The LEADr & COMPASSr:

Development, Evidence, and Readiness for Implementation

International Society for Prosthetics and Orthotics

Learn. Exchange. Support. Innovate.

International Society for Prosthetics and Orthotics

February 2026

Project Funded by ATscale

EXECUTIVE SUMMARY

The LEAD (Lower Extremity Amputation Dataset) and COMPASS (Consensus Outcome Measures for Prosthetic and Amputation Services) were developed by the International Society for Prosthetics and ORthotics (ISPO) to enable standardised outcome data collection for prosthetic rehabilitation services in low- and middle-income countries (LMICs). The need for such data is well recognised. It is the absence of outcome evidence that underlies chronic underfunding of prosthetic rehabilitation in LMICs. Yet the tools had not been tested in the clinical environments where they would need to work. This report presents the findings of a pilot study conducted across three sites in the Asia Pacific region between August and December 2025, involving 121 people with lower limb absence and clinical staff across diverse service delivery settings.

The pilot confirmed that standardised outcome measurement is valued, feasible, and generates meaningful data. Satisfaction was high: 95.5% of participants rated the appropriateness of questions at 4 or 5 out of 5. The performance-based measures were enthusiastically received, with staff at several sites already incorporating elements into routine practice before the pilot concluded. The pilot also revealed that users of prostheses had rarely or never been asked about their experience of living with limb absence or a prosthesis. The act of being asked, and having their perspective formally recorded, held value for participants well beyond the data it produced.

The pilot also delivered unexpected benefits. Staff described the experience as professionally transformative, learning new approaches to assessment that they had not been taught or practised. At one site, the research raised the profile of the prosthetic and orthotic profession within the local health system, giving local services a visibility and legitimacy that routine clinical work alone had not achieved. All three sites participated without financial incentives, a deliberate design choice that demonstrated genuine engagement and supports the case for sustainability.

At the same time, the pilot identified clear areas requiring change. Administration of the full LEAD and COMPASS took one to three hours per participant, far exceeding the 30 to 40 minutes considered acceptable for routine use. The absence of validated local language versions meant every question required real-time translation, which was the single biggest contributor to time burden and data inconsistency. Content redundancy between outcome measures, particularly the TAPES-R and PEQ, caused frustration for both users and clinicians.

These findings have directly informed a revised tool. The LEADr and COMPASSr retains 23 data items and three core outcome measures (TUG, PEQ-UT, PEQ-RL), with the PSFS and EQ-5D-5L available as optional measures. By excluding the AMP, 2MWT, and TAPES-R, the core COMPASSr outcome measures have a combined administration time of under 14 minutes, reduced from over 70 minutes for the full original COMPASS. Technical prosthetic details have been separated for clinician completion outside of the appointment, further reducing the direct contact time. Two data items, ability to work and social and community support, have been retained with revised question formats to address comprehension difficulties identified during the pilot.

The LEADr and COMPASSr are ready for staged implementation. The next phase will focus on validated local language translations, application development for broader device compatibility, user self-completion of the self-reported components, and site preparation including training, organisational integration, and local ethics submissions. At least twelve ISPO member societies have expressed early interest in adoption. With these steps in place, the LEADr and COMPASSr have the potential to provide the standardised outcome evidence that prosthetic services in LMICs need to demonstrate their value, secure sustainable funding, and improve the quality of care for people with limb absence.

TABLE OF CONTENTS

1. INTRODUCTION: The LEAD and COMPASS and the need for pilot testing

1.1 Why does outcome data matter for prosthetic services?

Prosthetic and rehabilitation services for people with lower limb absence remain critically underfunded in low- and middle-income countries (LMICs), despite the significant impact these services have on individual function, participation in society, and economic productivity. A fundamental reason for this underfunding is the absence of evidence; not evidence that the services are provided, but evidence that they make a difference.[1]

Currently, most data collection in LMIC prosthetic services focuses primarily on outputs: the number of prostheses manufactured and the number of times people attend at the clinic to be seen. What is largely missing is outcome data: evidence of functional improvement, enhanced quality of life, or return to productive activities. Without this, service providers are unable to demonstrate the tangible benefits of their work and policymakers lack the evidence base needed to prioritise prosthetic services within competing health sector demands.[1]

The need for outcome-focused data collection has been well recognised within the international prosthetics and orthotics community, yet significant barriers have prevented its implementation in LMICs. These include the lack of consensus on which outcomes to measure, concerns about the time burden of comprehensive assessments, uncertainty about the cultural appropriateness of existing tools, and questions about whether routine data collection is feasible in resource-constrained clinical environments. The vision, a standardised dataset collected across services and countries, enabling the field to answer questions that have so far been unanswerable, has remained an aspiration.

1.2 What is the LEAD and COMPASS?

The LEAD (Lower Extremity Amputation Dataset) and COMPASS (Consensus Outcome Measures for Prosthetic and Amputation Services) were developed to turn the aspiration of standardised outcome data collection into something practical: a framework that services in LMICs could realistically use as part of routine clinical care. The LEAD is a standardised dataset developed between 2020 and 2021, designed to collect consistent, comparable data about people with lower limb absence and the services they receive across prosthetic and rehabilitation settings. Embedded within the LEAD is the COMPASS: a set of outcome measures that capture the difference those services make. While technically a single integrated framework, both names are commonly used to describe the project, reflecting the two functions the tool serves: collecting data about people and services (LEAD) and measuring outcomes (COMPASS). Throughout this report, "LEAD and COMPASS" is used to refer to the complete framework.

The LEAD contains 25 data items covering demographics, details about the amputation, rehabilitation interventions provided, and health status. The COMPASS comprises three patient-reported outcome measures and three performance-based outcome measures, supplemented by

measures of individual functional goals and general health-related quality of life. The idea is straightforward: if services in different countries collect the same data in the same way, the field builds an evidence base that can answer questions about the value and impact of prosthetic services that have so far been unanswerable.

1.3 How were the LEAD and COMPASS developed?

The LEAD and COMPASS were developed by the International Society for Prosthetics and Orthotics (ISPO), with support from USAID, UNOPS, and ATscale, between 2020 and 2021 through an international consensus process specifically designed to ensure relevance and feasibility in LMIC contexts.[1]

The development process was deliberately broad, involving people with limb absence, clinicians, policymakers, managers, representatives from industry, researchers and staff from international humanitarian organisations. It considered measurement properties as well as practical factors including ease of administration, appropriateness across different cultural contexts, and relevance to stakeholders' decision-making needs.

The COMPASS outcome measures were identified through a systematic review of outcome measures relevant to people with lower limb absence,[2] followed by expert panel evaluation of psychometric evidence.[3] The LEAD dataset was developed through a parallel process: a scoping review of existing registries and databases, and semi-structured interviews with those who had managed similar data collection systems. Both were then refined through similarly structured international consensus groups.[4]

Throughout this process, there was an acknowledged tension between comprehensiveness and practicality. Participants recognised that while detailed data collection would be valuable, an overly burdensome assessment would not be sustainable in routine clinical practice. The consensus group worked to identify what they considered a realistic time burden while simultaneously including data items and outcome measures deemed essential for demonstrating service value.[1] However, this balance was based on expert opinion and theoretical considerations rather than empirical evidence from real-world implementation. The tools had been developed, published in peer-reviewed journals, and supported by a comprehensive implementation guide, but they had not yet been tested in the clinical environments where they would need to work.[5]

1.4 Why did the LEAD and COMPASS need to be piloted?

The LEAD and COMPASS were developed through a rigorous consensus process, but the balance between comprehensiveness and practicality was based on expert opinion and theoretical considerations rather than empirical evidence from real-world implementation. Without testing, there was a risk that the tools would prove too burdensome for routine use in resource-constrained settings, or that items would be poorly understood across different linguistic and cultural contexts. Implementing untested tools at scale would risk wasting the time of clinicians and prostheses users, and a failed rollout could undermine confidence in outcome data collection more broadly. Best practice in registry development also emphasises the importance of pilot

testing in authentic clinical environments before proceeding to full implementation (Gliklich et al., 2020). The pilot therefore needed to answer several key questions about the practicality, appropriateness, and time burden of the LEAD and COMPASS in routine clinical care.

1.5 What questions did the pilot set out to answer?

The pilot was designed to address five key research questions. The first two relate to the content of the tools themselves: **(1) Are the contents of the LEAD and COMPASS culturally appropriate?** Specifically, are participants in diverse LMIC settings willing to provide the data requested? And **(2) are the contents linguistically appropriate and understandable?** Can the questions and assessment tasks be effectively translated and comprehended by participants and clinicians in different linguistic and cultural contexts? These questions are core to the success of a data collection process, because if people are unwilling or unable to engage with the tools, the data collection will not succeed.

The third question addresses what was a major element of the consensus process**: (3) What is the actual time burden of the LEAD and COMPASS and its component parts?** How does the real-world time required for data collection compare with the consensus group's estimates of acceptable time burden?

The fourth question explores the experiences of those involved: **(4) What are participants' reactions to the LEAD and COMPASS?** How do people with limb absence, clinicians administering the assessments, and centre managers view the practicality, appropriateness, and time burden of the data collection?

Finally, **(5) which data items in the LEAD and COMPASS are redundant or repetitive?** Are there items that provide overlapping information or items that could be removed without significant loss of meaningful data? This question directly addresses the tension between comprehensiveness and practicality that was acknowledged but unresolved during the development process.

By addressing these questions through pilot testing, the pilot aimed to achieve two goals: to produce a refined version of the LEAD and COMPASS that has been streamlined based on real-world evidence, and to establish confidence that the revised framework is something prosthetic services would realistically adopt as part of routine clinical care.

1.6 How did the pilot answer these questions?

These questions could not be answered from a single perspective. Understanding whether the LEAD and COMPASS would work in routine practice required an examination of the experiences of everyone involved — the people with limb absence who completed the assessments, the clinicians who administered them, the managers responsible for their services, and the implementation team who trained and supported data collection.

The pilot was organised into several components to explore the different participant views: quantitative analysis of response rates and score distributions from the LEAD and COMPASS

data itself; a structured feedback survey completed by prosthesis users after their assessments; focus groups with clinicians and separately with managers across all three sites; and the reflections and observations of the implementation team throughout the process. Each of these contributes a different perspective on the same core questions, and together they provide both the formal evidence base and the practical context needed to refine the LEAD and COMPASS for routine use.

The following section describes the methodology of the pilot in detail. The main findings are then presented thematically, organised around the research questions, drawing on all of these perspectives.

2. METHODOLOGY: How the pilot was conducted

2.1 Study Design & Sites

This pilot employed a descriptive, mixed-methods design to test the LEAD and COMPASS in routine clinical settings. Data collection occurred between August 2025 and December 2025 across three countries representing different economic contexts, health systems, and service delivery models - from a national prosthetic service operating across geographically remote regions in Papua New Guinea, to a provincial rehabilitation centre in Cambodia, to an established academic clinical centre in Thailand.

The three participating sites were:

Papua New Guinea: Data were collected across three branches of the National Orthotics and Prosthetics Services (NOPS), Health Facilities Standards Branch, National Department of Health:

- Port Moresby General Hospital, National Capital District, Southern Region
- Western Highlands Provincial Health Authority, Mt Hagen, Western Highlands Province
- Morobe Provincial Health Authority, Lae, Morobe Province, Momase Region

Together, these sites span the country's southern coastal, highlands, and northern coastal regions, providing prosthetic and orthotic services to geographically and culturally diverse populations across Papua New Guinea.

Cambodia: Kampong Cham Physical Rehabilitation Centre (KPRC), Kampong Cham. This centre is one of several provincial physical rehabilitation centres in Cambodia's national rehabilitation network, supported by Humanity and Inclusion (HI). It provides prosthetic and orthotic services, physiotherapy, and social support to the provinces of Kampong Cham and Tbong Khmum, serving over 2,000 people with disabilities each year.

Thailand: Sirindhorn School of Prosthetics and Orthotics (SSPO), Faculty of Medicine Siriraj Hospital, Mahidol University, Bangkok. SSPO is Thailand's first and only dedicated school of

prosthetics and orthotics, and holds ISPO Prosthetist/Orthotist level (former ISPO Category I) recognition. Operating for over 20 years, this academic centre combines clinical service provision with undergraduate and postgraduate education and research.

All three sites had sufficient volume of users to recruit the target sample within the data collection period, established prosthetic services with experienced clinical staff, and had participated in the original LEAD and COMPASS consensus development. Efforts to include sites beyond the Asia Pacific region were constrained by the timelines required for local ethics approval processes.

2.2 Participants

Answering the pilot's research questions required formal input from three distinct groups, each group bringing a different perspective on whether the LEAD and COMPASS would work in practice.

People with lower limb absence were the focus of the data collection itself. Their willingness and ability to complete the assessments was central to understanding whether the tools were culturally and linguistically appropriate, and whether the time burden was acceptable. Adults aged 18 years and older who were receiving routine prosthetic rehabilitation services at the participating centres during the data collection period were eligible to participate. This included individuals at any stage of prosthetic rehabilitation, from initial assessment through to established prosthesis users receiving follow-up care or replacement devices, with lower limb absence due to any cause and at any level of amputation. The target sample was 30 participants per site, for a total of approximately 90 participants across the three countries.

Clinicians were the people who would need to administer the LEAD and COMPASS as part of their daily work. Their views on feasibility, clinical utility, and practical challenges were essential. All clinical staff involved in administering the LEAD and COMPASS assessments at each site were invited to participate in focus group discussions. This typically included prosthetists, physiotherapists, and rehabilitation professionals who interact directly with people receiving prosthetic services.

Managers were responsible for the services where the LEAD and COMPASS would need to be sustained. Their perspective on organisational and resource implications was critical for understanding whether routine implementation was realistic. The managers of each participating centre were invited to participate in a separate focus group discussion to provide perspectives on the organisational and resource implications of implementing the LEAD and COMPASS.

Participation was entirely voluntary for all groups, with informed consent obtained prior to any data collection or focus group participation.

2.3 The original LEAD and COMPASS: what was tested

The original LEAD and COMPASS, as tested in this pilot, comprised 25 LEAD data items and 8 COMPASS outcome measures. In practice, this meant a clinician working through a structured

assessment with each person with lower limb absence at different points in their care, collecting a combination of data about the person, their care, and the outcomes achieved.

Some information was collected once, at enrolment. This included demographic details, the cause and level of amputation, and whether the person could walk prior to their amputation. Other items were collected at each episode of rehabilitation care, such as details about the prosthetic device provided, the rehabilitation services involved, and health indicators. The COMPASS outcome measures were administered both before and after an episode of care, allowing comparison over time.

The COMPASS outcome measures fell into three groups:

Patient-reported outcome measures (PROMs) which has three questionnaires completed by the person with lower limb absence: the Trinity Amputation and Prosthesis Experience Scales-Revised (TAPES-R),[6] which examines adjustment to amputation, prosthesis satisfaction, activity restriction, and social support; the Prosthesis Evaluation Questionnaire - Residual Limb Health (PEQ-RL),[7] which assesses residual limb symptoms including pain, skin problems, and changes in limb volume; and the Prosthesis Evaluation Questionnaire - Utility (PEQ-UT),[7] which evaluates how well the prosthesis enables participation in daily activities.

Performance-based outcome measures (PerfOMs) which has three physical assessments administered by the clinician: the Timed Up and Go (TUG),[8] which measures the time to stand, walk three metres, turn, and return to sitting; the 2-Minute Walk Test (2MWT),[9] which records the distance covered walking at a comfortable pace; and the Amputee Mobility Predictor (AMP),[10] a comprehensive assessment designed specifically for people with lower limb absence covering sitting and standing ability, standing balance, dynamic balance through reach and turn tasks, and gait characteristics including step length, cadence, and the ability to navigate obstacles and stairs.

Adjunct and Quality of Life measures include the Patient-Specific Functional Scale (PSFS),[11] where individuals identified up to five activities important to them and rated their current level of difficulty; and the EQ-5D-5L,[12] a widely used measure of health-related quality of life assessing mobility, self-care, usual activities, pain, and anxiety/depression.

Taken together, the original LEAD and COMPASS represented a comprehensive but substantial assessment. Whether it was too substantial for routine clinical use was a central question of this pilot.

2.4 How the pilot evaluated the LEAD and COMPASS

To evaluate the LEAD and COMPASS from the multiple perspectives described in Section 1.6, the pilot used three complementary approaches: a purpose-built application for delivering the assessments and capturing quantitative data; a participant feedback survey administered immediately after each assessment; and focus group discussions with clinicians and managers at each site after the conclusion of data collection.

The data collection application

The LEAD data items and COMPASS outcome measures were delivered through a purpose-built iPad application developed by Orthocare Innovations and administered by Modus Ltd. The app integrated all LEAD data items and COMPASS outcome measures into a single interface that guided clinicians through the assessment process, with built-in timers for the timed measures, on-screen instructions for administering each assessment, and automatic score calculation. Given the sensitivity of health data collected across multiple countries, data security was a priority throughout. Data were encrypted and transmitted securely to cloud storage using the BioT Multi Region platform, and stored exclusively within Australia in compliance with data protection requirements.

Beyond its role in delivering the assessments, the application also provided the quantitative evidence base for evaluating the LEAD and COMPASS itself. The data captured, including response rates, score distributions, and completion patterns, allowed analysis of which items were practical to administer, where comprehension problems arose, and where there was redundancy or overlap between measures.

The participant feedback survey

Understanding whether the LEAD and COMPASS assessment was acceptable to the people completing it was essential. Response rates and completion data could show where problems occurred, but not why, or how participants felt about the experience. Following completion of the LEAD and COMPASS assessment, participants were invited to complete a brief anonymous feedback survey about their experience. The survey was delivered via Google Forms and covered four domains: time burden, appropriateness of questions, relevance to the respondent's situation, and comfort with data storage. Each domain was rated on a scale from 0 (least satisfied) to 5 (most satisfied), with an open comment field for each domain and a final open-ended question.

Focus groups with clinicians and managers

The quantitative data and participant feedback could show whether the tools worked from the perspective of the person being assessed, but not from the perspective of those who would need to administer and sustain them. To capture these views, focus group discussions were conducted at each site, one with clinicians and another with managers. Separate groups were used to allow clinicians to speak freely about their day-to-day experience of administering the tools, while managers could address organisational, resource, and sustainability considerations without their staff present. Semi-structured interview guides ensured consistent coverage of key topics across sites, including time burden, redundancy, cultural and linguistic issues, clinical utility, and views on sustainability. This approach structured the focus group while allowing flexibility to explore issues specific to each context.

2.5 How the data were collected

The study received ethical approval from the UNSW Human Research Ethics Committee (HC240347) and from local ethics committees in each participating country. Organisational letters

of support were obtained from senior leadership at each participating site prior to commencement of data collection.

Prior to data collection at each site, a member of the ISPO research team visited in person to train clinical staff on the use of the iPad application and the reliable administration of each outcome measure, and the safety protocols developed to protect participant wellbeing throughout the assessment process. The research team remained available remotely throughout the data collection period. The safety protocol was followed in all cases where participant wellbeing concerns arose during data collection, including signs of emotional distress, fatigue, or discomfort.

Data collection occurred primarily during routine clinical appointments. Participants provided written consent before completing the LEAD and COMPASS assessment, which could take place at both initial and follow-up visits where applicable, allowing examination of the measures' sensitivity to change following an intervention. Responses to all sections of the LEAD and COMPASS were voluntary; participants were able to skip any item. Timing for each component was recorded automatically by the application.

Focus group discussions were conducted via Zoom after data collection was completed. Discussions lasted 1-2 hours and were recorded (audio) with participants' consent.

2.6 How the data were analysed

Data Preparation and Quality Control

Prior to analysis, participant records underwent systematic cleaning and validation. Records were categorized as complete cases with analyzable data; practice/”Test” profiles; and enrolled participants without assessment data. Practice / “Test” profiles were excluded from analysis through unique identifier filtering. Duplicate records were identified by examining identical data entries with different timestamps; in all cases there was an obvious complete record and an empty record. Functional outcome measures were reviewed to identify physiologically implausible values.

Quantitative Analysis

The quantitative analysis served two purposes: understanding how the LEAD and COMPASS performed as a data collection tool, and examining the data itself for evidence of redundancy or overlap between items.

Data collected through the LEAD and COMPASS was analysed using descriptive statistics and these summaries were presented back to the managers at the data collection sites. Completion rates were calculated for each data item and outcome measure, with items having completion rates below 85% flagged for detailed review. Time burden was analysed by calculating mean and range for the overall assessment and for each data item, both overall and by site.

Each COMPASS outcome measure was scored according to its standardised protocol. There were not enough participants who had data at both initial and follow-up appointments to allow

further analysis of change over time. Rasch analysis and correlation analyses were performed to further identify and support items which could be eliminated from the data collection.

Qualitative Analysis

The qualitative analysis aimed to capture the experiences and perspectives of those directly involved in the data collection, questions explored what worked, what didn't, and why.

Audio recordings from the focus group discussions were transcribed and de-identified. Thematic analysis was conducted following a three-stage approach; transcripts were first read closely and coded line-by-line to capture participants' experiences and perspectives in their own words; these initial codes were then systematically organised into descriptive groupings reflecting key topic areas relevant to the research questions; and finally iterative review was used to develop overarching themes and sub-themes to represent the patterns and relationships within the data. Particular attention was paid to identifying specific items that were consistently described as being clinically valuable, and other items identified as challenging or redundant.

Data from the online participant survey was analysed descriptively, with open-text responses coded thematically.

Bringing the evidence together

The integration of quantitative and qualitative findings allowed triangulation of evidence regarding which items should be removed, modified, or retained in a revised version of the LEAD and COMPASS. This triangulation, drawing on quantitative data, participant feedback, clinician and manager perspectives, and the direct experience of the implementation team, forms the basis of the findings presented in the following section.

3. RESULTS: the findings and the LEADr and COMPASSr they produced

3.1 The revised tool

This pilot identified areas of the LEAD and COMPASS that required refinement, resulting in a revised tool that is shorter, more focused, and designed for routine clinical use. The revised version, the LEADr and COMPASSr retains 23 data items and three core outcome measures, with two further measures available as optional.

Table 1 presents the complete LEADr and COMPASSr, what it explores and when each item is collected.

Table 1: LEADr and COMPASSr

Item	What it captures
Collected once, at initial registration	
1) Date of birth	Demographics
2) Sex	Demographics
3) Country/jurisdiction	Demographics
4) Unique identifier	Administrative function
Collected at the start of each course of prosthetic rehabilitation	
5) Cause of amputation	Amputation details
6) Amputation details	Amputation details
7) Ability to walk prior	Baseline function
Collected by the clinician at the end of each course of prosthetic rehabilitation	
8) Professions involved	Services provided
9) Rehabilitation services	Services provided
10) Prosthetic intervention	Services provided
11) Type of socket	Device specification
12) Prosthetic foot/ankle	Device specification
13) Prosthetic knee	Device specification
14) Prosthetic hip	Device specification
15) Tobacco use	Health
16) Body mass index	Health
17) Education level	Socioeconomic proxy
Collected before and after each course of prosthetic rehabilitation care	
18) Date of commencement/completion	Episode timing
19) Use of mobility devices	Function
20) Ability to work	Work capacity
21) Social and community support	Support used
22) Ambulatory activity level	Function

23) Falls	Safety
COMPASS Revised Outcome Measures	
PEQ-RL	Residual limb health
PEQ-UT	Prosthetic utility
TUG	Mobility
Optional Outcome Measures	
PSFS	*Individual goals*
EQ-5D-5L	*Quality of life*

Technical prosthetic and rehabilitation service details (items 8–17) have been separated for clinician completion outside the appointment, reducing the time burden during the clinical encounter. The rationale for each decision, including items removed from the original tool, is set out in Section 3.3.

3.2 Participant details - People with LLA Sample Characteristics

Table 2. Participants with lower limb absence demographics and amputation characteristics

	Total (N=121)
Participants Demographics	
Participants, n	121
Sex, n (%)	
Male	85 (70.2%)
Female	36 (29.8%)
Age at enrolment, years	
Mean (SD)	48.4 (16.8)
Median (IQR)	50.2 (35.2–62.2)
Range	14–86
Amputation Characteristics	
Participants with amputation data, n	119
Total amputations, n	130
Bilateral amputees, n	11
Cause of amputation, n (%)	
Trauma	74 (56.9%)
Diabetes	29 (22.3%)
Other	27 (20.8%)
Amputation level, n (%)	
Transtibial (BK)	95 (73.1%)
Transfemoral (AK)	24 (18.5%)
Other	11 (8.5%)
Amputation type, n (%)	
Major amputation	122 (93.8%)

Revision (bone or soft tissue)	8 (6.2%)
Amputation side, n (%)	
Left	60 (46.2%)
Right	70 (53.8%)
Able to walk prior to amputation, n (%)	
Yes	84 (64.6%)
No	46 (35.4%)

Notes for Table 2:
Amputation characteristics are reported per amputation (n = 130), not per participant with lower limb absence, as 11 participants had bilateral/multiple amputations. 2 participants had no amputation record. Cause of amputation “Other” includes vascular, cancer, infection, congenital, and other causes (collapsed to protect identifiability of small subgroups). Amputation level “Other” includes toe, ray resection, ankle disarticulation, and knee.

Focus Groups

Clinical staff involved in administering the LEAD and COMPASS and managers/senior staff responsible for organising and supporting data collection each participated in separate focus group discussions across the three sites. Due to the small numbers involved, further demographic details are not reported.

3.3 How these decisions were made

Each item in the original LEAD and COMPASS was evaluated using completion rates, score distributions, correlation and Rasch analyses, participant feedback, and focus group discussions with staff across all three sites. Full psychometric detail will be presented in forthcoming peer-reviewed publications. This section summarises the key findings that informed each decision.

LEAD Items retained

Items 1–4 (date of birth, sex, country/jurisdiction, unique identifier) all achieved 100% completion with no comprehension or administration difficulties. Items 3 and 4 are automated.

Items 5–7 (cause of amputation, amputation details, ability to walk prior) each achieved 100% completion. Some confusion between the diabetes and vascular categories in item 5 was identified; clearer definitions have been included within the response options, supported by additional training guidance. The type of amputation sub-question within item 6 achieved 94%

major amputation, with the classification terminology requiring clarification; this question has been reordered and will be completed by the clinician.

Items 8–17 capture technical prosthetic details, rehabilitation services provided, and health and demographic information. These are collected at the end of each course of care. Few participants reached this point during the pilot period, and implementation training had not adequately covered these items, both of which contributed to low response rates. These items are retained for clinician completion outside the appointment. Items 13 and 14 (prosthetic knee and hip) had no responses, consistent with the absence of higher-level amputations in the pilot sample; both remain clinically relevant.

Items 18–23 are collected alongside the COMPASS outcome measures before and after each course of care to enable comparison over time. Item 18 (date of commencement/completion) achieved 50% and has been changed to be collected by clinicians (without the prosthesis user present) or automation. Item 19 (use of mobility devices) achieved 95% and was well understood with meaningful variation. Item 22 (ambulatory activity level) achieved 94% with meaningful variation in both standing and walking time. Item 23 (falls) achieved 95% for the initial question, with lower completion for the follow-up questions on frequency (47%) and injury (58%); 63% of respondents reported falling, and of those, 73% reported a fall-related injury.

LEAD Items retained with refinement

Item 20 (ability to work) achieved 93% completion but encountered significant comprehension difficulties across all three sites. The underlying construct, work capacity, has clear relevance, but the questioning requires simpler, more direct language. A visual Likert scale linked to ICF qualifiers has been adopted as well as rewording of the question. There was also a significant correlation between ability to work responses and TUG results ($r = -0.282$, $p = .004$), suggesting some overlap with functional mobility, but the construct captures information beyond what physical measures alone can provide.

Item 21 (social and community support) achieved response rates ranging from 54% to 94% across its sub-questions. The topic has clear relevance but the question structure required simplification. The question has therefore been revised to ask directly whether the person receives assistance and from what source.

LEAD Items excluded

Two LEAD data items were also excluded. The participation item asked people to rate their level of participation relative to their own expectations and relative to someone of the same XXXX without limb loss. Comprehension was universally poor across all three sites, including among participants with strong English; the concept of "participation" and the comparison framing were not understood by either participants or staff.

The morbidities/conditions item required manual entry of ICD diagnostic codes and received no responses in the pilot; this level of clinical coding is impractical in most settings where the

LEADr and COMPASSr will be used, though the information could potentially be extracted from electronic medical records where these exist.

Retained outcome measures

The Prosthesis Evaluation Questionnaire - Residual Limb Health (PEQ-RL) achieved 82.8% completion and the Prosthesis Evaluation Questionnaire - Utility (PEQ-UT) achieved 84.4%. Both were well understood across all three sites and address aspects of prosthetic experience that are under-explored in routine care.
The Timed Up and Go (TUG) achieved 86.9% and was the most straightforward performance-based measure to administer. It was well received by both clinicians and people with lower limb absence across all three sites. An open response field has been added to support clinical note-taking.

Optional outcome measures

The Patient-Specific Functional Scale (PSFS) achieved 51.6% completion. It was highly valued by participants who engaged with it — particularly those who appreciated identifying their own functional goals, but the measure requires people with lower limb absence to identify specific activities they are unable to perform or have difficulty with, and some people could not conceptualise or articulate these. This difficulty was compounded by the need for translation and explanation. The PSFS is retained as an optional measure, available for sites that choose to adopt it with appropriate training.

The EQ-5D-5L achieved 93.4% but the concept of rating general health-related quality of life felt out of place alongside prosthetic-specific instruments and was difficult for some participants to comprehend. It is retained as optional for services wishing to include a quality of life measure, such as for economic evaluation; services should select a QoL measure that is used locally.

Excluded outcome measures

The Amputee Mobility Predictor (AMP) achieved 56.6% completion, the lowest of any outcome measure. It was valued by clinicians as a comprehensive assessment designed specifically for people with lower limb absence, covering sitting and standing ability, balance, gait characteristics, and obstacle navigation. However, it was consistently identified as long and repetitive across all three sites, with an average administration time of 13.41 minutes. It contributed significantly to overall assessment burden and its low completion rate reflected both its length and the fatigue it created for participants. The AMP demonstrated significant correlation with several items in the LEAD and COMPASS including the two minute walk test, PEQut, standing time in hours, walking time in hours and community participation.

The 2-Minute Walk Test (2MWT) achieved 81.1% completion and was well accepted by both clinicians and prosthesis users. However, it was significantly correlated with the TUG at 0.01, and both measures assess functional mobility. The 2MWT also required additional time, space, and set-up that was not always available in clinical settings, including access to a long corridor and

challenges around counting laps. The TUG, being the most straightforward to administer, is retained in place of both the 2MWT and AMP.

The Trinity Amputation and Prosthesis Experience Scales-Revised (TAPES-R) achieved 67.2% completion with an average administration time of 22.22 minutes. It correlated strongly with the PEQ-UT ($r = 0.631$, $p < .001$) and moderately with the PEQ-RL ($r = 0.474$, $p < .001$), indicating substantial overlap with retained measures. Clinicians across all three sites identified it as too long, with elements already captured elsewhere. Research protocols were followed to provide support in cases where a participant experienced any discomfort. These protocols are detailed in section 2.5. It should be noted for the purposes of this decision to exclude that in most locations there was limited professional support available at pilot sites to follow up on the psychosocial issues the measure raised, meaning data were being collected without the capacity to fully act on it.

3.4 Participant Feedback Survey

The participant feedback survey captured how the assessment experience felt for the people completing it. A total of 89 participants responded across the three sites.

Table 3. Participant satisfaction ratings by domain (n = 89)

Domain	No. of Respondents	Mean	0	1	2	3	4	5
			← Less satisfied			→ More satisfied		
Time to complete	88	4.41	—	1.1%	3.4%	12.5%	19.3%	63.6%
Appropriateness of questions	89	4.65	—	1.1%	—	3.4%	23.6%	71.9%
Performance-based measures	89	4.62	1.1%	1.1%	1.1%	3.4%	18.0%	75.3%
De-identified data storage	89	4.75	1.1%	—	—	1.1%	16.9%	80.9%

Satisfaction was high across all four domains, with medians of 5.0 throughout. Comfort with de-identified data storage received the strongest endorsement, while time to complete showed the widest spread of responses. Open-text comments confirmed that the questionnaire length, rather than the physical assessments, was the source of concern about time. For appropriateness of questions, 95.5% scored 4 or 5, though a small number of respondents identified limitations, including a person with a congenital limb difference and a bilateral wheelchair user for whom

some items were inapplicable, highlighting that not all parts of the tool may suit everyone with lower limb absence without adaptation.

The open comments also revealed a recurring theme: participants had rarely or never been asked about their experience of living with a prosthesis, and the act of being asked held value beyond the data it produced. This theme is explored in detail in Section 4.4.4.

3.5 Focus Groups

Focus group discussions were conducted separately with clinicians and managers at each of the three sites after data collection was completed. Discussions covered the practicality of administering the tools, clinical utility of the measures, cultural and linguistic issues, redundancy, and views on sustainability.

The focus groups provided the qualitative evidence that, alongside the quantitative analysis and participant feedback, informed the revision decisions described in Section 3.3. Their findings are presented in detail through Sections 4.1 to 4.5, organised by research question.

3.6 Administration times

Administration times were recorded automatically by the application for each outcome measure. Times were calculated after excluding cases with no time data and trimming high-end outliers using the standard interquartile range method (values above Q3 + 1.5 × IQR). Mean times are reported below in Table 4.

Table 4. Mean administration time by COMPASS outcome measure

Outcome measure	Status in COMPASSr	Mean time	Cases (trimmed)
TUG	Core	4:13	130
PEQ-UT	Core	6:04	116
PEQ-RL	Core	3:14	110
Core COMPASSr total		**13:31**	
PSFS	Optional	4:50	105
EQ-5D-5L	Optional	3:35	113
AMP	Excluded	13:41	103
TAPES-R	Excluded	22:22	112
2MWT	Excluded	12:07	115
Excluded total		**48:10**	

Notes for Table 4: Times are mean values after trimming high-end outliers (values above Q3 + 1.5 × IQR). Cases reflect the number remaining after excluding profiles with no time data and trimming outliers.

4. DISCUSSION: what the pilot found

The pilot study confirmed that standardised outcome measurement for prosthetic rehabilitation is both valued and realistic across diverse clinical settings. This section presents the findings organised around the five research questions set out in Section 1.5, drawing together evidence from the quantitative data, participant feedback survey, and clinician and manager focus groups.

4.1 Are the contents culturally appropriate?

The contents of the LEAD and COMPASS were broadly accepted across all three sites. Senior staff reported high acceptance rates, with consent at or near 100%, and the types of questions asked were generally considered appropriate for a health and rehabilitation context. As one clinician put it, people were "very kind to give us the information to use." The anonymous participant feedback survey confirmed this, with 95.5% of respondents rating appropriateness of questions at 4 or 5 out of 5.

Some cultural considerations were identified that are relevant to future implementation. Those involved in data collection noted possible response bias (a tendency for people to give socially desirable rather than honest answers) when the treating clinician conducted the interview. As one senior staff member observed, when the treating clinician conducts the interview, "sometimes the

patient may not give us the actual opinion or experience." It was noted that self-completed versions could address this, noting that people "could complete it without other people hearing their answer or without having the clinician involved that might change their response." The anonymous feedback section at the end of the assessment was specifically praised by clinicians, who gave the tablet directly to people and told them "you can read it and you just type whatever your honest feedback is."

At some sites, people were reluctant to discuss specific topics, such as skin conditions or employment, unless clinicians first explained the clinical rationale. Once the purpose was explained, people were generally willing to respond. A small number of the more sensitive or personal questions were associated with the TAPES-R, which has been excluded from the revised COMPASS, reducing this concern.

Privacy in the physical environment also mattered. At one site, consent and interviews at times occurred in congested areas of the clinic. One focus group participant noted that the issue was "about the setting … rather than the question, because I think the question is appropriate. But it's the way that we collect the data." Implementation guidance will address the physical environment for data collection, including access to a quieter space for self-reported sections or self-completion on a personal device.

4.2 Are the contents linguistically appropriate?

The LEAD and COMPASS were tested in English, with clinicians translating in real time during administration. This was the single biggest contributor to time burden and was consistently identified as the most important issue to resolve before routine implementation.

All three sites required real-time translation, which clinicians found challenging and time-consuming. Clinicians at times found real-time translation "quite confusing" for participants, often needing to explain questions multiple times. Beyond translation, some questions lacked local relevance. For example, distances expressed in imperial measurements needed to be reframed in terms people could relate to, and some questions about daily activities did not match the local context.

Each site developed its own informal approach to managing this. At one site, a printed translation was created that staff consulted during interviews. At another, an informal translation was prepared at the start, but different clinicians would "give the different example to the patient," leading to variation. The lack of standardisation meant that each person involved in data collection "did their own translation" and "there was no consistency."

Clinicians reported a significant learning curve that improved over time. After the fourth or fifth time administering the tool, clinicians were able to translate directly and keep the assessment in a conversational flow rather than reading and re-reading each question.

Some question formats worked better than others. Likert-style response scales were frequently confusing for people. In contrast, the Visual Analogue Scale, placing a mark on a line within the

app, was well received because people "can make a decision directly" without needing to understand verbal anchors. The LEAD comparison questions, which asked respondents to compare themselves to peers of the same age without limb loss, were problematic across all sites. One senior staff member explained that people would "just nod their head" and go along, "they don't really understand at all." This difficulty extended to staff, suggesting the issue was not simply translation but a deeper difficulty with the comparison concept itself. The LEADr addresses this directly: the participation item has been removed, and the work capacity question has been restructured with a visual Likert scale and simpler, more direct language that no longer relies on peer comparison.

The unanimous view was that validated local language versions are essential for routine use. Fewer validated outcome measures in the COMPASSr also means the majority of translation work now concerns the LEADr data items rather than complex validated instruments, making multilingual deployment substantially more feasible.

4.3 What is the actual time burden?

Time burden was the most frequently discussed barrier to routine implementation. The full LEAD and COMPASS took between one and three hours to complete per participant, against a target of 30 to 40 minutes that consensus participants in the original development identified as acceptable for routine use. As one clinician summarised, "one, two hours, it's very long." A single clinician with five people to see "cannot go all day doing that."

The duration had visible effects on people. Clinicians observed that "some patients were uneasy and moving a lot," visibly affected by the length of the assessment. Some people needed bathroom breaks during administration, and at one site, people who were told the process would take one to two hours declined to participate. The time burden also affected workflow, particularly for unscheduled prosthesis users, where clinicians couldn't "miss that opportunity to interview them" even though it disrupted planned work.

However, the majority of this time burden arose from sources that have been directly addressed. The exclusion of the AMP (mean 13:41), TAPES-R (mean 22:22), and 2MWT (mean 12:07) from the COMPASSr removes approximately 48 minutes of administration time. The core COMPASSr outcome measures, the TUG (mean 4:13), PEQ-UT (mean 6:04), and PEQ-RL (mean 3:14), have a combined administration time of under 14 minutes. The PSFS (mean 4:50) and EQ-5D-5L (mean 3:35) are available as optional measures for sites that choose to use them, but not all sites are expected to use them. A substantial proportion of the remaining time in the pilot was spent translating and explaining questions in real time from English, an issue that local language versions will largely resolve. Clinicians inputting technical data outside the appointment will further reduce the time people spend in assessment. More broadly, the reduction in the number of validated outcome measures in the COMPASSr decreases the burden of interpretation and cross-cultural validation at each new site, strengthening the case for the LEADr and COMPASSr as an international multilingual tool.

The learning curve was itself a significant factor. Staff who initially took two to three hours were completing assessments much more efficiently after several administrations. As one staff member explained, "once staff became familiar with the content, "it was less time to work through it." Some staff could complete data collection in approximately 30 minutes once familiar with the content, demonstrating that the target time is achievable.

The most frequently suggested additional strategy was participant self-completion of the questionnaire components. Clinicians envisioned people completing self-reported measures during waiting times, "for instance while waiting for device adjustments or appointments." Self-completion would also address the possible response bias discussed in Section 4.1. In settings where literacy levels are lower, a hybrid model may be appropriate, with literate users completing sections independently and clinicians supporting others.

The cumulative effect of these changes is substantial. Reduction from over 70 minutes of outcome measures to under 14 minutes for the core COMPASSr, local language versions, participant self-completion, clinician input independently for technical items, and application improvements mean the LEADr and COMPASSr should be well within the 30–40 minute target. Administration times will be confirmed during the initial phase of implementation.

4.4 What are participants' reactions?

The fourth research question explored how people with limb absence, clinicians, and managers experienced the LEAD and COMPASS in practice. The findings are presented by type of measure, followed by technology and a cross-cutting theme about the value of engagement with people accessing services. Clinician and manager results have been combined to protect identities of respondents.

4.4.1 Performance-based outcome measures

The performance-based measures were the standout positive finding. The TUG, 2MWT, and AMP were widely enthusiastically received by both prosthesis users and clinicians across all sites. People were motivated to demonstrate their abilities when told they were being assessed on strength and function. As one clinician observed, "we told them that we are testing your strength, your muscle. So, if you're able to do this, do a certain type of thing, and then they are like, oh, I will show you how."

The physical activity provided welcome relief from the lengthy questionnaire component. Clinicians described the practical assessments as "a bit of fun" that helped both staff and participants when placed within the longer interview. Staff confirmed this enthusiasm: "we are more interested in doing this rather than asking questions. It was quite fun as well."

Clinicians found the measures clinically valuable. The AMP was considered especially useful for new users, enabling clinicians to observe the person’s capacity and inform device selection. One senior staff member reflected that structured performance testing reminded clinicians of best

practices: "to be honest, all of those is like, hey, come on, we've never done this with our clients. We fit the prosthetics and give it to them and off they go."

Despite the exclusion of the AMP and 2MWT from the COMPASSr due to their length and overlap with the TUG, the clinical utility of all three measures was clearly demonstrated. Several clinicians had already begun incorporating elements such as the TUG into daily practice, and one clinician had used assessment results to support a person's job application by providing documented evidence of functional capability. Practical challenges included space constraints for the 2MWT and the length of the AMP, both of which informed the decision to retain the TUG as the sole performance measure in the COMPASSr.

4.4.2 Patient-reported outcome measures

The PROMs addressed a fundamental gap in current prosthetic care by providing structured space for participant expression. Clinicians recognised the need to "give that space, that opportunity" for people to share how they "feel about their prosthetics," seeing this as directly connected to improving how they "deliver services...and communicate...with our service users."

Some questions created vulnerability for people who had experienced limb amputation. When asked about feelings around limb loss, one person "could not express how she felt". Another "had to reflect and re-live those painful moments." The type of amputation influenced sensitivity: trauma-related amputations were perceived as "a sensitive question for people to answer" while those with diabetes or infection-related amputations found it "easy for them to talk about." In all cases where participants showed signs of any discomfort, the safety protocol described in Section 2.5 was followed. The exclusion of the TAPES-R from the COMPASSr reduces the number of potentially emotionally sensitive questions, while the retained PEQ measures focus on the more practical experience of using a prosthesis rather than emotional experiences of limb loss.

At the same time, some questions helped people feel acknowledged. Participants told staff they had "never been asked about this...before," and staff observed that "they...feel seen when we ask." Another participant reflected that "he never thought about...this way before" when asked about quality of life, suggesting the tools helped people reflect on their experiences in ways that routine care does not facilitate.

Despite recognising the value of PROMs, some expressed uncertainty about how to apply the data in clinical practice. Without guidance on "how can we utilize this information" and a clear structure for integrating findings into clinical workflows, PROMs risked being collected without being acted upon. The contrast with PerfOMs was clear: performance-based results were "directly linked with the prosthetic objectives" and their application was intuitive. Training and implementation guidance for the LEADr and COMPASSr will need to address how PROM data should be used, and at what level: clinical, service, or advocacy.

The PSFS presented particular challenges. The measure requires people to identify specific activities they are unable to perform, but some people could not conceptualise or articulate

these. This difficulty was compounded by the need for translation and contributed to the PSFS having the lowest response rate of any outcome measure (51.6%). While those who engaged with it valued the ability to define their own goals, these challenges support the decision to make it optional rather than core.

An important limitation identified by clinicians was that many PROM questions assumed people were already using a prosthetic device. This created problems for new clients in early gait training who had not yet been fitted. The LEADr and COMPASSr implementation guidance should address how to manage assessment for people with limb absence that have not yet been fitted with a prosthesis.

4.4.3 Technology and digital data collection

The shift from paper-based to digital data collection was viewed positively across all sites. Clinicians who previously relied on manual data entry found the tablet-based approach a significant improvement: "I think using a tablet and having the questionnaires built into an app makes our work easy... it's really, really good and it's like a game changer when it comes to doing research work."

However, the choice of Apple iPad as the data collection device presents challenges for broader implementation. iPads are expensive, fragile, and not widely used in many LMIC settings. Sites with limited resources operated with a single device, creating workflow bottlenecks when multiple clinicians needed to collect data. As one staff member explained, "just having one tablet to do the work was, made challenging to run between a number of clients that came in." The suggested solution was to "either have issue more tablets to do the survey or load the app to our phones." Future implementation of the LEADr and COMPASSr should prioritise lower-cost Android tablets or smartphones, which are more widely available and would also support participant self-completion.

Reliable internet connectivity cannot be assumed in the settings where the LEADr and COMPASSr will be used. Interruptions to Wi-Fi and mobile data disrupted data collection at multiple sites. Offline capability with automatic synchronisation when a connection becomes available was the unanimous recommendation.

The specific application changes required for the next phase are set out in Section 5.1.3.

4.4.4 The value of being asked

A recurring and powerful theme across all sites was that participants had rarely or never been asked about their experience of living with limb absence or using a prosthesis. The act of being asked was itself significant. People expressed that they felt seen and acknowledged — as one staff member observed, "some of the patients mentioned that 'Oh, I've never been asked about this matter before.' It seems like they also feel seen when we ask." A prosthetic user reflected that "oftentimes I see that we are treated like, you know, like objects. He's an amputee, let's just fix the leg and send him on his way."

This finding extended beyond data collection into clinical practice. Clinicians reported that the research experience taught them to ask people about their emotional state before beginning treatment, an approach they had not previously considered. The implication for routine implementation is that the LEADr and COMPASSr may have value not only as a measurement tool but as a structured way of ensuring peoples' perspectives are heard as part of their care.

4.5 Which items are redundant or repetitive?

Question repetition frustrated both clinicians and participants across all sites. Clinicians identified significant overlap between the PROMs, particularly the PEQ and TAPES-R. As one clinician observed, "some question is pretty similar. So it seems like we repeat ourself. But we know that it's not the exact the same, but it's in the same set." People found this irritating, with questions "repeated over and over" and some explicitly requesting to skip items they felt they had already answered.

At one site, staff also identified overlap between the LEAD and COMPASS and existing clinical data collection tools. Staff emphasised that they could not maintain two parallel systems: "the tool has been used in our center, so we have to remove another tool," and that integration with or replacement of existing tools would require engagement with senior leadership.

The exclusion of the TAPES-R from the COMPASSr addresses the most significant source of content redundancy. The TAPES-R correlated strongly with the PEQ-UT ($r = 0.631$, $p < .001$) and moderately with the PEQ-RL ($r = 0.474$, $p < .001$), confirming that it substantially overlapped with retained measures. Future versions of the application could further reduce perceived repetition through design features. Clinicians suggested that software auto-population could carry forward similar responses across instruments, so that where a person has already answered a question in one measure, the response is pre-filled in related items.

Beyond reducing repetition for participants and clinicians, the exclusion of redundant measures simplifies any cross-cultural validation required at each new site, strengthening the case for the LEADr and COMPASSr as an international multilingual tool.

4.6 Caveats and limitations

The pilot was designed to answer practical questions about whether the LEAD and COMPASS could work in routine clinical settings, and the evidence presented in this report supports clear answers to those questions. However, some caveats should be noted. The in-app timer did not function reliably for all pages, and there was no mechanism to detect idle time when the device was set down mid-assessment, meaning the administration times reported are likely upper estimates. The touchscreen format also introduces the possibility of unintentional data inputs. All data were collected through real-time translation rather than validated local language versions, and the three participating sites, while diverse, were all in the Asia-Pacific region. These factors do not change the substance of the findings, but they should be kept in mind when interpreting specific figures.

5. IMPLICATIONS: the next steps

5.1 The LEADr and COMPASSr: from pilot to routine use

The pilot confirmed that prosthetic services want standardised outcome data, but it also showed what needs to be in place for routine collection to work. The tools need to be in the language people actually use. The application needs to fit the clinical environments where it will be deployed. Participants with limb absence who can complete sections independently should be able to. And the data that comes back needs to be useful to those who collected it, informing clinical decisions, supporting service planning, and building the case for investment. The following sections describe how each of these conditions will be met during implementation.

5.1.1 Reporting and data use

The LEADr and COMPASSr data serve three purposes. First, at the individual level, it informs clinical decisions about the person being assessed and informs the person with limb absence about their health status. Second, at the service level, it gives senior staff the information they need to plan, allocate resources, and account for outcomes. Third, at the advocacy and international level, it builds the evidence base that prosthetic services need to demonstrate their value and secure funding. Informing advocacy and policy is discussed in Section 5.3.

At the individual level, the application will provide participant reports that track progress over time, with visual indicators of change, similar to routine health check results, and the ability to compare results with previous assessments. Clinicians should be able to see at a glance whether a person's residual limb health, prosthetic utility, or mobility has improved, deteriorated, or remained stable between visits. These reports should be shareable with prosthesis users and saveable to the medical record. As mentioned, the pilot already produced a compelling example of individual-level data use in supporting a person's job application.

At the service level, senior staff need aggregated reports showing user volumes, outcome trends, and patterns across groups. This information supports service planning, resource allocation, and accountability to funders. The pilot showed that outcome data can also inform procurement decisions, helping services select materials and components better suited to their local context and climate. Standard service-level report templates will be developed as part of the application, designed to answer the questions that senior staff and funders most commonly ask.

Equally important is structured guidance on interpreting and acting on the data. The pilot identified a gap between collecting self-reported outcome data and knowing what to do with it. Performance-based results were intuitive: a TUG score relates directly to a person's ability to function with their prosthesis. Self-reported data were less straightforward, and staff were often uncertain how to integrate results into clinical decisions. Standard analyses and interpretation guidance will be developed alongside the reporting features, ensuring that clinicians and senior staff can each see what the data means at the level that matters to them.

5.1.2 Local language versions

Local language versions are essential for routine use. Full psychometric revalidation of each outcome measure in every language would be the gold standard, but this would be time-consuming and expensive. Translations should rather leverage the possibility of artificial intelligence and be led by committees of local experts, simplified for accessibility, and integrated directly into the application so that users can select their preferred language at the point of use. The reduction in COMPASSr outcome measures means that the majority of translation work now concerns the LEADr data items rather than complex validated instruments, making this task substantially more feasible than it would have been with the original tool.

5.1.3 Application refinement and device compatibility

The application should be extended to Android tablets and smartphones in addition to the current iPad version, with offline data collection and automatic synchronisation when connectivity is available. Priority usability improvements include a participant search function, clear display of unique identifiers after registration, the ability to edit data before final submission, direct navigation to specific outcome measures, and a mechanism to distinguish refusal to answer questions from missing data. Further the possibility of question logic could be better applied, for instance to those without a prosthesis in order to skip irrelevant questions.

5.1.4 Participant self-completion

Participant self-completion of the self-reported components should be supported within the application, enabled by the local language versions. This would reduce clinician time burden and improve data quality by removing the possible response bias associated with clinician-administered interviews. A hybrid model should accommodate people with varying literacy levels.

5.2 Preparing sites for implementation

Moving from pilot to routine use requires preparation at site level. The pilot demonstrated that with adequate training and familiarisation, staff could complete assessments efficiently and effectively, but it also showed that training needs to go beyond the technical use of the application.

5.2.1 Training

Training should cover the clinical purpose and interpretation of each measure, ensuring that staff understand what the data means and how it can be used, not only how to collect it. Peer mentoring approaches, where more experienced staff support colleagues through early adoption, should be built into implementation planning. The pilot showed that the learning curve is steep but short: staff who initially took two to three hours were completing assessments within much shorter timeframes after several administrations.

5.2.2 Organisational integration

Sustainable implementation would require organisational preparation at each site: revision of clinical protocols to incorporate the LEADr and COMPASSr into routine workflows, resolution of overlap with existing data collection tools, engagement of senior leadership, and compliance with local data protection legislation. These steps should be planned collaboratively with each site and supported by ISPO.

5.2.3 Local ethics

Local ethics submissions would likely be required at each implementing site. Experience from the pilot indicates that this is a significant lead-time item and should be initiated early in the implementation process.

5.3 Making the case: outcome data as evidence for investment in prosthetic services

The LEADr and COMPASSr generate evidence about what prosthetic services achieve: functional improvement, return to work, and quality of life. As a standardised core dataset, the LEADr and COMPASSr can link with other data collection efforts at site, national, or regional level, with the common data items enabling pooling and comparison across services and countries. This would build an evidence base that supports funding decisions, informs service planning, and enables international benchmarking.

The pilot demonstrated a strong appetite for this evidence generation. Senior staff identified that having "concrete evidence to show to the donor" would strengthen the case for funding sustainability, and that standardised outcome data provided a credible basis for demonstrating the value of prosthetic services. One participant recognised the broader potential: "if everyone is completing the database, then you can also compare. You have a comparison basis across different countries."

The data also had practical implications for service delivery. Participant feedback suggested that outcome data could inform clinical decisions and procurement alike, helping services select materials and components better suited to their context and climate.

One of the most powerful examples from the pilot was outcome data supporting a person's job application, replacing what had previously been an informal letter with documented evidence of functional capability. This demonstrates the potential of standardised outcome measurement to contribute directly to the economic participation of people with limb absence.

5.4 Staged implementation

The LEADr and COMPASSr are ready for deployment. The core COMPASSr outcome measures have a combined administration time of under 14 minutes, and with local language versions, participant self-completion, and clinician input of technical items outside the appointment, the full

assessment should be well within the 30–40 minute target. All pilot sites expressed genuine enthusiasm for continued involvement, and approximately twelve ISPO member societies have expressed early interest in adoption. A detailed implementation plan accompanies this report, setting out the phased approach to deployment, including information about the technology platform, translation process, site preparation, training, data use and reporting, governance and timeline.

6. CONCLUSION

This pilot set out to answer five questions about the LEAD and COMPASS. It answered all five. The contents are broadly culturally appropriate, with specific issues identified and addressed in the revision. The content was comprehensible when supported by real-time translation, but cannot be delivered effectively in English alone. The time burden of the original tool was too high for routine use, but the sources of that burden have been identified and addressed. Participants, people with lower limb absence, clinicians, and senior staff, valued both the process and the data it produced. And there was significant redundancy that could be removed without losing meaningful information.

These findings directly resulted in a revised tool, the LEADr and COMPASSr: a shorter, more focused tool designed for routine clinical use.

The LEADr and COMPASSr retain the core of what the original set out to achieve: standardised, comparable outcome data for prosthetic rehabilitation, while addressing the practical barriers that would have prevented routine adoption. The core COMPASSr outcome measures can be administered in under 14 minutes, compared with over 70 minutes for the original COMPASS. The revisions were driven directly by evidence: statistical analysis confirmed that excluded measures substantially overlapped with retained ones, while focus group discussions and participant feedback identified where content was too long, too repetitive, or too difficult to translate effectively.

Several findings went beyond what the pilot was designed to test. People with lower limb absence across all sites expressed that they had rarely, if ever, been asked about their experience of living with a prosthesis. The structured assessment created a form of engagement that many had not previously experienced, one that clinicians recognised as changing how they understood their own client's needs. Staff described the process as professionally enriching, with some already incorporating assessment elements into daily practice. At one site, the research raised the profile of the prosthetic and orthotic profession within the health system, providing professional development opportunities rarely available to local practitioners and generating strong interest in continued collaboration with ISPO. These outcomes suggest that standardised outcome measurement has value both as a data collection exercise and a catalyst for more person-centred, evidence-based prosthetic care.

All three sites expressed genuine enthusiasm for involvement and participated without financial incentives, a deliberate design choice that tested whether engagement would be sustained

without external funding. The answer was clearly affirmative, and this finding is itself evidence for the sustainability of routine implementation.

The path forward is staged implementation. Local language versions will address the single biggest contributor to time burden. Application refinement will extend device compatibility beyond the iPad, support participant self-completion of the self-reported components, and provide the reporting functionality that clinical and senior staff identified as essential for integrating outcome data into routine workflows. Site preparation, including training, organisational integration, and local ethics submissions, will be planned collaboratively with each implementing site. With approximately twelve ISPO member societies already expressing early interest, the LEADr and COMPASSr are well positioned to build the international evidence base that prosthetic services in LMICs have long needed.

REFERENCES

1. Kohler F. LEAD and COMPASS: tools to map the way forward for people with lower limb absence. Prosthet Orthot Int. 2021;45(5):369-372. doi:10.1097/PXR.0000000000000045
2. Tan JM, Halford GRJ, McBain BK, Lukin M, Kohler F. Psychometric properties of outcome measures for lower-limb absence: a protocol for a systematic review. Prosthet Orthot Int. 2023;47(3):293-299. doi:10.1097/PXR.0000000000000220
3. Kohler F, Halford GRJ, Lukin M, et al. Categorization and recommendations for outcome measures for lower limb absence by an expert panel. Prosthet Orthot Int. 2023;47(6):565-574. doi:10.1097/PXR.0000000000000307
4. Tan JM, Halford GRJ, Lukin M, Kohler F. Recommendations from the ISPO lower-limb COMPASS: patient-reported and performance-based outcome measures. Prosthet Orthot Int. 2023;47(1):13-25. doi:10.1097/PXR.0000000000000197
5. Gliklich RE, Leavy MB, Dreyer NA, editors. Registries for evaluating patient outcomes: a user's guide. 4th ed. Rockville (MD): Agency for Healthcare Research and Quality; 2020. AHRQ Publication No. 19(20)-EHC020. doi:10.23970/AHRQEPCREGISTRIES4
6. Gallagher P, MacLachlan M. Development and psychometric evaluation of the Trinity Amputation and Prosthesis Experience Scales (TAPES). Rehabil Psychol. 2000;45(2):130-154. doi:10.1037/0090-5550.45.2.130
7. Boone DA, Coleman KL. Use of the Prosthesis Evaluation Questionnaire (PEQ). J Prosthet Orthot. 2006;18(6):P68-P79.
8. Shumway-Cook A, Brauer S, Woollacott M. Predicting the probability for falls in community-dwelling older adults using the Timed Up & Go Test. Phys Ther. 2000;80(9):896-903. doi:10.1093/ptj/80.9.896
9. Brooks D, Parsons J, Hunter JP, Devlin M, Walker J. The 2-minute walk test as a measure of functional improvement in persons with lower limb amputation. Arch Phys Med Rehabil. 2001;82(10):1478-1483. doi:10.1053/apmr.2001.25153
10. Gailey RS, Roach KE, Applegate EB, et al. The Amputee Mobility Predictor: an instrument to assess determinants of the lower-limb amputee's ability to ambulate. Arch Phys Med Rehabil. 2002;83(5):613-627. doi:10.1053/ampr.2002.32309
11. Stratford P, Gill C, Westaway M, Binkley J. Assessing disability and change on individual patients: a report of a patient specific measure. Physiother Can. 1995;47(4):258-263.
12. Herdman M, Gudex C, Lloyd A, et al. Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). Qual Life Res. 2011;20(10):1727-1736. doi:10.1007/s11136-011-9903-x