid	
		FAC (Functional Ambulation Categories)	Gen	Valid	
	5	FAI (Frenchay Activities Index)†	Gen	Valid	
		FIM (Functional Independence Measure), total score	Gen	Valid	
		FSST (Four Square Step Test)	Gen	Valid	
		LEMOCOT (Lower-Extremity Motor Coordination Test)	Gen	Valid	
		OPCS (Office of Population Censuses and Surveys Scale)	Gen	Valid	
		PROS (Prosthetist's Perception of Client's Ambulatory Abilities)	Gen	Valid	
	6	SIGAM (Special Interest Group of Amputation Medicine)†	Gen	Valid	
	3	SF-12 and SF-36 (Short Form Health Surveys 12 and 36) Specifically: SF-12 total score and SF-36 PF (Physical Functioning subscale, PF 15 performed better than PF-10)	Gen	Valid	
Instruments with Evidence of Reliability (Only) Generalizable to the Medicare Population		OPUS (Office of Population Censuses and Surveys Scale) Specifically: subscales Quality of Life, Lower Limb Function, and Satisfaction	Gen		Reliable
	2	PEQ (Prosthetic Evaluation Questionnaire)‡ Specifically: the overall scale and each of the items, including PEQ MS 13/7, except the items shower and bathe safely (version with 7 categories, 1 to 7)	Gen		Reliable
		PGI (Patient Generated Index)	Gen		Reliable
		PSFS (Patient-Specific Functional Scale)	Gen		Reliable
		SAT-PRO (Satisfaction with Prosthesis Questionnaire)	Gen		Reliable
	3	SF-36V (Short Form Health Survey 36 for use with veterans) Specifically: SF-36V subscales General Health, Physical Functioning, and Role Physical	Gen		Reliable
		Walking Speed, 10 meters	Gen		Reliable

Abbreviations: Gen = generalizable to the Medicare population, MCare = Medicare (generalizability), Rep = repeated (see note for explanation).

Note: Instruments are organized across Tables A to D by whether the studies that evaluated them were generalizable to the Medicare (MCare) population (indicated by "Gen" in the MCare column) and by whether there was evidence of both validity and reliability, validity only, reliability only, and evidence of lack of reliability (indicated by "No" in the Reliability column of the relevant tables). Across tables, blank cells in the Validity or Reliability columns indicate that there was not evidence to support validity or reliability (not that the studies were found to be not valid or not reliable). In the Rep column, instruments that had parts that were variably generalizable, validated, or found to be reliable, and are thus repeated in different tables (or sections of tables), are noted with unique indicators (unique numbers) for each instrument. Instruments with blank cells in the Rep column are presented only in this table.

*Instruments that are included in multiple sections (repeated [Rep]) are indicated by unique numbers.

†BBS, FAI, and SIGAM have evidence of validity among studies generalizable to the Medicare population, but evidence of both validity and reliability among studies not generalizable to the Medicare population.

‡PEQ MS 13/11 has been reported to be both valid and reliable in studies generalizable to the Medicare population. The overall PEQ scale and each of the items, except shower and bathe safely (version with seven categories, 1 to 7) were reported to have reliability but were not evaluated for test validity in a study generalizable to the Medicare population.

Table C. Instruments with evidence of both validity and reliability not generalizable to the Medicare population

Rep*	Instrument	MCare	Validity	Reliability
4	BBS (Berg Balance Scale)†	No	Valid	Reliable
5	FAI (Frenchay Activities Index)†	No	Valid	Reliable
	L Test (L Test of Functional Mobility)	No	Valid	Reliable
1	LCI (Locomotor Capabilities Index)‡ Specifically: LCI-4 Basic and Advanced (Basic and Advanced components, separately, with a 4 point ordinal scale), LCI 5 (LCI with a 5 point ordinal scale), and LCI10 4 (10-item scale which combined two of the response levels from LCI 5)	No	Valid	Reliable
7	Patient Activity Monitor Specifically: Walking Velocity	No	Valid	Reliable
2	PEQ (Prosthetic Evaluation Questionnaire)§ Specifically: the Function subscale Residual limb health, the Mobility subscale Ambulation, the Psychosocial subscales Frustration and Social Burden, and the Global subscale Well-Being; and PEQ MS 12/5 (the Mobility Subscale with 12 items and 5 categories)	No	Valid	Reliable
	PFI (Physical Function Index) Including the overall instrument and the four subscales Squat to Pick Up Object, Walk at Steady Pace, Run at Steady Pace, and Climb Stairs	No	Valid	Reliable
8	PLUS-M (Prosthetic Limb Users Survey of Mobility) Specifically: the form version SF-12#	No	Valid	Reliable
9	PROMIS-29 (Patient-Reported Outcomes Measurement Information System 29-Item Profile) Specifically: the Physical Function subscale	No	Valid	Reliable
10	Q-TFA (Questionnaire for Persons with a Transfemoral Amputation) Specifically: the subscales Prosthetic Use, Prosthetic Mobility, and Problem	No	Valid	Reliable
6	SIGAM (Special Interest Group of Amputation Medicine)†	No	Valid	Reliable
	SIP-PD (Sickness Impact Profile-Physical Dimension) Including the overall instrument and the three subscales Ambulation, Body Care and Movement, and Mobility	No	Valid	Reliable
11	TAPES (Trinity Amputation and Prosthesis Experience Scales) All subscales of TAPES and TAPES-R except Weight Satisfaction (from the original TAPES) and Satisfaction with Prosthesis Subscale 1 (esthetics, from TAPES-R)	No	Valid	Reliable

Abbreviations: Gen = generalizable to the Medicare population, MCare = Medicare (generalizability), Rep = repeated (see note for explanation).

Note: Instruments are organized across Tables A to D by whether the studies that evaluated them were generalizable to the Medicare (MCare) population (indicated by "Gen" in the MCare column) and by whether there was evidence of both validity and reliability, validity only, reliability only, and evidence of lack of reliability (indicated by "No" in the Reliability column of the relevant tables). Across tables, blank cells in the Validity or Reliability columns indicate that there was not evidence to support validity or reliability (not that the studies were found to be not valid or not reliable). In the Rep column, instruments that had parts that were variably generalizable, validated, or found to be reliable, and are thus repeated in different tables (or sections of tables), are noted with unique indicators (unique numbers) for each instrument. Instruments with blank cells in the Rep column are presented only in this table.

*Instruments that are included in multiple sections (repeated [Rep]) are indicated by unique numbers.

†BBS, FAI, and SIGAM have evidence of validity among studies generalizable to the Medicare population, but evidence of both validity and reliability among studies not generalizable to the Medicare population.

‡LCI-4 (the total instrument) has been reported to be both valid and reliable in studies generalizable to the Medicare population. LCI-4 Basic, LCI-5, and LCI1-4 were not evaluated among studies generalizable to the Medicare population.

§PEQ MS 13/11 has been reported to be both valid and reliable in studies generalizable to the Medicare population. The overall PEQ scale and each of the items, except shower and bathe safely (version with seven categories, 1 to 7) were reported to have reliability but were not evaluated for test validity in a study generalizable to the Medicare population.

#The form version SF-12 (not to be confused with the Short Form Health Survey SF-12) has evidence for both test validity and reliability.

Table D. Instruments with evidence of either validity or reliability not generalizable to the Medicare population

	Rep*	Instrument	MCare	Validity	Reliability
Instruments With Evidence of Validity (Only) Not Generalizable to the Medicare Population		AMPSIMM (Amputee Single Item Mobility Measure)	No	Valid	
		Employment Questionnaire	No	Valid	
		TWT (Timed Walking Test)	No	Valid	
		WHOQOL-BREF (World Health Organization Quality of Life-Brief Version) Specifically: the Physical Health, Psychological Health, Social Relations, and Environmental subscales	No	Valid	
Instruments With Evidence of Validity But Not Reliability Not Generalizable to the Medicare Population	7	Patient Activity Monitor Specifically: Step Count and Step Length	No	Valid	No
	10	Q-TFA (Questionnaire for Persons with a Transfemoral Amputation) Specifically: Global Health subscale	No	Valid	No
	11	TAPES (Trinity Amputation and Prosthesis Experience Scales) Specifically: the Satisfaction with Prosthesis Subscale 1 (esthetics) from TAPES R	No	Valid	No
Instruments With Evidence of Reliability (Only) Not Generalizable to the Medicare Population		ADAPT (Assessment of Daily Activity Performance in Transfemoral Amputees) Specifically: Items 10 to 18; items 1 to 9 were not evaluated	No		Reliable
		NQ-ACGC (Quality of Life in Neurological Conditions – Applied Cognition / General Concerns)	No		Reliable
	8	PLUS-M (Prosthetic Limb Users Survey of Mobility) Specifically: the form versions CAT (Computer Adaptive Test) and SF-7 (a short form version)	No		Reliable
	9	PROMIS-29 (Patient-Reported Outcomes Measurement Information System 29-Item Profile) Specifically: the Anxiety, Depression, Fatigue, Pain Intensity, Pain Interference, Physical Function, Sleep Disturbance, and Social Role Satisfaction subscales	No		Reliable

Abbreviations: Gen = generalizable to the Medicare population, MCare = Medicare (generalizability), Rep = repeated (see note for explanation).

Note: Instruments are organized across Tables A to D by whether the studies that evaluated them were generalizable to the Medicare (MCare) population (indicated by "Gen" in the MCare column) and by whether there was evidence of both validity and reliability, validity only, reliability only, and evidence of lack of reliability (indicated by "No" in the Reliability column of the relevant tables). Across tables, blank cells in the Validity or Reliability columns indicate that there was not evidence to support validity or reliability (not that the studies were found to be not valid or not reliable). In the Rep column, instruments that had parts that were variably generalizable, validated, or found to be reliable, and are thus repeated in different tables (or sections of tables), are noted with unique indicators (unique numbers) for each instrument. Instruments with blank cells in the Rep column are presented only in this table.

*Instruments that are included in multiple sections (repeated [Rep]) are indicated by unique numbers.

Key Question 4. LLP Comparative Effectiveness by Subgroup

It should be noted that this review makes no attempt to make conclusions about the overall effects of different LLP components or configurations. Key Question 4 addressed whether there is evidence regarding heterogeneity of treatment effects (whether outcomes with specific devices vary across individuals based on different characteristics such as age or health status) in the field of LLP research and whether studies used validated measures.

A relatively small percentage of comparative studies report sufficient data to allow subgroup analysis and evaluation of heterogeneity of treatment effect (14%, 15 of 104 otherwise eligible articles). These studies either address or provide sufficient data that allowed us to address the focused question of whether the relative effect of different components or configurations differs across different subgroups of lower limb amputees.

Twelve of the 14 studies included between 5 and 168 users of LLPs, one included 899 amputees, and one 1013. Seven studies evaluated microprocessor knees (compared to mechanical knees), two evaluated other knee components, three evaluated ankle/foot components, and one each evaluated pylons or sockets. One large study developed a regression model to evaluate the predictive ability of a wide range of participant characteristics. Another study (Hahn 2015) conducted correlation and regression analyses but did not fully report the results of these analyses. Overall, studies that investigated subgroup effects did not identify participant characteristics that predict which lower limb amputees would benefit most or least from any given component (low strength of evidence).

Of the 13 studies, only seven used validated predictor and outcome measures. Only one of the eligible studies was a randomized trial, but it evaluated atypical, nonvalidated predictor variables (subgroups of the Medicare Functional Classification Level K2). Only two studies explicitly evaluated heterogeneity of treatment effect; others reported individual participant level data without conducting their own subgroup analyses. Across studies, a scattering of statistically significant differences in relative effects of different components were found based on different subgroup comparisons. However, these findings were not consistent across, and often within, studies. Only one study, which compared a specific microprocessor knee (Genium™) to any prior used knee (mostly another microprocessor knee, C-Leg™), analyzed the most important aspect of the KQ, namely whether any study participant characteristics (or set of characteristics) could accurately and effectively predict which patients would benefit most or least from a given component. However, there were methodological and analytic concerns with this study. Despite finding numerous statistically significant associations between participant characteristics and functional outcomes, the study concluded that no model accurately predicted relative effect (between the Genium microprocessor knee and, mostly, the C-Leg microprocessor knee).

Overall, studies that investigated subgroup effects did not identify participant characteristics that predict which lower limb amputees would benefit most or least from a given LLP component or configuration. Based on the methodology used to assess strength of evidence, the studies warrant a low strength of evidence that patient characteristics evaluated in the studies do not predict which patients would benefit most or least from a given LLP component or configuration (Table E). Although one large study attempted to develop a model to predict success with microprocessor knees, the study did not use a validated outcome and had several methodological and analytic issues. It, therefore, provided insufficient additional evidence regarding who would be more likely or less likely to benefit from a microprocessor knee. An additional issue across almost all studies was that study participants were in general not likely to be representative of the Medicare population, being both mostly young and with amputations due to trauma, with relatively few people with dysvascular disease.

Key Question 5. Expectations of Ambulation

We found no study that addressed this key question.

Key Question 6. Patient Satisfaction With Process

Two studies addressed this Key Question. Note that this Key Question did not address satisfaction with the LLP itself. Studies addressing satisfaction with LLPs (or function with the prosthesis) would have been eligible for Key Question 4 if they reported subgroup analyses. One study surveyed individuals about satisfaction with upper or lower prosthetic limbs and related services. The second study, designed to assess the reliability and construct validity of the Orthotics and Prosthetics National Office Outcomes Tool in clients with LLPs, reported data about satisfaction with the prosthetist appointments.

A moderate risk of bias study (of generally younger adults about one-third of whom had dysvascular disease) found that at least three-quarters of people receiving an LLP were

satisfied with the process of accessing their LLP and a high risk of bias study (in which about half had Medicare or Medicaid insurance) found that on average clients were satisfied with their visits to their prosthetists' offices (average score about 83 of 100). Together, the studies provide low strength evidence that people are satisfied with their encounters with their prosthetists (Table F).

Table E. Key Question 4 evidence profile

Outcome	No. Studies (N)	Study Limitations	Consistency	Precision	Reporting Bias	Directness*	Other Issues	Findings	SoE Grade
Validated predictors and outcomes (univariable)	8 (1096, 1013 in 1 study)	Medium†	Consistent	Imprecise	Undetected	Indirect‡	High degree of multiple testing; mostly evaluations of knee components; mostly K2 or K3 level, unilateral transfemoral amputations due to traumatic etiologies	Some trends of differences in relative effect based on participant characteristics, however, none statistically significant after correcting for multiple testing	Low
All outcomes (univariable)	13 (1328, 1013 in one study)	Medium†	Consistent	Imprecise	Undetected	Indirect‡	Nonvalidated outcomes, high degree of multiple testing; mostly K2 to K4 level, unilateral transfemoral amputations due to traumatic etiologies	Some trends of differences in relative effect based on participant characteristics, however, none statistically significant after correcting for multiple testing	Low
Ambulatory and functional outcomes, nonvalidated (multivariable model)	1 (899)	High§	NA	Precise	Undetected	Indirect#	K2 to K4 (mostly K3) level, mostly traumatic etiologies. Study does not directly address Key Question.	A large set of variables individually were associated with better outcomes with the microprocessor knee. No model predicted who would most benefit from knee.	Insufficient

Abbreviations: NA = not applicable, RoB = risk of bias, SoE = strength of evidence.

*Representative of either (or both) older adults (≥65 years old) or those with dysvascular amputations.

†Nonrandomized studies, univariable analyses (mostly individual participant data reports), generally lack of evaluation of heterogeneity of treatment effect, mostly small studies.

‡Both relatively young age amputees and primarily people with amputations due to trauma in most studies. Almost all (that reported) had unilateral transfemoral amputations.

§Nonrandomized, likely biased sample of participants, nonvalidated outcomes, unclear which outcome(s) used in final models. See text.

#Highly selected participants who had been assessed as likely to benefit from a microprocessor knee, possibly biased dropouts, relatively young and two-thirds had trauma etiology.

Table F. Key Questions 5 and 6 evidence profile

Outcome	No. Studies (N)	Study Limitations	Consistency	Precision	Reporting Bias	Directness*	Other Issues	Findings	SoE Grade
Alignment of outcomes with expectations (KQ 5)	0	NA	NA	NA	NA	NA	NA	None	Insufficient
Satisfaction with process (KQ 6)	2 (~1663)	Medium	Consistent	Precise	Undetected	Direct †	Nonvalidated outcomes	Clients generally satisfied with their encounters with their prosthetists	Low

Abbreviations: KQ = Key Question, NA = not applicable, SoE = strength of evidence.

*Representative of either (or both) older adults (≥65 years old) or those with dysvascular amputations.

†One study included a wide range of prosthetics practices; about half the participants had Medicare or Medicaid as a primary payer. The other study was less representative.

Key Question 7. Long-Term Outcomes

We found eight studies with at least 100 participants who were followed for at least 6 months after prescription of an LLP. The studies analyzed data from 109 to 555 participants followed for 1 to 7 years (except for two studies that implied long-term followup, but did not report a timeframe). The studies only sparsely covered the subquestions pertaining to specific outcomes, particularly related to questions about different outcomes in different subgroups of amputees. Studies did not explicitly account for intervening mortality or subsequent surgeries or injuries.

Table G summarizes the strength of evidence for each outcome and subgroup analysis with data. For all outcomes of interest, there is low or insufficient strength of evidence because evidence is sparse, most studies were conducted in the 1990s or earlier, and only one of the studies was conducted in the United States, with its unique healthcare system and standards for prosthesis prescription. Also, most studies had methodological limitations, most populations analyzed were not directly applicable to the Medicare population, and some study findings were inconsistent with each other. Subgroup analyses in single studies tended to be underpowered to detect differences, mostly leading to determinations that the evidence was insufficient.

We found a low strength of evidence, based on six studies, that about 11 to 22 percent of lower limb amputees who receive an 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. However, only one of the studies was conducted in the United States and it used hospital data as of 1998; most other studies were also old. 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.

Based primarily on two generally representative studies, there is low strength of evidence that 24 to 29 percent of LLP recipients use their prostheses only indoors at 1 year. 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. There is insufficient evidence about the rates of failure to maintain bipedal ambulation (1 study, 7% at 7 years), use of prostheses only for transfer (1 study, 4% at 1 year), and why people abandon their prostheses. No study reported on “major problems” with prostheses.

Table G. Key Question 7 evidence profile

Outcome	Subgroup	No. Studies (N)	Study Limitations	Consistency	Precision	Reporting Bias	Directness*	Other Issues	Findings	SoE Grade
Failure to maintain bipedal ambulation	All participants	1 (148)	High	NA	Precise	Undetected	Indirect	Unclear outcome, old study	7% at 7 years	Insufficient
Use of prosthesis only for transfers	All participants	1 (196)	High	NA	Precise	Undetected	Indirect	Single 25 year old study	4% at 1 year	Insufficient
	TF vs. TT	1 (196)	High	NA	Precise	Undetected	Indirect	see above	No significant difference	Insufficient
	Age	1 (196)	High	NA	Precise	Undetected	Indirect	see above	Nonsignificantly higher limited used with older age	Insufficient
Use of prosthesis only indoors	All participants	4 (1040)	Medium	Inconsistent	Imprecise	Undetected	Direct	Mostly old, non-U.S.	24-29% at 1 year	Low
	TF vs. TT	2 (337)	High	Inconsistent	Precise	Undetected	Direct	see above	Twice as many TF use only indoors (1 study, P=0.008)), no difference (1 study)	Insufficient
	Age	1 (196)	High	NA	Precise	Undetected	Direct	see above	Older more likely to use only indoors (P=0.042)	Insufficient
	Bilateral vs. unilateral	1 (141)	High	NA	Precise	Undetected	Direct	see above	Bilateral more than twice as likely to use only indoors (P=0.0006)	Insufficient
Abandonment of prosthesis	All participants	6 (1153)	Medium	Consistent †	Precise	Undetected	Direct	Mostly old, non-U.S.	11-22% at 1 year (or undefined)†	Low
	TF vs. TT	3 (538)	High	Consistent	Precise	Undetected	Direct	see above	TF more likely to abandon prosthesis than TT	Low

Table G. Key Question 7 evidence profile (continued)

Outcome	Subgroup	No. Studies (N)	Study Limitations	Consistency	Precision	Reporting Bias	Directness*	Other Issues	Findings	SoE Grade
	Bilateral vs. unilateral	3 (452)	High	Inconsistent	Precise	Undetected	Direct	see above	Nonsignificant, but conflicting directionality	Insufficient
	Age	2 (397)	High	Inconsistent	Precise	Undetected	Direct	see above	Older nonsignificantly more likely to abandon (1 study), no difference in age (1 study)	Insufficient
	Multiple	1 (201)	High	NA	Precise	Undetected	Indirect	Multiple testing	No significant associations	Insufficient
Major problems with prosthesis	All participants	0	NA	NA	NA	NA	NA	NA	None	Insufficient
Reasons for poor outcomes	All participants	1 (201)	High	NA	Imprecise	Undetected	Indirect	Single non-U.S. study	Various general categories of reasons reported	Insufficient

Abbreviations: NA = not applicable, RoB = risk of bias, SoE = strength of evidence, TF = transfemoral amputation, TT = transtibial amputation.

*Applicability to the Medicare population (based on mean age and percent with dysvascular amputations).

†Except that one outlier study from Taiwan found that only 0.9% of study participants abandoned their prostheses at a mean of 28 months.

Discussion

A large number of studies have evaluated LLP for people with major lower limb amputations. We found over 100 studies that compared at least two LLP components or configurations that reported ambulatory, functional, or other patient-centered outcomes. We found many additional studies that evaluated only biomechanical properties of the components (which this review does not evaluate) and likely several hundred studies that evaluate just a single component. However, we found few studies that evaluated (or at least provided data to allow us to evaluate) heterogeneity of treatment effect. Overall, the evidence is currently sparse and fails to adequately address whether different subgroups of amputees are more likely or less likely to benefit from specific LLP components or configurations. We also found generally sparse evidence regarding patient expectations, patient satisfaction with care, and long-term outcomes.

From the amputee's and the clinician's perspective, among the most important questions is which LLP configuration (comprised of which prosthetic components) would best enable maximal health, function, and quality of life for a given individual. Given the large number of component types (knee, foot/ankle, socket, liner, etc.) and the range of features for each of these, the process of determining which LLP component or configuration is best for individuals is quite complex. However, the majority of the evidence addresses the question of which LLP component or configuration maximizes ambulation and function in the average patient, as opposed to which LLP component or configuration would best suit the needs of a given individual. In other words, few studies address the issue of heterogeneity of treatment effect. 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 prosthesis, and unnecessarily increase health care expenditures.

We found evidence to enable the evaluation of the psychometric properties of 50 instruments (many containing evaluated subscales and items) in people with lower limb amputations. Many of the studies that evaluated instrument psychometric properties, however, were conducted in samples of participants who were arguably different than typical lower limb amputees with Medicare insurance, many of whom have dysvascular conditions including diabetes and peripheral vascular disease, or who are older and are, thus, more typical of lower limb amputees with Medicare insurance. We found that 39 of the 50 instruments have been evaluated in studies deemed generalizable to the Medicare population. Seventeen of these instruments were found, as a whole or in part, to have evidence supporting both reliability and validity. However, we recommend that researchers who are using this report to determine which instruments to use for their own studies also review the primary studies to determine whether the instruments have been sufficiently validated for their needs, are responsive to clinically important change, and have been evaluated in a sample of people representative of their study population.

Notably, no study has evaluated psychometric properties of the Medicare Functional Classification Level (MFCL or K level) system. Furthermore, the reader is reminded that lack of evidence regarding the psychometric properties of instruments does not imply that these measures 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 earlier studies, may not have psychometric property evaluations more commonly reported now.

Nevertheless, we strongly encourage future researchers to maximize the use of instruments with evidence of validity and reliability in the population of interest. Where such measures are lacking, the validity of the instruments being used as pivotal outcomes should be examined before use in future studies. We also encourage journal editors to require use of validated and reliable instruments when appropriate and feasible. However, we recognize that it will remain common that unvalidated measures may be appropriate in select instances (e.g., when measures to assess a particular trait or construct do not exist).

Evidence Limitations

Despite the large literature base for research on LLP, 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.

The applicability of these studies to the general population of people with LLPs may be somewhat limited, as the studies mostly evaluated prosthetic knees and were mostly conducted in younger men with unilateral transfemoral amputations due to trauma. Furthermore, implicitly or explicitly, most of these studies included only people who were deemed (by their prosthetists) to be likely to benefit from their new (generally more complex) device. 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. Studies evaluating heterogeneity of treatment effect also evaluated a limited set of patient characteristics such as age, amputation level, or amputation etiology. None analyzed differences in treatment effect by subgroups based on any assessment techniques, prediction tools, or outcome measures. Eligible studies reporting long-term LLP use after prescription were almost all conducted outside the United States and were mostly more than a decade old. Additional evidence limitations are discussed in the full report.

Analysis Limitations

Assessment of reliability, validity, and other psychometric properties is open to interpretation. By the strictest definition, an instrument would be considered to be valid and appropriate for use in a given study only if there is good evidence regarding the multiple aspects of validity for the specific population, conditions, and outcomes under evaluation. That an instrument demonstrates convergent validity with a given related measure does not imply that it also can distinguish differences related to subgroups of patients or an intervention effect. That an instrument has predictive validity regarding one outcome, such as future successful use of an LLP, does not imply predictive validity for other ambulatory outcomes, such as speed of walking or community ambulation. Despite these challenges, and the lack of a universal gold standard for determining absolute validity, we took a liberal approach in our literature synthesis. We considered an instrument to have evidence of validity if there was evidence of any type of validity (other than face/content). We, thus, categorized the evidence and dichotomized data so that instruments were classified as valid or not. It is incumbent on each study's researchers to determine whether given instruments and measures have sufficient evidence of validity and are appropriate for their study purposes. Additional evidence limitations are discussed in the full report.

Future Research Recommendations

Future research is needed to adequately address most of the questions in this review. While numerous instruments have evidence of validity, at least in part, additional studies are needed to confirm their psychometric properties and to better understand specific aspects of validity. Well-conducted studies, using validated predictors and outcomes, are needed to evaluate which devices would be most effective to achieve successful outcomes for which patients. To as great an extent as possible, studies should assess validated, patient-centered outcomes related to ambulation, function, quality of life, and related outcomes. Continued use of ad hoc and nonvalidated measures greatly limits the interpretability, usability, representativeness, and overall value of the studies. Ideally, studies should use a core set of validated, patient-centered outcomes that incorporate the perspectives of patient and other key stakeholders (a core outcome set); in addition, studies may measure other specific outcomes, as needed. This would allow comparability across studies and pooling of study findings (e.g., meta-analysis). Creation of such a core outcome set would likely require a consensus development process among a range of stakeholders. More specific recommendations for studies of heterogeneity of treatment effect and studies on expectations, satisfaction with services, and long-term followup are provided in the full report.

Conclusions and Clinical Implications

Numerous instruments that assess ambulation, function, quality of life, and other patient-centered outcomes exist for people with lower limb amputations and LLPs. Researchers should minimize the use of nonvalidated or ad hoc measures. Those who wish to use new or previously unvalidated instruments should validate these measures before using them. Researchers with an interest in assessing LLPs for the Medicare population would be best served to focus on those instruments with evidence of reliability and validity for this population or validate the measures in this population. The majority of the evidence on LLPs addresses the question of which LLP component or configuration maximizes ambulation and function in the average patient, as opposed to which LLP would best suit the needs of a given individual. In other words, few studies address the issue of heterogeneity of treatment effect. A small evidence base does not provide data to guide LLP selection for a specific patient to maximize their ambulation, function, and quality of life or to minimize abandonment or limited use. However, this does not imply that the evidence suggests patient characteristics cannot effectively predict which patients would benefit most or least from one or another specific component; only that the current evidence does not support use of any given predictor. There is low strength of evidence that patients are generally satisfied with the prosthetic services they receive. Further high-quality research is needed to better assess the psychometric properties of instruments (whether assessment techniques, prediction tools, or outcome measures) and to answer the Key Questions addressed in this systematic review.

References

1. Boyle JP, Thompson TJ, Gregg EW, et al. Projection of the year 2050 burden of diabetes in the US adult population: dynamic modeling of incidence, mortality, and prediabetes prevalence. Popul Health Metr. 2010 Oct 22;8:29. doi: 10.1186/1478-7954-8-29. PMID: 20969750.
2. Ziegler-Graham K, MacKenzie EJ, Ephraim PL, et al. Estimating the prevalence of limb loss in the United States: 2005 to 2050. Arch Phys Med Rehabil. 2008 Mar;89(3):422-9. doi: 10.1016/j.apmr.2007.11.005. PMID: 18295618.
3. Etter K, Borgia M, Resnik L. Prescription and repair rates of prosthetic limbs in the VA healthcare system: implications for national prosthetic parity. Disabil Rehabil Assist Technol. 2014 May 22:1-8. doi: 10.3109/17483107.2014.921246. PMID: 24852068.
4. Resnik L, Borgia M. Predicting prosthetic prescription after major lower-limb amputation. J Rehabil Res Dev. 2015;52(6):641-52. doi: 10.1682/jrrd.2014.09.0216. PMID: 26562228.
5. Borrenpohl D, Kaluf B, 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 Jul;97(7):1053-63. doi: 10.1016/j.apmr.2016.02.024. PMID: 27016261.
6. AHRQ Methods for Effective Health Care. Methods Guide for Effectiveness and Comparative Effectiveness Reviews. Rockville (MD): Agency for Healthcare Research and Quality (US); 2011.
7. Dite W, Connor HJ, Curtis HC. Clinical identification of multiple fall risk early after unilateral transtibial amputation. Arch Phys Med Rehabil. 2007 Jan;88(1):109-14. doi: 10.1016/j.apmr.2006.10.015. PMID: 17207685.

Full Report

This evidence summary is part of the following document: 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.

AHRQ Pub. No.18-EHC017-1-EF
September 2018
www.ahrq.gov