LEAD and COMPASS:

Defining Outcome Measures and a Core Dataset for the Lower Limb Prosthetics Sector

International Society of Prosthetics and Orthotics

September 2021

Project Funded by USAID in support of ATscale and managed through

United Nations Office for Project Services (UNOPS)

Table of Contents

List of Tables and Figures

List of Abbreviations

COMPASS	**(Consensus of Outcome Measures for Prosthetics and Amputation ServiceS)**
LEAD	**(Lower Extremity Amputation Dataset)**
AAS	Amputee Activity Score
ABC	Activity Balance Confidence Scale
ABIS	Amputee Body Image Scale
AMPSIMM	Amputee Single Item Mobility Measure
ADAPT	Assessment of Daily Activity Performance in Transfemoral Amputee
AMP	Amputee Mobility Predictor – inc AMPPRO and AMPnoPRO
BBS	Berg Balance Scale
CAPP-FSI	Child Amputee Prosthetics Project – Functional Status Inventory
CDS	Core Dataset
CES-D	Centre for Epidemiologic Studies-Depression Scale
CHAMP	Comprehensive High-Level Activity Mobility Predictor
CLASS	Comprehensive Lower-limb Amputee Socket Survey
COMPASS	Consensus of Outcome Measures for Prosthetics and Amputation ServiceS
COSMIN	Consensus-based Standards for the selection of health Measurement Instruments
CRPD	Convention on the Rights of Persons with Disabilities
CTT	Classical Test Theory
EEARB	Engagement in Everyday Activities involving Revealing the Body scale
EMR	Electronic Medical Record
EQ-5D-5L	EuroQol five dimension five levels
FAI	Frenchay Activities Index
FATO	La Fédération Africaine des Techniciens Orthoprothésistes
F8WT	Figure 8 Walking Test
FIM	Functional Independence Measure
FMA	Functional Measure for Amputees
FSST	Four Square Step Test
FRT	Functional Reach Test
GDPR	General Data Protection Regualtion
HAI	Hill Assessment Index
HI	Humanity and Inclusion
HICs	High Income Countries
HRQoL	Health Related Quality of Life
IAG	Industry Advisory Group
IC2A	International Confederation of Amputee Associations
ICC	Intra-class Correlation Coefficient
ICF	International Classification of Functioning, Disability and Health
ICRC	International Committee of the Red Cross
IRT	Item Response Theory
ISPO	International Society of Prosthetics and Orthotics
LCI	Locomotor Capabilities Index
LEAD	Lower Extremity Amputation Dataset
LLA	Lower Limb Absence
LMICs	Low and Middle Income Countries
MBSRQ	Multidimensional Body-Self Relations Questionnaire
NBWT	Narrowing Beam Walking Test
OPCS	Office of Population Censuses and Surveys Scale
OPOT	Orthotics and Prosthetics National Office Outcomes Tool

OPUS	Orthotic Prosthetic User Survey
OPUS – LLFM	Orthotic Prosthetic User Survey -Lower Limb Functional Measure
OPUS – SDS	Orthotic Prosthetic User Survey -Satisfaction with Device Scale
PDDQ	Prosthesis Donning and Doffing Questionnaire
PEmbS-LLA	Prosthesis Embodiment Scale - Lower Limb Amputees
PerfOM	Performance Based Outcome Measure
PEQ	Prosthesis Evaluation Questionnaire
PGI	Patient Generated Index
PLUS-M	The Prosthetic Limb Users Survey of Mobility
PMQ	Prosthetic Mobility Questionnaire including the PMQ 2.0
PPA	Prosthetic Profile of the Amputee
PREM	Patient Reported Experience Measure
PROM	Patient Reported Outcome Measure
PROMIS-29	Patient Reported Outcomes Measurement Information System
PROSPERO	International Prospective Register of Systematic Reviews
PSFS	Patient Specific Function Scale
QALY	Quality Adjusted Life Year
Q-TFA	Questionnaire for Persons with a Transfemoral Amputation
QR&S	Questionnaire Rising and Sitting Down
QEHFT	Questionnaire to Explore Human Factors and their Technical potential
RoB	Risk of Bias
RMI	Rivermead Mobility Index
SAI	Stair Assessment Index
SCS	Socket Comfort Score
SDGs	Sustainable Development Goals
SF-36V	Short Form 36 Items - Veteran population
SIGAM	Special Interest Group in Amputee Medicine Mobility Scale
TAPES-R	Trinity Amputation and Prosthesis Experience Scales-Revised
TUG	Timed Up and Go (TUG)
UASPO	United Arab Society of Prosthetics and Orthotics
UN	United Nations
USAID	United States Agency for Inernational Development
WFOT	World Federation of Occupational Therapists
WHO	World Health Organisation
2MWT	2-minute walk test
3D GAGDI	3D Gait Analysis Gait Deviation Index
6MWT	6-minute Walk Test
10mWT	10-metre (timed) walk test

Executive Summary

Numerous international efforts to improve services for people with disability are progressing and have significant positive impact. Challenges and impediments for continued improvement in providing high quality prosthetic services for people with disabilities need to be addressed.

This report focusses on people with lower limb absence (LLA). One of the impediments to expansion of services is a lack of standardised data and a lack of information about outcomes. Standardised data and standardised information on outcomes can support decision making at all levels, (service user/clinician, management, and policy) and help make an effective business case for investment in services for people with LLA. LEAD (Lower Extremity Amputation Dataset) and COMPASS (Consensus of Outcome Measures for Prosthetic and Amputation ServiceS) are two tools which have been developed to be used in conjunction with each other and to address this lack of basic information.

COMPASS

Outcome measures are standardised instruments that can assist to quantify the current health status and evaluate change in health status of an individual over time, allowing the change to be attributed to the use of an intervention. Outcome measures must be reliable, sensitive to any changes of the properties being measured, and accurately reflect the function of the individual. Collectively, these features are commonly described as 'adequate psychometric properties.' A large number of outcome measures are used in various settings in people with LLA, creating confusion as to which outcome measures should be used routinely in clinical practice. A systematic review was designed (Section 1) and performed (Section 2) for published work that details psychometric properties of outcome measures in people with LLA. 108 studies detailing the psychometric properties of 60 outcome measures were found. From each article the details of the psychometric evidence was extracted and the strength of evidence appraised with the Consensus-based Standards for the selection of health Measurement Instruments (COSMIN) risk of bias tool. An expert panel was convened (Section 3) to weigh both the psychometric evidence and the COSMIN rating of that evidence. The expert panel recommended twelve patient reported outcome measures (PROMs) and eleven performance-based outcome measures (PerfOMs) along with eleven PROM subscales; in total 34 outcome measures and subscales.

These outcome measures advanced from the expert panel were considered by a consensus process (Section 4). Nominees were sought primarily through institutional invitations with additional focused invitations to ensure representation from all regions of the world and balance from high-

middle- and low-income countries. Focus was given to including people with LLA, clinicians, managers, researchers and policy makers who had experience with prosthetics and related rehabilitation services.

The consensus participants weighted the usability, time and equipment burden, appropriateness, and safety of the outcome measures, as well as the value of the information the outcome measure provides for the user, clinician, manager and policy maker. Ultimately the COMPASS was defined with three PerfOMs (the Amputee Mobility Predictor (AMP), the Timed Up and Go (TUG), and Two-minute Walk Test (2MWT)), and three PROMs (the Prosthesis Evaluation Questionnaire (PEQ) – Utility, PEQ – Residual Limb Health and the Trinity Amputation and Prosthesis Experience Scale – (TAPES-R)). For high functioning users an optional additional pair of PerfOMs, the Comprehensive High-Level Activity Mobility Predictor (CHAMP) and the Six Minute Walk Test (6MWT) define the COMPASS+. The COMPASS Adjunct contains the Patient Specific Function Scale (PSFS) due to its significant clinical utility, despite its limitations with comparability between individuals. In addition, use of a generic health related quality of life (HRQoL) measure is recommended, such as the Patient Reported Outcome Measures Information System – 29 item profile (PROMIS-29) or EuroQoL (EQ-5D-5L), depending on what is locally applicable and recommended, to enable broader comparison of HRQoL across different health conditions.

The time burden of the COMPASS is around 45 minutes; about 20 minutes for PerfOMs which require clinical involvement, and about 25 minutes for the PROMs which could be administered without the direct involvement of the clinician. Extra time for scoring and interpretation of the measures is required. There is potential for streamlining the scoring with online or application-based methods of administration. Minimal equipment and space is needed to conduct these outcome measures: two hard chairs with arm rests, a stopwatch, cones for the floor and adequate walking space to conduct the walking tests.

Other recommendations from the COMPASS consensus process include the need for development and psychometric measurement of outcome measures appropriate for activities common in Low- and Middle-Income Countries (LMICs), such as cross-legged sitting, squatting and kneeling, as well as ensuring outcome measures are used in their published and validated forms without editing or changes which reduce comparability.

Implementation activities to support the COMPASS will be conducted by ISPO including publication of a COMPASS User Guide and production of instructional videos.

LEAD

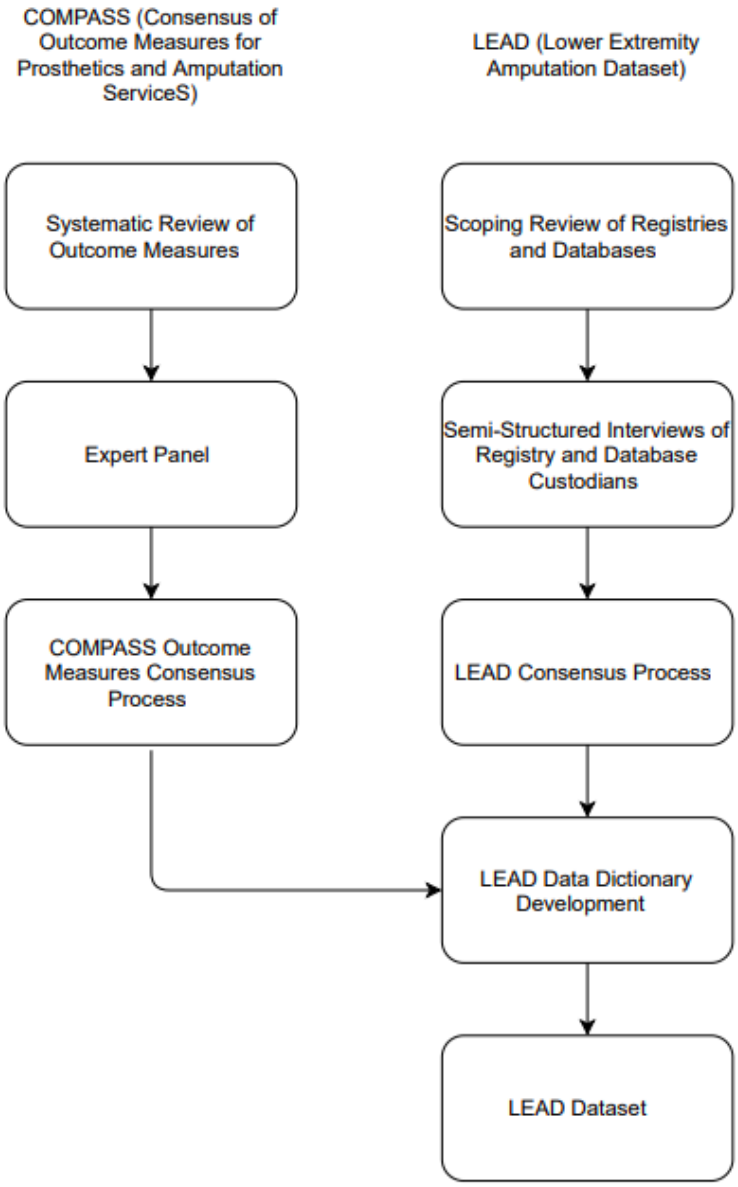
The LEAD consensus process to develop a core dataset (CDS) for people with LLA (Section 6) was conducted through a scoping review of peer reviewed and grey literature that revealed information about registries and databases which collect information about people receiving prosthetic or orthotic deceives or other assistive technology. Semi structed interviews were performed with registry and database custodians to cross check published information and obtain information about unpublished initiatives. From these sources summaries of data items and themes about registries and databases were compiled.

A separate diverse group of consensus attendees, with a similar selection process to the outcome measures consensus process, was convened to consider data elements for a CDS with the categories of demographics, surgical/limb absence, rehabilitation interventions, confounders, and outcomes. From this process a data dictionary was developed with twenty five categories of data items. These data items are grouped as per time of data collection; at enrolment, for each limb absence/amputation, once per episode of rehabilitation care and pre and post episode of rehabilitation care. Themes associated with registry design and function were also discussed and recommendations formed. These themes were purpose and objectives of a registry, eligibility criteria including inclusion and exclusion criteria, recruitment methods, data collection methods, data handling, data privacy, security and access, stakeholders, and funding. The establishment of a defined CDS will allow capture of routine data in a wide variety of clinical settings to inform policy and management decisions.

The COMPASS is envisaged to be administered within the LEAD. The process of refining the LEAD and its corresponding data dictionary considered the outcome measures included in the COMPASS. A flow chart describing the entire process of development is included in Figure 1.

Development of data collection, storage and reporting platform(s) is required to fully realise the potential of the LEAD, including solutions that would allow for the LEAD to function as a multinational registry with appropriate data security and privacy safeguards. Both the LEAD and COMPASS can be used to standardise data collection and reporting in standalone databases and registries, this would allow direct comparisons to be made and the possibility of aggregation of data to demonstrate effects of rehabilitation of people with LLA.

Figure 1.0 Flowchart of the LEAD and COMPASS development process.



1.0 Background

The provision of meaningful services for people with disabilities to promote inclusion and improved quality of life has been an increasing focus of health and social services worldwide. This development has been progressing rapidly over the last few decades. The major international achievements include the development of the International Classification of Functioning, Disability and Health (1), which was developed to create a common language in the assessment of functioning and its limitation and restrictions and the Convention on the Rights of Persons with Disabilities (CRPD) (2), a framework to protect the rights and dignity of persons with disabilities. The broad actions to bring this development together were expressed in the Sustainable Development Goals (SDGs) (3), with a significant focus on disability. Initiatives such as Rehab 2030 (4) and ATscale (5) seek to operationalise these goals.

At a practical level, operationalising the provision of assistive technology faces multiple hurdles which include a lack of appreciation of the benefits of assistive technology, a lack of understanding of what technology is available, lack of awareness of the number of persons with disabilities and subsequently no proper quantification of the need for assistive technology.

The manufacture and fitting of prostheses for individuals with lower limb absence (LLA), amenable to prosthetic provision, has major impacts on improving functioning and quality of life and this has been demonstrated by numerous studies. It provides job satisfaction for many clinicians who work with persons with lower limb absence and is the motivation for many researchers and the prosthetic industry to develop prosthetic components to further enhance performance and functioning.

Prosthetic services are however not universally available, and particularly in some low- and middle-income countries (LMICs) prosthetic service provision is not a priority, with limited if any services being available. Many of the services available in such settings are provided by charitable or non-government organisations rather than the government and there is often a separation between rehabilitation services and other mainstream health services that limits accessibility.

Prosthetic and broader assistive technology service provision and rehabilitation services are often attached to the Ministry of Social Affairs (MoSA) or its equivalent and given low priority. Where these services are attached to the Ministry of Health (MoH) they are often ranked with lower priority and importance than other health issues. Commonly these services are not directly provided but rather regulated by MoSA or MoH which sometimes results in further restriction rather than facilitation of service provision. It has become a focus of the United Nations (UN) and the World

Health Organisation (WHO) to prioritise rehabilitation and assistive technology service provision at the country level. This is in line with CRPD article 26.3: States Parties shall promote the availability, knowledge and use of assistive devices and technologies, designed for persons with disabilities, as they relate to habilitation and rehabilitation (2). To develop sustainable services worldwide requires coordinated investment and service development by governments at all levels, with or without external support. To assist governments to make wise investments for the provision of prosthetic and orthotic services, the WHO in collaboration with the International Society for Prosthetics and Orthotics (ISPO) and the United States Agency for International Development (USAID), developed the Prosthetic and Orthotic Standards (6).

Development and expansion of assistive technology services and prosthetic provision has however continued to be slow. One of the impediments for further development is a lack of standardised data and information about outcomes. Many high-income countries have limited data about people with LLA. Even very basic information such as the number of people with LLA, particulars of details of the nature of the limb absence and demographic or aetiological information is not universally available. In LMICs reliable data is even less common and there are numerous publications which refer to the paucity of data for people with LLA (7,8). Information about outcomes of people with LLA is available in many scientific publications, but numbers included in these studies are generally small, there is a bias towards high income countries and there is no standardised outcome measure used to allow comparisons to be made, or to enable the synthesis of a plausible convincing argument for the development of further services for individuals with LLA or for the improvement of quality of the services which are available.

The publication bias towards high income countries intrinsically also biases the publications towards more expensive and more technologically complex prostheses (usually referred to as ‘advanced prostheses’). A person with LLA will however achieve significant improvements in their function and quality of life even with a relatively simple prostheses at a small fraction of the investment. Jurisdictions planning to introduce better access to prosthetic services, or for that matter any assistive technology or rehabilitation services are challenged by this discrepancy of information available. It is possible that the complexity of the limited information available is a further impediment to prioritising investment in these services.

The responsibility for appropriate data collection is also clearly outlined in the CRPD Article 31$^{(d)}$ on Statistics and Data Collection (2):

1. *States Parties undertake to collect appropriate information, including statistical and research data, to enable them to formulate and implement policies to give effect to the present Convention. The process of collecting and maintaining this information shall:*
    a. *Comply with legally established safeguards, including legislation on data protection, to ensure confidentiality and respect for the privacy of persons with disabilities;*
    b. *Comply with internationally accepted norms to protect human rights and fundamental freedoms and ethical principles in the collection and use of statistics.*
2. *The information collected in accordance with this article shall be disaggregated, as appropriate, and used to help assess the implementation of States Parties' obligations under the present Convention and to identify and address the barriers faced by persons with disabilities in exercising their rights.*
3. *States Parties shall assume responsibility for the dissemination of these statistics and ensure their accessibility to persons with disabilities and others.*

Developing a standardised data collection, based on a consensus approach of determining what data and outcomes to measure, will provide a reliable and consistent source of information. This information can then be used by any jurisdiction to assist them to make informed and balanced decisions on how to rationalise their scarce resources to achieve the maximum possible benefit for the largest number of people with a view to equitable distribution of services and investment in people with disabilities. The more reliable data that is available, the more support there will be for decision makers to make an economic and social argument for the planning and establishment of services. Data on patient outcomes will allow comparisons to be made about different models of service provision and the effectiveness of different components and methods of prosthetic manufacture and prosthetic design. If cost data is added to the outcome measures, cost-benefit ratios or economic efficiency of interventions can be estimated which in turn will further enhance the case for investment in services. More detailed aetiological, demographic and outcome data will enhance the opportunities for such comparisons to be made for subgroups of the prosthetic users to further facilitate individualised focussed service developments.

Estimates based on modelling suggests that up to 42 million people in LMICs need a prosthetic device. According to estimates of the WHO only approximately 5-15% are fitted with a prosthesis (6). In order to increase the attractiveness to potential donors and investors better data is required. A prosthetics market and sector analysis undertaken as part of a product narrative for ATscale has demonstrated that in the LMIC public sector, capacity and financing for prosthetic services is low (9). Government investment requires data on the need and gaps in provision of services and without such data, the case for investment is unlikely to be successful. An example of a successful project in

a low resource setting was undertaken in the Philippines, where people with LLA were screened with the assistance of mobile phones to outline the needs of the individuals affected. Subsequently government investment in the sector was established and a system of reimbursement for prostheses was initiated.

The International Society for Prosthetic and Orthotics (ISPO) was established in November 1970. Its mission statement (10) is:

'To improve the quality of life for persons who may benefit from the rehabilitation practice of prosthetic, orthotic, mobility and assistive technology by:

- *Promoting multidisciplinary practice*
- *Facilitating professional education to provide quality care*
- *Promoting research and evidence-based practice*
- *Facilitating innovative and appropriate technology*
- *Fostering international collaboration and consensus, and*
- *Facilitating knowledge exchange'*

Promoting collection of appropriate data and outcome measures was a challenge which ISPO by itself would not be able to successfully bring to fruition. The Industry Advisory group, which is a group to which all members involved in the P&O and broader assistive technology industry are welcome, worked closely with the board of ISPO to develop an outline of how a project to shape future data collection could proceed. Momentum was gained when discussions with ATScale and the Clinton Health Access Initiative (CHAI) established that there was a common goal between these organisations; namely to develop a consensus on a CDS and outcome measures to be used for people with LLA which could be used to enable improved and rational planning and resource allocation, outcome-driven reimbursement systems, a more harmonized regulatory environment, clear linking of costs and benefits, and alignment of industry's efforts with the national governments and policy-makers and most importantly, with the real needs of the end-users taken into consideration.

Funding to carry out this project was successfully secured from USAID as part of their support for ATScale and administered by the United Nations Office for Project Services (UNOPS.)

The aims and justification of the project were agreed as:

1. Define, achieve global convergence, build consensus and disseminate recommendations for Prosthetic Outcome Measures (OM): Defining and aligning OM with the needs of the international community will help to:

a. Inform the investment case for prosthetics, including defining the benefits of the service and assistive technology to both the individual and the community;
b. Inform research protocols and standards for new technologies and solutions;
c. Inform design of a core data set; and
d. Open opportunities for the development and evaluation of new service delivery models including manufacturing processes such as increased digital transformation.

2. Define, build consensus and disseminate recommendations for a Core Data Set (CDS) to standardize data collection worldwide: Defining the CDS useful to industry, implementers and policy-makers will help to:
    a. Inform the investment case for prosthetics; and
    b. Guide the set-up of an international, national or local amputee registry to support rehabilitation programs. The core components will include data items for the CDS, key outcome measures (see above), data quality, and data collection approaches including identifying impediments to data collection, data ownership, data security, data sharing opportunities and challenges and other issues related to extracting useful reports from the CDS.

References

1. World Health Organization. The International Classification of Functioning, Disability and Health (ICF). Geneva, Switzerland: WHO; 2001 Sep.
2. UN General Assembly. Convention on the Rights of Persons with Disabilities: resolution / adopted by the General Assembly [Internet]. United Nations; 2007 Jan [cited 2021 Aug 9]. Available from: https://www.un.org/disabilities/documents/convention/convoptprot-e.pdf
3. Sustainable Development Goals [Internet]. unfoundation.org. [cited 2021 Aug 9]. Available from: https://unfoundation.org/what-we-do/issues/sustainable-development-goals/
4. World Health Organization. Rehabiliation2030: A Call for Action [Internet]. [cited 2021 Aug 9]. Available from: https://cdn.who.int/media/docs/default-source/documents/health-topics/rehabilitation/call-for-action/keyforhealth21stcentury.pdf?sfvrsn=43cebb7_5
5. Fineberg A, Savage M, Austin V, Boiten S, Droop J, Allen M, et al. 9789240000261-eng.pdf. In: ATscale: Establishing a cross-sector partnership to increase access to assistive technology [Internet]. Geneva, Switzerland: WHO; 2019 [cited 2021 Aug 9]. p. 428–39. Available from: https://apps.who.int/iris/bitstream/handle/10665/330372/9789240000261-eng.pdf
6. Standards for Prosthetics and Orthotics. [Internet]. Geneva: World Health Organization; 2017. Available from: https://apps.who.int/iris/handle/10665/259209
7. McDonald CL, Westcott-McCoy S, Weaver MR, Haagsma J, Kartin D. Global prevalence of traumatic non-fatal limb amputation. Prosthet Orthot Int. 2021 Apr;45(2):105–14.
8. Moxey PW, Gogalniceanu P, Hinchliffe RJ, Loftus IM, Jones KJ, Thompson MM, et al. Lower extremity amputations - a review of global variability in incidence: Lower extremity amputations-a global review. Diabet Med. 2011 Oct;28(10):1144–53.
9. ATscale Product Narrative: Prostheses April 2020 [Internet]. [cited 2021 Aug 9]. Available from: https://at2030.org/static/at2030_core/outputs/Prostheses_Product_Narrative_a11y_20200827.pdf
10. ISPO Introduction [Internet]. [cited 2021 Aug 10]. Available from: https://www.ispoint.org/news/

12. Mokkink LB, de Vet HCW, Prinsen CAC, Patrick DL, Alonso J, Bouter LM, et al. COSMIN Risk of Bias checklist for systematic reviews of Patient-Reported Outcome Measures. Qual Life Res. 2018 May;27(5):1171–9.

2.0 Outcome Measures Systematic Review

2.1 Background

Outcome measures are standardised instruments used in a clinical or research setting, that can assist to quantify current health status, or evaluate changes in the health status of an individual, group, or population, which can be attributed to the use of an intervention (1). Outcome measures are broadly categorised as patient-reported (PROM) or performance-based (PerfOM). Usually administered as questionnaires, PROMs measure the individual's own perspective of their clinical presentation and/or treatment. Conversely, PerfOMs require clinicians or researchers to collect information (data) in a pre-defined standardised manner whilst the patient undertakes a particular task, such as the time taken to rise from a seated position.

Historically, outcome measures (both patient-reported and performance-based) have been used to a variable degree in clinical practice in individuals with lower-limb absence (LLA). Choice of appropriate outcome measures and their use within clinical and research settings, is largely determined by prior experience and ease of implementation (2, 3) and not necessarily driven by current evidence. The majority of evidence surrounding outcome measures commonly used in LLA populations has largely been tested in "similar" populations of people with a disability, such as stroke, Parkinson's disease, falls risk, or the elderly, however individuals with LLA have differing needs and capabilities to these population groups. This therefore limits the ability of clinicians to use the results from outcome measures to make accurate and informed clinical judgements with respect to rehabilitation plans for individuals with LLA.

Various parties (e.g. clinicians, users, administrators, funders, etc.) increasingly acknowledge and understand the value of functional and health related quality of life (HRQoL) improvements that can be demonstrated with the use of outcome measures. Completion of outcome measures that demonstrate functional HRQoL improvements, can be potentially used to influence both clinician remuneration and patient funding. However, there are many different factors that determine remuneration and funding, including the capability of the practicing clinician and possibly their experience and skill at choosing and using outcome measures. Ultimately, it is most likely the patient who will be disadvantaged when outcome measures are not accurately completed.

While some international organisations and individual clinicians use outcome measures, standardised outcome measures are less commonly used in LMICs, which significantly limits the ability of governments and aid funding organisations to make informed decisions regarding the

distribution of services, particularly for prosthetic components. This highlights the need to specify key outcome measures which can be applied in a standardised manner in the multitude of cultural, economic and geographic settings around the globe for both clinical practice and research. Standardised utilisation of specific outcome measures will help to inform investment cases for prosthetic devices and service delivery, facilitate an increase in world-wide access to prosthetic devices and services, highlight the priority of access to rehabilitation services in LMICs, and assist to improve research protocols and standards in LLA populations. Collectively, this can aid in maximising outcomes for all individuals with LLA.

The aim of this systematic review was to identify and critically appraise, compare, and summarise the quality of PROMs and PerfOMs that have been psychometrically tested in LLA populations. The results of this review will inform the development of the Consensus of Outcome Measures for Prosthetic and Amputation ServiceS (COMPASS) and provide clinicians with a reference guide to inform clinical practice and help to set future research agendas.

2.2 Methods

2.2.1 Registration

This systematic review conforms to the guidelines of the Preferred Reporting Items for Systematic Review and Meta-Analysis (PRISMA) (4), and was registered prospectively with the International Prospective Register of Systematic Reviews (PROSPERO) (ID: CRD42020217820).

2.2.2 Population

The eligibility criteria were developed using elements of the PICO framework (population: individuals with LLA; intervention/exposure: outcome measures; comparator: non-disabled individuals [where applicable] or populations demonstrating similar characteristics [e.g. stroke]; and outcome: psychometric properties and are reported in detail below.

Inclusion criteria

This systematic review included full-text peer-reviewed journal articles published in English which clearly report on the development or psychometric properties of outcomes measures in individuals with LLAs. No date limit was set to ensure all available evidence was captured.

Exclusion criteria

The following exclusion criteria were applied: prediction and risk factor studies, studies which investigated the use of electronic devices to measure outcomes (e.g. step counters, wearable accelerometers, etc.), mixed-populations in which data relating to individuals with LLA could not be extracted in isolation, systematic reviews, literature reviews, correspondences, posters, abstracts, newsletters, reports, not peer-reviewed, and non-English language. Articles which did not explicitly state that their aim was to investigate the development or the psychometric properties of outcome measures and/or the statistical analysis did not incorporate elements of the Consensus-based Standards for the selection of health Measurement Instruments (COSMIN) checklist, were excluded.

2.2.3 Outcome Measures

To answer the aim of this systematic review, the primary outcomes of interest were the psychometric properties of PROMs, PerfOM's, or combined (comprising of both PROM and PerfOM) outcome measures that have been psychometrically tested within an LLA population. Psychometric properties include (but are not limited to): reliability (presented as intra-class correlation coefficient [ICC] and kappa statistics [*k*]); validity (correlation coefficient [Pearson's or Spearman's], sensitivity, specificity, and positive and negative likelihood ratios); internal consistency (Cronbach's alpha); measurement error estimates (standard error of the mean [SEM], minimal detectable change [MDC] or smallest detectable change [SDC], and limits of agreement [LOA]); and responsiveness estimates (correlations, minimal clinically important difference [MCID], area under the curve [AUC], standardised response mean [SRM], mean change, and effect size [ES]).

2.2.4 Information sources

Between October and December 2020, the following databases were searched: CINAHL, Embase, Medline (Pubmed), and PsychInfo. The reference lists of included articles were hand-searched to ensure all relevant articles were captured and included. Finally, a search was conducted through Google Scholar to ensure all articles that were not yet indexed through Medline were captured.

2.2.5 Search strategy

Table 2.1 shows the search string used in Medline, with similar search strings used in the other databases.

Table 2.1. Sample Medline search string for literature review of outcome measures for LLA

("lower limb amput*" OR "lower-limb amput*" OR "lower extremity amput*" OR "lower limb prosthe*" OR "lower-limb prosthe*" OR "lower extremity prosthe*" OR "lower limb loss" OR "lower-limb loss" OR "hip disarticulation" OR "through hip amput*" OR "through-hip amput*" OR "hindquarter amput*" OR "hemipelvectomy" OR "above knee amput*" OR "above-knee amput*" OR "transfemoral amput*" OR "trans-femoral amput*" OR "knee disarticulation" OR "through knee amput*" OR "through-knee amput*" OR "transtibial amput*" OR "trans-tibial amput*" OR "below knee amput*" OR "below-knee amput* OR "ankle disarticulation" OR "through ankle amput*" OR "through-ankle amput*" OR "syme*" OR "pirogoff" OR "partial foot amput*" OR "partial-foot amput*" OR "midtarsal amput*" OR "mid-tarsal amput*" OR "chopart amput*" OR "lisfranc amput*" OR "transmetatars* amput*" OR "trans-metatars* amput*" OR "transtarsal amput*" OR "trans-tarsal amput*" OR "metatarsophalangeal disarticulation" OR "ray resection" OR "digital amput*" OR "toe amput*" OR "partial toe amput*" OR "partial-toe amput*" OR "forefoot amput*" OR "midfoot amput* OR "hindfoot amput*") **OR** ((prosthe* OR "prosthe* user" OR "artificial limb" OR "leg prosthe*" OR "prosthe* limb" OR "limb prosthe*" OR "bone anchored prosthe*" OR osseointegration OR "residual limb" OR stump) **AND** ("lower limb" OR "lower-limb" OR "lower extremity" OR leg)) **OR** (amputation [MeSH: exp] OR amputation stumps [MeSH: exp] OR amputation, traumatic [MeSH: exp] OR lower extremity deformities, congenital [MeSH: exp] OR Artificial limbs [MeSH: exp] OR Bone-Anchored Prosthesis [MeSH: exp] OR Osseointegration [MeSH: exp]) **AND** Leg [MeSH: exp])
AND
("measurement tool*" OR "patient-report* outcome measure*" OR "patient report* outcome measure*" OR "patient-report* experience measure*" OR "patient report* experience measure*" OR "clinician-report* outcome measure*" OR "clinician report* outcome measure*" OR "observer-report* outcome measure*" OR "observer report* outcome measure*" OR "patient-report* outcome*" OR "patient report* outcome" OR "observer-report* measure*" OR "observer report* measure*" OR "self-report* outcome measure*" OR "self report* outcome measure*" OR "self-report* measure*" OR "self report* measure*" OR "performance-based measure*" OR "performance based measure*" OR "functional outcome measure*" OR "outcome instrument*" OR "outcome measure*" OR "clinical outcome measure*" OR "measurement instrument" OR "rehabilitation outcome measure" OR "functional assessment" OR "measurement propert*" OR "outcome assessment*" OR "outcome tool" OR "satisfaction" OR "assessment tool*" OR "predict*" OR "clinical outcome*" OR "performance-based assessment*" OR "ambulatory outcome*" OR "health survey" OR "outcome variable" OR "health status indicators" OR "participation" OR instrument* OR survey* OR questionnaire* OR "treatment outcome*" OR "abandonment and prosthe*" OR "rejection* and prosthe*") **OR** (Surveys and questionnaires [MeSH: exp] OR Quality of Life [MeSH: exp] OR Treatment outcomes [MeSH: exp] OR Patient Reported Outcome Measures [MeSH: exp] OR Outcome assessment, health care [MeSH: exp] OR Health Status [MeSH: exp] OR Patient Outcome Assessment [MeSH: exp])
AND
(valid* OR reliab* OR associat* OR relationship OR development* OR practical* OR "internal consistency" OR "ceiling effect*" OR "floor effect*" OR feasibl* OR evaluation OR quality OR scal* OR responsiveness OR interpretability OR bias OR precision OR acceptability OR psychometr* OR generali*) **OR**

(Reproducibility of results [MeSH: exp] OR Psychometrics [MeSH: exp])
Limited to English, peer-reviewed
MeSH: Medical Subject Heading; **exp:** MeSH term exploded in database search

2.2.5 Study Records

Data management

All articles found within the database searches were imported into Covidence (systematic review software management program, Melbourne, Australia) with duplicate citations removed. Data from included articles was extracted into a Microsoft Excel (Microsoft Corporation, Washington, USA) spreadsheet with statistical analyses undertaken using IBM SPSS Statistics Version 26 for Windows (IBM Corporation, NY, USA). All data was stored on a secure server.

Selection process

Two authors independently screened the title and abstract of each article identified from the initial search to determine whether a full-text review was warranted. Any discrepancies were resolved by a third author.

Data extraction

Two authors independently completed the data extraction, each completing half of the included articles. Ten percent of the data extraction were cross checked by both authors to ensure agreement. Any discrepancies were resolved by a third author.

2.2.6 Data items

The following data was extracted from each of the included articles: name of the outcome measure, the author, year, study setting, sample size, level of amputation, cause of amputation, time since amputation, sex, age, ethnicity, comorbidities, previous use of prosthesis, use of assistive devices/technology, confounding variables (i.e. contamination and/or cointervention), outcome measure(s), comparator outcome measure(s), time to complete outcome measure(s), method of completion (i.e. paper or electronic), dropouts, psychometric properties evaluated (e.g. validity, reliability, responsiveness, etc.), retest-interval, and the key statistics related to psychometric

properties. Key statistics related to psychometric properties were interpreted according to the parameters set out by Balk and colleagues (5) as shown in Table 2.2.

2.2.7 Risk of bias

Methodological quality of included articles was evaluated using COSMIN (6). Where applicable, the COSMIN 2020 update for PerfOMs was also used (7). In accordance with the COSMIN guidelines, each study was scored on a four-point scale as follows: 'very good', 'adequate', 'doubtful', or 'inadequate', for the relevant psychometric properties (e.g. reliability, validity, responsiveness, etc.). To determine the overall score for each psychometric property tested within each included study, the lowest score from the checklist criteria is taken as the overall score. Two authors independently completed the risk of bias for all included articles. This was done simultaneously to the data extraction phase. Any discrepancies were discussed during a consensus meeting and if consensus was not reached, a third author was employed to resolve the difference of opinion.

Table 2.2. Parameters for interpreting psychometric properties of included studies (adapted from Balk et al.(5))

Reliability										
Psychometric property	Internal consistency		Test-retest, inter-rater, and intra-rater							
Reporting method	Cronbach α	Rasch analysis person-separation reliability index	ICC (continuous data)	Kappa (categorical data)	Requirements					
Cut-offs	Excellent: ≥0.80 Adequate: 0.60 to 0.79 Poor (not reliable): <0.60	Excellent: ≥0.90 Good: 0.80 to 0.89	NR	Excellent: ≥0.80 Good: 0.60 to 0.79 Poor (not reliable): <0.60	Test-interval be defined, large enough, and well justified Defined training of testers and test administration					
Validity										
Psychometric property	Content (face) validity	Criterion validity	Convergent (concurrent) validity			Divergent (discriminant) validity	Predictive validity	Construct validity	Structural validity	Cross-cultural
Reporting method	Content of instrument either has: face validity (e.g. steps per day) is based on evidence-based or consensus-based process (e.g. expert panel, Delphi process, etc) or well-documented decision process Not sufficient for "overall" validity	Criterion standard scores (for norm-based scores, cited age-matched normative values, etc) Well-defined and justified criterion standard	Strength and direction of a priori correlation (p or ps [standardized])	ICC (continuous data)	Statistically significant association (*p*-value) of a priori hypothesis in regression analysis		Diagnostic test accuracy (e.g. sensitivity, specificity, area under the ROC curve) Correlation (Pearson's or Spearman's) or regression strength with future outcome	Factor analysis Principal component analysis	Factor analysis Rasch analysis IRT methods	NR

Cut-offs		Criterion validity is largely theoretical for the instruments of interest since there are not criterion standards to compare with	Large: ≥0.5 Moderate: 0.3 to 0.5 Small: 0.1 to 0.29	Excellent: ≥0.80 Good: 0.60 to 0.79 Poor (not reliable): <0.60	Weak evidence if only this analysis is reported	Low correlation (<0.1) in testing different constructs	Correlations with future events is a weaker form of evidence for predictive validity than diagnostic test accuracy. Specific cut-offs NR	N ≥10 per item Root mean square error of approximation ≤0.05 to 0.08 Standardised response means ≤0.08 Model fit measures ≥0.95	Fit statistics are between 0.5 and 1.5 (i.e. items fit the model)	NR
Responsiveness				**Measurement error**						
Psychometric property	Responsiveness statistics	MDC / MCID	Floor/ceiling effect	SEM	LOA					
Reporting method	ES with pooled SD ES with baseline SD SRM Guyatt responsiveness index ROC curve	Test-retest analyses 90% or 95% CI	≥15% of sample within the margin of error of the minimum or maximum value	NR	NR					
Cut-offs	NR	NR	NR	NR	NR					

CI: confidence interval; **ES:** effect size; **ICC:** intraclass correlation coefficient; **IRT:** item response theory; **LOA:** limits of agreement; **MCID:** minimum clinical important difference; **MDC:** minimal detectable change; **NR:** not reported; **ROC:** receiver operating characteristics; **SD:** standard deviation; **SEM:** standard error of the mean; **SRM:** Standardized response mean

2.2.8 Data synthesis, analysis, and cumulative evidence

Data synthesis and analyses was completed by two authors. Quantitative synthesis was used to summate the participant characteristics of the included studies, and presented as means, standard deviations, and percentages where appropriate. Kappa statistics were used to determine agreement between authors for both study selection and the application of the risk of bias, with results interpreted using the following cut-offs: 0.00 = no agreement, 0.01 to 0.20 = slight agreement, 0.21 to 0.40 = fair agreement, 0.41 to 0.60 = moderate agreement, 0.61 to 0.80 = substantial agreement, 0.81 to 1.00 = almost perfect agreement (8).

Qualitative synthesis was conducted to report on both the risk of bias of included studies and the psychometric properties of outcome measures. The quality appraisal of included studies was presented individually and grouped by outcome measure, using the COSMIN guidelines (6,7). Psychometric properties of included outcome measures was presented by consolidating all studies that had investigated a single psychometric property and presenting these results as a range according to the criteria defined by Balk et al.(5) Where Balk et al. has not identified psychometric cut-offs (e.g. cross-cultural validity) or there are missing statistical reporting methods for psychometric results (e.g. the use of Mann Whitney U or rs values for construct validity), extracted data was presented in its entirety to an expert panel (a group of individuals who have expertise in the development and/or validation of outcome measures commonly used in LLA populations) for input regarding the interpretation of psychometric results (See Section 3.0).

2.3 Results

2.3.1 Search results

Our search yielded 10,110 articles. After removal of duplicates 7,798 studies remained and were screened for inclusion using their title and abstract. This resulted in 294 articles moving forward to full-text review. Another 121 articles were included from hand searches of reference lists and Google Scholar. This resulted in a total of 415 articles which underwent full-text review. Of these, 108 studies were considered eligible (9-117). A total of 60 different outcome measures were identified from this search (41 patient-reported, 18 performance-based, and two hybrid outcome measures – Functional Measure for Amputees [FMA] (118) for adolescents (114) and Rivermead Mobility Index [RMI] (119)) (see Table 2.1). For study selection there was a fair agreement between the two authors (k = 0.579). The flow of the study selection process is presented in Figure 2.2.

2.3.2 Study characteristics

The 108 eligible studies included 15,134 participants with LLA and were conducted between 1981 and 2020. Sample sizes ranged from five to 3256, with majority of studies being conducted in the United States of America (n = 22), Canada (n = 15), Italy (n = 7), The Netherlands (n = 6), and the United Kingdom (n = 5). A summary of the study characteristics can be viewed in Table 2.3.

Table 2.1. Outcome measures identified from the systematic review

PROMs (n = 40)	PerfOMs (n = 18)	Hybrid (n = 2)
1. Amputee Activity Score (AAS) (9, 10) 2. Activity Balance Confidence (ABC) Scale (11-14) 3. Amputee Body Image Scale (ABIS) (15-19) 4. Amputee Single Item Mobility Measure (AMPSIMM) (20) 5. Child Amputee Prosthetics Project – Functional Status Inventory (CAPP-FSI) (111-113) 6. Centre for Epidemiologic Studies-Depression (CES-D) Scale (21) 7. Comprehensive Lower-limb Amputee Socket Survey (CLASS) (22) 8. Climbing Stairs Questionnaire (23) 9. Functional and social performance checklist for lower limb individuals with amputation (DSF-84) (24) 10. Engagement in Everyday Activities involving Revealing the Body scale (EEARB) (25) 11. Functional Independence Measure (FIM) (9, 26, 27) 12. Frenchay Activities Index (FAI) (29) 13. Harold Wood-Stanmore Mobility Scale / Amputee Medical Rehabilitation Society (AMRS) Scale (30) 14. Houghton Scale (31, 32, 120)	1. 10-metre (timed) walk test (10mWT) (90, 106, 107) 2. Two-minute walk test (2MWT) (46, 103, 104) 3. 3D Gait Analysis Gait Deviation Index (3D GAGDI) (105) 4. Six-minute Walk Test (6MWT) (46, 99, 100) 5. 180 degree turn test (38) 6. Assessment of Daily Activity Performance in Transfemoral Amputee (ADAPT) (79) 7. Amputee Mobility Predictor (AMP) – inc AMPPRO and AMPnoPRO (46, 80, 81) 8. Berg Balance Scale (BBS) (82-85) 9. Comprehensive High-Level Activity Mobility Predictor (CHAMP) (86-88) 10. Figure 8 Walking Test (F8WT) (89) 11. Four Square Step Test (FSST) (38, 90) 12. Functional Reach Test (91) 13. Hill Assessment Index (HAI) (92, 93) 14. L-Test (94-96) 15. Narrowing Beam Walking Test (NBWT) (90, 97) 16. Physiological Cost Index (PCI) (98) 17. Stair Assessment Index (SAI) (93)	1. Functional Measure for Amputees (FMA) (9, 26, 27, 114, 118) 2. Rivermead Mobility Index (RMI) (108, 109)

15. Locomotor Capabilities Index (LCI) (34-42, 110)
16. Limb Laterality Recognition Score (43)
17. Multidimensional Body-Self Relations Questionnaire (MBSRQ) (44)
18. Office of Population Censuses and Surveys (OPCS) Scale (9)
19. Orthotics and Prosthetics National Office Outcomes Tool (OPOT) (45)
20. Orthotic Prosthetic User Survey (OPUS) - Lower Limb Functional Measure (LLFM) (46, 47)
21. OPUS - Health Related Quality of Life (HR-QoL) (46, 47)
22. OPUS - Satisfaction with Device Scale (SDS) (46, 47)
23. OPUS - Satisfaction with Services (47)
24. Prosthesis Donning and Doffing Questionnaire (PDDQ) (48, 49)
25. Prosthesis Embodiment Scale - Lower Limb Amputees (PEmbS-LLA) (50)
26. Prosthesis Evaluation Questionnaire (PEQ) (13, 34, 36, 46, 51-56)
27. Patient Generated Index (PGI) - modified for amputees (28)
28. The Prosthetic Limb Users Survey of Mobility (PLUS-M) (13, 57-59)
29. Prosthetic Mobility Questionnaire (PMQ) including the PMQ 2.0 (60, 61)
30. Prosthetic Profile of the Amputee (PPA) (33, 36, 62)
31. Patient Reported Outcomes Measurement Information System (PROMIS-29) (13, 63)
18. Timed Up and Go (TUG) (38, 46, 90, 101, 102)

32. Patient Specific Function Scale (PSFS) (46)
33. Questionnaire for Persons with a Transfemoral Amputation (Q-TFA) (64)
34. Questionnaire Rising and Sitting Down (QR&S) (65)
35. Questionnaire to Explore Human Factors and their Technical potential (QEHFT) (66)
36. Socket Comfort Score (SCS) (13, 67)
37. Short Form 36 Items - Veteran population (SF-36V) (46)
38. Special Interest Group in Amputee Medicine (SIGAM) Mobility Scale (68-71)
39. Trinity Amputation and Prosthesis Experience Scales (TAPES) including the TAPRES-Revised (TAPES-R) (72-77)
40. Walking questionnaire (78)

PROMs: patient-reported outcome measures; **PerfOMs:** performance-based outcome measures; **Hybrid:** combination of a PROM and a PerfOM

Figure 2.2. Study selection process

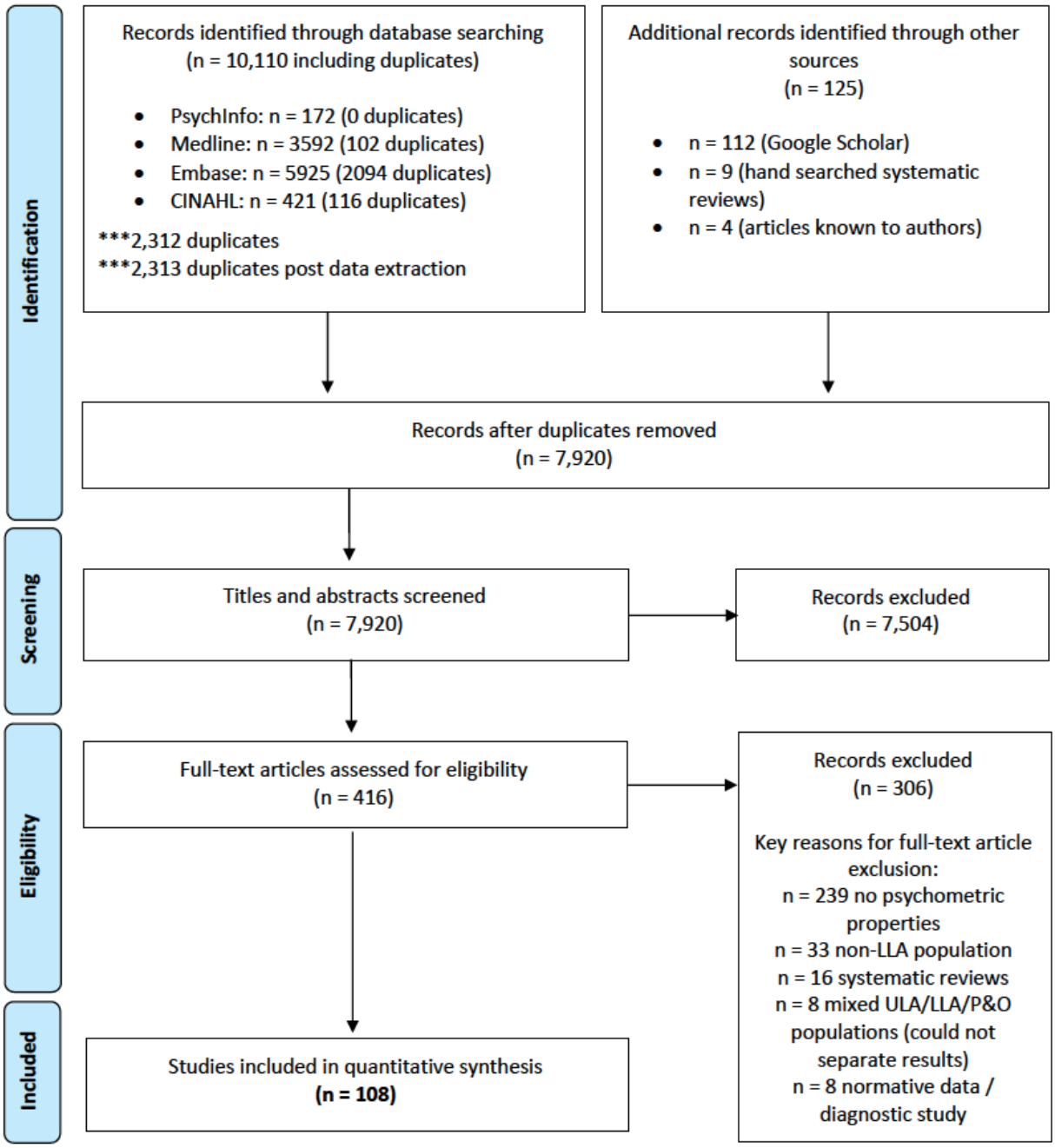


Table 2.3. Study characteristics

Study characteristics	
Sex, n	
Male	8024
Female	3140
Age, mean (SD) in years	55.3 (13.7)
Setting (country)	Australia = 2; Brazil = 1; Belgium = 1; Canada = 15; Denmark = 1; Finland = 3; France = 3; Germany = 3; Iran = 2; Ireland = 4; Italy = 7; Japan = 1; Jordan = 1; Malaysia = 1; The Netherlands = 6; Norway = 1; Slovenia = 1; Sweden = 3; Switzerland = 2; Turkey = 3; United Kingdom = 5; United States of America = 22; NR = 26
Outcome measures	
PROM	40
PerfOM	18
Hybrid	2
Number of outcome measures assessed	
One	85
Two	15
Three	4
Four	2
Five	1
Six	0
Seven	0
Eight	1
Time since amputation, mean (SD) in months	95 (87)
Level of amputation	
Partial foot amputation	51
Symes	30
Transtibial amputation	5659
Knee disarticulation	320
Transfemoral amputation	3126
Hip disarticulation	46
Other	146
Unilateral	8965
Bilateral	525

Psychometric properties assessed	
PROM development	21
Content validity	7
Structural validity	18
Internal consistency	44
Cross-cultural validity	3
Reliability	
PROM	38
PerfOM	18
Hybrid	2
Measurement error	
PROM	11
PerfOM	8
Hybrid	0
Criterion validity	2
Construct validity	
Convergent validity	65
Discriminative validity	41
Responsiveness	9

PerfOM: performance-based outcome measure; **PROM:** patient-reported outcome measure; **NR:** not reported

2.3.3 Methodological quality assessment

Quality appraisal of included studies reporting on the psychometric properties of outcome measures ranged from 'inadequate' through to 'very good' according to COSMIN (6, 7). No outcome measure scored 'very good' for all psychometric properties assessed. Internal consistency was the psychometric property which demonstrated the best methodological quality of the included studies, with 28 out of 34 outcome measures scoring 'very good'. All PerfOMs which assessed outcome measure development, content validity, structural validity, and responsiveness, scored 'inadequate'. Assessment of the methodological quality of included studies resulted in a kappa coefficient of moderate to almost perfect agreement (k = 0.566 to 0.859), with the overall kappa coefficient demonstrating substantial agreement (k = 0.710). Discussion for solving and addressing sources of conflicts using COSMIN resulted in the harmonisation of agreement between reviewers (k = 1.000).

2.3.4 Synthesis of results and inception of an expert panel

Qualitative synthesis was undertaken to report on both the quality appraisal of included studies and the psychometric properties of outcome measures. Psychometric properties of included outcome measures were synthesised by consolidating all studies that had investigated a single psychometric property and presenting these results as a range according to a pre-defined criterion by Balk and

colleagues (5). As the outcome measure consensus arm of the project aimed to include a variety of participants ranging from prosthetic-users (i.e. individuals with LLA who have experience using prosthetic devices and/or services), clinicians, health policy makers, health clinic managers and researchers, it was anticipated that many participants would not have the sound knowledge or background understanding required to adequately interpret the methodological quality of included studies (i.e. the COSMIN results) or the psychometric properties of the 60 outcome measures identified, and incorporate these findings into their consensus votes. Therefore, it was decided than an expert panel was required to assist with improving the interpretation of the results to facilitate the incorporation of the psychometric strengths and weaknesses into the consensus participants votes. Individuals were deemed experts if they had experience in the development and/or validation of outcome measures commonly used in LLA populations. Of the 60 outcome measures identified in the systematic review, the goal of the expert panel was to identify which of these 60 contained adequate psychometric properties and present them in a clear and translatable format to the consensus participants. It was expected that the expert panel process would significantly reduce the number of outcome measures to be considered during the consensus meetings. The process and outcomes of the expert panel are described in Section 3.0 below.

References

1. Robinson C, Fatone S. You've heard about outcome measures, so how do you use them? Integrating clinically relevant outcome measures in orthotic management of stroke. Prosthet Orthot Int. 2012;37(1):30-42.
2. Gaunaurd I, Spaulding SE, Amtmann D, Salem R, Gailey R, Morgan SJ, et al. Use of and confidence in administering outcome measures among clinical prosthetists: Results from a national survey and mixed-methods training program. Prosthet Orthot Int. 2015;39(4):314-21.
3. Miller LA, McCay JA. Summary and Conclusions From the Academy's Sixth State-of-the-Science Conference on Lower Limb Prosthetic Outcome Measures. J Prosthet Orthot. 2006;18(6):2-7.
4. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. Int J Surg. 2010;8(5):336-41.
5. Balk EM, Gazula A, Markozannes G, Kimmel HJ, Saldanha IJ, Trikalinos TA, et al. Psychometric Properties of Functional, Ambulatory, and Quality of Life Instruments in Lower Limb Amputees: A Systematic Review. Arch Phys Med Rehabil. 2019;100(12):2354-70.
6. Mokkink LB, De Vet HC, Prinsen CA, Patrick DL, Alonso J, Bouter LM, et al. COSMIN risk of bias checklist for systematic reviews of patient-reported outcome measures. Qual Life Res. 2018;27(5):1171-9.
7. Mokkink LB, Boers M, Vleuten Cvd, Bouter L, Alonso J, Patrick DL, et al. COSMIN Risk of Bias tool to assess the quality of studies on reliability or measurement error of outcome measurement instruments: a Delphi study. BMC Med Res Methodol. 2020;20(1):1-13.
8. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics. 1977:159-74.
9. Panesar BS, Morrison P, Hunter J. A comparison of three measures of progress in early lower limb amputee rehabilitation. Clin Rehabil. 2001;15(2):157-71.
10. Day HJ. The assessment and description of amputee activity. Prosthet Orthot Int. 1981;5(1):23-8.
11. Miller WC, Deathe AB, Speechley M. Psychometric properties of the Activities-specific Balance Confidence Scale among individuals with a lower-limb amputation. Arch Phys Med Rehabil. 2003;84(5):656-61.
12. Sakakibara BM, Miller WC, Backman CL. Rasch analyses of the Activities-specific Balance Confidence Scale with individuals 50 years and older with lower-limb amputations. Arch Phys Med Rehabil. 2011;92(8):1257-63.

13. Hafner BJ, Morgan SJ, Askew RL, Salem R. Psychometric evaluation of self-report outcome measures for prosthetic applications. J Rehabil Res Dev. 2016;53(6):797-812.
14. Fuller K, Omaña Moreno HA, Frengopoulos C, Payne MW, Viana R, Hunter SW. Reliability, validity, and agreement of the short-form Activities-specific Balance Confidence Scale in people with lower extremity amputations. Prosthet Orthot Int. 2019;43(6):609-17.
15. Breakey JW. Body image: The lower-limb amputee. J Prosthet Orthot. 1997;9(2):58-66.
16. Lai FH-y, Wong E, Wong SK-m, Soo AK-w, Yip KT-f, Chan CK-c, et al. Development and validation of a Body Image Assessment for patient after lower limb amputation—the Chinese Amputee Body Image Scale-CABIS. Asian J Occup Ther. 2005;4(1):1-11.
17. Gallagher P, Horgan O, Franchignoni F, Giordano A, MacLachlan M. Body image in people with lower-limb amputation: a Rasch analysis of the Amputee Body Image Scale. Am J Phys Med Rehabil. 2007;86(3):205-15.
18. Bumin G, Bayramlar K, Yakut Y, Sener GY. Cross cultural adaptation and reliability of the Turkish version of Amputee Body Image Scale (ABIS). J Back Musculoskelet Rehabil. 2009;22(1):11-6.
19. Safaz I, Yilmaz B, Goktepe AS, Yazicioglu K. Turkish version of the Amputee Body Image Scale and relationship with quality of life. Bulletin of clinical psychopharmacology. 2010;20(1):79-83.
20. Norvell DC, Williams RM, Turner AP, Czerniecki JM. The development and validation of a novel outcome measure to quantify mobility in the dysvascular lower extremity amputee: The amputee single item mobility measure. Clin Rehabil. 2016;30(9):878-89.
21. Rybarczyk BD, Nyenhuis DL, Nicholas JJ, Schulz R, Alioto RJ, Blair C. Social discomfort and depression in a sample of adults with leg amputations. Arch Phys Med Rehabil. 1992;73(12):1169-73.
22. Gailey R, Kristal A, Lucarevic J, Harris S, Applegate B, Gaunaurd I. The development and internal consistency of the comprehensive lower limb amputee socket survey in active lower limb amputees. Prosthet Orthot Int. 2019;43(1):80-7.
23. de Laat FA, Rommers GM, Geertzen JH, Roorda LD. Construct validity and test-retest reliability of the climbing stairs questionnaire in lower-limb amputees. Arch Phys Med Rehabil. 2010;91(9):1396-401.
24. Monteiro RP, Pfeifer LI, Soares I, Dos Santos Ade A, Sousa N. Validation of the functional and social performance - DSF-84 checklist: preliminary study. Disabil Rehabil. 2013;35(18):1527-33.
25. Donovan-Hall MK, Yardley L, Watts RJ. Engagement in activities revealing the body and psychosocial adjustment in adults with a trans-tibial prosthesis. Prosthet Orthot Int. 2002;26(1):15-22.
26. Leung EC, Rush PJ, Devlin M. Predicting prosthetic rehabilitation outcome in lower limb amputee patients with the functional independence measure. Arch Phys Med Rehabil. 1996;77(6):605-8.
27. Stineman MG, Shea JA, Jette A, Tassoni CJ, Ottenbacher KJ, Fiedler R, et al. The Functional Independence Measure: tests of scaling assumptions, structure, and reliability across 20 diverse impairment categories. Arch Phys Med Rehabil. 1996;77(11):1101-8.
28. Callaghan BG, Condie ME. A post-discharge quality of life outcome measure for lower limb amputees: test-retest reliability and construct validity. Clin Rehabil. 2003;17(8):858-64.
29. Miller WC, Deathe AB, Harris J. Measurement properties of the Frenchay Activities Index among individuals with a lower limb amputation. Clin Rehabil. 2004;18(4):414-22.
30. Gardiner MD, Faux S, Jones LE. Inter-observer reliability of clinical outcome measures in a lower limb amputee population. Disabil Rehabil. 2002;24(4):219-25.
31. Devlin M, Pauley T, Head K, Garfinkel S. Houghton Scale of prosthetic use in people with lower-extremity amputations: Reliability, validity, and responsiveness to change. Arch Phys Med Rehabil. 2004;85(8):1339-44.
32. Wong CK, Gibbs W, Chen ES. Use of the Houghton Scale to Classify Community and Household Walking Ability in People With Lower-Limb Amputation: Criterion-Related Validity. Arch Phys Med Rehabil. 2016;97(7):1130-6.
33. Gauthier-Gagnon C, Grise MC. Prosthetic profile of the amputee questionnaire: validity and reliability. Arch Phys Med Rehabil. 1994;75(12):1309-14.
34. Miller WC, Deathe AB, Speechley M. Lower extremity prosthetic mobility: a comparison of 3 self-report scales. Arch Phys Med Rehabil. 2001;82(10):1432-40.
35. Franchignoni F, Orlandini D, Ferriero G, Moscato TA. Reliability, validity, and responsiveness of the locomotor capabilities index in adults with lower-limb amputation undergoing prosthetic training. Arch Phys Med Rehabil. 2004;85(5):743-8.
36. Ferriero G, Dughi D, Orlandini D, Moscato T, Nicita D, Franchignoni F. Measuring long-term outcome in people with lower limb amputation: cross-validation of the Italian versions of the Prosthetic Profile of the Amputee and Prosthesis Evaluation Questionnaire. Eura Medicophys. 2005;41(1):1-6.
37. Franchignoni F, Giordano A, Ferriero G, Munoz S, Orlandini D, Amoresano A. Rasch analysis of the Locomotor Capabilities Index-5 in people with lower limb amputation. Prosthet Orthot Int. 2007;31(4):394-404.
38. Dite W, Connor HJ, Curtis HC. Clinical identification of multiple fall risk early after unilateral transtibial amputation. Arch Phys Med Rehabil. 2007;88(1):109-14.
39. Larsson B, Johannesson A, Andersson IH, Atroshi I. The Locomotor Capabilities Index; validity and reliability of the Swedish version in adults with lower limb amputation. Health Qual Life Outcomes. 2009;7(1):1-9.

40. Salavati M, Mazaheri M, Khosrozadeh F, Mousavi SME, Negahban H, Shojaei H. The Persian version of Locomotor Capabilities Index: Translation, reliability and validity in individuals with lower limb amputation. Qual Life Res. 2011;20(1):1-7.
41. Becker P, Repo JP, Piitulainen K, Tukiainen EJ, Roine RP, Järvenpää S, et al. Validity and Reliability of the Finnish Version of the Locomotor Capabilities Index-5 in Patients Fitted with a Prosthesis after Major Lower Extremity Amputation. J Prosthet Orthot. 2019;31(1):43-50.
42. Franchignoni F, Traballesi M, Monticone M, Giordano A, Brunelli S, Ferriero G. Sensitivity to change and minimal clinically important difference of the Locomotor Capabilities Index-5 in people with lower limb amputation undergoing prosthetic training. Ann Phys Rehabil Med. 2019;62(3):137-41.
43. Wong CK. Limb laterality recognition score: A reliable clinical measure related to phantom limb pain. Pain Medicine (United States). 2018;19(4):753-6.
44. Wetterhahn KA, Hanson C, Levy CE. Effect of participation in physical activity on body image of amputees. Am J Phys Med Rehabil. 2002;81(3):194-201.
45. Hart DL. Orthotics and Prosthetics National Outcomes Tool (OPOT): initial reliability and validity assessment for lower extremity prosthetics. J Prosthet Orthot. 1999;11(4):101-11.
46. Resnik L, Borgia M. Reliability of outcome measures for people with lower-limb amputations: distinguishing true change from statistical error. Phys Ther. 2011;91(4):555-65.
47. Jarl G, Heinemann AW, Lindner HY, Norling Hermansson LM. Cross-Cultural Validity and Differential Item Functioning of the Orthotics and Prosthetics Users' Survey With Swedish and United States Users of Lower-Limb Prosthesis. Arch Phys Med Rehabil. 2015;96(9):1615-26.
48. Abu Osman NA, Eshraghi A, Gholizadeh H, Wan Abas WAB, Lechler K. Prosthesis donning and doffing questionnaire: Development and validation. Prosthet Orthot Int. 2017;41(6):571-8.
49. Giray E, Eshraghi A, Gündüz OH. Validity and Reliability of the Turkish Version of the Prosthesis Donning and Doffing Questionnaire for Persons with Transtibial Amputations. J Prosthet Orthot. 2021;April 29.
50. Bekrater-Bodmann R. Perceptual correlates of successful body-prosthesis interaction in lower limb amputees: psychometric characterisation and development of the Prosthesis Embodiment Scale. Sci Rep. 2020;10(1):14203.
51. Legro MW, Reiber GD, Smith DG, del Aguila M, Larsen J, Boone D. Prosthesis evaluation questionnaire for persons with lower limb amputations: assessing prosthesis-related quality of life. Arch Phys Med Rehabil. 1998;79(8):931-8.
52. Franchignoni F, Giordano A, Ferriero G, Orlandini D, Amoresano A, Perucca L. Measuring mobility in people with lower limb amputation: Rasch analysis of the mobility section of the prosthesis evaluation questionnaire. J Rehabil Med. 2007;39(2):138-44.
53. Christensen J, Doherty P, Bjorner JB, Langberg H. Reliability and construct validity of a new Danish translation of the Prosthesis Evaluation Questionnaire in a population of Danish amputees. Prosthet Orthot Int. 2017;41(5):469-75.
54. Repo JP, Piitulainen K, Hakkinen A, Roine RP, Kautiainen H, Becker P, et al. Reliability and validity of the Finnish version of the prosthesis evaluation questionnaire. Disabil Rehabil. 2018;40(17):2081-7.
55. Franchignoni F, Ferriero G, Giordano A, Monticone M, Grioni G, Burger H. The minimal clinically-important difference of the Prosthesis Evaluation Questionnaire - Mobility Scale in subjects undergoing lower limb prosthetic rehabilitation training. Eur J Phys Rehabil Med. 2020;56(1):82-7.
56. Repo JP, Häkkinen AH, Barner-Rasmussen I, Lahdenperä N-I, Kautiainen H, Roine RP, et al. Comparison of Measurement Properties of the Mobility Section of the Prosthesis Evaluation Questionnaire on a Visual Analogue Scale and Its Modified Version with Five Response Categories. J Prosthet Orthot. 2020;32(3):188-93.
57. Morgan SJ, Amtmann D, Abrahamson DC, Kajlich AJ, Hafner BJ. Use of cognitive interviews in the development of the PLUS-M item bank. Qual Life Res. 2014;23(6):1767-75.
58. Hafner BJ, Morgan SJ, Abrahamson DC, Amtmann D. Characterizing mobility from the prosthetic limb user's perspective: Use of focus groups to guide development of the Prosthetic Limb Users Survey of Mobility. Prosthet Orthot Int. 2016;40(5):582-90.
59. Karatzios C, Loiret I, Luthi F, Leger B, Le Carre J, Saubade M, et al. Transcultural adaptation and validation of a French version of the Prosthetic Limb Users Survey of Mobility 12-item Short-Form (PLUS-M/FC-12) in active amputees. Ann Phys Rehabil Med. 2019;62(3):142-8.
60. Franchignoni F, Monticone M, Giordano A, Rocca B. Rasch validation of the Prosthetic Mobility Questionnaire: A new outcome measure for assessing mobility in people with lower limb amputation. J Rehabil Med. 2015;47(5):460-5.
61. Burger H, Giordano A, Bavec A, Franchignoni F. The Prosthetic Mobility Questionnaire, a tool for assessing mobility in people with lower-limb amputation: validation of PMQ 2.0 in Slovenia. Int J Rehabil Res. 2019;42(3):263-9.
62. Grise MC, Gauthier-Gagnon C, Martineau GG. Prosthetic profile of people with lower extremity amputation: conception and design of a follow-up questionnaire. Arch Phys Med Rehabil. 1993;74(8):862-70.
63. Amtmann D, Morgan SJ, Kim J, Hafner BJ. Health-related profiles of people with lower limb loss. Arch Phys Med Rehabil. 2015;96(8):1474-83.
64. Hagberg K, Branemark R, Hagg O. Questionnaire for Persons with a Transfemoral Amputation (Q-TFA): initial validity and reliability of a new outcome measure. J Rehabil Res Dev. 2004;41(5):695-706.
65. de Laat FA, Rommers GM, Geertzen JH, Roorda LD. Construct validity and test-retest reliability of the questionnaire rising and sitting down in lower-limb amputees. Arch Phys Med Rehabil. 2011;92(8):1305-10.

66. Herbst J, Polanski-Schrader LMT, Vogt J, Schurmann T, Beckerle P. Validation and revision of the questionnaire to explore human factors and their technical potential for lower limb prosthetics. Prosthet Orthot Int. 2020:1-8.
67. Hanspal RS, Fisher K, Nieveen R. Prosthetic socket fit comfort score. Disabil Rehabil. 2003;25(22):1278-80.
68. Ryall N, Eyres S, Neumann V, Bhakta B, Tennant A. The SIGAM mobility grades: a new population-specific measure for lower limb amputees. Disabil Rehabil. 2003;25(15):833-44.
69. Joussain C, Laroche D, Casillas JM, Paysant J, Ader P, Bastable P, et al. Transcultural validation of the SIGAM mobility grades in French: The SIGAM-Fr. Ann Phys Rehabil Med. 2015;58(3):161-6.
70. Yilmaz H, Gafuroglu U, Ryall N, Yuksel S. Establishing the Turkish version of the SIGAM mobility scale, and determining its validity and reliability in lower extremity amputees. Disabil Rehabil. 2018;40(3):346-52.
71. de Laat FA, Roorda LD, Geertzen JH, Rommers C. Test-retest reliability of the Special Interest Group on Amputation Medicine/Dutch Working Group on Amputations and Prosthetics Mobility Scale, in persons wearing a prosthesis after a lower-limb amputation. Disabil Rehabil. 2020;42(12):1762-6.
72. Gallagher P, MacLachlan M. Development and psychometric evaluation of the trinity amputation and prosthesis experience scales (TAPES). Rehabil Psychol. 2000;45(2):130-54.
73. Gallagher P, Maclachlan M. The Trinity Amputation and Prosthesis Experience Scales and quality of life in people with lower-limb amputation. Arch Phys Med Rehabil. 2004;85(5):730-6.
74. Mazaheri M, Fardipour S, Salavati M, Hadadi M, Negahban H, Bahramizadeh M, et al. The Persian version of Trinity Amputation and Prosthetics Experience Scale: translation, factor structure, reliability and validity. Disabil Rehabil. 2011;33(19-20):1737-45.
75. Gallagher P, Franchignoni F, Giordano A, MacLachlan M. Trinity amputation and prosthesis experience scales: a psychometric assessment using classical test theory and rasch analysis. Am J Phys Med Rehabil. 2010;89(6):487-96.
76. Massarweh R, Sobuh MMD. The Arabic Version of Trinity Amputation and Prosthetic Experience Scale - Revised (TAPES-R) for Lower Limb Amputees: Reliability and Validity. Disability, CBR & Inclusive Development. 2019;30(1):44-56.
77. Luthi F, Praz C, Leger B, Vouilloz A, Favre C, Loiret I, et al. Cross-cultural adaptation and measurement properties of the French version of the Trinity Amputation and Prosthesis Experience Scales-Revised (TAPES-R). PLoS ONE [Electronic Resource]. 2020;15(2):e0229084.
78. de Laat FA, Rommers GM, Geertzen JH, Roorda LD. Construct validity and test-retest reliability of the walking questionnaire in people with a lower limb amputation. Arch Phys Med Rehabil. 2012;93(6):983-9.
79. Theeven P, Hemmen B, Stevens C, Ilmer E, Brink P, Seelen H. Feasibility of a new concept for measuring actual functional performance in daily life of transfemoral amputees. J Rehabil Med. 2010;42(8):744-51.
80. Gailey RS, Roach KE, Applegate EB, Cho B, Cunniffe B, Licht S, 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):613-27.
81. Côté-Martin M-É, Tremblay A, Couture M, Roy J-S. Translation, Reliability, and Validity of the French Version of the Amputee Mobility Predictor. J Prosthet Orthot. 2020;32(2):101-6.
82. Wong CK, Chen CC, Welsh J. Preliminary assessment of balance with the Berg Balance Scale in adults who have a leg amputation and dwell in the community: Rasch rating scale analysis. Phys Ther. 2013;93(11):1520-9.
83. Major MJ, Fatone S, Roth EJ. Validity and reliability of the Berg Balance Scale for community-dwelling persons with lower-limb amputation. Arch Phys Med Rehabil. 2013;94(11):2194-202.
84. Wong CK. Interrater reliability of the Berg Balance Scale when used by clinicians of various experience levels to assess people with lower limb amputations. Phys Ther. 2014;94(3):371-8.
85. Azuma Y, Chin T, Miura Y. The relationship between balance ability and walking ability using the Berg Balance Scale in people with transfemoral amputation. Prosthet Orthot Int. 2019;43(4):396-401.
86. Gailey RS. Construct validity of Comprehensive High-Level Activity Mobility Predictor (CHAMP) for male servicemembers with traumatic lower-limb loss. J Rehabil Res Dev. 2013;50(7):919-30.
87. Gailey RS. Development and reliability testing of the Comprehensive High-Level Activity Mobility Predictor (CHAMP) in male servicemembers with traumatic lower-limb loss. J Rehabil Res Dev. 2013;50(7):905-18.
88. Anton A, Legault Z, Dudek N. Validity of the Comprehensive High-Level Activity Mobility Predictor in a heterogeneous population with lower extremity amputations. Prosthet Orthot Int. 2020;44(2):60-5.
89. Schack J, Mirtaheri P, Steen H, Gjovaag T. Assessing mobility for persons with lower limb amputation: the Figure-of-Eight Walk Test with the inclusion of two novel conditions. Disabil Rehabil. 2019;43(9):1323-32.
90. Sawers A, Kim J, Balkman G, Hafner BJ. Interrater and Test-Retest Reliability of Performance-Based Clinical Tests Administered to Established Users of Lower Limb Prostheses. Phys Ther. 2020;100(7):1206-16.
91. Gremeaux V, Damak S, Troisgros O, Feki A, Laroche D, Perennou D, et al. Selecting a test for the clinical assessment of balance and walking capacity at the definitive fitting state after unilateral amputation: a comparative study. Prosthet Orthot Int. 2012;36(4):415-22.
92. Highsmith MJ, Kahle JT, Miro RM, Mengelkoch LJ. Ramp descent performance with the C-Leg and interrater reliability of the Hill Assessment Index. Prosthet Orthot Int. 2013;37(5):362-8.
93. Highsmith MJ, Kahle JT, Kaluf B, Miro RM, Mengelkoch LJ, Klenow TD. Psychometric evaluation of the hill assessment index (HAI) and stair assessment index (SAI) in high-functioning transfemoral amputees. Technol Innov. 2016;18(2-3):193.

94. Deathe AB, Miller WC. The L test of functional mobility: measurement properties of a modified version of the timed "up & go" test designed for people with lower-limb amputations. Phys Ther. 2005;85(7):626-35.
95. Rushton PW, Miller WC, Deathe AB. Minimal clinically important difference of the L Test for individuals with lower limb amputation: A pilot study. Prosthet Orthot Int. 2015;39(6):470-6.
96. Hunter SW, Frengopoulos C, Holmes J, Viana R, Payne MW. Determining Reliability of a Dual-Task Functional Mobility Protocol for Individuals With Lower Extremity Amputation. Arch Phys Med Rehabil. 2018;99(4):707-12.
97. Sawers A, Hafner B. Validation of the Narrowing Beam Walking Test in Lower Limb Prosthesis Users. Arch Phys Med Rehabil. 2018;99(8):1491-8.e1.
98. Hagberg K, Tranberg R, Zügner R, Danielsson A. Reproducibility of the Physiological Cost Index among Individuals with a Lower-Limb Amputation and Healthy Adults. Physiother Res Int. 2011;16(2):92-100.
99. Lin SJ, Bose NH. Six-minute walk test in persons with transtibial amputation. Arch Phys Med Rehabil. 2008;89(12):2354-9.
100. Reid L, Thomson P, Besemann M, Dudek N. Going places: Does the two-minute walk test predict the six-minute walk test in lower extremity amputees? J Rehabil Med. 2015;47(3):256-61.
101. Schoppen T, Boonstra A, Groothoff JW, de Vries J, Goeken LN, Eisma WH. The Timed "up and go" test: reliability and validity in persons with unilateral lower limb amputation. Arch Phys Med Rehabil. 1999;80(7):825-8.
102. Clemens SM, Gailey RS, Bennett CL, Pasquina PF, Kirk-Sanchez NJ, Gaunaurd IA. The Component Timed-Up-and-Go test: the utility and psychometric properties of using a mobile application to determine prosthetic mobility in people with lower limb amputations. Clin Rehabil. 2018;32(3):388-97.
103. 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):1478-83.
104. Brooks D, Hunter JP, Parsons J, Livsey E, Quirt J, Devlin M. Reliability of the two-minute walk test in individuals with transtibial amputation. Arch Phys Med Rehabil. 2002;83(11):1562-5.
105. Kark L, Odell R, McIntosh AS, Simmons A. Quantifying prosthetic gait deviation using simple outcome measures. World J Orthop. 2016;7(6):383.
106. Boonstra A, Fidler V, Eisma W. Walking speed of normal subjects and amputees: aspects of validity of gait analysis. Prosthet Orthot Int. 1993;17(2):78-82.
107. Datta D, Ariyaratnam R, Hilton S. Timed walking test - an all-embracing outcome measure for lower-limb amputees? Clin Rehabil. 1996;10(3):227-32.
108. Franchignoni F, Brunelli S, Orlandini D, Ferriero G, Traballesi M. Is the Rivermead Mobility Index a suitable outcome measure in lower limb amputees?--A psychometric validation study. J Rehabil Med. 2003;35(3):141-4.
109. Ryall NH, Eyres SB, Neumann VC, Bhakta BB, Tennant A. Is the Rivermead Mobility Index appropriate to measure mobility in lower limb amputees? Disabil Rehabil. 2003;25(3):143-53.
110. Gauthier-Gagnon C, Grise M, Lepage Y. The locomotor capabilities index: content validity. J Rehabil Outcomes Meas. 1998;2(4):40-6.
111. Pruitt SD, Varni JW, Setoguchi Y. Functional status in children with limb deficiency: development and initial validation of an outcome measure. Arch Phys Med Rehabil. 1996;77(12):1233-8.
112. Pruitt SD, Varni JW, Seid M, Setoguchi Y. Functional status in limb deficiency: Development of an outcome measure for preschool children. Arch Phys Med Rehabil. 1998;79(4):405-11.
113. Pruitt SD, Seid M, Varni JW, Setoguchi Y. Toddlers with limb deficiency: conceptual basis and initial application of a functional status outcome measure. Arch Phys Med Rehabil. 1999;80(7):819-24.
114. Pierce S, Fergus A, Brady B, Wolff-Burke M. Examination of the functional mobility assessment tool for children and adolescents with lower extremity amputations. Pediatr Phys Ther. 2011;23(2):171-7.
115. Hafner BJ, Gaunaurd IA, Morgan SJ, Amtmann D, Salem R, Gailey RS. Construct Validity of the Prosthetic Limb Users Survey of Mobility (PLUS-M) in Adults With Lower Limb Amputation. Arch Phys Med Rehabil. 2017;98(2):277-85.
116. Topuz S, Ulger O, Yakut Y, Gul Sener F. Reliability and construct validity of the Turkish version of the Trinity Amputation and Prosthetic Experience Scales (TAPES) in lower limb amputees. Prosthet Orthot Int. 2011;35(2):201-6.
117. Callaghan BG, Sockalingam S, Treweek SP, Condie ME. A post-discharge functional outcome measure for lower limb amputees: Test-retest reliability with trans-tibial amputees. Prosthet Orthot Int. 2002;26(2):113-9.
118. Marchese VG, Rai SN, Carlson CA, Hinds PS, Spearing EM, Zhang L, et al. Assessing functional mobility in survivors of lower-extremity sarcoma: reliability and validity of a new assessment tool. Pediatr Blood Cancer. 2007;49(2):183-9.
119. Collen FM, Wade DT, Robb G, Bradshaw C. The Rivermead mobility index: a further development of the Rivermead motor assessment. Int Disabil Stud. 1991;13(2):50-4.
120. Miller WC, Deathe AB, Speechley M. Lower extremity prosthetic mobility: A comparison of 3 self-report scales. Arch Phys Med Rehabil. 2001;82(10):1432-40.

3.0 Expert Panel

3.1 Background

The systematic review found 60 outcome measures that have had their psychometric properties evaluated in people with LLA. A wider array of outcome measures is often used in both clinical and research settings that have not had their psychometric properties reported in this population.

Ensuring that outcome measures considered for selection within the COMPASS had adequate psychometric properties was imperative. If an outcome measure has poor psychometric properties, there will be concerns about the validity and reliability of the test results. Validity is the degree to which the outcome measure tests what it proports to. Reliability reflects the degree to which the score or result is accurate compared to the actual construct being measured. All outcome measures will have less than perfect validity and reliability but outcome measures with poor psychometrics will result in an increased amount of error and hence uncertainty in interpreting results. Decisions based on data of poor quality will be of lesser quality at all levels, the clinical encounter, management, and policy. There is no standard way to determine if the psychometric properties of an outcome measures are adequate for routine use.

Reaching consensus on which outcome measures should be routinely used for people with LLA is imperative so that these outcome measures can be actively promoted, used regularly in clinical care, included in quality improvement programs and research studies, used for benchmarking across population groups and services, and applied to answer policy questions about efficacy and cost effectiveness of rehabilitation interventions. Given that numerous researchers have made significant contributions to the development, validation and implementation of a number of outcome measures in this population it is not surprising that publications alone have not yet resulted in a consolidated list of recommended outcome measures. To collate a consolidated agreed list of outcome measures is best done by seeking broad agreement among the major stakeholders. Researchers who have developed and/or tested outcome measures have extensive skills and are therefore a logical first group to advise on the psychometric properties of outcome measures. If experts who have worked with, developed and evaluated different outcome measures can work together and achieve consensus in an evaluation of the adequacy of psychometric properties of different outcome measures this would be a very useful initial step in achieving a broader consensus on the use of outcome measures as it would underpin a minimum standard of psychometric properties for any recommended outcome measures. In turn this would guide and help others in developing a basic appreciation of the importance of good psychometric properties of various

outcome measures. Expert panels have been implemented in numerous circumstances to help gain consensus on a variety of topics and demonstrate promising results, even in an online format (1).

The aim of this phase of the project was to assemble an expert panel to review the available psychometric evidence on outcome measures and categorise them into three distinct categories (A: recommend; B: recommend with qualification; and C: unable to recommend). The results of the expert panel were used in further work to achieve a wider consensus from clinicians, health policy makers, health clinic managers, researchers, and end-users (i.e. individuals with LLA) on outcome measure selection at the ISPO lower-limb COMPASS consensus meeting.

3.2 Methods

3.2.1 Expert panel

Expert panel members were selected based on their expertise in the development and/or validation of outcome measures commonly used in LLA populations, and publication record (as demonstrated by their H-index (2)). Where possible, a mix of experts by gender and geographical location was sought. Only one researcher from a given research centre or collaborative team was invited to participate in the consensus meeting in order to reduce potential bias during the categorisation process. Given LLA outcomes measurement research is a relatively small field globally, the minimum number of experts for this process was determined using a 70% consensus cut-off (3). With seven experts, a minimum of five individuals were required to reach 70% consensus, which eliminated the risk of one individual being able to veto consensus.

3.2.2 Categorisation of outcome measures

In order to reach consensus on the categorisation of outcome measures, a three-stage process was implemented as described below.

Stage I

Blinded to the composition of the expert panel, each panel member was asked to categorise each outcome measure according to three categories:

> **A: Recommended** – better than adequate psychometric properties demonstrated at this time

B: Recommended with qualification – adequate psychometric properties demonstrated at this time

C: Unable to recommend – inadequate psychometric properties demonstrated at this time

Each outcome measure (N = 60; 41 PROM, 18 PerfOM, and one hybrid (4)) was categorised based on the 108 eligible publications included in the systematic review (Section 2.0). As previously detailed the Consensus-based Standards for the selection of health Measurement Instruments (COSMIN) risk of bias (RoB) tool (5, 6) was applied to each publication and where applicable, the COSMIN 2020 update for performance-based outcome measures (7) was also applied. In accordance with the COSMIN guidelines, each criterion was rated using a four-point ordinal scale (i.e., "very good", "adequate", "doubtful" or "inadequate") for the relevant psychometric properties (e.g. reliability, validity, responsiveness, etc.). The lowest rating from the specified criteria for each property was used to establish an overall psychometric property rating (5-7). Expert panel members were then asked to consider both the psychometric properties of the outcome measures whilst considering the COSMIN RoB when categorising each outcome measure. By convention, if the same psychometric property for an outcome measure was examined in multiple publications, the superior result was accepted as an indication of the outcome measures' strength (1) for the COSMIN ratings. Experts were also able to provide general comments on each outcome measure, which would be discussed further during the online meeting in Stage II. Lastly, each expert was asked to vote on all outcome measures, including the ones they were involved in developing or testing. The rationale for this, was that with the number of experts and the threshold for voting specified in Section 3.2.1 any bias introduced was adequately mitigated by votes from the other panel members. Once completed, each expert panel member was asked to send their categorisations to a member of the project team. A summary of the results and panel comments was produced, with calculations of percentage of votes by category for each outcome measure.

Stage II

An online meeting was convened for expert panellists to discuss the results of the individual categorisation assessment (i.e. Stage I). A summary of the Stage I results was sent to all expert panellists in advance of the meeting. Experts were blinded to each other's votes and comments to reduce bias in Stage II. Consensus was deemed to be achieved when 70% (or more) (3) of the expert panel agreed on a single category for each outcome measure. Outcome measures that did not reach consensus in Stage I formed the focus of discussions in Stage II. After discussions, the panellists were asked whether they wanted to revise their vote for each outcome measure's category. During the discussion the experts agreed that some of the outcome measures were better viewed as being a

collection of subscales and it would be appropriate to categorise the individual subscales of (Trinity Amputation and Prosthesis Experience Scales – Revised [TAPES-R], Prosthesis Evaluation Questionnaire [PEQ], and the Questionnaire for persons with a Transfemoral Amputation [Q-TFA]).

Stage III

For outcome measures which did not reach consensus after Stage II, a second online meeting was conducted for the expert panel to further review the COSMIN risk of bias ratings and individual psychometric properties and discuss in detail the remaining outcome measures' strengths and limitations, with the aim of trying to reach consensus on the category for all outcome measures. Additionally, the initial categorisation results for outcome measure subscales were viewed, discussed, voted on in order to establish consensus.

3.3 Results

Eighteen experts were invited to take part in the categorisation of LLA outcome measures. Nine responded and agreed to take part (Table 3.1).

Table 3.1. Characteristics of the expert panel

Characteristics of the expert panel	N = 9
Gender, n (%)	Female = 3 (33.3) Male = 6 (66.7)
Geographic location, n (%)	Australia = 1 (11.1) United States of America = 5 (55.6) Ireland = 1 (11.1) Italy = 1 (11.1) Sweden = 1 (11.1)
H-index, mean (SD) [range]	20.3 (4.9) [12 to 27]

Stage I of the expert panel resulted in categorised consensus on 20 (33%) outcome measures. Of those that achieved consensus, zero were categorised as A (recommended), two were categorised as B (recommended with qualification), and 18 were categorised as C (unable to recommended) and thus removed from the list of outcome measures to be discussed at the ISPO lower-limb COMPASS consensus meeting. This left 40 outcome measures that were yet to reach consensus.

Twenty-two of the 60 reviewed outcome measures received at least one vote for category A in Stage I, however only 53 out of a possible 540 votes (9.8%) were categorised as A. Following a discussion

at the beginning of Stage II, panellists agreed that it was unlikely that any outcome measures would reach the ≥70% consensus threshold for category A given the published psychometric properties for outcome measures evaluated in this population. Therefore, the panellists decided that outcome measures with a combined total of ≥70% consensus across categories A and B would be appropriate to include in the lower-limb COMPASS process. Based on this revised combined category, A and B together, the expert panel reached 70% consensus on 14 of the 40 outcome measures that did not achieve consensus after Stage I when categories A plus B together were used as the standalone category. From Stage II onwards, all A votes were combined with B and thus designated a B categorisation (recommended with qualifications). This left 26 outcome measures that did not reach consensus and were moved into Stage II for additional discussion.

Stage II of the expert panel involved presenting the results of the blinded individual expert votes and a discussion of the 26 outcome measures which were yet to reach consensus. Discussion and re-voting during the meeting resulted in an additional 21 (81%) of the remaining outcome measures reaching consensus. Six were categorised as B and 15 were categorised as C and eliminated. The second half of Stage II involved presenting the outcome measure subscale votes to the panel members. By the end of Stage II, panel members were unable to reach consensus on five outcome measures (Functional Reach Test, Orthotic Prosthetic User Survey [OPUS]: lower limb functional measure [LLFM], OPUS: satisfaction with services [CSS], Patient Reported Outcomes Measurement Information System – 29-items [PROMIS–29], and socket comfort score [SCS]) and four PEQ subscales (Residual Limb Health, Sounds, Perceived Response, and Transfers). This was primarily due to the limited psychometric properties of these outcome measures in an LLA population.

Stage III of the expert panel involved in an in-depth discussion of the five outcome measures and four subscales which did not reach consensus in Stage II. Of the five outcome measures yet to reach consensus, consensus was achieved for the Functional Reach Test (categorised as C), OPUS: LLFM (categorised as C), OPUS: CSS (categorised as C), and the PROMIS-29 (categorised as B). Consensus was also reached on the four PEQ subscales (Residual Limb Health and Sounds categorised as B; Perceived Response and Transfers categorised as C). Consensus could not be reached on the SCS. The final 23 outcome measures (12 patient-reported, 11 performance-based) are displayed in Table 3.2.

Table 3.2. Outcome measures to be included in the wider consensus process (N = 23)

Patient-reported		Performance-based	
A + B combined: total of 70% consensus (n = 7)	B: recommend with qualification (n = 5)	A + B combined: total of 70% consensus (n = 7)	B: recommend with qualification (n = 4)
Stage I	Stage II onwards	Stage I	Stage II onwards
Activity Balance Confidence (ABC) scale (8)	Orthotic Prosthetic User Survey (OPUS): Health-related quality of life (HR-QoL) (9)	Amputee Mobility Predictor (AMP) (10)	Berg Balance Scale (BBS) (11)
Amputee Body Image Scale (ABIS) (12)	OPUS: Satisfaction with device (OPUS: CSD) (9)	Comprehensive High-Level Activity Mobility Predictor (CHAMP) (13)	Figure 8 walking (F8W) test (14)
Houghton Scale (15, 16)	Prosthetic Mobility Questionnaire (PMQ) (including the PMQ 2.0) (17, 18)	L-Test (19)	Four Square Step Test (FSST) (20)
Locomotor Capabilities Index – 5-point scale (LCI–5) (21)	Patient Reported Outcomes Measurement Information System – 29-item (PROMIS-29) (22)	Six-minute walk test (6MWT) (23)	Narrowing Beam Walking Test (NBWT) (24)
Prosthesis Evaluation Questionnaire (PEQ) and associated subscales (25) − Residual limb health − Utility − Wellbeing − Mobility − Appearance − Frustration − Social burden − Sounds	Special Interest Group in Amputee Medicine (SIGAM) Mobility Scale (26)	Timed Up and Go (TUG) (27)	
Prosthetic Limb Users Survey of Mobility (PLUS-M) (including the PLUS-M 12-item short form) (28, 29)		Two-minute walk test (2MWT) (23)	

Trinity Amputation and Prosthesis Experience Scales-Revised (TAPES-R) (including consideration of the original TAPES) and associated subscales (30, 31) – Activity restriction – Satisfaction – Psychological adjustment		10-metre walk test (10mWT) (32)	

3.4 Discussion of Expert Panel

The aim of the expert panel was to determine which outcome measures should be included in a stakeholder consensus process based on the evidence of each outcome measures' psychometric properties in people with LLA. The results have provided an initial list of PerfOMs and PROMs that have been found to have adequate psychometric properties for the ISPO lower-limb COMPASS consensus meeting attendees to consider.

3.4.1 Key outcomes from the categorisation process

Expert panellists recommended 12 PROM and 11 PerfOMs for inclusion in the ISPO lower-limb COMPASS process (see Table 3.2). The expert panel did not reach consensus on one outcome measure, the SCS, and as such it was not moved forward into the ISPO lower-limb COMPASS process. Experts were unable to agree on whether socket comfort could be adequately quantified by a single item question, given the multi-dimensional factors which can influence socket comfort. There were also concerns regarding this outcome measures ability to capture and assess a true change in socket comfort given its poor reliability in people with LLA (33, 34). However, experts also recognised that poor reliability may be related to the high variability in socket comfort from day-to-day or hour-to-hour depending on the activities of the individual. Nevertheless, the experts recognised this outcome was an important construct and advocated for further development and testing of outcome measures which assess socket comfort.

Of the 36 outcome measures that the panel was unable to recommend (Category C), 28 were PROMs, seven were PerfOMs, and one hybrid (Rivermead Mobility Index (4)). Consensus dictated that these outcome measures be excluded from the ISPO lower-limb COMPASS process on the basis that they currently demonstrated poor and/or limited psychometric properties in an LLA population.

During the Stage II there was also consideration of generic outcome measures that had established psychometric properties in other populations with disability (e.g. stroke, Parkinson's disease, elderly falls-risk groups, etc). However, given the diversity of disability profiles with many categories of the functional profiles in these population groups not being transferrable to LLA populations, the expert panel decided to only focus on studies that currently demonstrated psychometric properties within an LLA population.

3.4.2 Outcome measure subscales

Two outcome measures recommended for inclusion, the PEQ and the TAPES-R (including consideration of the original TAPES), are comprised of multiple subscales. The panel decided to assess and then vote on each subscale separately. The rationale for suggesting subscales be reviewed and voted on an individual basis was that subscales are developed and tested to assess different constructs, which results in varying psychometric properties. Hence, it was more precise to evaluate the subscales individually, rather than looking at the outcome measure as a whole. Given that the majority of these subscales demonstrated adequate psychometric properties, this can assist with reducing administration burden of outcome measures, particularly if being combined with other outcome measures. Further, it can facilitate outcome measure selection to span a larger range of disability categories as demonstrated by the number of ICF domains covered.

3.4.3 Benefits of the expert panel process

The primary benefit of the expert panel was the ability to review a large body of published literature with different psychometric methods and properties reported for each outcome measure, weigh the relative strength of reported psychometric properties, COSMIN ratings, and synthesise this information to be presented in a simple format. This expert panel process also helped to increase the rigour of the ISPO lower-limb COMPASS process by making more concrete recommendations to the diverse group of project participants. The ISPO lower-limb COMPASS process participants were expected to have little-to-no knowledge of psychometric proprieties, but their input was essential for reaching a meaningful consensus on which outcome measures might be recommended for routine use. As the expert panel were only presented with outcome measures which had been psychometrically evaluated, concerns of upvoting outcome measures with poor or unknown psychometric properties in the ISPO lower-limb COMPASS process have been avoided. Lastly, this process is deemed extremely effective via the high rate (98.3%) of consensus achieved by the expert panel (only one outcome measure, out of 60, did not reach consensus) and the efficiency of the process (two, two-and-a-half hour meetings).

The majority of studies included in the systematic review investigating the psychometric properties of outcome measures in an LLA population, received a relatively poor methodological quality rating according to the COSMIN guidelines. This prevented the majority of outcome measures being categorised as A by the expert panel, with less than 10% of initial votes by the panellists being allocated to category A in Stage I. Twelve outcome measures (20%) had votes cast by the experts in both criteria A and C during Stage I. Given experts were able to categorise the same outcome measure as either A or C based on the same information emphasises the lack of standardisation in assessing psychometric properties, with no current fixed method to weigh-up psychometrics across any published quality appraisal tool (35-37). The divergence of ratings was only evident in stage one of the expert panel process. At this stage the panellists were not able to discuss any of the issues related to the psychometric properties but the subsequent discussions in the process enhanced the ability to reach consensus and underpin the importance of a consensus process used in this project.

Given the stringency of COSMIN and other quality appraisal tools, in combination with the challenges associated with assessing psychometric properties, this highlights the need of an expert panel for this process. In addition to considering the psychometric properties and quality of published studies, decisions on how the expert panel categorised the outcome measures often came down to an outcome measure demonstrating sound evidence of content and construct validity, good test-retest reliability (i.e. low measurement error in longitudinal applications) as established in reasonably-sized studies that included a diverse group of people with LLA; the importance of the construct being measured, the ease of administration, and current/ongoing work that was not yet published. Despite the challenges, 20 (out of 60) outcome measures achieved consensus in Stage I, and a further 14, reached consensus when category A was merged with category B, demonstrating moderate agreement (56.6%) between the experts before any group discussion in Stage II.

3.4.4 Strengths and limitations of the Expert Panel

The results of this categorisation process need to be viewed in light of both its strengths and limitations. The principal strength of this process was that leading academics and researchers in outcome measures for people with LLA were able to work collaboratively to evaluate the outcome measures and reach a very high degree of consensus. A further strength was that the group of expert panellists included gender, limited geographical and professional diversity, as well as individual and collective vast experience in psychometric testing of outcome measures. Although not the first time a process like this has been implemented (38), it was important to undertake, as it has provided clearer guidance on outcome measures and can be considered as a useful adjunct for use within clinical practice, research, and the broader use of outcome measures in the future.

Conducting this process in an online forum enabled an opportunity for doubt or discrepancy regarding the categorisation of outcome measures to be extensively debated among experts. Feedback from the expert panel identified that focussed moderation of the consensus meetings positively contributed to the efficiency of the process, while simultaneously giving the opportunity to discuss any concerns and issues at length.

A key limitation to this effort was the choice and application of the COSMIN RoB tool. While it was generally agreed that COSMIN was an appropriate tool to use for this purpose, the standards set by the COSMIN tool challenged expert panellists to recommend outcome measures that did not meet the rigorous standards. Further, while the expert panel noted that no RoB tool is perfect, some elements of COSMIN may have impeded the interpretation and representation of psychometric properties of outcome measures for use in people with LLA. These included: the heavy focus on appraising the strength of the study design and not the outcome measure itself; the 2020 COSMIN checklist containing less detail than the 2018 COSMIN checklist, thus having to use a combination of the two for performance-based outcome measures, the recommended quantifiable test re-test reliability timing (it can be debated that the two-week interval between test and re-retest assessments recommended by COSMIN 2018 may be excessive when research has showed that retest periods of two and 14 days produce similar results (39), and it is more important that the investigators appropriately justify their interval), lack of appropriate recognition of the fundamental measurement requirements and procedures as stated by modern psychometric techniques, including Rasch models as a prerequisite for further statistical analysis (e.g. COSMIN lacks detailed and pertinent extension for studies using Rasch methods (35)), the development criteria (COSMIN Box 1) not being well adapted for PerfOMs or accounting for outcome measures developed from prior instruments; no weighting regarding the superiority of psychometric properties (e.g. reliability should be viewed more favourably than any type of validity, given validity becomes meaningless if you cannot rely on a measurement); the 'worst score counts system' of COSMIN being very blunt, even for well justified deviations; floor and ceiling effects not being assessed; and complexities in interpreting COSMIN guidelines.

It is important to highlight that some individuals included in the expert panel were involved in the development and/or validation of outcome measures included in this project. However, a maximum of two experts were involved in the development and/or validation of any single outcome measure, which did now allow for a majority vote during the categorisation process, hence, the above methodology can still be deemed rigorous.

Whilst an effort was made to include a range of expert panellist, geographic location was limited to high-income countries (HICs). This reflects the disparity of research between HICs and LMICs.

3.4.5 Future directions

The expert panel recommended further research with improved designs to enhance evidence of psychometric properties of included and excluded outcome measures. Specifically, the expert panel recommended: improvement and an increase in paediatric specific LLA outcome measures, testing of common generic outcome measures within an LLA population (e.g. EuroQol (EQ) 5D-5L, Short Form 12-item), constructs such as socket comfort, satisfaction with services, and devices of interest, should be the focus of further outcome measure development, testing (including head-to-head comparisons) of existing outcome measures should take priority over the development of entirely new outcome measures, consideration of the evolution of outcome measures (e.g. TAPES to the TAPES-R), and scoring systems should not be changed without re-testing reliability.

3.5 Conclusion of the Expert Panel

The expert panel reached consensus on 59 (out of 60) outcome measures identified in a systematic review, grouping them into one of two categories (B: recommend with qualification or C: unable to recommend). Twenty-three outcome measures (12 patient-reported, 11 performance-based) and 11 subscales were identified as currently having adequate psychometric properties in LLA populations. The outcomes from the expert panel has made a significant contribution to the ISPO lower-limb COMPASS process, by enabling all attendees to use this information in combination with their own experiences (clinical or personal) to place their consensus vote, thereby consolidating a final list of outcome measures which should be routinely used in people with LLA.

References

1. Khodyakov D, Hempel S, Rubenstein L, Shekelle P, Foy R, Salem-Schatz S, et al. Conducting Online Expert panels: a feasibility and experimental replicability study. BMC Med Res Methodol. 2011;11(1):1-8.
2. Hirsch JE. An index to quantify an individual's scientific research output. Proceedings of the National academy of Sciences. 2005;102(46):16569-72.
3. Boers M, Kirwan J, Tugwell P. The OMERACT Handook. Chapter 7 - Methods for reaching consensus: OMERACT; 2018 [updated 2021; cited 14 April 2021]. Available from: http://www.omeract.org/pdf.
4. Collen FM, Wade DT, Robb G, Bradshaw C. The Rivermead mobility index: a further development of the Rivermead motor assessment. Int Disabil Stud. 1991;13(2):50-4.
5. Mokkink LB, De Vet HC, Prinsen CA, Patrick DL, Alonso J, Bouter LM, et al. COSMIN risk of bias checklist for systematic reviews of patient-reported outcome measures. Qual Life Res. 2018;27(5):1171-9.
6. Prinsen CA, Mokkink LB, Bouter LM, Alonso J, Patrick DL, De Vet HC, et al. COSMIN guideline for systematic reviews of patient-reported outcome measures. Qual Life Res. 2018;27(5):1147-57.

7. Mokkink LB, Boers M, Vleuten Cvd, Bouter L, Alonso J, Patrick DL, et al. COSMIN Risk of Bias tool to assess the quality of studies on reliability or measurement error of outcome measurement instruments: a Delphi study. BMC Med Res Methodol. 2020;20(1):1-13.
8. Powell LE, Myers AM. The activities-specific balance confidence (ABC) scale. J Gerontol A Biol Sci Med Sci. 1995;50(1):M28-M34.
9. Heinemann AW, Bode RK, O'Reilly C. Development and measurement properties of the Orthotics and Prosthetics Users' Survey (OPUS): a comprehensive set of clinical outcome instruments. Prosthet Orthot Int. 2003;27(3):191-206.
10. Gailey RS, Roach KE, Applegate EB, Cho B, Cunniffe B, Licht S, 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):613-27.
11. Berg K, Wood-Dauphine S, Williams J, Gayton D. Measuring balance in the elderly: preliminary development of an instrument. Physiother Can. 1989;41(6):304-11.
12. Breakey JW. Body image: The lower-limb amputee. J Prosthet Orthot. 1997;9(2):58-66.
13. Gailey RS. Development and reliability testing of the Comprehensive High-Level Activity Mobility Predictor (CHAMP) in male servicemembers with traumatic lower-limb loss. J Rehabil Res Dev. 2013;50(7):905-18.
14. Hess RJ, Brach JS, Piva SR, VanSwearingen JM. Walking skill can be assessed in older adults: validity of the Figure-of-8 Walk Test. Phys Ther. 2010;90(1):89-99.
15. Houghton A, Allen A, Luff R, McColl I. Rehabilitation after lower limb amputation: a comparative study of above-knee, through-knee and Gritti—Stokes amputations. Br J Surg. 1989;76(6):622-4.
16. Houghton A, Taylor P, Thurlow S, Rootes E, McColl I. Success rates for rehabilitation of vascular amputees: implications for preoperative assessment and amputation level. Br J Surg. 1992;79(8):753-5.
17. Franchignoni F, Monticone M, Giordano A, Rocca B. Rasch validation of the Prosthetic Mobility Questionnaire: A new outcome measure for assessing mobility in people with lower limb amputation. J Rehabil Med. 2015;47(5):460-5.
18. Burger H, Giordano A, Bavec A, Franchignoni F. The Prosthetic Mobility Questionnaire, a tool for assessing mobility in people with lower-limb amputation: validation of PMQ 2.0 in Slovenia. Int J Rehabil Res. 2019;42(3):263-9.
19. Deathe AB, Miller WC. The L test of functional mobility: measurement properties of a modified version of the timed "up & go" test designed for people with lower-limb amputations. Phys Ther. 2005;85(7):626-35.
20. Dite W, Temple VA. A clinical test of stepping and change of direction to identify multiple falling older adults. Arch Phys Med Rehabil. 2002;83(11):1566-71.
21. Gauthier-Gagnon C, Grise M, Lepage Y. The locomotor capabilities index: content validity. J Rehabil Outcomes Meas. 1998;2(4):40-6.
22. Cella D, Riley W, Stone A, Rothrock N, Reeve B, Yount S, et al. The Patient-Reported Outcomes Measurement Information System (PROMIS) developed and tested its first wave of adult self-reported health outcome item banks: 2005–2008. J Clin Epidemiol. 2010;63(11):1179-94.
23. Butland R, Pang J, Gross E, Woodcock A, Geddes D. Two-, six-, and 12-minute walking tests in respiratory disease. Br Med J (Clin Res Ed). 1982;284(6329):1607.
24. Sawers A, Hafner BJ. Narrowing beam-walking is a clinically feasible approach for assessing balance ability in lower-limb prosthesis users. J Rehabil Med. 2018;50(5):457-64.
25. Legro MW, Reiber GD, Smith DG, del Aguila M, Larsen J, Boone D. Prosthesis evaluation questionnaire for persons with lower limb amputations: assessing prosthesis-related quality of life. Arch Phys Med Rehabil. 1998;79(8):931-8.
26. Ryall N, Eyres S, Neumann V, Bhakta B, Tennant A. The SIGAM mobility grades: a new population-specific measure for lower limb amputees. Disabil Rehabil. 2003;25(15):833-44.
27. Podsiadlo D, Richardson S. The timed “Up & Go”: a test of basic functional mobility for frail elderly persons. J Am Geriatr Soc. 1991;39(2):142-8.
28. Amtmann D, Abrahamson DC, Morgan S, Salem R, Askew RL, Gailey R, et al., editors. The PLUS-M: item bank of mobility for prosthetic limb users. Qual Life Res; 2014: Springer van Godewijckstraat 30, 3311 GZ, Dordrecht, Netherlands. .
29. Morgan SJ, Amtmann D, Abrahamson DC, Kajlich AJ, Hafner BJ. Use of cognitive interviews in the development of the PLUS-M item bank. Qual Life Res. 2014;23(6):1767-75.
30. Gallagher P, MacLachlan M. Development and psychometric evaluation of the trinity amputation and prosthesis experience scales (TAPES). Rehabil Psychol. 2000;45(2):130-54.
31. Gallagher P, Franchignoni F, Giordano A, MacLachlan M. Trinity amputation and prosthesis experience scales: a psychometric assessment using classical test theory and rasch analysis. Am J Phys Med Rehabil. 2010;89(6):487-96.
32. Datta D, Ariyaratnam R, Hilton S. Timed walking test - an all-embracing outcome measure for lower-limb amputees? Clin Rehabil. 1996;10(3):227-32.
33. Hanspal RS, Fisher K, Nieveen R. Prosthetic socket fit comfort score. Disabil Rehabil. 2003;25(22):1278-80.
34. Hafner BJ, Morgan SJ, Askew RL, Salem R. Psychometric evaluation of self-report outcome measures for prosthetic applications. J Rehabil Res Dev. 2016;53(6):797-812.
35. McKenna SP, Heaney A. Setting and maintaining standards for patient-reported outcome measures: can we rely on the COSMIN checklists? J Med Econ. 2021;24(1):502-11.

36. Balk EM, Gazula A, Markozannes G, Kimmel HJ, Saldanha IJ, Trikalinos TA, et al. Psychometric Properties of Functional, Ambulatory, and Quality of Life Instruments in Lower Limb Amputees: A Systematic Review. Arch Phys Med Rehabil. 2019;100(12):2354-70.
37. Rosenkoetter U, Tate RL. Assessing features of psychometric assessment instruments: A comparison of the COSMIN checklist with other critical appraisal tools. Brain Impair. 2018;19(1):103-18.
38. Miller LA, McCay JA. Summary and Conclusions From the Academy's Sixth State-of-the-Science Conference on Lower Limb Prosthetic Outcome Measures. J Prosthet Orthot. 2006;18(6):2-7.
39. Marx RG, Menezes A, Horovitz L, Jones EC, Warren RF. A comparison of two time intervals for test-retest reliability of health status instruments. J Clin Epidemiol. 2003;56(8):730-5.

4.0 Outcome Measures Consensus Process

4.1 Methods

A modified Delphi consensus process, consisting of an initial survey followed by several virtual meetings, was used to select from the measures recommended by the Expert Panel (described in Section 3.0). Virtual, rather than in-person meetings, were held due to travel restrictions associated with the COVID-19 pandemic. The virtual meetings were used to discuss results from the pre-meeting survey, results from the Expert Panel meetings, and to reach consensus on decisions regarding the final list of outcome measures.

4.1.1 Modified Delphi and Consensus Process

Selection of consensus participants

The project team met in December 2020 with input from the ISPO IAG to establish selection criteria for consensus participants, as well as to identify specific potential participants. It was important to have broad representation in the consensus group so as to increase the diversity of opinions in the process and increase the buy-in from a broad range of stakeholders. Organisations who are significant contributors to prosthetic service provision around the world were invited to nominate consensus conference participants. This included professional organisations (World Federation of Occupation Therapists [WFOT], World Confederation for Physical Therapy [WCPT], International Society of Physical and Rehabilitation Medicine [ISPRM], Pedorthists [IVO]); international organisations (World Health Organisation [WHO], International Committee of the Red Cross [ICRC], Humanity and Inclusion [HI]), industry (including members of the ISPO IAG); scientific experts (Cochrane, members of the expert panel), funders (ATscale, United Stated Agency for International Development [USAID]), regional organisations (La Fédération Africaine des Techniciens Orthoprothésistes [FATO], ISPO Asia, United Arab Society for Prosthetics and Orthotics [UASPO], ISPO Amesur, Range of Motion Project [ROMP], Exceed Worldwide), and consumer representatives (International Confederation of Amputee Associations [IC2A]). Participants were also selected based on their clinical experience with individuals with LLA (i.e. prosthetists & orthotists, physiotherapists, occupational therapists, psychologists, and medical practitioners), being a health clinic manager or health policy maker, and active researchers in the fields of LLA outcome measures and psychometrics. Prosthetic-users were invited and included in the process to capture their

perspectives and ensure their needs were represented. Their viewpoints were deemed essential to the process. All expert panel members (see Section 3.0) were also invited to take part in the consensus process. Potential consensus participants were invited via email between December 2020 to April 2021. Participants were required to be available to complete both the pre-consensus survey and attend the four virtual consensus meetings. Potential consensus participants were also able to nominate other participants for invitation that the project team and IAG may not have identified. Where possible, participants and nominations from under-represented regions (e.g. South America) were actively sought.

Pre-consensus survey and voting

Based on the results of the systematic review and outcomes of the expert panel, the aforementioned 23 individual outcome measures (12 PROM and 11 PerFOM) and 11 subscales were introduced into the consensus process (see Section 3.2). During the expert panel process, it was identified that two PROMs (Prosthesis Evaluation Questionnaire [PEQ] (1) and Trinity Amputation and Prosthesis Experience Scales-Revied [TAPES-R] (2, 3)) should be presented in their entirety as well as by individual subscale (PEQ subscales n = 8, TAPES-R subscales n = 3).

The project team then developed the following set of four questions to be asked for each individual outcome measure and/or subscale: (i) “Based on your experience and the information presented, is this outcome measure one you would recommend for use in routine *clinical practice*?” with a ‘yes’/‘no’ response option; (ii) “Based on your experience and the information presented, is this outcome measure one you would recommend for use in *research*?” with a ‘yes’/‘no’ response option; (iii) “Please rate your experience using this outcome measure” with response options ‘I have no experience and am not familiar with it’, ‘I have heard of it and/or read about it, but have not used it’, ‘I have some experience using it’, or ‘I consider myself an expert and have used it extensively in clinical practice or research’; and (iv) ”Do you have any other comments you would like to add about your recommendation or experiences with this outcome measure?”. The pre-consensus survey was accompanied by a manual which contained detailed information on each outcome measure (i.e. outcome measure description, instructions, video resources where applicable, chapter-level International Classification of Function [ICF] categorisation, and references). For efficiency and focus, the survey was divided into two – PerfOMs (survey one) and PROMs (survey two).

Each consensus participant was sent a link to an anonymised survey (Survey Monkey, San Mateo, California, USA) at least one-week before the virtual meeting. Two project team members compiled the results of the survey to present to the consensus participants at the start of the first and second

virtual consensus meetings, demonstrating which outcome measures had achieve consensus (70% or higher agreed or disagreed (4, 5).

Virtual meetings

Four, 2.5-hour virtual meetings were conducted to accommodate the multiple time zones where consensus participants were located. Virtual meetings were conducted via Zoom teleconference software (San Jose, California, United States of America) and facilitated by CollaborateUp (an 8(a) SBA Certified Small Business, Washington DC, United States of America) with assistance from the project team.

At the first virtual meeting, PerfOM survey results were presented, discussed, and re-voted (as 'include' or 'exclude') in a blinded process using the Zoom poll function, with the results announced after each vote. No option was provided for 'Don't know/Not sure' in order to reach a final resolution on outcome measures. Given the diversity of consensus attendees, the purpose of the discussion and live re-vote was to allow individuals the chance to share experiences from a variety of contexts, raise questions, or obtain clarification about specific outcome measures. Outcome measures achieved consensus when 70% of the attendees agreed to include the measure in the live re-vote. Those that did not reached consensus (≤69%) from the re-vote but achieved a vote between 31 to 69% were placed into a 'yet to reached consensus' list, on the premise that they may contain elements (e.g. ICF categorisation) required to achieve a well-rounded list of measures for the ISPO COMPASS. Outcome measures which achieved a vote ≤30% were removed entirely.

The second virtual meeting focused on PROMs, using the same process as the PerfOMs. Post the second virtual meeting, consensus attendees were asked to: (i) recommend outcome measures that had not yet been included in the consensus process and that they felt were important; (ii) suggest ideas for consolidating the outcome measures into the ISPO lower-limb COMPASS list; and (iii) in order of importance, rank the chapter-level ICF categories which the remaining outcome measures covered.

The third virtual meeting was used as an opportunity for open discussion amongst the consensus attendees regarding the outcome measures which had reached consensus (≥70%) and those which achieved a vote of 31 to 69%, as per the re-votes in the first and second virtual meetings. The outcome measures were presented by type (PROM or PerfOM), chapter-level ICF category and importance (as voted by consensus attendees following the second virtual meeting). Additional outcome measures which were considered important and nominated by consensus attendees, were also presented to the consensus group for further discussion and voting. Attendees were asked to consider the micro (clinician and patient), meso (clinic manager and research), and macro (research

and policy) utility of these outcome measures during the open discussion. At the conclusion of the third virtual meeting, attendees were asked to provide a recommended list of outcome measures (based on the remaining lists of 'included' and 'yet to reach consensus') with a justification for their choices. The lists were consolidated and the outcome measures which reoccurred the most, were placed into the final ISPO lower-limb COMPASS.

During the fourth and final virtual meeting, the final ISPO lower-limb COMPASS and the associated chapter-level ICF categories were presented to the consensus attendees for a concluding group discussion (Table 4.2).

4.2 Results of Outcome Measure Consensus process

This consensus process (May and June 2021) involved 39 participants (18 females, 21 males; 12 from low-middle income countries, and 27 from high-income countries). A full demographic breakdown of those that attended at least one meeting can be viewed in Table 4.1.

Following the pre-consensus survey and live re-voting, 14 outcome measures achieved the predetermined threshold for consensus of ≥70% agreement for use in clinical practice and 20 for use in research, with only one outcome measure (The Narrowing Beam Walking Test) reaching ≥70% agreement to be excluded from use within clinical practice.

Table 4.1. Demographic data of Outcome Measures Consensus participants

Consensus participants demographics	N = 39
Gender, n	Female = 18 Male = 21
Nationality, n	American= 4; Australian = 2; Austrian = 1; Argentinian = 2; Brazilian =1; British = 7; Cambodian = 1; Danish = 2; Dutch = 2; El Salvador = 1; German = 1; Indian = 2; Iraqi = 1; Irish = 3; Madagascan = 1; Malaysian = 1; South African = 2; Spanish =1; Swiss = 3; Swedish = 1
Profession, n	End-user = 5; Occupational Therapist = 2; Pedorthotist = 1; Physical Rehabilitation Medicine = 4; Physiotherapist = 7; Policy Maker = 3; Prosthetist/Orthotist = 13; Public Health = 1; Researcher = 3

After an in-depth discussion and re-voting during the first two virtual meetings, seven outcome measures and three subscales achieved consensus to be included, 11 outcome measures and four subscales achieved a vote of 31 to 69%, and five outcome measures and four subscales achieved consensus to be excluded. Two additional outcome measures that were not included in the initial survey (N = 34), were identified by the consensus participants as being important to include – EuroQoL (EQ)-5D-5L (6, 7) and Patient Specific Function Scale (PSFS) (8, 9). The outcome measures which reoccurred most frequently when individuals complied their recommended list after the third virtual meeting where combined into the final COMPASS.

At the conclusion of the consensus process, the *COMPASS* included three PerfOMs (Amputee Mobility Predictor [AMP] (10), Timed Up and Go [TUG] (11), and Two-minute Walk Test [2MWT] (12)) and three PROMs (PEQ – Residual Limb Health (1), PEQ – Utility (1), and TAPES-R (2, 3)), which collectively span 10 chapter level ICF categories (see Table 4.2). They are recommended for use routinely for episodes of rehabilitation care and should be administered before and after to measure change in functioning.

Additionally, the *COMPASS+* included two PerfOMs (Comprehensive High-Level Activity Mobility Predictor [CHAMP] (13) and Six-minute Walk Test [6MWT] (12)). The *COMPSS+* for high functioning individuals with LLA was recommended to overcome the ceiling effects of the *COMPASS* and can be optionally offered.

Further, the *COMPASS Adjunc*t contains one generic PROM (Patient Specific Functional Scale (PSFS (8, 9)) chosen because of its high clinical relevance.

It was recommended that a locally relevant health-related quality of life (HRQoL) measure be used such as the EQ-5D-5L (6, 7, 14, 15) or Patient Reported Outcomes Measurement Information System – 29 item (PROMIS-29©) (16).

Table 4.2. COMPASS ICF Chapter Level Mapping

				Chapter level ICF categories in order as ranked by consensus participants (left considered most important)											Not in ICF		
	Name of Outcome Measure	Time (min)	Languages currently available (other than English)	Mobility (d4)	Sensory Functions & Pain (b2)	Self-Care (d5)	General Tasks & Demands (d2)	Functions of the Skin & related structures (b8)	Domestic Life (d6)	Community, Social & Civil Life (d9)	Major Life Areas (d8)	Products & Technology (e1)	Mental Functions (b1)	Interpersonal Interactions & Relationships (d7)	Support & Relationships (e3)	Socket Comfort	Satisfaction with Prosthesis	Don and doffing
PerfOMs	AMP	10-15	1	x	x							x						
	TUG	1-2	-	x														
	2MWT	<5	-	x														
	Total time	**16**																
PROMs	PEQ-Residual Limb Health	<5	5			x		x										
	TAPES-R	15	3	x	x		x			x	X	x		x	x	x	x	
	PEQ-Utility	<5	5			x						x				x	x	x
	Total time	**<20**																
COMPASS +	CHAMP	15	-	x														
	6 MWT	<10	-	x														
	Total time	**25**																
Generic HRQoL	EQ-5D-5L	<5	130	x	x	x			x	x	X		x					
	PROMIS-29®	<5	48	x	x		x		x	x			x	x	x			
	Total time	**<1**																
COMPASS Adjunct	PSFS	5-10	5															

Six key recommendations regarding the use of outcome measures for individuals with LLAs arose from the consensus process. These recommendations are as follows:

1. Recommendation 1: That the AMP, TUG, 2MWT, PEQ – Residual Limb Health, PEQ – Utility, and TAPES-R make up the ISPO lower-limb COMPASS.
2. Recommendation 2: That the CHAMP and 6MWT, two additional PerfOMs recommended for high functioning individuals with LLA make up the COMPASS+.
3. Recommendation 3: That the PSFS make up the COMPASS Adjunct.
4. Recommendation 4: That a generic HRQoL outcome measure, such as the EQ-5D-5L or PROMIS© Brief Profile (PROMIS-29) be used to supplement the COMPASS.
5. Recommendation 5: That outcome measures suited to low- and middle-income countries (LMICs) are developed with a focus on activities such as sitting cross-legged, kneeling, squatting, and other culturally important mobility related activities.
6. Recommendation 6: That translation, validation, and open sharing of translated outcome measures included in the COMPASS, COMPASS+, and COMPASS Adjunct occurs.

4.3 Discussion of Consensus Process Recommendations

4.3.1 Discussion of Consensus process recommendations

Recommendation 1: COMPASS

It was recommended that the AMP, TUG, 2MWT, PEQ – Residual Limb Health, PEQ – Utility, and TAPES-R make up the ISPO COMPASS.

Collectively, these six outcome measures take approximately 45 minutes to complete, with about 20 mins required for PerfOMs, which require in-person administration, and about 25 for PROMs, which can be done in a waiting area or online. The consensus process participants agreed these times are feasible within the time that majority of clinicians spend on a patient's first patient clinician interaction, while noting there might be variability about what is feasible in various contexts. This combination of outcome measures span 10 different chapter level ICF categories plus important non-ICF categories like socket comfort, satisfaction with prosthesis, and prosthesis donning and doffing. These outcome measures will enable detailed assessments and description of functional

status at various time points throughout the rehabilitation process, noting that as devices need to be replaced and renewed there will be multiple episodes of care.

Recommendation 2: COMPASS+

It was recommended that the CHAMP and 6MWT, two additional PerfOMs recommended for high functioning individuals with LLA make up the COMPASS+.

The CHAMP (13) and 6MWT (12) are two PerfOMs that are recommended for high-functioning individuals with LLA and help to mitigate the ceiling effect of outcome measures included in the core list. Many prosthetic users will not be high functioning individuals, so it is not warranted that the COMPASS+ is routinely administered to all people with LLA. Mobility is extremely important for highly active individuals with LLA and is linked with prosthetic prescription and modifications (17, 18) given the differing components to those required in an everyday prosthetic device. These two outcome measures take approximately 25 minutes to complete, and cover the mobility (d4) ICF category.

Recommendation 3: COMPASS Adjunct

It was recommended that the PSFS (a generic PROM) make up the COMPASS Adjunct.

Despite the PSFS not being deemed psychometrically adequate by the expert panel process due to poor reliability and measurement error (19) (Section 3.0) the high clinical utility of this outcome measure resulted in it being re-introduced and recommended by the consensus participants. The PSFS takes between 5 to 10 minutes to complete, demonstrates very good internal consistency, but is unable to be categorised using the ICF due to the personalised nature of the activities chosen by the respondent. This outcome measure is especially valuable within a clinical setting due to the ability of individuals to nominate any activities they feel are important and thus it focuses clinical discussions on achievement of individually identified goals that may not be adequately captured using standardised outcome measures. Further, the ability of individuals to change their nominated activities over time accurately reflects individual changes that occur throughout rehabilitation. The highly individualised nature of the PSFS should also enable increased specificity of rehabilitation plans that hopefully translate to further improved outcomes for individuals with LLA.

Recommendation 4: Use of a generic HRQoL instrument

It was recommended that a generic HRQoL outcome measure such as the EQ-5D-5L or PROMIS-29© be used to supplement the COMPASS.

The EQ-5D-5L (14, 15) and PROMIS-29© (16) are two generic HRQoL PROMs which take less than five minutes each to complete and were deemed important to include as an optional additional outcome measures, given their ability to facilitate health economic evaluations. It was recognised by the consensus participants that both outcome measures were deemed equally important due to differing government and organisational requirements, which may already have certain health economic outcome measures in place. For example, within the United Kingdom, the EQ-5D has been widely used throughout the National Health Service since 2009 (20) compared to the United States, where the PROMIS-29© is often used for health economic evaluation (21, 22). Of note, are the potential costs associated with the use of these two outcome measures, with the PROMIS being free to use while the EQ-5D-5L incurs some costs, which may impact on their use within LMICs. Further, despite its widespread use, the EQ-5D-5L has not been psychometrically tested within an LLA population. Other HRQoL measures in common use in a particular jurisdiction should also be considered for ease of comparability. Where there is no routine collection of HRQoL it would limit the additional benefits of collecting this data, as one of the major benefits is to compare HRQoL across different conditions.

Recommendation 5: Outcome Measure development for LMICs

It was recommended that outcome measures suited to low- and middle-income countries (LMICs) are developed with a focus on activities such as sitting cross-legged, kneeling, squatting, and other culturally important mobility related activities.

Whilst the outcome measures included within the COMPASS span 10 different chapter level ICF categories, it was highlighted that no outcome measure currently exists to accommodate culturally important mobility activities such as sitting cross-legged, kneeling, or squatting. These activities are an important part of everyday living in many LMICs, and if they are unable to be effectively executed, may prevent an individual from being able to perform basic living functions such as going to the bathroom or sharing a meal with family. Further, it was clear from the systematic reviews (Section 2.0) and emphasised during the consensus meeting discussions, that not many outcome measures have been translated into languages commonly used in LMICs. As such, it is recommended that future studies focus on the development of outcome measures that are able to capture a greater diversity of culturally important mobility activities, are psychometrically tested and are translated into languages appropriate for LMICs.

Recommendation 6: Translation and Open Sharing of Outcome Measures

It was recommended that translation, validation, and open sharing of translated outcome measures included in the COMPASS, COMPASS+ and COMPASS Adjunct occurs.

Outcome measures included in the COMPASS should be translated into multiple languages and their psychometric properties evaluated. These outcome measures should then be made openly available at no charge so they are able to be used routinely. Psychometric testing of translated outcome measures is important to ensure that the validity and reliability of the translated measures is adequate, and thus enable comparability between outcome measures administered in different languages. Outcome measures that have not been psychometrically evaluated in the target language cannot be assumed to have the same validity and reliability as the instrument presented in the original (i.e. source) language.

4.3.2 Impact on funding bodies

Funding bodies can consider using the ISPO COMPASS to evaluate research proposals, improve data collection and reporting, and maximise the potential that their investments will be effective.

4.3.3 Strengths

A key strength of this consensus process was the diversity of individuals involved. Consensus attendees included various users, experts and professionals who had extensive experience in the field of LLA. Further, there was representation from multiple high-, middle-, and low-income countries, as well a diversity in the gender of those present at the consensus meetings (see Table 4.1). The ability to conduct this consensus meeting virtually enabled a larger number and greater diversity of individuals to be involved, thus increasing the world-wide representation and rigour of this process. Lastly, benefits of a list of agreed-upon outcome measures not only include improved comparability and ability to aggregate outcome measure data internationally, but also the potential to inform clinical care at the patient level (micro level impact), decision-making at a centre level (meso level impact), and policy making at the national level (macro level impact).

4.3.4 Limitations

Psychometric properties are critical to ensure that outcome measures accurately measure the construct within the population of interest. However, they may also be difficult for some stakeholders, particularly those without scientific or clinical training, to understand. Thus, it may

have been challenging for some stakeholders to appropriately weigh and consider each measure's psychometric properties when casting their votes.

The systematic review that helped to inform the consensus process were restricted to the English language. This was a limitation from the historical as well as the implementation perspectives. Non-English outcome measures would have been missed and adoption of these outcome measures in routine practice is restricted to the English language, with limited validated translated outcome measures currently available. This may impede the implementation of outcome measures in LMICs.

There is a vast array of outcome measures which have been developed and tested in other clinical populations and may also be appropriate for use in LLA. However, further psychometric testing would need to be undertaken to ensure these outcome measures function appropriately in the LLA population.

A clinical decision must be taken for each service user that it is safe to perform the COMPASS with regard to balance ability, cardiovascular endurance and lists of contraindications should be carefully considered. In areas where there is significant unmet need and clinical time is the major constraint to meeting that need, time spent on outcome measure data collection may result in further limiting clinical service provision.

4.3.5 Future directions

Regular review of the outcome measures available and the associated literature to support their use will be needed as new research evidence emerges. The COMPASS should be reviewed and revised as new measures and evidence becomes available.

The availability of normative outcome measure data for the LLA population was not included in this consensus process. Known normative data (although somewhat limited in this population), would be highly valuable for clinicians and researchers with regards to benchmarking users, formulating rehabilitation and management plans, and tracking improvements over time.

The use of the COMPASS will facilitate future comparisons and pooling of data, which is highly useful with regards to identifying areas for funding and establishing real-world effectiveness of interventions upon outcomes. Organisational (e.g. government, regulatory bodies, hospitals, etc.) requirement of routine automated collection of outcome measures will help to facilitate this process.

Further linkage of the included outcome measures to cover more ICF categories at the second and third level is required. Additional development of outcome measures with diverse consumer involvement to enable broader coverage of the ICF categories relevant to people with LLA is needed.

Refinement of outcome measures to allow better evaluation of technical and/or environmental components that contribute to patient outcomes would allow better evaluation of the benefits of environmental/technological contributions to the functioning of the individual with lower limb absence.

Infographics, instructional videos, outcome measure instructions, and access to data collection sheets will be developed and made available as the COMPASS User Guide through ISPOLearn porthole to assist with world-wide dissemination and implementation.

4.4 Conclusion

The consensus process successfully developed a short list of outcome measures recommended for routine clinical use worldwide which has been formulated into the COMPASS, the COMPASS+, and the COMPASS Adjunct. Participants of the ISPO lower-limb COMPASS process recommend the use of six core outcome measures in individuals with LLA (AMP, TUG, 2MWT, PEQ – Residual Limb Health, PEQ – Utility, and TAPES-R) as part of the COMPASS. Collectively, these outcome measures span 10 chapter-level ICF categories. Additional outcome measures have been recommended to supplement the core ISPO lower-limb COMPASS and include the CHAMP and 6MWT which make up the COMPASS+, the PSFS included in the COMPASS Adjunct, and the use of a generic HRQoL outcome measure (e.g. EQ-5D-5L and PROMIS-29©). Outcome measures suited to LMICs need to be developed with a focus on culturally important mobility activities such as sitting cross-legged, kneeling, and squatting. To ensure global uptake, outcome measures included in the ISPO COMPASS should be translated into multiple languages, tested for psychometric performance in the target language, and shared free of charge openly to facilitate widespread use.

References

1. Legro MW, Reiber GD, Smith DG, del Aguila M, Larsen J, Boone D. Prosthesis evaluation questionnaire for persons with lower limb amputations: assessing prosthesis-related quality of life. Arch Phys Med Rehabil. 1998;79(8):931-8.
2. Gallagher P, MacLachlan M. Development and psychometric evaluation of the trinity amputation and prosthesis experience scales (TAPES). Rehabil Psychol. 2000;45(2):130-54.
3. Gallagher P, Franchignoni F, Giordano A, MacLachlan M. Trinity amputation and prosthesis experience scales: a psychometric assessment using classical test theory and rasch analysis. Am J Phys Med Rehabil. 2010;89(6):487-96.

4. Fink A, Kosecoff J, Chassin M, Brook RH. Consensus methods: characteristics and guidelines for use. Am J Public Health. 1984;74(9):979-83.
5. Boers M, Kirwan J, Tugwell P. The OMERACT Handook. Chapter 7 - Methods for reaching consensus: OMERACT; 2018 [updated 2021; cited 14 April 2021]. Available from: http://www.omeract.org/pdf.
6. Rabin R, Charro F. EQ-5D: a measure of health status from the EuroQol Group. Ann Med. 2001;33(5):337-43.
7. EuroQoL Research Foundation. EQ-5D [updated 2021; cited 4 June 2021]. Available from: https://euroqol.org/eq-5d-instruments/eq-5d-5l-about/.
8. Sterling M, Brentnall D. Patient Specific Functional Scale. Aust J Physiother. 2007;53(1):65.
9. 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;47:258-63.
10. Gailey RS, Roach KE, Applegate EB, Cho B, Cunniffe B, Licht S, 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):613-27.
11. Podsiadlo D, Richardson S. The timed "Up & Go": a test of basic functional mobility for frail elderly persons. J Am Geriatr Soc. 1991;39(2):142-8.
12. Butland R, Pang J, Gross E, Woodcock A, Geddes D. Two-, six-, and 12-minute walking tests in respiratory disease. Br Med J (Clin Res Ed). 1982;284(6329):1607.
13. Gailey RS. Development and reliability testing of the Comprehensive High-Level Activity Mobility Predictor (CHAMP) in male servicemembers with traumatic lower-limb loss. J Rehabil Res Dev. 2013;50(7):905-18.
14. Brooks F, Paringe V, Tonge A, Lewis J, Mohanty K. Patient estimates of healthcare costs in trauma and orthopaedics. Eur J Orthop Surg Traumatol. 2013;23(6):639-42.
15. Group TE. EuroQol-a new facility for the measurement of health-related quality of life. Health Policy. 1990;16(3):199-208.
16. Cella D, Riley W, Stone A, Rothrock N, Reeve B, Yount S, et al. The Patient-Reported Outcomes Measurement Information System (PROMIS) developed and tested its first wave of adult self-reported health outcome item banks: 2005–2008. J Clin Epidemiol. 2010;63(11):1179-94.
17. Burger H. Can the International Classification of Functioning, Disability and Health (ICF) be used in a prosthetics and orthotics outpatient clinic? Prosthet Orthot Int. 2011;35(3):302-9.
18. Schaffalitzky E, Gallagher P, MacLachlan M, Wegener ST. Developing consensus on important factors associated with lower limb prosthetic prescription and use. Disabil Rehabil. 2012;34(24):2085-94.
19. Resnik L, Borgia M. Reliability of outcome measures for people with lower-limb amputations: distinguishing true change from statistical error. Phys Ther. 2011;91(4):555-65.
20. Devlin NJ, Parkin D, Browne J. Patient-reported outcome measures in the NHS: new methods for analysing and reporting EQ-5D data. Health Econ. 2010;19(8):886-905.
21. Craig BM, Reeve BB, Brown PM, Cella D, Hays RD, Lipscomb J, et al. US valuation of health outcomes measured using the PROMIS-29. Value Health. 2014;17(8):846-53.
22. Craig BM, Reeve BB, Cella D, Hays RD, Pickard AS, Revicki DA. Demographic differences in health preferences in the United States. Med Care. 2014;52(4):307.

5.0 Qualitative Discussion of Outcome Measures

Quotes from the consensus process participants are included in *blue italics*.

5.1 Levels of use of Outcome Measures

Outcome measures have various possible functions. Three broad levels of outcome measure functionality were identified throughout the consensus process and their utility discussed, including:

- Service user/ clinician level (Micro)
- Service level/ Manager (Meso)
- Policy level (Macro)

5.1.1 Service User/ Clinician level (micro)

Outcome measures can be used before, during and after rehabilitation interventions to ascertain both baseline function and progress and help identify key areas of concern that joint rehabilitation efforts of the clinician and service user should focus on. Many Patient Reported Outcome Measures (PROMs) allow for the identification of key 'limiting factors' or issues that people with lower limb absence face, and can yield important information that may be overlooked by even the most thorough subjective assessment. This can be particularly useful in supporting less experienced or time pressured clinicians in determining the most important factors to a person with rehabilitation needs, and to ensure their individual circumstances and priorities are properly incorporated into the treatment plan. Furthermore, individual clinicians can track their own performance and improve their treatments over time through reflection and consistent use of outcome measures.

Outcome Measure benefits: Users

Outcome measurement can benefit individual users in several ways. Objective measurement of outcomes can help to develop and refine personal rehabilitation goals as well as provide individuals with feedback on their progress as a way of motivating them to actively participate in their rehabilitation. *"Seeing the improvement over time can be very good for users, it helps them stay motivated"*

Education and empowerment of users was identified as a means to promote active participation of people with lower limb absence in their own care and ensure ownership and direction is provided by the service user towards outcomes that are meaningful and valued by the individual. Users

themselves expressed that the ability to interact with professionals as a full member of the rehabilitation team and on an equal basis was important. *"I went on this journey, (of self-education and empowerment) to be able to be in control of my destiny. I can discuss eye to eye and make joint decisions."*

The person with lower limb absence is the key member of the rehabilitation team. They bring expertise about their own life circumstances, including their experience using a prosthesis, and thus need to be educated and empowered over time to fulfil this role in their rehabilitation team. *"How do I become expert on my own fitting?"*

Outcome measures can educate both directly and through clarifying conversations with treating clinicians. For example, the "Prosthesis Evaluation Questionnaire Residual Limb Health Subscale (PEQ-RLH)" was considered useful in gathering detailed information from a service user and encouraging reflection on aspects within the outcome measure so that, during follow up appointments, a more wholistic picture of residual limb health can be obtained. For instance, pain might start to be expressed in a more nuanced way as skin pain, residual limb pain or phantom pain or as being experienced while standing, walking or sitting, thus helping the user input to the clinical encounter in a way that drives towards resolution of the clinical problem. *"It is a good sub scale for patients to pay more attention to their residual limb"*

Users can further reflect on constructs or questions within outcome measures, which can challenge any self-limiting beliefs that may exist. For example, the TAPES-R asks people to rate their agreement with the question "Does having a prosthesis limit you in going to work" which may open up thought and conversation about the specific limiting factors for employment opportunities. Efforts to improve outcome measure results by individuals and their treating clinicians can then focus on the specific prosthetic solutions needed to be able to rate this more highly. *"If others expect things from me, I expect them from myself"*

5.1.2 Service Level/Manager (meso)

When managing rehabilitation services, it is important to review the performance of the service as a whole, to ensure the quality of care and plan and review quality improvement programs. The results of outcome measures can help identify areas where improvement is needed and review efforts to develop them. Investments in areas such as training can be tracked, and a point of diminishing return identified. General areas of poor outcomes can be identified, and corrective action planned that is more appropriate for the whole service than for individual clinicians e.g. improving referral mechanisms to local psychological services or incorporating physiotherapy services into the

treatment pathway. Benchmarking of a service can be undertaken to compare results over time or to best practice and identify areas of improvement needed. Analysis of levels of functionality at admission can be examined to determine opportunities for improvement of immediate post-operative rehabilitation. Analysis of user characteristics at time of admission can allow service development to cater more adequately to the population accessing services, for example the need for pre-prosthetic rehabilitation. Comparison of function at admission and discharge from rehabilitation can establish the degree of functional improvement that the rehabilitation service provides.

5.1.3 Policy Level (macro)

Health policy planners and decision makers need information about the outcome of investments in order to plan future service needs and expenditure. Measurement of outcomes improve the basis for the investment case for rehabilitation and can inform broad policy questions such as service-wide unmet need, usage and reimbursement of particular prosthetic components, incorporation of wheelchairs and/or other services. HRQoL measurement before and after routine care can be used in economic evaluations to compare across areas of health care to determine comparative cost effectiveness and determine the broader effects of rehabilitation on social, community and vocational participation. Whilst it is admirable and often necessary to seek to increase available funds for rehabilitation, it is often a prerequisite of additional funding to demonstrate that present funds are allocated efficiently. Both demonstrating the effectiveness of current expenditure and building a case for further spending is supported by outcome measures data.

5.2 COMPASS: Consensus of Outcome Measures for Prosthetics and limb Absence ServiceS

As described in Recommendation 1 (Section 4.3.1) the COMPASS consists of six outcome measures recommended for use in routine clinical practice:

1. Amputee Mobility Predictor (AMP)
2. Timed Up and Go (TUG)
3. Two Minute Walk Test (2MWT)
4. Prosthesis Evaluation Questionnaire Residual Limb Health Subscale (PEQ-RLH)

5. Prosthesis Evaluation Questionnaire Utility Subscale (PEQ-UT, and
6. Trinity Amputation and Prosthesis Experience Scales – Revised (TAPES-R)

5.2.1 Time burden COMPASS

Three of the outcome measures in the COMPASS are Performance Based (AMP, TUG and 2MWT) and it is anticipated that they will be collected by a member of the treating clinical team. They take around 20 minutes in aggregate to complete, which was seen as realistic to do at the beginning and end of each episode of care.

Three of the outcome measures are PROMs and can be administered in a variety of ways that does not require time from the treating clinical team such as by tablet or paper in the waiting area or through an online platform, with a link sent via email. The PROMs take the service user a total of about 25 minutes, which was seen as realistic although the treating clinicians should review and discuss the results with users in order to derive some of the benefits of these tools. Time will be also needed for scoring of PROMs, be this clinical or administration time. This time will not be insignificant with the PROMs in their current paper-based form. This time can be reduced with automation on online or application based (app) platforms if developed (Section 7.1.13).

5.3 COMPASS: Descriptions and discussion of each Outcome Measure

The outcome measures selected for inclusion in the COMPASS, the COMPASS+ and the COMPASS Adjunct are discussed below.

5.3.1 COMPASS Performance Based Outcome Measures

Amputee Mobility Predictor (AMP)

The AMP is based on Tinetti's Performance-Oriented Assessment of Mobility Problems (POMA) and the Duke Mobility Skills Profile (DMSP) (1), and is designed to assess the functional capacity and ambulatory potential of individuals with LLA (1-3). It involves activities of progressively increasing difficulty, including static and dynamic sitting and standing, transfers, and gait.

There are two possible configurations of the AMP, each consisting of 21 items. The AMPPRO is used for individuals with a prosthesis, and the AMPnoPRO for those without a prosthesis. It has been

noted that AMPnoPRO has better potential to assist prosthetic prescription and to predict functional mobility after fitting (1, 3). Possible scores range from 0 to 42 for the AMPPRO and 0 to 38 for the AMPnoPRO (due to elimination of one item measuring standing ability on the prosthetic side) with the scores from each not being directly comparable (1). Higher scores on either configuration of the AMP indicate better mobility.

The scoring form and protocol can be found in the publication by Gailey et al. 2002 (1). The AMP is free to use if printed for clinical purposes, however licensing fees may apply for use of the electronic format, commercial use, and some research purposes. The AMP is copyright of ART (Advanced Rehabilitation Therapy, LLC).

***ICF Category*:** Mobility (d4) and Products and Technology (e1)

Resources / equipment: Timer or stopwatch, 2 hard chairs with arms, a 12-inch (30cm) ruler, a 4-inch (10cm) high obstacle, pencil. There should be access to a set of stairs with 3 steps, and a 15-ft (4.5m) walkway. The AMP also requires a walker (1).

Administration time: approximately 10 – 15 minutes, with additional time for setup.

Languages other than English: French (3).

Commentary on the AMP:

The AMP (including the AMPnoPRO) was acknowledged as a somewhat time-consuming measure with the official 10-15 mins being challenged. Participants reported it taking 15-25 minutes and even 30 minutes with older users, although this was for occasional use and it was agreed that regular practice would reduce the implementation time. Another barrier to occasional use was identified as the long list of instructions, although once learning on the part of the administrators has occurred it was acknowledged that faster implementation would be possible.

The AMPs clinical utility was noted in defining rehabilitation goals around a variety of relevant physical movements included in the outcome measure. This utility was extended to defining timing of prosthetic rehabilitation with use of the AMPnoPRO and focusing pre-prosthetic rehabilitation goals on activities needed for successful prosthetic fitting. It was felt that the range of activities and movements assessed in the AMP allowed prediction of real-world function and successful prosthetic rehabilitation better than many other PerfOMs. *"We use it to determine if someone is a viable candidate for prosthetic rehabilitation."*

AMP was found to be usable in "basic settings", requiring only simple equipment such as 2 hard chairs, a walker, and three stairs. This was questioned in some clinics where moving into the

physiotherapy area would be needed to complete the AMP as the prosthetic clinic itself had no available space for this outcome measure to be performed.

People with lower activity levels, including those with new amputations and comorbidities, might find some of the AMP activities challenging, making the application of the outcome measure difficult but this was noted as beneficial to allow both the person with limb absence and the clinical team to track improvements over time. People with higher activity levels were likely to encounter ceiling effects, which would precipitate a need to use the CHAMP or another high-activity specific outcome measures to overcome

Timed Up and Go (TUG)

The TUG is designed to assess most of the manoeuvres that comprise basic mobility including balance, transfers, walking and turning (4).

(IMAGE SOURCED : SIONS ET AL. 2020)

The TUG involves standing up from a standard chair, walking a distance of 3 metres, turning 180 degrees, then walking back to the chair and sitting down (4, 5). The entire test is conducted at a comfortable walking speed (2, 6), with the time (in seconds) taken as the overall score. The use of assistive devices should be documented. A reduction in the overall score indicates an improvement

in basic mobility. A ceiling effect of around 7 – 10 seconds has been reported for fit elderly, or younger Individuals with LLA (8).

ICF Category: Mobility (d4)

Resources / equipment: Stopwatch or timer, measuring tape / wheel and a standard chair with arms (height between 40 – 50 cm / 15 – 20 inches (2, 6)). The test may be conducted in parallel bars should more assistance be required.

Administration time: Less than 1 minute, with additional time for set up (5).

Commentary on the TUG:

The TUG was discussed as being widely used in clinical practice across many populations including those with LLA, with the advantage of allowing comparability across health conditions. The incorporation of standing, sitting and turning was considered an advantage over the 10MWT while others preferred the L-Test due to the addition of a turn in both directions (not a requirement in the TUG). It was noted that some knee joints, including microprocessor knees, have features that purposively slow sitting and may increase the TUG score (thus indicating poorer result) when they may be more functional than other alternatives.

Being a relatively simple and short test, ceiling effects were noted for more active people with LLA, and it was discussed that the TUG should be completed in conjunction with a longer walking test as they would measure different aspects of mobility.

The lack of standardisation of the TUG was discussed as a limiting factor in terms of the comparability of scores. Different administration methods exist with some timing the trials from the moment of leaving the chair while others start the time when someone says go. Various objects are also used to demarcate the turning point, some clinics use a chair or a cone on the floor which requires the user to walk around, while other clinics place a mark on the floor allowing the user to turn at a single location and quickly move back along the walkway. This point of needing to standardise directions for implementation in order to allow comparability across settings was noted as relevant to all PerfOMS, *"Administration instructions would need to be well defined to make sure it is being performed consistently."*

Two Minute Walk Test (2MWT)

The 2MWT is the shorter version of the six and twelve-minute walking tests (7). Its psychometric properties have been found to correlate closely with both tests (7), the main advantage being the

shorter walking time making it more appropriate for persons, like those with LLA, who may have limited mobility (8).

The 2MWT is a single-task test; the only requirement is to walk for 2 minutes and cover as much distance as possible (7). Distance walked is the primary outcome of interest, but average walking speed can also be calculated (7). An increase in the distance walked over the 2-minute test period indicates an improvement in mobility. The use of assistive devices should be documented (7).

***ICF Category**:* Mobility (d4)

Resources / equipment: Stopwatch or timer, measuring tape / wheel to determine distance walked, a chair (for rest breaks if needed). The test is typically conducted indoors, along a quiet, uncarpeted hallway or corridor.

Administration time: Less than 5 minutes, with additional time for set up

Commentary on the 2MWT:

The 2MWT was widely favoured as a less time-consuming test than the 6MWT while still being somewhat predictive of the longer tests results. Issues with the 6MWT including adequate space that is free of interference and the number of turns needed during the testing period were still noted but acknowledged as somewhat less acute with the shorter 2MWT. Problems of possible variable administration of walking tests such as the 2MWT were noted with walkways in different locations likely to have variability in the number of turns, obstacles encountered etc. Such variability undermines the comparability of results across locations.

While it was noted that the 2MWT could be used to observe gait symmetry and identify prosthetic issues, it would be more challenging to identify issues related to endurance in a shorter walking test. In such cases, the 6MWT may be a better alternative. It was also noted that due to the shorter administration time, the 2MWT could be more suitable than longer PerfOMs for routine visits. A ceiling effect for highly active users was noted when compared to the 6MWT but overall its reduced time burden and clinical utility made it widely favourable. *"The 2MWT is easy, quick and provides reliable measure of change through rehabilitation."*

5.3.2 COMPASS Patient Reported Outcome Measures

Prosthetic Evaluation Questionnaire (PEQ) subscales

- ***Prosthesis Utility subscale***

- ***Residual Limb Health subscale***

The PEQ is a multidimensional questionnaire designed to assess prosthetic function and aspects of prosthesis-related quality of life in individuals with LLA (9, 10).

It consists of 82 items, divided into 9 subscales addressing 4 major areas of concern including prosthetic function, mobility, psycho-social aspects, and well-being (11, 12). There are also 40 additional items that are scored individually, covering concerns such as satisfaction, pain, prosthetic care and experience with the prosthesis (10-12).

Most questions begin with the common stem 'Over the past 4 weeks…' (12). Respondents are required to indicate their score on a 100mm Visual Analogue Scale (13), framed by anchors such as 'never' and 'all the time' on either end. Analysis and interpretation of scores requires careful measurement of each response on the VAS using a ruler, followed by calculations with reference to a scoring guide (2, 12). Scores are calculated per subscale, as average of the individual item scores and range from 0 to 100 with higher scores indicating better health.

The PEQ may be completed as a postal or email survey. The English version has recently been made available in electronic format as an easy-to-use application through Google Play or the Apple Store. The guide for scoring and analysis of the PEQ is available online: http://analisedemarcha.com/papers/o_p/peq/EN/peq-Evaluation_Guide.pdf

While the PEQ itself was excluded, two out of the nine PEQ subscales were included in the COMPASS:

- *Utility subscale*: 8 items assessing the fit, weight, comfort, and general use of the prosthesis
- *Residual Limb Health subscale:* 6 items assessing the state of skin and other functions of the skin while wearing the prosthesis

Key themes (based on ICF framework):

- *Utility subscale:* Self-care (d5), Products and Technology (e1), Mental Functions (b1)
- *Residual Limb Health subscale:* Functions of the skin and Related structures (b8), Self-care (d5), Products and Technology (e1) and Mental Functions (b1)

Administration time: Not reported (Estimated less than 5 minutes for each of these subscales)

Languages: English, Arabic, Spanish, Italian (11), Turkish, and Danish (12)

Commentary on PEQ Utility:

The PEQ Utility was noted as including items of interest such as the ability to don the prosthesis, satisfaction with the prosthesis and socket comfort that, while not mapped to the ICF, were critical to successful prosthetic fitting. These domains were noted as inherently important in gathering information from users and it was noted that they are factors that all types of prosthetic services should be able to address. These questions were seen by many as empowering for users as they would encourage conversation with service providers about areas that are essential to successful prosthetic fitting for people with LLA.

Commentary on PEQ Residual Limb Health:

Residual limb health was seen as a critical and relevant construct to examine routinely and one which had significant bearing on the overall outcomes of people with LLA. It was noted that asking these questions in a systematic way could help educate users to observe any skin changes and become more active participants in their own healthcare. As a result this would also enable users to provide better information at subsequent clinical appointments. *"The PEQ Residual Limb Health is a good scale for patients to pay more attention to their residual limb"*

The direct relevance for prosthetic socket design was noted and the ability for rehabilitation services to impact this domain was seen as an added reason to collect this information in order to measure and improve outcomes. While it was noted that this information should be collected in routine practice anyway, the benefit of it being collected in a systematic way to prompt less experienced clinicians was acknowledged. *"This is a better way to do it in an organized and routine way."*

The Trinity Amputation and Prosthesis Experience Scale - Revised (TAPES-R)

The TAPES-R is a revised version of the original TAPES; a multidimensional questionnaire designed to examine the psychosocial process involved in adjusting to amputation and the use of a prosthesis (14, 15). It consists of 64 items divided into four sections:

- *Section 1, Psychosocial Adjustment Scale*: Divided into 3 subscales with 5 items each including general adjustment, social adjustment and adjustment to limitation.

- *Section 2, Activity Restriction Scale*: consists of 10 items addressing various types of restriction including athletic activity restriction, social restriction, mobility restriction and occupational restriction (15).
- *Section 3, Satisfaction with Prosthesis Scale*: Comprises 8 items divided into two scales assessing the aesthetic satisfaction (3 items) and functional characteristics (5 items) of the prosthesis.
- *Section 4:* Consists of various items exploring the experience of phantom limb pain, residual limb pain, and other medical conditions not related to the amputation (15).

Each scale in Sections 1 to 4 are scored individually. Section 4, along with additional items for general information and demographic data do not contribute to scoring (15).

TAPES-R may be completed in person or postal survey. There is a user guide that should be followed and a scoring sheet for interpretation. The TAPES- R is available online: www.psychoprosthetics.ie

Key themes (based on ICF framework):

- *Psychosocial Adjustment Scale:* General tasks and demands (d2), Interpersonal Interactions and Relationships (d7), Major Life Areas (d8), Community, Social and Civic Life (d9), Support and Relationships (e3)
- *Activity Restriction Scale:* Mobility (d4) and Social and Civic Life (d9)
- *Satisfaction with Prosthesis Scale:* Sensory Functions and Pain (b2) and Products and Technology (e1)

Administration time: Approximately 15 minutes

Languages: English, French (16), Persian (17) and Turkish (18)

Commentary on the TAPES-R:

The full version of the TAPES-R was considered to be broad ranging and good overall outcome measure that in contrast to most other outcome measures, examines a wide range of constructs beyond simple mobility. It was commented that people who had experienced LLA more recently would be the primary target of this outcome measure due to a focus on adjustment to amputation in one scale. It was also noted that "adjustment" does not follow a fixed timeline *"This should be used primarily with patients who have been recently amputated."* Discussion followed that the Part 2 questions that asked about pain (residual limb and phantom) and other health conditions should not be limited to people with recently acquired LLA.

The usability and time burden of the TAPES-R was questioned, though it was noted that a lot of constructs are covered in a relatively short outcome measure which yields significant clinical utility.

5.3.3 COMPASS+ – Additional Outcome Measures for High Functioning Individuals

As described in Recommendation 2 (Section 4.3.1), two additional PerfOMs have been recommended for high-functioning individuals with LLA that can be used to supplement the COMPASS, these include:

- Comprehensive High-Level Activity Mobility Predictor (CHAMP) and
- Six Minute Walk Test (6MWT)

The value of a single universally recommended list of outcome measures is that all participants perform the same outcome measures to allow comparison and possible aggregation of results. Throughout the process, the suggestion of using different outcome measures for people in different circumstances, countries and phases of rehabilitation was resisted to meet this objective. However, for those with high activity levels and superior functioning it was felt that the ceiling effect of the COMPASS would prevent a true picture of their positive outcomes being meaningfully measured.

All people who are high functioning should complete the COMPASS and have the option of performing the COMPASS+ (6MWT and CHAMP) to overcome ceiling effects and measure improvements not captured by the COMPASS. The cut off score for the AMP used in research to progress to the CHAMP is a score of 37.

Time burden of COMPASS+

The aggregate time taken to complete both the COMPASS+ is about 25 minutes. Both tests are PerfOMs and therefore require at least one member of the clinical team to be involved. This was considered realistic, as in many settings a minority of people would be completing these additional measures and it is likely they would be intrinsically motivated to do so.

Required space for the COMPASS+

The COMPASS+ requires significant physical space to perform with a minimum area of 13 metres by 10 metres needed for the CHAMP, and an adequate walking track to walk unimpeded for 6 minutes for the 6MWT. While some rehabilitation facilities have the requisite space for these tests, it is acknowledged that many do not, with space being severely limited especially where rehabilitation services are co-located within a hospital.

Comprehensive High-Level Activity Mobility Predictor (CHAMP)

The CHAMP was originally developed to measure high-level mobility in Service Members with traumatic LLA, but has since also been applied successfully to all active individuals with LLA (quantified in research as a score of 37 or above on the AMP) (19).

The CHAMP consists of four advanced physical tests providing an overall assessment of balance, coordination, speed, power and multidirectional agility. Specifically:

- *Single-limb Stance Test:* length of time (in seconds) standing on each leg unassisted, with opposite foot raised at least 15.2 cm off the floor.
- *Edgren side step test*: Stepping sideways back and forth along 5 cones, placed in a line 1 metre apart, crossing as many cones as possible within a 10 second period (19).
- *T-Test*: time taken (in seconds) to complete a course moving forward, sideways and backward in a 10 metre x 10 metre T-shaped pattern (19); and
- *Illinois Agility Test*: a fast-paced test measuring the time taken (in seconds) to complete a course beginning with prone to stand transfer, moving forward, turning and weaving through several points placed 3.3 metres apart (19).

Overall scores for the CHAMP range from 0 to 40 and are determined by adding the best score of each of the four items (converted into a 0 to 10 scoring system). Higher scores indicate better performance.

ICF Category: Mobility (d4)

Resources / equipment: Stopwatch or timer and 19 cones to lay out the course. It should be noted that significant amount of space is required (minimum 13 metres by 10 metres). Minimal training is required to administer the CHAMP, with a training guide and instructions accessible online through the publication by Gailey et al. 2012 (19).

Administration time: 15 minutes

Commentary on the CHAMP:

The main advantage of the CHAMP was noted as being able to measure outcomes for people with high levels function. The CHAMP could therefore be used to measure and justify the effects of *"high activity products"* and other rehabilitation interventions intended to facilitate function beyond basic mobility. It was highlighted that the CHAMP was the only outcome measures that has this important characteristic. *"This is the only tool among the tools measuring high level function and thus I think important to be included."*

The CHAMP requires little equipment and what is needed was felt to be easily obtainable in most contexts with an experienced practitioner in LMIC's calling it "easy and cheap to implement". One constraint that was highlighted was physical space, with the 13 metre by 10 metre floor space being described as *"prohibitive in some settings"*.

The relevance for people with lower activity levels was questioned although it was pointed out that while many people run while performing the CHAMP, this is not necessary and it can be performed while walking as well. Despite this it was acknowledged that the CHAMP would not be the first-choice outcome measure for people with lower activity levels. *"Considering recommendation of minimum standard level OM, CHAMP might not be used for different contextual clinical rehabilitation set up(s)"*

Six Minute Walk Test (6MWT)

The 6MWT was developed in 1982 (20) as a modified version of the 12-minute walk test, designed to measure aerobic capacity and endurance (2).

The 6MWT is a single-task test; the only requirement is to walk as far as possible in 6 minutes, taking standing rest breaks as needed (21). The score used to compare changes in performance is the distance covered. An increase in the distance walked indicates an improvement in mobility. Use of assistive devices as well as the duration and number of rests taken (if any) should be recorded (21).

ICF Category: Mobility (d4)

Resources / equipment: Stopwatch or timer, measuring tape / wheel to determine distance walked and a suitable object (e.g. a cone or tape) to mark turnaround points along the walkway. Official ATS guidelines (21) suggest the test be completed on a 100-ft (30 metre), quiet, indoor, flat and straight hallway with markings every 3 metres however some studies have successfully used 65-ft (20 metre) and 165-ft (50 metre) walkways (22, 23).

Administration time: less than 10 minutes, with additional time for set up

Commentary on the 6MWT:

The 6MWT was the longest walking test included in this process, and was seen as useful and important for active users and persons with LLA high activity levels. The main utility of the measure was seen as being in the metric of measurement, ie the distance covered in 6 minutes of walking.

An additional possible benefit of the 6MWT is that while observing the administration, the observer can note gait performance and learn about socket comfort after a longer period of sustained walking, outcomes which are not documented within the test, but which have clinical utility. *"We get more information than how far someone can walk…we observe and view gait after some minutes of walking, this is when we see the problems of the socket and the prosthesis, including signs of asymmetry".*

The 6MWT protocol allows people to stop and rest as needed throughout. Allowing people to stop was supported by many as making the test more appropriate for older and less active people with LLA This inability for some users to finish the 6MWT without numerous breaks led to it being discussed as an outcome measure more suited to those with a higher activity level. *"We only do 6MWT with people with higher mobility grades."* It was agreed that a shorter walking test was more appropriate for newly amputated individuals or the elderly.

This and the other walking tests were acknowledged as being easy to use in most settings and not requiring resources other than space and time. The space for the 6MWT was a problem in some places, especially in LMICs where rehabilitation facilities may be small spaces and/or within a hospital compound. It was noted that the space also requires the absence of other people and obstacles, which was considered often hard to find. While doing this test consistently in one facility was considered somewhat possible it was thought there would be variability between facilities in the number of turns taken given the lay out of the walking space, making the results less reliable for direct comparison.

Risks of such a test to be recommended for all people with limb absence was noted due to problems with aerobic capacity and endurance, with the contraindications for the test including heart conditions and other comorbidities (21).

Some questioned the relevance of this test for every day functioning noting that most activity bursts are briefer than six minutes in order to complete activities of daily living. *"I am not sure it tells us anything meaningful from a patient's perspective."*

Others praised the utility stating it has *"excellent utility"* and that it may differentiate between prosthetic components better than the 2MWT.

5.3.4 COMPASS Adjunct

As described in Recommendation 3 (Section 4.3.1) the COMPASS Adjunct comprises the PSFS due to its significant clinical relevance.

A wholistic yet individualised interpretation of rehabilitation was encouraged with acknowledgement that each individual's situation and ability to cope with this situation is unique. *"What a test doesn't measure is the person's situation* (life circumstances) *and how they cope with situations that are important to them."*

Whilst outcome measures included in the COMPASS may be useful in determining goals through questions and constructs they include, open ended questions about personal goals were thought to better reflect good clinical practice.

A structured way of goal setting through use of the PSFS was advocated as an adjunct to the COMPASS. *"Quality of life is the amount of activities to have a fulfilled life, but this* (specific activities of importance) *are different for everyone".* Identification of individual goals allows the user and clinician to work together to identify a realistic vision of success. This also allows both the user and clinician to have agreed-upon and specific goals aligned with that definition.

While not as useful for between-user comparisons or pooled analysis of data from groups of users, the PSFS was viewed as encouraging the type of personalised conversations that might happen in a high-quality routine clinical encounter. Goal setting through the PSFS was noted as an opportunity for realistic goals to be set, discussed and achieved, which could be highly motivating for the user. Qualitative research could be conducted from the PSFS responses through content analysis which could allow for better understanding of a general population to be obtained in a structured way.

Patient Specific Function Scale (PSFS)

The PSFS is a questionnaire designed to identify important activity limitations, and to measure functional outcomes achieved by individuals with any orthopaedic condition (24, 25). It has been psychometrically tested for reliability in persons with LLA (2).

To complete the PSFS, individuals are asked to nominate up to five activities that they are currently having difficulty with or are unable to do, and are then asked to rate the current level of difficulty associated with each activity using an 11 – point scale where 0 = "unable to perform" and 10 = "able to perform the same level as before injury or problem" (24, 25). Data is collected at two or more time-points (e.g. at baseline/initial assessment and at follow-up).

Following the intervention, or at the follow-up appointment(s), individuals are again asked to rate the activities previously identified from 0 to 10 (and are given the chance to nominate new problematic activities that might have arisen during that time) (24, 25). Individuals are to select a value that best describes their current level of ability on each activity assessed, with the possibility of selecting and scoring new goals once they are achieved.

Key themes (based on ICF framework): Depends on the individual activities chosen

Administration time: 5 to 10 minutes

Languages: English, Danish, Turkish, Italian, French and Norwegian

Commentary on the PSFS:

The PSFS was noted as an outcome measure which makes it difficult to compare two people so has limited utility for management or policy makers from the perspective of being able to compare functional improvement. However, the primary advantage of the PSFS is that it openly invites people with LLA to define their own goals and review them at a future time. Through identification and planning to achieve goals, it focuses both service user and rehabilitation staff on attainment of these goals. This may also be used by health service planners and policy makers to indicate what people with LLA are seeking to achieve, so that this can be taken into consideration when planning and reviewing services.

5.3.5 COMPASS - Generic Health Related Quality of Life Outcome Measures

As per recommendation 4 (Section 4.3.1), a generic HRQoL outcome measure such as the EQ-5D-5L or PROMIS-29© can be used to supplement the ISPO COMPASS.

The PROMIS-29 was included by the systematic review and advocated by the expert panel as it had been psychometrically tested in people with LLA. The EQ-5D-5L was nominated despite its psychometric properties not having been examined in LLA, as it has been extensively tested in similar populations.

There was general agreement that a generic HRQoL measure was critical for policy-level decision-making as it allows for calculation of Quality Adjusted Life Years (QALYs) which is a common metric of health-related utility and allows multiple health conditions to be compared in economic evaluations. Generic HRQoL measures are good at capturing the health state of people with comorbidities, as is often the case for many with LLA.

Rather than recommend a single generic HRQoL measure within the COMPASS, it was decided by the consensus participants to make a general recommendation that a generic measure relevant to the context of use should be adopted. This allows within-country comparisons to be made more easily to other parts of the local health service with the results being readily understandable to policy makers.

It was noted that crosswalks may exist to convert values from one HRQoL measure to another for international comparison, should the need arise.

Concerns were expressed that a generic HRQoL measure is likely to be less responsive to change than condition-specific measures, so while they allow comparisons across health states, they may be less sensitive at discerning differences in outcomes within a single health condition.

It was recognised that outcome measures that examine broad aspects of health, such as Quality of Life (QoL) were important to users. More limited tests, like those PerfOMS that quantify an aspect of mobility performance (e.g., walking for a short period) were viewed as a useful proxy for important domains of rehabilitation, but acknowledged as not the end goal in itself. For example, successful mobilisation only acts to open up possibilities for a fulfilled life. *"Yes, I can walk this distance in this amount of time, but what does this mean?"*

Examination of HRQoL was encouraged by various participants in the process in order to open up rehabilitation to achievement of improvements that are meaningful to rehabilitation users. *"What are the criteria for each person to have a good quality of life, then comes the next question, what do I need to achieve this?"*

It was highlighted by prosthetic users that their 'technology' was only one part of why their rehabilitation was successful and rather that success was much more multifactorial.

Patient Reported Outcome Measures Information System – 29 item profile (PROMIS-29®)

PROMIS-29 item Profile consists of a set of short-form measures from select PROMIS® item banks (26-28); the most recent version is the PROMIS-29 v2.1. All PROMIS measures, or short-forms, have been developed using rigorous methodology including techniques from both Classical Test Theory (CTT) and Item Response Theory (IRT).

The PROMIS-29® is made up of 29 items specifically designed to assess HRQoL in the domains of anxiety, depression, fatigue, pain, physical function, sleep disturbance and ability to participate in social roles and activities. Each section consists of 4-items, most of which begin with the phrase "in the past 7 days…". It has a five-option response format which produces a score from 1 to 5 for each

item. The final section in the PROMIS-29 is a single item measure of pain intensity, which is rated on a 11-point numerical scale with 0 indicating 'no pain' and 10 equating to the 'worst pain imaginable'.

Raw scores from each section (i.e., domain) of the PROMIS-29® are used to derive T-scores using provided lookup tables, with higher T-scores indicating higher levels of the construct being measured. For example, while a higher T-score for fatigue equates to worse health, a higher T-score for physical function indicates better function. PROMIS T-scores all have a mean of 50 and a standard deviation of 10 to facilitate interpretation. The mean of 50 reflects the average T-score for the calibration sample (typically a large sample representative of the US general population). T-scores derived from PROMIS® instruments are standardised and can be compared across other PROMIS® instruments provided they are derived using the appropriate scoring table. PROMIS-29® is available in both paper and electronic formats.

Key themes (based on ICF framework): Mental Functions (b1), Sensory Functions and Pain (b2), General Tasks and Demands (d2), Mobility (d4), Domestic Life (d6), Interpersonal Interactions and Relationships (d7), Community Social and, Civic Life (d9), Support & Relationships (e3)

Administration time: Less than 5 minutes

Languages: Available in 48 languages other than English, including Spanish, Portugese, Arabic, French and Hindi. A full list of languages can be viewed via the PROMIS® online portal: https://www.healthmeasures.net/explore-measurement-systems/promis/intro-to-promis/available-translations

Commentary on the PROMIS-29:

This generic HRQoL instrument was advocated due to the ability to compare across health conditions and having been psychometrically tested in a population of people with LLA. The extensive list of available languages (48) was felt to be a distinct advantage of the PROMIS-29 along with the existence of crosswalks to other generic measures of quality of life such as the EQ-5D. Furthermore, the ability to calculate a Quality adjusted Life Year (QALY) which can then be used in economic evaluations, was considered a key aspect in order to allow its use in wider health policy and building an investment case for rehabilitation. It was also highlighted that the more positive framing of the questions in the PROMIS were considered an advantage over other generic HRQoL measures. *"Policy value of the PROMIS seems great, is also more balanced in framing. As a user, EQ5D is negatively framed and borderline depressing"*

EuroQol (EQ) 5D-5L

The EQ 5D-5L also known as the EuroQoL (29-32), consists of five single-item questions encompassing five dimensions, including:

1. Mobility
2. Self-care
3. Usual activity
4. Pain and discomfort
5. Anxiety and depression

Under each dimension, individuals are required to select the best of five possible responses. The EQ 5D-5L Index is calculated from a predefined algorithm, with a score of 1 representing the best imaginable health state, 0 representing death, and negative scores indicating a state worse than death (30, 32). The EQ 5D-5L has not yet been psychometrically tested in persons with LLA.

The EQ 5D-5L has an accompanying visual analogue scale (VAS) where individuals can self-report their current health state (0 = worst imaginable health state, 100 = best imaginable health state) (32). User guides and the EQ 5D-5L are available online, and in various modes of administration: https://euroqol.org/publications/user-guides/

It is free for non-commercial organisations but registration for its use is required. Commercial users are charged a fee. Registration for the use of the EQ 5D-5L can be accessed online: https://euroqol.org/eq-5d-registration-form/

Key themes (based on ICF framework): Mobility (d4), Self-care (d5), Major Life Areas (d8), Domestic Life (d6), Community Social and Civic Life (d9), Sensory Functions and Pain (b2), Mental Functions (b1)

Administration time: Less than 5 minutes

Languages: Available in more than 130 languages

Commentary on the EQ-5D:

Nominated as an additional outcome measure without psychometric properties reported for people with LLA, the EQ-5D was suggested to be included as a common HRQoL measure that is commonly used in economic evaluations.

Similar to the PROMIS-29, the EQ-5D-5L can be used to generate a Quality Adjusted Life Year (QALY) that is a common unit of well-being often used to compare to other health conditions and current or future interventions/treatments.

It was indicated that the EQ-5D's formulation of questions may be perceived as negative from a user perspective. *"As a user the framing of questions is quite negative and can be demoralising or 'triggering.' We have also had some issues providing this to other users for similar reasons. It is a shame that there is not (yet) a more neutral or positively-phrased version."*

It should be noted that the EQ-5D has not been psychometrically tested in populations with LLA and this may lead to results that are not valid or reliable in this population. This was defended due to its widespread use and testing across multiple regions and health conditions. *"Has been examined in so many places to examine its cultural translatability and relevance."*

5.4 Discussion

5.4.1 Lack of Outcome measures appropriate to examine activities in Low- and Middle-Income Countries (LMICs)

As detailed in Recommendation 5 (Section 4.3.1), further development of outcome measure more sensitive to the user's ability to perform activities relevant for LMICs is required, with a focus on activities such as sitting cross-legged, kneeling, squatting, and other culturally important activities related to mobility.

As each outcome measure was discussed, the appropriateness of the instrument was reviewed to ensure relevance in all parts of the world. During this process it was noted that outcome measures developed in High Income Countries (HICs) tend to focus on activities relevant in HICs such as standing from a seated position (i.e. from a chair) or performing other activities of daily living which are less relevant in many LMICs. The absence of psychometrically tested outcome measures that assess cultural activities and behaviours specific to those of LMICs, therefore limits the applicability of outcome measures developed in/for HICs in some situations.

Activities relevant to LMICs that could be incorporated into outcome measures in the future include:

- Rising to stand from a seated position on the floor
- Ability to sit or kneel on the floor (e.g. Cross-legged sitting)
- Ability or restriction in squatting
- Ability or restriction in kneeling, and
- Ability to donn/doff shoes

"I think it's becoming clear that there isn't a good outcome measure that takes into account squatting toilets, people living in stilted houses with access up steep ladders, people who kneel on the

floor for prayer and worship, people who sit cross legged to eat with the family… those things are not captured in a meaningful way, and if this process identifies these things that are not captured perhaps that can go forward as a call for more research and better understanding and more listening to the people we serve."

A suggestion was made for development of outcome measures that were similar in form to those used in HICs, which incorporate relevant seating positions, such as a cross-legged sitting version of the TUG that could be used *in addition* to the standard TUG.

5.4.2 Consistency of use of outcome measures

Consensus participants noted that outcome measures should be completed in their published and validated formats to yield comparable results. It was noted throughout the consensus process that various outcome measures have been changed at times by practitioners implementing them in clinical settings. This is problematic and may undermine the reliability of the outcome measure.

PerfOMs should be administered in a manner described by the original developers and with a standardised protocol. Inconsistency in the application of the test, such as turning 180 degrees at the end of a walkway by reaching a mark on the floor versus walking around a cone or a chair will introduce error and affect reliability. Such error may produce a slightly different result for the same participant and introduce increased variability in results, leading to challenges when trying to combine or pool data across sites.

PROMs questions should not be changed from their original format, including in translations into a different language. Inconsistency can occur through clinicians seeking to improve the relevance of the measure to their specific context, for example changing reference to crossing icy ground to crossing slippery, muddy ground. Whilst changing the wording of a question in an outcome measure might seem innocuous, it should be noted that the altered measure cannot be assumed to still possess the same psychometric properties of the original version.

All outcome measures in the COMPASS have been considered by the consensus process to be relevant, understandable, and appropriate for use in clinical practice worldwide and their known psychometric properties have been reviewed. All outcome measures should be performed in their validated published form using the protocols described in the COMPASS Users Guide. In addition, the discussion and commentary of each outcome measure can yield insights into the utility of each measure.

5.4.3 PROMs for Mobility

The COMPASS includes no PROMs for measuring mobility, despite significant research and development of mobility-specific outcome measures in recent years. Whilst mobility was found to be the most important ICF domain by consensus participants, discussions revealed a lack of certainty that such outcome measures would be appropriate in terms of the questions asked for all people with LLA across the world. It should be noted that consensus participants were not as familiar with these PROMs as they were with PerfOMs and that the PerfOMs assessing mobility were easier to conceptualise from a written description. Given the consensus participants were a sample of people involved in rehabilitation of people with LLA, along with users, this lack of familiarity likely reflects a general lack of familiarity with PROMs in routine clinical practice.

Measuring one's perceived mobility, as well as one's demonstrated mobility could be important (especially when these two things do not align) and present an opportunity for clinical intervention. It should be noted that at least one outcome measure, the PLUS-M overcomes the problem of broad appropriateness with a bank of short forms, enabling selection of different items for different regions of the world, thereby still measuring mobility but doing it in different ways (with culturally applicable items). A lack of familiarity with the PROMs for mobility by consensus participants extended to the PLUS-M and the banks of questions were not discussed in depth. Use of PROMs to measure perceived mobility may become more common with time and any update of the COMPASS should examine inclusion of such an instrument carefully.

5.4.4 Psychological Related Outcome Measures

There was some reluctance to explore domains considered unimprovable in all contexts, due to a lack of resources in the form of appropriate services to deal with issues that might be revealed by the measure. This led to some outcome measures being excluded such as the Amputee Body Image Scale - Revised (ABIS-R) as participants in the consensus process felt it could build expectations for psychological services that do not universally exist or present clinicians with clinical issues that were outside their scope of practice or expertise.

While the entire TAPES-R was included in the COMPASS, it should be noted that there was some disagreement about including the sub scale on Psychological Adjustment. This scale was seen by some to lead towards areas that some prosthetists and other practitioners, without the support of a multidisciplinary team, were not equipped to deal with. It was felt that the questions in the TAPES-R Psychological Adjustment subscale are important and general enough to be informative to any professional even if specialist psychological services are not available. Where possible, linkage to any

available psychological support services should be encouraged. *"A better connection between prosthetic services and psychological support systems is important – this issue is significant"*

5.4.5 International Classification of Functioning (ICF) - Ranking

Ranking of the ICF domains was undertaken to ensure the most important domains covered by the outcome measures within this process were captured within the final COMPASS.

ICF Categories / Themes: Ranking results from most important to least important

1. Mobility (d4)
2. Sensory Functions & Pain (b2)
3. Self-care (d5)
4. General Tasks & Demands (d2)
5. Functions of the skin and Related Structures (b8)
6. Domestic Life (d6)
7. Community, Social & Civic Life (d9)
8. Major Life Areas (d8)
9. Products & Technology (e1)
10. Mental Functions (b1)
11. Interpersonal Interactions & Relationships (d7)
12. Support & Relationships (e3)

Three categories not covered by the ICF were later added and not included in the ranking. These were:

1. Satisfaction with Prosthesis
2. Socket Comfort, and
3. Donning Ability

Ultimately it was felt that all domains were important enough to include within the COMPASS by selection of outcome measures that explore each domain identified. The ICF rankings were not disaggregated based on region or background but were the overall rankings given from all participants.

The three domains ranked most important, mobility (d4), sensory function and pain (b2) and self-care (d5) are examined with multiple measures in the final COMPASS.

5.4.6 Translations of the COMPASS

As detailed in recommendation 6 (Section 4.3.1) from the results the COMPASS outcome measures should be translated, assessed for psychometric performance in the translated form, and shared broadly. Psychometric testing of translated outcome measures is important to ensure that the validity and reliability of the translated measures is adequate and that comparability between measurements conducted in different languages is possible.

Particular attention should be given to the PROMs in the COMPASS outcome measure list to improve their usability worldwide as PerfOMs may be used more readily by clinical staff who can understand the language they are published in. ISPO national member societies and regions will be encouraged to promote translation into relevant regional languages and will be encouraged to collaborate with researchers to ensure validity of the translated outcome measures.

5.4.7 Limitations

Limitations of the COMPASS, COMPASS+, COMPASS Adjunct and the recommendation to use a generic HRQoL measure include the lack inclusion of a PROM for mobility to determine differences between perceived and actual function and the advantages of a PROM in terms of clinical time of administration and automated scoring. The current lack of an online or application-based system to automatically score and record results may add to the overall time burden, and this was not explored as fully within the consensus process as the time burden of application of the outcome measures. The overall time burden may not be realistic in all locations, especially LMICs, given the lack of trained staff to provide clinical services and the significant unmet need that exists in many parts of the world, which may lead to managers and clinicians seeking to minimise the time spent with each user in order to allow clinical time for others, in order to improve overall access.

The comparability of outcome measure results is dependent on consistent application using protocols in their published and validated forms, something that the COMPASS User Guide can address if adopted widely. Consistent use of outcome measures in their standard form will be challenging to monitor or evaluate.

5.4.8 Implementation and suggested use of the COMPASS

To support routine use of the COMPASS, various activities have and will be undertaken by ISPO:

- The COMPASS User Guide will be available on the ISPO web site and instructional videos will be completed to facilitate consistent use of the outcome measures within routine clinical practice worldwide.

- Promotional videos will be produced to highlight the existence of COMPASS, its usability and utility.
- Short courses will be available on ISPOLearn and Physiopedia to educate clinicians on the COMPASS and how it can be used.

Publication of the entire process in peer reviewed publications is planned to allow understanding and critique of the process and its outcomes.

References

1. Gailey RS, Roach KE, Applegate EB, Cho B, Cunniffe B, Licht S, 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):613-27.
2. Resnik L, Borgia M. Reliability of outcome measures for people with lower-limb amputations: distinguishing true change from statistical error. Phys Ther. 2011;91(4):555-65.
3. Côté-Martin M-É, Tremblay A, Couture M, Roy J-S. Translation, Reliability, and Validity of the French Version of the Amputee Mobility Predictor. J Prosthet Orthot. 2020;32(2):101-6.
4. Podsiadlo D, Richardson S. The timed “Up & Go”: a test of basic functional mobility for frail elderly persons. J Am Geriatr Soc. 1991;39(2):142-8.
5. Condie E, Scott H, Treweek S. Lower limb prosthetic outcome measures: a review of the literature 1995 to 2005. J Prosthet Orthot. 2006;18(6):P13-P45.
6. Schoppen T, Boonstra A, Groothoff JW, de Vries J, Goeken LN, Eisma WH. The Timed "up and go" test: reliability and validity in persons with unilateral lower limb amputation. Arch Phys Med Rehabil. 1999;80(7):825-8.
7. Gremeaux V, Damak S, Troisgros O, Feki A, Laroche D, Perennou D, et al. Selecting a test for the clinical assessment of balance and walking capacity at the definitive fitting state after unilateral amputation: a comparative study. Prosthet Orthot Int. 2012;36(4):415-22.
8. Heinemann AW, Connelly L, Ehrlich-Jones L, Fatone S. Outcome instruments for prosthetics: clinical applications. Phys Med Rehabil Clin. 2014;25(1):179-98.
9. Legro MW, Reiber GD, Smith DG, del Aguila M, Larsen J, Boone D. Prosthesis evaluation questionnaire for persons with lower limb amputations: assessing prosthesis-related quality of life. Arch Phys Med Rehabil. 1998;79(8):931-8.
10. Boone DA, Coleman KL. Use of the prosthesis evaluation questionnaire (PEQ). J Prosthet Orthot. 2006;18(6):P68-P79.
11. Ferriero G, Dughi D, Orlandini D, Moscato T, Nicita D, Franchignoni F. Measuring long-term outcome in people with lower limb amputation: cross-validation of the Italian versions of the Prosthetic Profile of the Amputee and Prosthesis Evaluation Questionnaire. Eura Medicophys. 2005;41(1):1-6.
12. Christensen J, Doherty P, Bjorner JB, Langberg H. Reliability and construct validity of a new Danish translation of the Prosthesis Evaluation Questionnaire in a population of Danish amputees. Prosthet Orthot Int. 2017;41(5):469-75.
13. Franchignoni F, Giordano A, Ferriero G, Orlandini D, Amoresano A, Perucca L. Measuring mobility in people with lower limb amputation: Rasch analysis of the mobility section of the prosthesis evaluation questionnaire. J Rehabil Med. 2007;39(2):138-44.
14. Gallagher P, Maclachlan M. The Trinity Amputation and Prosthesis Experience Scales and quality of life in people with lower-limb amputation. Arch Phys Med Rehabil. 2004;85(5):730-6.
15. Gallagher P, Franchignoni F, Giordano A, MacLachlan M. Trinity amputation and prosthesis experience scales: a psychometric assessment using classical test theory and rasch analysis. Am J Phys Med Rehabil. 2010;89(6):487-96.
16. Luthi F, Praz C, Leger B, Vouilloz A, Favre C, Loiret I, et al. Cross-cultural adaptation and measurement properties of the French version of the Trinity Amputation and Prosthesis Experience Scales-Revised (TAPES-R). PLoS ONE [Electronic Resource]. 2020;15(2):e0229084.
17. Mazaheri M, Fardipour S, Salavati M, Hadadi M, Negahban H, Bahramizadeh M, et al. The Persian version of Trinity Amputation and Prosthetics Experience Scale: translation, factor structure, reliability and validity. Disabil Rehabil. 2011;33(19-20):1737-45.
18. Topuz S, Ulger O, Yakut Y, Gul Sener F. Reliability and construct validity of the Turkish version of the Trinity Amputation and Prosthetic Experience Scales (TAPES) in lower limb amputees. Prosthet Orthot Int. 2011;35(2):201-6.
19. Gailey RS. Development and reliability testing of the Comprehensive High-Level Activity Mobility Predictor (CHAMP) in male servicemembers with traumatic lower-limb loss. J Rehabil Res Dev. 2013;50(7):905-18.

20. Butland R, Pang J, Gross E, Woodcock A, Geddes D. Two-, six-, and 12-minute walking tests in respiratory disease. Br Med J (Clin Res Ed). 1982;284(6329):1607.
21. Laboratories ACoPSfCPF. ATS statement: guidelines for the six-minute walk test. Am J Respir Crit Care Med. 2002;166:111-7.
22. Lipkin D, Scriven A, Crake T, Poole-Wilson P. Six minute walking test for assessing exercise capacity in chronic heart failure. Br Med J (Clin Res Ed). 1986;292(6521):653-5.
23. Troosters T, Gosselink R, Decramer M. Six minute walking distance in healthy elderly subjects. Eur Respir J. 1999;14(2):270-4.
24. Sterling M, Brentnall D. Patient Specific Functional Scale. Aust J Physiother. 2007;53(1):65.
25. 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;47:258-63.
26. Cella D, Riley W, Stone A, Rothrock N, Reeve B, Yount S, et al. The Patient-Reported Outcomes Measurement Information System (PROMIS) developed and tested its first wave of adult self-reported health outcome item banks: 2005–2008. J Clin Epidemiol. 2010;63(11):1179-94.
27. Amtmann D, Morgan SJ, Kim J, Hafner BJ. Health-related profiles of people with lower limb loss. Arch Phys Med Rehabil. 2015;96(8):1474-83.
28. Health Measures: Transforming How Health is Measured. PROMIS: Northwestern University; 2008 [updated 2021; cited 14 April 2021]. Available from: https://www.healthmeasures.net/explore-measurement-systems/promis.
29. Rabin R, Charro F. EQ-5D: a measure of health status from the EuroQol Group. Ann Med. 2001;33.
30. EuroQoL Research Foundation. EQ-5D [updated 2021; cited 4 June 2021]. Available from: https://euroqol.org/eq-5d-instruments/eq-5d-5l-about/.
31. Brooks F, Paringe V, Tonge A, Lewis J, Mohanty K. Patient estimates of healthcare costs in trauma and orthopaedics. Eur J Orthop Surg Traumatol. 2013;23(6):639-42.
32. The EuroQol Group. EuroQol-a new facility for the measurement of health-related quality of life. Health Policy. 1990;16(3):199-208.

6.0 LEAD (Lower Extremity Amputation Dataset)- Development of a Core Dataset

The LEAD (Lower Extremity Absence Dataset) is a core dataset (CDS) intended to standardise and improve data collection efforts for people with LLA. The LEAD CDS can be used in a data base or registry (with data collected at designated time points) and allows for collection of participant data and demographics, rehabilitation intervention details and outcomes. The COMPASS makes up part of the LEAD, with data from the outcome measures collected before and after rehabilitation services are offered. The LEAD has been developed in parallel with the COMPASS to minimise overlap and reduce the overall time burden of data collection.

To develop the LEAD, firstly a scoping review of published work of registries and databases relevant to people with LLA was completed. In addition, a series of semi structured interviews were conducted with the authors of papers found in the scoping review and through ISPOs contacts. Secondly, a consensus process was held with a diverse group to determine the data elements felt to be important for inclusion.

6.1 Scoping Review

6.1.1 Background

Rationale

A scoping review was conducted to establish the names and details of databases and registries related to prosthetics, orthotics, wheelchairs and other externally worn or used assistive devices. Through the scoping review, information was also collected about registry design, registry usage, data items collected within the registry, the target population and other relevant details.

The rationale for this research was to understand the diversity of current and past practice and to gain an understanding of what has been done previously, with particular respect to data elements within a registry. Thus, the ISPO LEAD consensus process could consider data items that had been used previously in similar registries and databases in addition to getting ideas from the consensus participants themselves. Information that could be determined about existing registries and their data items was also explored by reviewing annual reports and other associated grey literature.

Further, conversation and debate of the relevance of data items in past or current efforts was seen as important to spurn ideas and possible usage of data items within the consensus process. In many instances an existing registry will have gone through a long process of initial development and subsequent refinement, resulting in data items that are both relevant for use and practical to collect in their specific settings.

Objective

The objective was to establish which current and past registries or databases have been developed worldwide for people with LLA and other health conditions that detail externally worn or used assistive technology, and establish details of the registry's or database's design, usage, data collection methods, data items and data protection methods.

6.1.2 Methods

Eligibility Criteria

The eligibility included peer reviewed and published work in the English language with no time restrictions. The rationale for this criterion was that establishing even the name of a registry or database would allow a grey literature search to be completed in addition to the work found through the scoping review. It was also considered that past efforts, even if discontinued, would be valuable sources of information.

Information Sources

The search was executed on the 14th January 2021 in Medline, Embase and CINAHL databases. Registries that were mentioned briefly within articles found in the scoping review were targeted with grey literature searches using an internet search (Google and Google Scholar) to find additional information about registries from sources such as annual reports and websites, including contact information and further information for data extraction.

Search

MESH terms were identified for each registry from the following key words. Population: Amputation*, Allied Health. Intervention: Prosthe***, Assistive Technology, Assistive device*, Mobility Device*, Wheelchair*, Adaptive technology. Outcome: Registry, Registries, Data set, Dataset, Database, Register.

An example of the search is presented in Table 6.1 for Embase, using only MESH terms:

Table 6.1. MESH search string for scoping review with terms adapted for Embase

limb Prosthesis OR assistive technology OR assistive technology devices OR ambulation aids OR wheelchairs
AND
databases OR Database design OR database management software

The rationale for using only MESH terms, and not keywords, was to restrict the number of returned results, with the full key word search of the Embase database yielding 13,178 articles while the MESH term search returned 223. Scanning the results of both the key word and MESH term searches revealed that numerous articles returned in the keyword search mentioned the word 'database' in passing only.

Selection of sources of evidence

Sources of evidence identified from the above searches were deemed eligible for inclusion if they referred to a registry or database related to prostheses and/or other assistive technology (e.g. prostheses, orthoses, wheelchairs, etc. Articles reporting information or reporting from a registry were included as were articles detailing the development and purpose of a registry. When articles mentioned the existence of a registry or database this was further explored in a grey literature search.

Data Charting Process and Data Items

Data was charted from each article related to the following topics:

- Purpose and Objectives
- Eligibility Criteria/Target Population
- Data Collection Methods
- Demographics Information
- Confounders
- Surgical Interventions

- Rehabilitation/Prosthetic Interventions
- Outcomes
- Data Protection and,
- Other important information.

Generic publications about health registries had their data charted within the following topics:

- Purpose and Objective
- Stakeholders
- Funding
- Type of Registry
- Data Elements
- Patient Population Identification
- Patient Population Recruitment
- Data handling
- Data privacy.

Data items in articles about broader registry design were extracted considering the current broad target population (people with LLA) and hence data items such as sexual history or past exposure to environmental risk factors were not extracted as they were seen as not relevant to LLA.

Efforts were made to contact each registry, if it still existed, to participate in semi-structured interviews intended to cross-check and expand the available information.

Critical Appraisal of individual sources of information

No critical appraisal was completed within the scoping review with respect to the articles found. The rationale for this was that all data items and other registry particulars charted from these articles (design, inclusion criteria, etc.) would be viewed within the consensus process, discussed and viewed in light of their contemporary relevance for a CDS for people with LLA.

Synthesis of results

Within each category of data charting, duplicate data items were removed or combined where relevant. For thematic descriptions of categories, such as purpose or target population, all sources that mentioned them were referenced and any diversity arising from the various sources was discussed within the final descriptions.

6.1.3 Results

Selection of sources of evidence

After 19 duplicates were removed, a total of 367 citations were identified from searches of electronic databases. Based on the title and abstract, 301 were excluded with 35 full text articles to be retrieved and assessed for eligibility. Of these, 22 were excluded for the following reasons:

- Registry listed only the prevalence of people receiving a type of assistive technology in broad categories (n=8).
- Not enough information in the article to extract data items but contact successfully made for Semi Structured Interview (n=4).
- Not enough information in article to extract but contact made for more information with no response (n=10).

The grey literature search revealed 7 sources of information including the annual report from an existing registry (SwedeAMP), the websites of two government departments which maintain registries (Department of Veterans Affairs USA and Department of Defence USA), a lower limb registry in development (Limb Loss and Preservation Registry USA), the reports of a humanitarian organisation which collects data in an internal database (International Committee of the Red Cross) and the websites of two existing registries (AMPROM, UK and Scottish Physiotherapy Amputee Research Group, UK).

Characteristics of the Sources of Evidence

The articles included provide details about; The Danish Amputation Register (1), SwedeAMP (2), Funtional Mobility Assessment (FMA) Uniform Dataset (UDS) for wheelchairs (3), International Society of Wheelchair Professionals (ISWP) Minimum Uniform Dataset (MUD) (4), Wheelchair Users Registry (5). General publications about registry design included were Balancing the Optimal and the Feasible: A Practical Guide for Setting Up Patient Registries for the Collection of Real-World Data for Health Care Decision Making Based on Dutch Experiences (6) and Registries for evaluating patient outcomes: A Users Guide Third Edition (7).

Semi Structured Interviews

Relevant registries identified in the scoping review and through ISPO's network were contacted and asked to participate in semi-structured interviews. In these interviews, information from the scoping review and grey literature was checked and other information sought.

Interviews were conducted with the registries or database owners listed in Table 6.1.

Table 6.1 Names of all registries/databases that participated in semi structured interviews

Name of Registry/Database	Country
AMPROM	UK
ASCENT	Philippines
Department of Defence	USA
International Committee of the Red Cross	Switzerland (international coverage)
Limb Loss and Preservation Registry	USA
OttoBock	Germany (International coverage)
Scottish Physiotherapist Amputee Research Group	Scotland
SwedeAMP	Sweden
Department of Veterans Affairs	USA

Synthesis of Results: Scoping review and semi-structured interviews

The synthesis of information from both the scoping review and the semi-structured interviews was formed into preambles and questions for the consensus participants. This was delivered in an online format (SurveyMonkey).

Limitations

The scoping review and subsequent semi-structured interviews should be viewed in light of their limitations. The scope was limited to assistive technology due to the similarity of design needed for a registry for people with LLA that needs to encompass prosthetic service provision. Given the relative paucity of work in this area, it is possible that a broader scope including registries which focused on other chronic diseases and implanted medical technology would yield registries with a wider variety of designs and data elements that might prove useful to a registry focused on people with LLA and their prosthetic treatment.

Only one registry Scottish Physiotherapist Amputee Research Group (SPARG), was successfully contacted by emailing authors or making requests through websites for semi-structured interviews,

with all other successful contacts made through ISPO. Five registries or databases were not discovered in the scoping review (LLPR, AMPROM, Otto Bock, ICRC and ASCENT), rather networks and personal contacts of ISPO and ATscale were leveraged to establish their existence and obtain information. Given this experience in searching for registries, it is likely that other registries and databases exist within governments and other organisations that were not contacted for interviews.

6.2 Consensus Process

6.2.1 Method

Selection of consensus participants

Selection criteria to identify potential consensus participants was set with input from the ISPO IAG. Requests for nominations were sent from December 2020 to various organisations deemed relevant, including international professional societies, international non-government organisations and international organisations working with people with LLA and representatives of various registries and databases including everyone who participated in the semi-structured interviews. Nominations were accepted based on the following criteria:

1. Individuals' clinical experience with persons with LLA (i.e. prosthetists, orthotists, physiotherapists, occupational therapists, psychologists, and medical doctors)
2. Being a health clinic manager
3. Being a health policy maker
4. Researches in the field of LLA
5. Patients/end-users
6. Experience setting up or maintaining a registry or database.

Attendees were required to complete the online surveys and be available for all online meetings. Select potential consensus participants were also asked to suggest other participants for nomination to improve overall representation.

Consensus participant demographics

All consensus participants were considered to be experts at the time of meeting with regards to being a clinician who regularly consults with individuals with LLA, a researcher with a sound knowledge of population research, or being a health clinic manager or health policy maker, or an end-user with lived experience. In total, 37 participants were involved in the consensus process (14

female, 23 male, 14 from low and middle income countries, and 23 from high income countries). A full demographic breakdown can be viewed in Table 6.2.

Table 6.2 Demographic data of LEAD Consensus participants

Consensus participants demographics	N = 37
Gender, n	Female = 14 Male = 23
Nationality, n	Australian = 4; American = 4; Austrian = 1; Afghan =1; Belgian =2; Brazilian =1; British = 4; Canadian =1; Colombian = 1; German = 1; Guatemalan =1; Indian = 2; Irish = 1; Italian =1; Jordanian =1; Kenyan =1; Malaysia = 1; Namibian =1; Papua New Guinean =1; South African = 2; Swiss = 1; Swedish = 3; Tanzanian =1; Thai =1
Profession, n	Bioengineer = 1; Endocrinologist = 1; End-user = 4; Engineer = 1; Occupational Therapist = 1; Physiatrist = 1; Physical Rehabilitation Medicine = 3; Physiotherapist = 6; Prosthetist/Orthotist = 15; Public Health = 4; Researcher = 1 (Some with multiple professions)

Meetings

Four 2.5-hour meetings were held over four weeks with the following overall agenda:

- Objectives
- Target Population
- Inclusion and exclusion criteria
- Maximum time burden
- Recruitment
- Data Collection Methods
- Data Dictionary
- Demographic data items
- Amputation and surgical data items
- Rehabilitation data items
- Outcome data items
- Provider's data items
- Other information/confounders
- Funding
- Stakeholders
- Registry Design and Access
- Multinational Registry
- Data Access
- Individual Identifiers
- Review of data items

In each meeting, the results of the survey were presented as well as any comments received. Conversations were facilitated and a revoting of meeting participants was held as needed.

6.2.2 Results of the Consensus Process

Objectives

Consensus has been reached that the ***main*** objectives considered most relevant at this time are:

1. Measuring and Improving the Quality of Care
    a) To evaluate the quality of clinical care in all settings
    b) To examine questions of quality with a focus on equity, accessibility, and safety
2. Determining clinical effectiveness and informing[1] cost effectiveness
    a) Determining the effectiveness of different types of prosthetic treatments with an examination of outcomes
    b) To examine the effectiveness of differing service delivery models
    c) To research user involvement and its effect on outcomes
3. Describing the natural history of lower limb absence
    a) Establish the etiological factors (causes) of amputation
    b) To determine the unmet need for lower limb prosthetic rehabilitation

[1] Cost effectiveness is best determined through use of an economic evaluation that includes modelling and is generally based upon a meta-analysis of multiple effectiveness trials or a single effectiveness trial. The CDS will be relevant for any future work to determine effectiveness and as such will contribute to and inform future efforts to determine cost effectiveness.

Target Population

Consensus has been reached that the target population should be people with lower limb amputation and lower limb absence (LLA) of all levels.

Inclusion/Exclusion Criteria

Consensus has been reached that a CDS for people with LLA should include all people with LLA. Specific categories considered and included in this definition were:

- People that have not had successful prosthetic fitting
- People under sixteen years old
- People with congenital limb difference and absence
- People with partial foot amputation/absence

Time

Consensus was reached that the maximum time available for activities related to data collection for each episode of care should be 8 to 12 minutes.

Recruitment

Consensus has not been reached on the preferred recruitment method with recruitment at the facility level being preferred and clinician level being also somewhat favoured.

Registry as Voluntary or Compulsory design

Consensus has been reached that a registry should adopt a voluntary design with the option for patients/users to opt in or out.

Data Collection

Consensus has been reached that the following recommendations should be adopted around data collection:

- Clinical data should be collected by the treating clinician
- Clinical data should not be collected by reviewing of case files by someone other than the clinician
- Patient Reported Outcome Measures (PROMs) should be collected independently of the treating clinician
- PROMs should ideally be collected independently of the treating facility
- PROMs should be collected where possible by tablet or computer where people enter their data directly, in the waiting area or another part of the facility
- Patients/Users should have the option of a paper-based questionnaire to provide data
- Patients/Users should not be phoned to have their PROMs collected by a third party not associated with their care

Demographics

Consensus has been reached the following demographic data should be included in a CDS for people with LLA:

- Sex/Gender
- Country
- Education Level

- Living Situation/Home circumstance
- Morbidities/conditions (from ICD)
- Health behaviours (alcohol, tobacco use, psychical activity, diet)
- Functional status (ability to perform activities of daily living)
- Employment (industry, job category)
- Social supports
- Economic Status/income
- Understanding of medical conditions and the risks and benefits of interventions
- Social environment (e.g. community services)

Demographic data that did not reach the level of consensus includes

- Partial zip code or geographic region

Demographic Data that reached consensus to exclude includes

- Veteran Status
- Recruitment method
- Social History
- Marital Status
- Work History
- Health Literacy
- Family History
- Enrolment in clinical trials

Surgical/Amputation/Limb Absence

Consensus has been reached the following surgical/amputation/limb absence data should be included in a CDS for people with LLA.

- Level of lower limb amputation/absence
- Side of lower limb amputation/absence (laterality)
- Level of contralateral amputation/absence
- Date of amputation surgery
- Underlying diagnosis and cause of amputation
- Surgical complications

Surgical/Amputation/Limb Absence data that did not reach the level of consensus includes

- Hospital

- Residual limb measurements
- Surgical technique
- Type of procedure
- Date of revision surgery
- Inpatient death and its cause

Surgical/Amputation/Limb Absence data that reached consensus to exclude includes

- Parenteral nutrition
- Antibiotics
- Length of Inpatient admission
- Post-operative dressing
- Re-amputation within same admission

Rehabilitation

Consensus has been reached the following rehabilitation data should be included in a CDS for people with LLA:

- Time of prosthetic fitting after amputation
- Type of prosthesis (first, new socket, new prosthesis)
- Type of socket
- Prosthetic Foot
- Prosthetic Knee
- Residual limb problems
- Date of fitting
- Physiotherapy treatments
- Gait training received
- Occupational therapy treatments
- Walking aids and other mobility devices
- Social work and Psychology support
- Peer support

Rehabilitation data that did not reach the level of consensus includes:

- Physical Rehabilitation Medicine
- Compression therapy used
- Load ability contralateral leg in the first prosthesis

Rehabilitation intervention data that reached consensus to exclude:

- Date of start of compression therapy

Outcomes

Consensus has been reached that the following outcomes data should be included in a CDS for people with LLA:

- Use of walking aids/wheelchair
- Participation
- Living situation
- Employment status
- Social functioning
- Independent don/doff (yes/no)
- Residual limb pain
- Phantom Limb pain
- Perception of general situation (very good/good/average/bad/very bad)
- Mortality
- Incidence of falls in last 3 months
- Skin Problems
- Prosthetic Use Score (Amount of normal prosthetic wear per week)
- Prosthesis comfort score (Users/ Patients rate the comfort of their socket on a 0 – 10 scale)

Outcomes data elements that did not reach the level of consensus include:

- K Level
- User's confidence and awareness about the assistive device

Outcomes data elements that reached consensus to exclude:

- Date of training commencement with first prosthesis
- Number of times per week leaving the home
- Means of transportation
- Residence same as baseline (yes/no)
- Hospitalizations in last 3 months
- Skin breakdown in last 3 months
- Drive to be active
- Economic status

Providers

Consensus has been reached that the following data from providers about their practice, should be included in a CDS for people with LLA:

- Geographical coverage
- Access barriers
- Quality improvement programs
- Case management
- Use of Information technology
- Name, nature and dates of formal quality improvement process that are being undertaken

Data elements about service providers and their practice that reached consensus to exclude:

- Disease management
- Compliance programs
- Breakdown of ISPO Education Category of staff at facility
- Number of years the facility has been providing lower limb prosthetic services
- Years of experience of the staff
- Quality of patient care
- Availability / Specific expertise of the different team professions
- Routine building of Lower Limb prosthesis (> 1 per month)
- Provider-client relationship

Other information / Confounding Factors

Consensus has been reached that the following data / confounding factors should be included in a CDS for people with LLA:

- Ability to walk prior to amputation
- Morbidities/ conditions

Data elements / confounding factors that reached consensus to exclude:

- Smoking habits and history
- Onset/duration
- Severity
- Treatment history
- Medications
- Adherence to treatment

- Diagnostic tests and results
- Procedures and outcomes
- Emergency room visits
- Long-term care or stays in skilled nursing facilities

Funding

Consensus has been reached that the perceived likelihood of receiving funding from Governmental departments is most likely, while receiving funding from Patient Organisations is unlikely.

Other potential sources of funding for a registry for people with lower limb absence that did not reach consensus:

- Manufacturers of prosthetic components
- Insurers
- Professional societies/organisations
- Private foundations
- Public foundations
- Advocacy groups
- Personal Funds

Stakeholders

Consensus has been reached that the following potential stakeholders should be engaged in the development and maintenance of a registry for people with LLA:

- Patients/service users
- Clinicians
- Providers
- Researchers
- Governmental departments
- Professional Societies

Other potential stakeholders that did not reach the level of consensus include:

- Insurers
- Manufacturers
- Training schools in P&O
- Grass roots NGOs

Registry Design

Consensus was reached that a favourable design registry that stores de-identified data centrally which is freely accessible to local health services in a re-identified format.

Data Access

Consensus was not reached on how access to a registry should be managed but the most favoured options were that only aggrege data should be provided after application though a relevant ethics committee and after agreement from the registry committee and that data access should be controlled through a process requiring application to the registry with an analysis plan.

It was well supported that users themselves should have access to meaningful data for their own information purposes (to monitor personal change, to compare outcomes to others in similar circumstances etc).

Multinational Registry

There was unanimous consensus that an international registry or platform that allowed for compliance with local data laws and was available in the local country or territory would be a good solution for improved data collection.

Individual Identifiers

Consensus was reached that identifiers should include

- Date of Birth
- Sex
- Unique identifier stored within the client file

The most preferred option was a Unique Identifier attributed at enrolment.

Other potential identifiers that did not reach the level of consensus include:

- Name
- National Identification number
- Contact information
- Another individual who can be reached for follow up
- Preferred language
- Unique identifier to which only the patient/user has access.

6.3 Discussion of themes in consensus process

Quotes from the consensus process participants are included in *blue italics*.

6.3.1 Purposes and Objectives

The possible purposes of a registry were presented to ensure participants considered the relationship between purpose and specific data items. Keeping this question open for consideration throughout the consensus process allowed participants to reflect on the main data issues they felt should be addressed in their work contexts. Whilst a risk in trying to achieve multiple purposes within one registry is that it fails to address any of them adequately, or that too many data items are added leading to a decrease in participation, an understanding of what are considered the main purposes of the registry was seen as important before consensus participants considered data elements. In addition, documentation of the main themes felt to be relevant for the registry was also considered important for subsequent progression of widespread routine data collection.

Whilst quality, effectiveness and natural history were broadly identified as the main purposes, there was diversity in what the focus should be within these categories of the database. Additionally, across these domains, information needs were identified as being potentially diverse with users, clinicians, managers, researchers and policy makers all requiring information relevant to them.

Quality can be improved in a variety of ways through a registry. At the individual level, a clinician's quality of outcomes can be measured and monitored. At the centre level quality improvement measures such as training (continuing professional development or more formal upgrading of qualifications), additional services added or a change in technology can be reviewed to ascertain if an expected change in outcomes has occurred. At the macro level, comparisons can be made between regions or countries and benchmarks established for the quality of care. There was wide support from participants that quality should be a main focus of any registry effort for people with LLA.

A broad interpretation of quality was discussed including the domains of equity, accessibility, and safety. It was noted during the consensus process that the huge unmet need for lower limb prostheses in the world is an imperative to address when considering equity.

Unmet need can be addressed directly and indirectly through the efforts of a registry. Adequate mapping of services within a country or region can suggest the presence of unmet need through highlighting the absence of data from a region or subgroup of the community. Alternatively, a

registry could directly collect information from people experiencing unmet need and seek to provide them with information on available treatment options. Accessibility, or lack thereof, can be highlighted in a similar way using registry data to establish unmet need, through highlighting underrepresented groups within the population and comparing their service utilisation. *"Equity and accessibility, what is it that prevents people from accessing services?"*

Effectiveness and cost effectiveness were identified as important policy areas to ensure real world data could guide investments. Effectiveness is related to quality in that the effect of the intervention is being measured. There was broad agreement that the field of lower limb prosthetics, which often sees trials published with small sample sizes, would benefit from larger scale studies of people in real world situations to establish if a meaningful change in outcome occurs. One participant noted that a focus then could be *"new and emerging technologies and how they offer improved function to LMICs"* with further comment that in resource scarce environments it is critical to ensure that funds are spent where they have the most impact. The use of categories for the possible accommodation of new and emerging technologies and the appraisal of their effectiveness was noted. *"New and emerging technology would ideally be able to feed into the CDS* (LEAD)*, built into the way the questions are proposed."*

As was noted in the results, cost effectiveness is best determined through use of an economic evaluation that includes modelling and is generally based upon a meta-analysis of multiple effectiveness trials, or a single effectiveness trial. The LEAD would be relevant for any future work to determine effectiveness, and as such will contribute to and inform future efforts to determine cost effectiveness. The cumbersome collection of cost data within a registry is not considered practical, instead average or modelled costs serve as a more realistic and flexible way to establish this data.

6.3.2 Target Population and Inclusion / exclusion criteria

Participants sought to maintain a broad focus for any registry by including all levels of LLA and people with and without a prosthesis. Significant thought needs to be put into this question including the practical considerations and divergent needs and purposes of information for a variety of subgroups within this broadly defined scope.

Those with digital amputations and other more minor partial foot amputations alone are unlikely to seek or obtain rehabilitation services in a dedicated prosthetics facility in many parts of the world, making accurate collection of this information challenging at the centre level. Likewise, those without a prosthesis are less likely to attend rehabilitation services. Collection of information in

hospitals is a challenge even in high income countries or military populations where dedicated health information management specialists can be found. In LMICs this barrier is likely to be significant where even obtaining a referral for rehabilitation can be a challenging.

6.3.3 Time

The overall time suggested at 8-12 minutes per patient/service user is very short, especially when viewed as being available per entire episode of care rather than for a single appointment. *"Minutes per service user. I.e. not every time they come for an appointment but per episode of care"*

This small amount of time that participants suggest reflects the challenges of busy caseloads and over-stretched health services in many parts of the world. Further comments highlighted the lack of time in clinical practice to complete PROMs and surveys. Likewise, a lack of support staff and reluctance on the part of many staff members to complete surveys was highlighted, unless they are part of the key performance indicators or produce valuable and relevant information. Balancing time spent inputting data, and the benefit and utility of information for those giving this time, was picked up on by one participant *"I think it is important to give back to clinicians and facilities who participate. Perhaps sending reports back that won't identify specific participants."*

With relation to the collection of PROMs, there was support for distancing this from the treating clinician as much as possible to improve accuracy of results. *"Clinicians should not be performing PROMs ideally – there is a knowledge asymmetry and power relationship, instead public health research agencies should do this."*

If not directly shared with the treating clinician however, information and data collected through PROMs may be counterproductive as it would result in the information not being able to be used to improve the quality of treatments.

6.3.4 Recruitment

No consensus was reached on the most favoured level at which to recruit participants, but it should be noted that most existing recruitment efforts have been focused at facility or health service levels rather than clinicians or through users directly. It was noted that while facilities might be recruited to participate in data collection, it would still be the clinician who enters and collects data from the patient directly, *'recruitment to participate in the registry, not who is going to enter the data. The data collection could be done at the clinician level, but we can recruit the facility and the clinicians will participate'*. There was broad agreement with the sentiment that involvement, support and

leadership from management was imperative, *"It's important that you have the director of each facility to join that registry should be used. If you don't have the 'bosses', it will not work"*.

Active involvement of the patient was also stressed as an important element, with another layer of recruitment being highlighted as necessary, that being recruitment of the user by the treating clinician and the importance of data checking to ensure errors are not recorded within the registry. *"The accuracy of the data or the completeness of the data is best checked by the user or the patient, so somehow I hear that there should be an ability for at least the patient to check the data or at least find a mechanism to correct or supplement any missing data."*

6.3.5 Participation as voluntary or compulsory

Voluntary participation (with the option to opt out) was favoured by the consensus participants due to the principle of people having control over their own data and exercising autonomy. *"I think that if people are being asked to provide their own information through anything that's self-report, you can't make it compulsory."* Further, the regulatory rules such as the General Data Protection Regulation (GDPR) in the European Union, and their equivalents in other jurisdictions, were highlighted as a constraint *"laws do not allow compulsory in many counties I suggest"* and *"With GDPR issues - it depends on the data we collect. in the current scenario - it has to be voluntary"* and "*Currently in* (Central American Country), *as a law, Informed consent is always required."* and "(Asian country) *does not allow collection of any patient data without ethical approval by the institution. Even if it is de identified. We have very strict PDPA* (Personal Data Protection Act) *rules."*

Voluntary participation of service users raises issues of data completeness or producing an incomplete dataset. *"If we have (a) voluntary contribution it will always only include a proportion of the population and therefore it will make it very difficult to make meaningful conclusions."* Some solutions to an incomplete dataset were offered including *"Can give incentive for user opt in , clinics must collect as matter of policy"* Another model suggested was one where management encourages or enforces complete participation by all staff who then check with users for consent for their data to be included *"I think it should be voluntary for the user, but compulsory for the clinician or the one collecting the data, if the user wants to opt in."*

The strength of a dataset that includes every service user was noted. "The powerful databases are the ones which are compulsory." Reports from such registries are generated in aggregated form, meaning no individual data is made public. Reporting limitations further ensure privacy where deidentified data is stored, with limits on reporting on data items where small numbers of

participants are in any one category. *"We have some minority groups which have data collected but we were not allowed to publish on those minority groups because potentially those minority groups could be identified"*

Consensus participants noted that data is collected in routine practice within the terms of a user accepting to receive services, and that this data is regularly pooled and analysed within the jurisdiction, meaning data collection is compulsory when receiving many health services.

6.3.6 Data Collection

Support was given to the treating clinician being responsible for collection of clinical information and implementation of PerfOMs as the most logical and proximal person to enter such clinical data.

On the other hand, support was also given to the idea that PROMs data should be collected independent of the treating clinician in order to reduce bias *"all of the public sector facilities score 80% and above in terms of patient satisfaction and patient feedback and so, from that experience, over the last 10 years, I can tell you that when patients have to report their own subjective experiences, if it is by the service provider itself, whether it's the clinician or a representative of the facility or even anonymised in the facility, it is definitely not accurate data that we will get"*.

A counterpoint to this was that PROM data has great clinical utility and would be valuable for clinician and patient to discuss. *"I believe the PROMs should be collected by the clinician. The PROMs should be incorporated into the care of the patient. Data from the outcomes can help with clinical planning".* A distinction was drawn between patient reported experience measures (PREMs) and PROMs "PREM and PROM are different kind of information. Being satisfied with service is PREM and should of course be collected independently." Significant agreement from other participants was expressed that this distinction is important.

Given the data must be collected, the question was raised *"If the neither the facility or the treating clinician are collecting the data, then how will it be collected?"* with the possibility being *"through a neutral third party".* This suggestion was not supported as it was felt to be impractical in many LMICs.

Where possible, use of technology in data collection was supported, although with some reservation *"Specifically, independent and electronic feedback may primarily be obtained of a population that is more literate as far as responses to such questionnaires are concerned."* While it was also highlighted that in a lot of LMICs such technology might not be available, or practical, particularly for people that

are elderly *"A lot of the population we are dealing with, is 60 years and older and there is a severe limitation in data literacy and electronic literacy when it comes to answering questionnaires and that is something I think one needs to understand as a practical limitation too."*

6.3.7 Providers

Information about providers can be particularly useful when quality improvement measures are conducted at a centre level to establish which types of providers obtain the best outcomes and which improvement measures enable better overall outcomes over time.

Some participants felt this whole category should be removed *"As I see it, the idea with a quality register is to focus on the patient treatment and its outcomes. Now you are into another area…and that is to quality assure a clinic. This as the education centres are quality assurance by ISPO > Cat 1 etc."*

Geographical coverage of providers was felt relevant because it allowed for determining country or region coverage to improve access to services.

Access barriers relevant in the centre catchment were discussed as important to understand when viewing centre data within the context of impediments to services.

Quality improvement programs were deemed important to be able to ascertain which improvement measures had the most impact on outcomes and the timeline for outcome improvement. It was supported that the name, nature and dates that have been/are being undertaken be collected from providers. *"Essential to measure quality improvement measures and their effectiveness it was supported along with combination with Quality Improvement programs. "This item does not mention accreditation explicitly. I like that addition."*

Case management is coordinated and integrated care and its availability was acknowledged as a possible important characteristic of providers in producing outcomes and to be included as a provider detail.

Use of IT was seen as relevant due to its ability to improve efficiency of the centre. *"Use of IT is dynamic and can change as providers invest more into their practice or as practices evolve".*

6.3.8 Funding

When seeking funding for a registry, it was felt that government funding is most likely given the benefit gained by governments from data to inform policy. This was acknowledged as challenging in LMICs. *"Realities prevail in LMICs and usually Ministries of Health need to see benefit before they recognise value and support a registry."* Funding a registry through a broad set of sources was seen as preferable with the limitation of an industry led effort being noted with perceived bias. *"If you have one group funding (like the private sector) the registry might be perceived as biased."* This concern was echoed and expanded to complete reliance on any single funding source. *"We need to be absolutely careful in what type of funding we get. Whether it's going to be from NGOs or government organizations or industry. Each one them have their own specific interests and those interests should not compromise the quality of data and the type of data that we collect."* Involvement of professional societies was encouraged. *"I think that professional societies might be willing to fund if they understand that this type of registry can build an investment case for more prosthetic care availability in their areas of operation."*

6.3.9 Stakeholders

A broad set of stakeholders was encouraged to give general support to data collection and data use, including service users, clinicians, providers, researchers, government and professional societies. It was noted that there must be benefit to all stakeholders in reported information for their participation to be maintained. Service users were seen as critical to engage and representative should be involved in registry leadership to ensure data is used in way that is both acceptable and provides benefit to the people whose data is being collected. Clinicians are important stakeholders who must remain committed to data collection and should be involved in a process of refinement to make the registry easy to use and to answer questions that are of value to them. One existing registry (SPARG) highlighted the value of annual meetings for review of data to exploit key learning opportunities and share ideas for improvement based on recently collected data. Providers and their management are key stakeholders as their support is pivotal to encourage or mandate data collection within the scope of services being provided. Usefulness of the data for providers to drive quality and service improvement is critical for the success of any registry. Researchers are ideally placed to use data to enhance reporting and answer questions relevant to policy makers and other stakeholders. Their engagement to ensure methodological factors are considered and accounted for in reporting is important. Governmental departments as ultimate health policy makers are a vital stakeholder. Whether funding directly or providing an enabling environment for service provision,

government needs to be engaged to ensure the data collected is utilised in productive ways. Professional societies can provide a link between clinicians, researchers and policy makers and engage their membership to drive participation.

6.3.10 Registry Design

Consensus was reached that the most favourable design is a registry that stores de-identified data centrally, and allows the data to be freely accessible to local health services in a re-identified format.

Discussion about design focused on the need to comply with local and regional data compliance laws, a topic expanded on below in the discussion on access (Section 6.3.11).

Two broad categories of design were discussed, firstly where identifiable data is owned and securely stored locally (to comply with data protection laws and regulations), but can be shared centrally in a fully de-identified manner and reported on only in aggregate form (e.g. total numbers, averages, standard deviations etc).

Secondly an open-access open-source design where data can be uploaded at a single time point and only the person providing the data (the individual with an LLA) has and records the unique identifier. All data is stored in a fully de-identified format in order to comply with data protection laws and regulations. *"If it's not open source and open access, then the future research will be jeopardised. The idea of open access and open source is basically the public domain so that future researchers can actually access to that and can research various aspects of the data".* The limitation of open-access open-source design as it was described in the consensus process is that if the individual does not record or forgets their unique identifier, longitudinal data on that individual can no longer be collected over various time points. It is possible with such a design to conceive of some basic pre-rehabilitation data being accessed from the medical file to make comparison. The inability to compare data at various time points was seen as a major limitation by many. *"Isn't the whole point of a registry to compare over time and track outcomes? It doesn't sound like a registry anymore if we take that approach"*

The importance of a registry design that allows longitudinal follow up and data collection at various time points was noted. Determining pre and post rehabilitation intervention outcomes without the impact of recall bias is important for decision makers at all levels to have confidence in the results. Further rehabilitation has long term effects not easily determined at the time of treatment and best measured over longer periods to best capture the effects of interventions. Whilst a longitudinal

design has various constraints in terms of design and the impact of data protection laws it was felt that the importance of long term follow-up was critical.

6.3.11 Data Access

The discussion concerning design and access centred around privacy and confidentiality on one hand, and the benefits and usage of data on the other. It was agreed that there must be a mechanism for access to data. *"If the data is not available then really no conclusions can be made from it, so it just sits in a data bank, which is a bit unfortunate because a lot of effort goes into the data collection"*

Consensus was not reached on how this access to a registry should be managed, but the most favoured options were that only aggrege data be provided and only after application through a relevant ethics committee, together with agreement from the registry committee. It was also agreed that data access should be controlled through a process requiring application to the registry with an analysis plan.

Routine reporting of aggregate data was suggested as a means of allowing general data for comparisons and indicating unmet need. *"I think open access is important; when we're talking about open access, we're not talking about individual datasets, we're taking about general datasets."* Aggregate data in non-sensitive fields might be reported freely such as the number of people with limb absence, the numbers of people successfully fitted with a prosthesis and the % breakdown of aetiology per country. This information would be useful for all people to know in order to plan and advocate for improved services.

It was well supported that users themselves should have access to meaningful data for their own information purposes (to monitor personal change, to compare outcomes to others in similar circumstances etc).

Within a country, data might be able to be examined by the owners in more detail to establish aetiological patterns and measure outcomes to plan investments.

6.3.12 Multinational registry

There was unanimous consensus that an international registry or platform that allowed for compliance with local data laws and was available in participants country or territory would be a good solution for improved data collection.

There is growing interest in working across countries to share deidentified and aggregated data. Data pooled across countries can lead to more generalizable evidence in terms of device performance and patient outcomes. A broader range of practice settings can capture higher variation in performance and provide more opportunity for discovering and analysing best clinical practices.

Any future international registry would have to overcome numerous challenges. The data harmonization required for pooling data must build on consensus from different stakeholders that bridges health systems, reimbursement structures, and cultures. The LEAD has accomplished this though it is yet to be field tested or piloted. Sharing of data is made more difficult by divergent national rules. Privacy laws, for example, vary by country and sometimes by region, allowing greater or lesser ease for global aggregation of de-identified patient-level data. An approach of selecting those countries with data privacy laws that will facilitative international collaboration is suggested. While many general principles are similar for participant enrolment and retention, there are different customs or regulations regarding contract language, requirements for ethics committee or other submissions, informed consent, and allowable approaches to patient retention or data deletion on request in different countries. An international registry will need to reference national and local regulations and norms to adopt a design that is widely compliant and acceptable.

A solution for an international registry could be to store only de-identified data in a cloud-based platform. This de-identified centrally stored data would be available in aggregate form, while local health services as data owners would maintain access their own specific data in a fully identifiable format. This would allow users/patients to be contacted, specific analysis to be performed and comparisons made using other data collected in the registry.

For any international effort there are some specific issues related to data protection and ensuring all information shared centrally is not identifiable. Most registries have reporting guidelines in place so that if only small number of people meet a specific category their results will not be reported since information related to their care may allow those viewing the data to discern an individual's protected health information. In the case of an international central repository, using deidentified data would preclude sharing of protected health information to this central location where there are small numbers of people in one jurisdiction with specific circumstances. Exact timings in terms of month and year can also be replaced with intervals between events to prevent any potential identification.

6.3.13 Individual Identifiers

Consensus was reached that identifiers should include a unique identifier stored within the client file.

The most preferred option was a unique identifier attributed at enrolment with sex and date of birth also included to provided valuable disaggregation of data.

6.3.14 General comments about design

Organisation and formatting of the data input platform or paper-based form was discussed with some useful specific hierarchical suggestions made to avoid repetitive questioning. For example, some early questions might avoid input of certain irrelevant subsequent fields. *"If we have the question about prosthetic use early on in the registry, then those wearing a prosthesis get one set of questions and those not wearing a prosthesis get another set of questions"*

6.4 Discussion on Specific Data Items

6.4.1 Demographic Data

Demographic data is primarily to establish different groups within a population to establish if their services or outcomes vary significantly. It was highlighted that such data is hence important for equity. *"If we want to look at equity, some of these items I think we need to include that we might otherwise think, well why should we include it. Things like education level – yes actually because education is very highly correlated with income, and no one answers income questions, but they will respond to education level questions. We should be mindful if one of the purposes of this registry is to look at equity, then we shouldn't discard some of the demographic questions just because we think they're a little less important."*

Focus was given to ensuring that questions were ones that participants were prepared to answer, and that clinicians and facilities were happy to ask.

Sex

Sex was considered more appropriate than gender due to the physiological differences between people born as a man or a woman.

Country

Country was considered uncontroversial and was adapted to include administrative jurisdictions where relevant.

Education level obtained

Education level was seen as a useful proxy for other demographic information such as income or employment which is a social determinate of health, that might be difficult to collect because of reticence to give sensitive information or through a lack of specificity of other proxies such as zip code. *"We know education is actually very highly correlated with health literacy. So, you get a big bang for your buck if you collect education…so then you can skip those two items because you never get people reporting their income, and then you don't have to ask health literacy."*

There was some discussion of education being a category that would be skewed against older people *"I think the education level is difficult. We are dealing with patients who are 70, 80 years old. At that time there was no PhDs and there was no Bachelor and Master gradings and different types of gradings, so we run into a problem there as to how to actually write that. I wouldn't ask it."* While others supported it as a simple and useful data point *"Education can be in a very simple way, three levels and no more, then it is not a problem"*

Living Situations/Home Circumstance

Living situations/home circumstances was considered important for understanding the circumstances of a person with LLA and an outcome in itself (e.g. increasing levels of care needed). Problems with establishing categories of housing that would be universally applicable were found when defining this. Care needed and care availability was ultimately included as part of social and community support.

Morbidities/Conditions (from ICD)

The health status of people with LLA was included and noted that this is required at multiple time points to understand the health progression of people with LLA. An open approach to allowing all ICD codes allows regionally relevant disease to be captured.

Health Behaviours (alcohol, tobacco use, physical activity, diet)

Such risk factors for health were considered important but the appropriateness of asking questions about alcohol led to its final exclusion. Complexities defining physical activity led to this being asked as a simple activity level question. Complexities in regard to diet were noted, as both over nourishment and under nourishment are problematic. Further challenges with asking diet related questions led to this being reduced to a question of Body Mass Index (BMI).

Functional Status (ability to perform activities of daily living)

Functional status was considered important to gather information about peoples' function. The COMPASS includes domains of activity restriction and this is further included as participation and activity level.

Employment (industry, job category)

Employment was noted as a sensitive topic as it may be considered to be obtaining information on ability to pay for services this was reformed into an ability to work question.

Social Supports available

Social supports available was felt to be important in order to establish the care needs for people with LLA and to establish the possible beneficial nature of social support on outcomes. Ultimately combined with community services.

Economic Status/Income

While income was supported as an important variable, there was broad acknowledgement that it would be sensitive and perhaps be in and of itself a reason for people to opt out of participation. *"I just know that If we put an income question, we'll have a lot of missing data. And so, I'm really not a fan of income questions just because if it's got a lot of missing data its sort of useless…I don't know why we would ask it…it is an off-putting question."* and because it might be poorly defined in places where participation in the informal economy is common. "Income is also difficult for some people to answer in general"

A compelling question was raised by a user about the purpose of Economic Status / Income as a data element: *"Why do you want data on anything related to my income or social way of living? Because you don't want to sell me a prosthesis, right? You want to take data for your keeping this data, right?*

Why do you require my income? …Why do you think that all this data are important as of persons processing data later and studying it?" The purpose of this question was clarified by the participants as a social determinate of health to gain understanding of how people are affected by their LLA. *"Because we want to (examine) the health disparities and look at how that's affecting health outcomes, so some of this information allows us to get some of the social determinants of health, It's not meant to target a particular individual but to get more an idea of, are there some health disparities that are affecting outcomes."*

Economic status/income was ultimately deemed too sensitive to ask directly and reformed into an ability to work question to capture labour market participation possibilities as an outcome.

Understanding of medical conditions and the risks and benefits of interventions

An end user advocated for inclusion of understanding of medical conditions and the risks and benefits of interventions as it seeks to measure how involved the user is in their own treatment, and to what degree they felt like an expert on their own health and rehabilitation outcomes. While supported in principle, the difficulty with defining this element has resulted in final exclusion.

Social environment (e.g. community services)

Social environment was combined with 'social supports.'

6.4.2 Surgical/Amputation/ Limb Absence

Proper definition of the various possible levels of amputation or limb absence, as well as timing and progression of subsequent revisions and new amputations on the contralateral limb drew significant support. Relating to the timing of revisions, there was comment that *"If it's the same admission then probably not important to capture, however if the revision is done at another admission, then it should be captured."* There was further comment about revision being important because they drive health care utilisation and consumption of scarce resources and timely rehabilitation might minimise these costs *"This is something that's important because anytime you revise, you're looking at utilisation of services. So, whether it's right away or down the road, it's going to potentially drive prosthetic need. Because if you have a revision, that's another prosthetic need."*

Concern about too much detail on surgical interventions reflected an acknowledgement that in many parts of the world, rehabilitation is not well integrated into wider health care service, and

information gathered directly from users might be inaccurate or lead to many missing data fields. *"I think that adding information about the type of amputation procedure or surgical technique might add a time burden to the registry. The person with amputation might not know this information and the person collecting the data might need to perform further research (contacting the surgical team) to know this. I'm not sure that this is critical to know."*

There was some discussion about omitting information on surgical revision and focusing the registry on the current presentation of the user *"I actually disagree – why should I care whether it's a revision amputation or it's an initial amputation? I care about the amputation surgery and I report it, then I'll deal with it as if it's a new one and I fit the prosthesis accordingly."* This approach would reduce the effort needed to track the course of the persons surgical history, but it was acknowledged as having a negative impact on the information within a registry. Simplification of registry design was highlighted as a reason to perhaps omit revision date.

"If we really look at the core dataset, then I think it is important to keep to essentials. I think there are research questions and those have been pointed out, specifically if one wants to go longitudinal where this aspect becomes very important, so I think that would then be a data point that would be collected in the additional data related to the specific research question of this specific registry and not necessarily be in the core dataset."

The potential importance of the history of previous amputations were also highlighted *"From the point of view of the actual treatment that you would give to someone with an amputation, or a second amputation, you would treat them, but from the point of view what they've been through, from the point of view of the emotional support they may or may not need, from the point of view of the lengths of disability and comorbidities they may have acquired there could be significant difference."* With some speakers even advocating for a longer medical history to be incorporated that included limb preservation *"The important thing to note…if we had patients who were losing partial feet and it almost always results in a less functional higher amputation much later, if that's of concern to you or to us, like if we're looking at limb preservation versus early amputation at a functional level, I think it's important to know."* In many LMICs and wherever there is a lack of integration of an electronic medical record (EMR) such data elements are likely have significant amounts of missing data.

There was broad agreement that each subsequent amputation was important to note and that if each surgical episode was recorded sequentially, with accurate information on side and level of amputation, this would be mostly realistic and allow for relevant information about amputation

history to be contained within the registry. *"If you think of every revision as an initial surgery amputation, then you lose that initial. So, I think that this is again, if this is a longitudinal database, we need to be able to annotate what was the original date which is the date of amputation that's how we would define it, and then any succeeding revision – those dates would be important because it would show what's happening over time and it may change procedure."*

It was highlighted that a registry should allow continuity of information with re-establishing identity to add to the history of previous entries. "Isn't that the point of a registry is to be able to follow people over time? "and "If we would be able to follow the patient over time? Because, what we are interested in is the final level, so we need to be able to follow the same patient over a period of time and then we get to know the final level."

6.4.3 Rehabilitation

There was a good level of agreement about rehabilitation categories, and it was suggested that many could be simplified using a drop-down menu or tick-box data collection method to make each data element easier to collect.

Presence of professionals and peer support was agreed to be collected just in terms of involvement or non-involvement *"Ok if we mention professions and members of the team then why don't we mention them all. Even if I know that there are not physicians available all over the world but it's also important to know whether they are and to assess the need, so for me it's quite crucial."*

Peer support was included with these categories of professional involvement given its relevance and potential impact in places where formal health care services are limited. *"Peer Support seems to be quite an important one in the field of amputation."*, *"WHO is making packages of rehabilitation, which include Peer Support."* and *"As we find it so important and we also may find it to be possibly not used to the extent that it may be used, my personal vote is to vote for yes to include it in order to make people aware that this is a very useful and important measure." "There is lots of good evidence about the importance of peer support in this population!"*

Interventions such as gait training or compression therapy were agreed to be added without mention of the professional that delivered them. Compression therapy for example was discussed *"I believe you can do it in a simple way. Like, did you start within 3 weeks or 6 weeks? Yes or no? It mustn't be that detailed"*

The diversity of practice worldwide, in terms of different rehabilitation categories, was noted with comment that the definitions would need to be broad to allow for data capture. *"We find it quite difficult to collect, and that's just looking at Scotland and because there's certainly quite a lot of different modalities used. When our colleagues down (south) collected it, they found it quite difficult as well because they used other things as well. So, if we've got variety in the UK, I can only imagine there must be vastly different variety."*

The diversity of experiences of users through their rehabilitation pathway (or lack thereof) worldwide was also noted "The vast majority of the world's amputees do not have a prosthesis and many have gone years without any sort of prosthetic care. If we are collecting this data about compression therapy in LMICs I'm not sure how it will help the bigger picture."

6.4.4 Outcomes

Use of walking aids/wheelchair

Use of walking aids or a wheelchair was agreed to be added as a drop-down or tick box, and the successful fitting of a prosthesis would be possible to add here also.

Participation in Community

Participation in the community was considered critical, and the most important outcome by users, participation was viewed as the ultimate goal of rehabilitation and therefore important to include. *"Participation for a user is very, very important".* However, challenges were noted in defining participation adequately *"I like the construct of participation, and I think it's really important, but I do worry how you measure it in a very timely and tight way".*

This data element was included as a question for individuals to rate their participation in their community relative to their own expectations and to their peers, thereby avoiding lengthy lists of activities or lengthy outcome measures.

Living Situation

Living situations was a duplicate item from demographics, this item was included as social and community supports.

Employment Status

Employment status was a duplicate item from demographics, this item was included as ability to work.

Social Functioning

Social functioning was combined with participation in community.

Independent Donn and Doff (yes/no)

Ability to independently donn and doff a prosthesis was considered important by consensus participants but ability to donn is an element of the PEQ Utility in the COMPASS so excluded.

Residual Limb pain and phantom limb pain

Residual limb pain and phantom limb pain was considered important by consensus participants but was excluded as it was already considered to be adequality covered by COMPASS outcome measures, PEQ Residual Limb Health and TAPES-R.

Perception of general situation (very good/good/average/bad/very bad)

Perception of general situation was voted for inclusion, the construct that is covered by perception was seen to be explored by the TAPES-R in the COMPASS so was not ultimately included in the LEAD.

Mortality

While Mortality was viewed as important, it was noted that this could be difficult to collect in many parts of the world. Given a lot of missing data and a lack of cause of mortality data it was seen as not relevant for a CDS.

Incidence of falls in last 3 months

Incidence of falls in last 3 months was viewed as an important outcome as services to prevent falls are often seen as cost effective.

Skin Breakdown / skin problems

Skin problems were highlighted as an important negative outcome and was therefore included as a data element. It was noted that it might be hard to detail and quantify skin problems, but a solution was offered suggesting to include skin problems that result in impaired prosthesis (or wheelchair)

use. *"Why not put these two together; skin breakdown and skin problems, and relate them to if the amount of problem makes you not being able to use your prosthesis?"* Hence it was agreed that *"Skin problems which affect prosthetic use or* potential *prosthetic use: that includes breakdown, it even includes red spots, which means you have to come off your limb for a day or two…and then we could probably leave out the skin breakdowns all together."*

The timing of the skin problems was noted as needing careful definition and the place of the skin problems. *"I personally feel skin breakdown, without 'in the last 3 months'; maybe six months can be taken into consideration. And skin problems according to me can be skipped and also skin breakdown should be for the residual limb only, because we have multiple skin problems and different issues."*

Ultimately this data element was deemed to be adequately covered by both the TAPES-R and PEQ Residual Limb Health in the COMPASS so was excluded.

Prosthesis Use Score

Seen as a better definition of use than days per week or hours per day this score allows more definitive understanding of usage than other measures.

Prosthesis comfort score/Socket comfort score

While reaching a consensus to include this simple one item outcome measure there was concern about its use. *"I'm doing quite a bit of research in this area at the moment, and I just find that the comfort score isn't a useful measure in any situation anyway, because it doesn't tell us anything about the patient's comfort. Comfort can be many things, it can be psycho-social, it can be functional, it can be related to well-being, it can be related to aesthetics…I feel that…it shouldn't actually be included here in this core dataset….it isn't perhaps the most detailed or useful outcome measure."*

Given concerns about the psychometrics raised in the outcome measures consensus process this was excluded as the construct is also covered in the TAPES-R and PEQ-Utility in the COMPASS.

Other information / Confounders

Ability to walk prior to amputation

Seen as important to better understand outcomes, noting that it must be carefully defined in terms of timeframe.

Morbidities and conditions

A duplicate item from demographics, this item was included.

6.5 LEAD Data Dictionary

The data dictionary was developed by a small group of consensus participants to ensure that all data elements and their response options were adequality defined. Several issues were considered when defining data elements such as time burden, simplicity, the availability of data, data usage, acceptability of data items to participants (people with LLA and clinicians), the ability to categorise response options in a way that is universally relevant or understood and the possibility of missing data which generally limits usage.

Time burden was considered when defining the data dictionary with short direct questions favoured where possible and an overall minimum of questions being asked. Simplicity of wording was favoured in both questions and answers possibilities and questions that could be easily recalled without needing to check the client file or other resources. Data availability was considered in terms of the likely setting being within rehabilitation services, with questions involving specifics of surgery being modified to avoid missing data. Data usage was considered to ensure each data element was valuable and could contribute to analysis, with the avoidance where possible of redundancy. Acceptability was considered broadly for a possible global population of people with LLA with the avoidance of some questions that might be sensitive even in some settings or some sub populations. An inability to categorise response options in a way that was universally relevant led to some data items being modified to allow the LEAD to be appropriate for use in a broad variety of contexts. Missing data due to items being hard to recall or establish was considered with data items being preferred that could be entered with simplicity by clinician or service user.

The categories of data elements considered within the consensus process of demographics, surgical/limb absence, rehabilitation, outcomes, confounders and unique identifiers were reformulated in terms of timing of collection. The new timing categories determined were at:

- Enrolment
- Amputation/Limb Absence
- Episode of rehabilitation
- Pre and post rehabilitation

The constructs deemed important in the consensus process have been included in the LEAD and data dictionary with few exceptions.

6.5.1 LEAD Data Dictionary

The following data dictionary lists the name of the data element, the question-and-answer possibilities and definitions of the variables where it was felt it was needed for clarity. In the future a more detailed list of definitions might be required including pictorial references to assist service users and clinicians to accurately identify correct response options. To overcome a primacy effect where response options at the top of a drop-down list get selected more often a yes/no format is encouraged, even if selection of one variable only is possible. Whilst this may lead to an increased time burden, possible digital solutions could be explored to minimise the burden and will be discussed in Section 7.1.13. Yes/No options are also more favoured for paper-based data collection.

Enrolment (Collected at initial enrolment within the registry)

Name of Data Element	Questions and Answers	Definitions of Variables
A) Date of Birth	Month and year of birth	
B) Sex	Sex: • Male • Female • Neither	Sex assigned at birth • Male • Female • Neither male or female
C) Country/jurisdiction	Country/Jurisdiction	Countries or recognised administrative jurisdictions
D) Unique Identifier stored within the client file	Each new person enrolled within the registry should be attributed a unique identifier; either by the facility/clinic who can hold a “bank” of these or automatically via an online portal.	Alphanumeric

Amputation/Limb Absence (Collected at enrolment or after a change of level of limb absence)

Name of Data Element	Questions and Answers	Definitions of Variables
E) Cause of amputation/limb absence	Primary cause of amputation • Trauma • Diabetes • Vascular • Cancer • Infection • Congenital • Other, please specify:	Primary cause is the main predisposing reason for this amputation or limb absence • Trauma includes acts of violence, accidents, injury etc. • Diabetes - micro-vascular • Vascular - macro-vascular • Cancer – Amputation to remove cancerous growth • Infection- local or systemic infection • Congenital difference- limb absence or difference from birth • Other- any other primary causes of amputation.
F) Amputation/limb absence details	Date of this amputation/limb absence Amputation Side: left/right Type of amputation 1. *Major amputation (changing level)* 2. *Bone revision at same level* 3. *Soft tissue revision at same level* Level of Amputation 1. *Toe amputation* 2. *Metatarsal Phalangeal disarticulation* 3. *Ray resection (meta-tarsal and toe)* 4. *Trans metatarsal amputation* 5. *Tarsal Metatarsal amputation (Lisfranc)* 6. *Trans tarsal amputation (Chopart)* 7. *Ankle disarticulation (Symes)* 8. *Transtibial (BK)* 9. *Knee disarticulation*	Date of limb absence: If congenital use DOB 1. Major amputation changes the bony level through or above a joint. 2. Bone revision does not progress to a higher named level through or above a more proximal joint and includes removal of bone spurs and changing of the contour of bone etc. 3. Soft tissue revision does not affect the bony anatomy but removes redundant or diseased soft tissue only. 6. Trans tarsal amputations include Chopart, Pigoroff and Boyd variations.

	10. Transfemoral (AK) *11. Hip Disarticulation* *12. Hemipelvectomy* *13. Hemicorporectomy*	
G) Ability to walk prior to amputation	Were you able to walk in the 3 months prior to your amputation?	Walking defined as moving with or without walking aids while bearing weight on lower limbs

Episode of rehabilitation care (collected once per episode of rehabilitation care)

Name of Data Element	Questions and Answers	Definition of variables
H) Professions involved in providing services (Rehabilitation)	Please answer yes or no for rehabilitation treatments received in this episode of care: 1. Prosthetist 2. Physiotherapist 3. Occupational Therapist 4. Social Worker 5. Psychologist 6. Rehabilitation Medicine Doctor 7. Peer support 8. Community Based Rehabilitation 9. Community health worker 10. Nurse	1 to 6. At least one formal appointment with a professional in which services or treatments were received, lasting at least 30mins 7. Peer support can be individual or group
I) Rehabilitation services	Please answer yes or no to the following rehabilitation interventions received in this episode of care: 1. Compression Therapy	1. Compression therapy

	Dropdown Yes/No: • shrinker • bandage • premade silicone/elastomer liner inflatable compression • Removable Rigid Dressing • Other Please specify 2. Gait Training 3. Adaptive Sport training	• Shrinker is an elasticated premade compression sock. • Bandage, roll on bandage applied with the purpose of compression. • Elastomer/Silicone liner (used as compression without prosthesis) • Inflatable compression -externally applied with or without standing device • Removable Rigid Dressing, custom made rigid dressing for compression and protection. 2. Gait training is occupational non-specific and includes at least 30min of instruction about walking with or without the use of a prosthesis in a safe and effective manner. 3. Adaptive sport training involves sport training or use of sport within rehabilitation
J) Prosthetic intervention	Type of Prosthetic Interventions 1. Prosthesis 2. Socket Replacement 3. Repair/Adjustments	1. Prosthesis: entire prosthesis with all componentry 2. Socket Replacement only 3. Repair/Adjustments of prosthesis including replacement of component parts (other than socket), servicing or changes to the fit or alignment of the prosthesis.
K) Type of socket	Date of delivery: Side: Left/Right Socket type/design Partial foot • Within shoe design	Date of final delivery of the prosthesis for use outside of supervised rehabilitation. Side: In case of Hemicorporectomy leave blank.

	• Ankle immobilised • Ankle immobilised and weight borne proximally Ankle disarticulation • Split socket • Window • Total Surface Bearing Trans tibial • Patella Tendon Bearing • Specific weight bearing • Total surface bearing • Hydrostatic • Thigh lacer • Osseointegration Knee Disarticulation • Window • No window Transfemoral • Quadrilateral (Ischial Support, Narrow AP) • Narrow ML • Plug Fit • Ischial containment • Sub Ischial • Osseointegration Hip Disarticulation Hemi-Pelvectomy Hemicorporectomy	Specific Weight bearing uses the basic design of the PTB socket but maintains total contact. Osseointegration is a prosthesis attached to a surgically implanted bone anchor.

	Socket type (can select multiple) • Fibreglass lamination • Carbon fibre lamination • Thermoplastic • 3D printed • Adjustable socket solution Liner type • No liner • Foam/pelite • Silicone Liner • Urethane Liner • Gel Liner Suspension • Self-suspending • Cuff/strap • Pin • Lanyard • Sleeve • Expulsion valve	An adjustable socket the user can adapt the fit themselves
L) Prosthetic Foot/ankle	Foot Type select all that apply to the foot on this prosthesis, and multiple categories in the case of crossover foot ankle combinations. • Powered foot • Microprocessor controlled foot • Hydraulic foot • Pneumatic foot • Special activity foot • Dynamic response foot • Multiaxial foot • Single axis foot	Special activity feet are designed and used not for general ambulation but only for the activity for which they are designed, such as running, surfing etc. and can include non-foot shape/type terminal device designs.

	• SACH foot • Hard rubber bare foot design	SACH (Solid Ankle Cushion Heel) Hard rubber bare foot design is for use without shoes or only with sandals/open toed shoes.
M) Prosthetic Knee	Knee type select all that apply to the knee on this prosthesis, multiple categories allowed where the knee has multiple properties on this list. • Single Axis • Multiaxial • Pneumatic • Hydraulic • Microprocessor • Externally powered knee	Multiple categories allowed
N) Prosthetic Hip	Hip type • Single axis • Multiaxial • Hydraulic • Pneumatic	
O) Tobacco use	In the past three months how often have you used tobacco-based products • Never • Once or twice • Monthly • Weekly • Daily or almost daily	Tobacco based products include cigarettes, chewing tobacco, cigars, snuff etc)
P) Body Mass Index	Height Weight	Options for metric or imperial Weight without prosthesis Height pre-amputation or calculated with anthropometric calculations if not available

Q) Morbidities/ conditions (from ICD)	ICD Codes of health conditions	ICD -International Classification of Disease
R) Maximum Education Level Obtained	What is the highest level of education that you have obtained? 1. No formal education attended 2. Early childhood education completed 3. Primary school completed 4. Lower secondary education completed 5. Secondary school completed 6. Post-secondary technical qualification 7. University degree completed 8. Postgraduate degree completed	2. Early Childhood: Education to support early development for children below age of 3 or from 3 to start of primary education 3. Primary: Programs typically designed to provide students with fundamental skills in reading, writing, and mathematics and to establish a solid foundation for learning 4. Lower Secondary: First stage of secondary education building on primary education, typically with more subject orientated curriculum 5. Secondary: Programs providing learning experiences that build on secondary education and prepare for labour market entry and/or tertiary education. The content is broader than secondary but not as complex as tertiary education. 6. Post-secondary Technical: Certificates, diplomas, and vocational trade qualifications in an institution other than a university. 7. University: Short first tertiary programs which are practically based, occupationally specific and prepare for labour market entry OR Bachelors or equivalent 8. Postgraduate: Completion of a full Master's degree or equivalent or Doctoral degree or equivalent

Collected pre and post Episode of Rehabilitation care

Name of Data Element	Questions and Answers	Definition of variables
S) Date of commencement or completion of rehabilitation care	Date of commencement or completion of rehabilitation episode of care	Commencement collected at beginning of rehabilitation episode and completion collected at the end.
T) Use of mobility devices	Please select the walking aids and other mobility devices that you have: List (Yes/No) 1. No walking aids 2. Single point stick 3. Quad base walking stick 4. Single Crutch 5. Pair of Crutches 6. Walking Frame/Walker 7. Wheeled walker 8. Manual Wheelchair 9. Powered Wheelchair or mobility scooter Follow up question for yes responses other than 1: How much do you use this walking aid in a normal day? • In a normal day I don't use it • Less than1 hour (a little) • 1-3 hours (some) • 3-6 (lots) • 6+ hours (mostly)	2. Single point stick or walking stick 3. Quad base walking stick has four points of contact with ground 4/5. Crutch, elbow or axilla (armpit) 6. Walking frame, four points of contact with ground, three sides, no wheels 7. Walking frame with at least two wheels so it can roll without full lifting 8. Manual wheelchair, seated wheelchair that requires propulsion from the user or another person. 9. Powered wheelchair, seated wheelchair or scooter with external powered propulsion, including as an extension of a manual wheelchair

U) Participation	Relative to your own expectations how do you currently rate your problems participating in the community? Relative to someone without limb loss, in similar circumstances but without lower limb absence how do you currently rate your problems participating in your community? Responses for both participation questions: • No problem (0-5%) • MILD problem (slight, low,...) 5-24% • MODERATE problem (medium, fair,...) 25-49% • SEVERE problem (high, extreme,...) 50-95% • COMPLETE problem (total,...) 96-100%	Response options are the ICF qualifiers
V) Ability to work	1. Relative to your own expectations how do you currently rate your problems participating in the formal or informal labour market if given the opportunity? 2. Relative to someone without limb loss in similar circumstances but without limb absence, how do you currently rate your problems participating in the formal or informal labour marker if given the opportunity? Responses for both ability to work questions: • No problem (0-5%) • MILD problem (slight, low,...) 5-24% • MODERATE problem (medium, fair,...) 25-49% • SEVERE problem (high, extreme,...) 50-95% • COMPLETE problem (total,...) 96-100%	Response options are the ICF qualifiers

W) Ambulatory Activity Level	1. How many hours do you spend standing in a normal day? 2. How many hours spend walking in a normal day? • In a normal day I don't • Less than 1 hour (a little) • 1-3 hours (some) • 3-6 (lots) • 6+ hours (mostly)	
X) Falls	1. Do you on occasion fall? Yes/No If yes to first question: • less than once in 6 months • a fall every 3-6 months • a fall every 1-3 months • in most months I would have a fall 2. If yes, a fall has resulted in an injury? Yes/No	1. A fall is a sudden loss of balance that leads to a complete fall to the ground 2. An injury is defined as a physical injury requiring medical assistance or not.
Y) Social and Community Support	1. Access to family and social support and assistance at home? Yes / No If Yes, do you utilise assistance of family and social support? Yes/No 2. Access to organised community services? Yes/No If Yes, do you utilise this from: • Government • Paid/private • Ngo /volunteer • Other	1. Access to family and social support and assistance at home = not paid 2. Community services include formal home help, home nursing care, NGO or government provided home help in which the care provider is paid or volunteers that are systematically organised

6.5.2 Discussion of Data items

6.5.2.1 At enrolment

At enrolment in the registry A) date of Birth, B) sex, C) country/jurisdiction are collected and a D) unique identifier is assigned.

Date of birth collecting only month and year allows for sufficient detail for analysis without collecting overtly personal data. Whilst gender, by its variation captures a more subjective and non-binary view and can better reflect prevailing societal views, it was felt that sex and its physiological consequences was most relevant given the aims of the registry. Sex assigned at birth was added to definition but left as 'sex' in question/answer as in many places, this simple description would be more readily understood.

6.5.2.2 For each Amputation/Limb Absence

For each amputation/limb absence E) cause of amputation, F) amputation details and ability to walk prior to amputation are collected.

The primary cause of the amputation was selected rather than all causes, with primary being defined as the main predisposing reason for the limb absence. The rationale for this is that whilst in many cases there is multiple reasons for a limb absence it is the primary cause that would be the target of prevention initiatives. An 'other' category was included with input of free text, recognising there is a balance between completeness and usability of data this was chosen in order to allow even rare causes of limb absence to be collected.

Onset was considered with different follow up questions for each aetiology, but ultimately it was considered too variable a term with many definitions based on first hospital admission. The variability of the data quality world-wide ultimately led to an onset question being removed. Amputation details allow accurate definition of the level and side and use standardised level-based descriptions.

Ability to walk prior to amputation was seen as a relevant confounder for future outcomes and as a general marker of prior health, with the time period of 3 months chosen to capture a medium-term perspective.

6.5.2.3 For each episode of Rehabilitation Care

For each episode of rehabilitation care H) Professionals involved, I) Rehabilitation services, J) Prosthetic interventions, K) Type of socket, L) Prosthetic foot ankle combination, M) Prosthetic Knee, N) Prosthetic Hip, O) Tobacco Use, P) Body Mass Index, Q) Morbidities/conditions and R) Maximum Level of Education Obtained are collected.

The professionals involved allows comparison of different outcomes for people with LLA who have access to various members of the multidisciplinary team, allowing for optimisation of team membership and to direct staffing priorities. Rehabilitation services provided is purposefully profession non-specific to account for the diversity of clinical roles and responsibilities. Compression therapy and gait training are mainstays of LLA rehabilitation, and it was felt sport is becoming more common even in LMICs.

Prosthetic intervention is a broad category to determine the degree of prosthetic treatment obtained. Each of socket, foot, knee and hip details broad categories which have been favoured over specific styles, makes or brands which was felt was unrealistic level of detail and would lead to significant errors or missing data. Multiple entries can be permitted for socket, foot, knee and hip as there is much variability and cross over in the categorisations.

Tobacco use was felt to be an important risk factor for vascular disease and therefore important to record its usage, it was felt not to be an excessively sensitive topic like some other risk factors such as alcohol use. The WHO ASSIST guideline was adapted and significantly simplified to develop this data item.

Body Mass Index was favoured over another suggested risk factor of diet, with problems defining diet and collecting reliable information in categories that were universally understood and given health problems are expected with both over and under nourishment. Pre-amputation height can be used for people with bilateral limb absence and if this is unknown an estimated arthrometric calculation can be performed.

Morbidities/conditions allows entry of International Classification of Disease codes. While in practice some common codes may have to be provided as options given knowledge of these codes is limited, it was felt this could be best done regionally or locally to offer increased flexibility.

Maximum education level obtained was based on the International Standard Classification of Education (ISCED) levels and are collected at each episode of rehabilitation as this may change over time.

6.5.2.4 Pre and Post Rehabilitation care

Before and after rehabilitation S) Date of both commencement and completion of rehabilitation, T) Use of Mobility Devices, U) Participation, V) Ability to Work, W) Ambulatory activity level, X) Falls and Y) Social and community support are collected.

Date of commencement and completion allows understanding about time intervals. Use of mobility devices has been defined both in terms of possession and usage as it was felt this might vary greatly, with the definitions not having significant detail about powered options but rather favouring more detail for more common devices worldwide.

Participation was strongly supported by users and clinicians as an important outcome and ultimate goal of rehabilitation. It has been defined using the ICF qualifiers and relative to the persons own expectations and to someone in similar circumstances but without limb absence. Ability to work is formulated in a similar way to participation and was favoured over direct questions of employment status, which were difficult to define in places where participation in the informal economy is common and were considered too sensitive in some places.

Ambulatory activity level was developed to have time intervals and descriptions for places where time is a more fluid construct. It was felt to be an important gauge of possible improvement as the result of rehabilitation.

Falls were included both in terms of frequency and effect and take into consideration the COMPASS with the PEQ Utility including a question on balance.

Social and community support was included as a domain that was felt to be important to outcomes. It examines both formal and informal care and its utilisation.

6.5.3 Implementation and usage of the LEAD

The LEAD is a CDS that can be utilised in databases and registries of various designs. Whilst an international registry has various challenges including country and regional data protection laws which are beyond the scope of this project to explore fully, efforts should be made to determine

solutions such as data being stored within the jurisdiction of collection and deidentified data being shared to an international registry to allow data combination and comparison.

The LEAD can be used in its current form within a jurisdiction without limitation and while it represents a CDS there is no limitation to collecting additional data elements. This report includes lists of excluded data items and items which did not reach consensus which should be considered if relevant in a particular context to improve local data useability and relevance.

When implementing the LEAD as part of a registry or database at any level (international, country, regional or centre) it is strongly encouraged not to modify or delete any data items as this will render the data incomparable or aggregable with data collected in other locations.

6.5.4 Future Directions

Determining the general usability of the LEAD will require piloting in various contexts. Whilst efforts have been made through the broad-based consensus process to ensure relevance, universality of language and definitions of categories that are applicable everywhere, registry datasets such as the LEAD often require adaptation and evolution over time through a process of testing and refinement. Publication of experiences using the LEAD and critique is encouraged.

Collection of the LEAD, including the COMPASS, could be greatly enhanced with use of a web or application-based portal which could prompt users and clinicians to enter data regarding their circumstances and care. Use of such a technological solution would be ideal in waiting areas for users and could significantly streamline data collection. The LEAD could also be integrated with existing electronic health records to allow relevant information to be automatically pulled and allow relevant data on outcomes to be stored for each user in their personal records.

Development of a paper-based data collection form based on the data dictionary is required for those unable to access digital technology.

6.5.5 Limitations

Limitations of the LEAD in its current form include a lack of representation from some countries and regions in the process, with acknowledgement that not all specific constraints or cultural norms will have been explored. As such the LEAD is based on the experiences and views of the participants and a different group may come to different conclusions about what is deemed core data and what is possible and realistic to collect. The total time burden is likely to exceed the amount that the

consensus process agreed of twelve minutes, although it should be noted in its current form data collection occurs at multiple time points and from different sources (user and clinician) no single time point of data collection is likely to exceed this. The lack of certainty over the timing of data collection reveals a further limitation, the lack of piloting of the LEAD in various contexts to determine its overall time burden, practicality, and universal appropriateness.

6.5.6 Unmet need

Addressing problems of unmet need was consistently mentioned throughout the consensus process including as an objective and purpose of data collection and concerning data collection methods that limit people not receiving rehabilitation services from being included. A registry needs to have data quality checking mechanisms in place to ensure that the decision makers are basing courses of action on reliable data. Opening up data entry broadly (e.g. to anyone with an internet connection) would lead to questions of the quality of the data collected.

It was felt that the objective of screening for people in need of rehabilitation services was a complementary but separate endeavour. Some of the data fields in the LEAD could be used in such a screening tool but with the specific goal of establishing the relevance of rehabilitation services. Such efforts are important given significant resources can be spent attending even an initial appointment for assessment.

Unmet need can be explored through registry data in a variety of other ways other than allowing screening of those not receiving services. Subgroups of the population can be compared to country wide averages to compare the rate of usage of rehabilitation services in order to establish underutilisation as a result of barriers to access. For example, if it is assumed that the rate of vascular and diabetes related limb absence is similar for people irrespective of sex, the rate of utilisation of services can be compared and it can be determined if barriers to access exist. Zip code/regional area and race/ethnicity were not added to the LEAD because of challenges of finding a universal way of defining these fields and due to concerns over privacy. Addition of these elements in local data collection would allow for unmet need and disparity of outcomes to be mapped extensively.

6.6 Conclusion

The LEAD is a CDS useful in clinical practice to ensure relevant information is collected in a consistent manner, making it comparable and aggregable with data collected in other locations. Effort has been made to make the LEAD appropriate for use in all parts of the world, with particular focus on LMICs. The data contained in the LEAD can become a powerful driver of change and improvement.

There is a need for piloting the LEAD to determine its overall appropriateness, the usefulness of each data item collected, and the universality of the language and categorisations used. There is a need to develop a standard but flexible data collection platform that might include options for a web-based platform and paper-based data collection in places where reliable internet is not available.

References

1. Ebskov B. The Danish Amputation Register 1972-1984. Prosthet Orthot Int. 1986 Apr;10(1):40–2.
2. Kamrad I, Söderberg B, Örneholm H, Hagberg K. SwedeAmp—the Swedish Amputation and Prosthetics Registry: 8-year data on 5762 patients with lower limb amputation show sex differences in amputation level and in patient-reported outcome. Acta Orthop. 2020;91(4):464–70.
3. Schmeler MR, Schein RM, Saptono A, Schiappa VJ. Development and Implementation of a Wheelchair Outcomes Registry. Arch Phys Med Rehabil. 2019 Sep;100(9):1779–81.
4. Toro-Hernández ML, Augustine N, Kankipati P, Karg P, Rispin K, Schein RM, et al. Preliminary steps of the development of a Minimum Uniform Dataset applicable to the international wheelchair sector. Rubinelli S, editor. PLOS ONE. 2020 Sep 11;15(9):e0238851.
5. Fitzgerald SG, Kelleher A, Teodorski E, Collins DM, Boninger M, Cooper RA. The development of a nationwide registry of wheelchair users. Disabil Rehabil Assist Technol. 2007 Jan;2(6):358–65.
6. de Groot S, van der Linden N, Franken MG, Blommestein HM, Leeneman B, van Rooijen E, et al. Balancing the Optimal and the Feasible: A Practical Guide for Setting Up Patient Registries for the Collection of Real-World Data for Health Care Decision Making Based on Dutch Experiences. Value Health. 2017 Apr;20(4):627–36.
7. Gliklich RE, Dreyer NA, editors. Registries for evaluating patient outcomes: a user's guide. Third edition. Rockville, MD: U.S. Department of Health and Human Services, Agency for Healthcare Research and Quality; 2014. 2 p. (AHRQ publication).

7.0 Challenges & Recommendations

The goals of this project were to:

1. Define, build consensus, and disseminate recommendations for prosthetic outcome measures; Defining and aligning outcome measures with the needs of the international community.
2. Define, build consensus and disseminate recommendations for a Core Data Set (CDS) for people with LLA to standardize data collection worldwide; Defining the CDS useful to industry, implementers and policymakers.

Two processes have been conducted; one intended to develop a recommended list of outcome measures for people with LLA has defined the COMPASS, a second process on a CDS for people with LLA, useful to industry, implementers and policy makers, has defined the LEAD.

The outcome measures consensus process proceeded with identification of outcome measures conducted through a systematic review of outcome measures with psychometric properties reported for people with LLA, extraction of data on those measures, critical review of the measures through an expert panel and final selection of the COMPASS through a multi stakeholder consensus process. In addition to the COMPASS, the COMPASS+ and the COMPASS Adjunct have been defined along with a recommendation about use of a generic HRQoL measure.

The CDS for people with LLA was developed through a process of identification of data items and key themes with a scoping review and a series of semi structured interviews, a multi stakeholder consensus followed by data dictionary development. The outcome of this process was the LEAD and a list of recommendations about various aspects of registry design.

7.1 Challenges and Recommendations for LEAD and COMPASS

Various challenges to implementation of the LEAD and COMPASS exist. Section 7 discusses these challenges and provides recommendations to address them. Where the project has already addressed the challenges, no further recommendations are made.

7.1.1 Data collection

Numerous challenges to data collection were identified and possible solutions proposed. A number of these data collection challenges are common to both the LEAD and COMPASS, including time burden, equipment and resources, data collection equipment and technology.

For the COMPASS a main constraint was the time burden discussed in Sections 5.2.1 and 5.3.3. A recognition of the usefulness and relevance of outcome measures for those using them needs to be evident to overcome this constraint. The consensus process emphasized the importance of clinical relevance and user usefulness throughout. Management support is also imperative to the successful implementation of the COMPASS including allocating time for outcome measure collection and utilisation of the results from outcome measures in managing the work of the department and planning and reviewing quality improvement measures. In contexts where funding is linked to outcome measures the usefulness of outcome measure data can be readily determined by end users and clinician.

The challenge of equipment and resources in many LMICs was highlighted throughout the COMPASS consensus process. This constraint has been addressed by choosing measures that require limited equipment (i.e., stairs with three or more steps, adequate space to perform walking tests and some form of paper or ICT solution to collect PROM data). Whilst the constraint of stairs and walking space may be a limiting factor still in some places around the world, it was felt that the equipment needs of the COMPASS were realistic.

For the LEAD time burden is also a major challenge to data collection. The use of the LEAD at various time points, at enrolment, for each amputation/limb absence, prior to each episode of rehabilitation and after each episode of rehabilitation minimises the time burden for each period of data collection. This in turn complicates the process by adding the requirement for systems and processes to prompt users for data at the appropriate time.

Appropriate interfaces to collect data for the LEAD is another constraint, with paper-based processes favoured by some older and less technological proficient users. Ideally a tablet, smart phone or computer web-based system would allow direct data entry by people with LLA and their clinicians with instant scoring and data checking performed.

A major way to overcome the challenges of data collection is to ensure stakeholders at all levels find the data useful. These various forms of usefulness have been highlighted throughout the process and include education and empowerment of end users, information for clinicians to make treatment

plans and review their outcomes, information for managers to direct workflows, manage quality and plan and review quality improvement processes and information for policy makers to make the investment case and allocate scarce resources efficiently.

Recommendation: The data should be used by and be useful to those that are involved in its collection; users, clinicians, and managers.

Recommendation: A data collection platform should be developed to allow collection of LEAD and COMPASS data that allows smart phone, tablet or computer input of data items directly.

7.1.2 Standardisation of Outcome Measures

It was noted throughout the COMPASS consensus process that many non-validated forms of outcome measures exist, including videos of PerfOMs modified by the test administrators and of PROMs being translated with questions changed to apparently improve local relevance, all without psychometric testing of the altered outcome measure. Comparisons between different forms of an outcome measure cannot be made as the tests are conducted in different ways. Standard protocols as described in the COMPASS User Guide should be used for all PerfOMs and translations of PROMs should be psychometrically tested to ensure optimal validity and reliability. To facilitate appropriateness and linguistic translation, all PROMs included in the COMPASS have a minimum of regionally-specific terms (such as reference to icy roads) that would limit their use in various contexts.

Recommendation: Outcome measures should be used as described in the COMPASS User Guide.

7.1.3 Data Security

Data security to prevent unauthorised access to stored data is critical. To provide appropriate data security, encryption both in transit and storage is required so it can only be accessed by system processes or users with a valid encryption key. Computer equipment used in all locations should be secure and any possibly identifiable data stored within the relevant jurisdiction. Automatic logouts from all web-based logins after several minutes should also be encouraged to prevent unauthorized users from gaining access in the event a user leaves a computer unattended. Role based access can ensure that secure data is only available to those with a pre-determined level of authority to view and use it.

Recommendation: That data is appropriately secured with encryption in transit and storage and role-based access is employed.

7.1.4 Data Ownership

Clear ownership of data needs to be established within the jurisdiction of data collection to comply with data protection laws and to assign responsibility for security, privacy and appropriate use of the data. The data owner needs to be a trusted organisation or entity that is seeking to provide an enabling environment for rehabilitation services and not subject to commercial or other competing interests. Responsibility can be broadened to include a board to allow formal stakeholder involvement, but ultimate ownership should reside with a single trusted party. The data owner should have adequate resources in an ongoing basis to undertake the various functions required of data ownership.

Recommendation: Data ownership should be established with a trusted independent organisation or government favoured.

7.1.5 Privacy, Access and Data Rights.

Privacy is a major challenge when collecting, storing, and manipulating health data. As well as formal compliance with various regional and country data protection laws, people with LLA who provide data need to be reassured that the data is kept and used in an appropriately private manner.

Collection in many jurisdictions requires agreement that informed consent is given for its collection and storage and use and that there is a possibility for deletion of data if requested.

All reporting of data should be in aggregate form with access to individual non-identifiable data being on the basis of an ethics committee approval and release by a registry oversight committee, with rules for data security, manipulation and destruction agreed before release. To avoid sharing data that could allow identification to the central registry, dates should be substituted with time periods between events (i.e. time period between amputation and rehabilitation treatments) and people with LLA who have rare or uncommon characteristics (e.g. a post graduate degree or particular type of componentry in a setting) should not have their data shared unless a sufficient number of similar people are included in the pooled dataset.

In order to address individual concerns about privacy potentially intrusive questions (e.g. income status or alcohol consumption) that might lead to an unwillingness to participate have been

eliminated from the LEAD. The data elements remaining in the LEAD should be evaluated during piloting to verify they are universally non-intrusive.

Authorised access is important to use the data. Strict role-based access determines who can access data and for what purposes, with an individual's access being on the basis of the function that they perform. Role-based access controls that grant user access to data and the functions they can perform on the data, such as modification and deletion, should be strictly governed by job function. Each user should be assigned a unique login credential and no user should have access to the system without proper authentication and authorization. Modern information technology solutions can also record each time a user logs in and what data is viewed, or functions performed.

Some broad categories of aggregate data can be freely available in an annual report or web based for broad policy questions such as the number of people with limb absence in a country.

Recommendation: That privacy of data be maintained in accordance with local and regional laws and regulations and all reporting done in annual reports or published work is deidentified and aggregated to preserve anonymity.

7.1.6 Language

Language challenges have been encountered within the development process of the LEAD and COMPASS with the variability of meanings and terms being noted across regions (e.g. orthopaedic technician being the most common and understood term for prosthetist in some regions). Efforts have been made to clearly define terms that might be ambiguous in the LEAD and to use simple and universally understood terms. Within the COMPASS development process, individuals identified terms in PROMs such as 'sidewalk' that may not be universally understood. Translatability of terms was considered in selection of the COMPASS, COMPASS+ and COMPASS Adjunct.

All material in the LEAD and COMPASS should be ultimately formally translated into a variety of languages to allow use in all parts of the world. The LEAD should be translated forward and back to ensure accuracy. For the COMPASS, as has been detailed in Section 7.1.2 a rigorous process of translating and psychometric testing should occur.

Recommendation: Translation of the LEAD and the COMPASS should occur and translated outcome measures should be psychometrically evaluated.

7.1.7 Cultural diversity

The variability of culture and norms around the world has been accommodated within both the LEAD and the COMPASS to the degree possible, noting for each recommended data element or outcome measure the perceptions of the consensus attendees on relevance and appropriateness. Within the LEAD development, the risk factor of alcohol use was excluded because it was felt to be not universally acceptable and housing status was excluded because no clear and universally understood categorisation was found.

The COMPASS consensus process participants noted a current lack of psychometrically tested outcome measures for constructs relevant in LMICs such as sitting cross legged, squatting, kneeling, or rising from these positions. Development and psychometric testing of outcome measures that examine these culturally important mobility related movements is encouraged. Outcome measures which are sensitive to the diversity of culture and norms and accommodate these with broad descriptions rather than specific references (like crossing at traffic lights) will be highly relevant for future efforts to update the COMPASS. PROMs that use an item bank of questions or utilise crosswalks may offer other solutions for more universal applicability.

Recommendation: Future outcome measures should where possible use terms and descriptions that are universally relevant or be made universally relevant with the use of item banks or crosswalks.

7.1.8 Limitations of outcome measures

Some outcome measures in the COMPASS were noted to have floor and ceiling effects which may limit their applicability for all people with LLA. Ceiling effects relate to an inability of the outcome measure to detect change above a certain level of function and vice versa for floor effects. Efforts have been made with the addition of the COMPASS+ to address this with reference to ceiling effects and the AMP contains elements with a low floor effect, with measurement possible even for people not ambulating with a prosthesis.

Most outcome measures fail to capture the full diversity of experiences people living with a health condition might have. As a result, the PSFS was included in the COMPASS Adjunct. Being an outcome measure that allows people with LLA to identify, nominate and measure achievement of their own realistic goals, it achieves a level of individualisation that most outcome measures lack. The individual nature of the PSFS prevents comparison between people so the data is not useful to aggregate or compare. It has been noted that qualitative research such as content analysis could be conducted on the results of the PSFS to identify key goals within a population of people with LLA.

Inclusion of a generic HRQoL recommendation was agreed in order to capture a broad perspective of the outcome of rehabilitation for people with LLA that avoids specifics, acknowledging that quality of life is the ultimate goal of rehabilitation and that specifics like ambulating or performing other activities of daily living are a means to achieve it.

7.1.9 Suitably trained clinicians

In many parts of the world a suitably trained workforce was seen as a challenge for reliable data collection for the COMPASS. PerfOMs require standardised administration for maximum reliability. An understanding of outcome measurement and the value of reliable data capture is important so as to encourage a disciplined adherence to established outcome measure protocols. Outcome measures are not routinely taught in all parts of the world.

Throughout the COMPASS process there was concern raised that certain PROMs with a focus on psychosocial aspects of adjustment might raise topics of conversation that individual clinicians with limited training may not be equipped to deal with. Without appropriate training or the support of a multidisciplinary team that includes the professions of psychology or social work it was felt that expectations might be raised for people with LLA but without possibility of adequately addressing this aspect of rehabilitation. Improved training and education of the workforce to address this issue is encouraged through both formal education programs and continuing professional development. As a result of the dual constraint of existing workforce training and a lack of multidisciplinary team support in many places, outcome measures were chosen within the COMPASS that did not focus in depth on psychological issues.

It is estimated that by 2050 a doubling of the number of trained prosthetists and orthotists is required to be able to meet the growing P&O needs of the population but at present the infrastructure for the training of the workforce is insufficient to meet these demands (1). The lack of sufficient training facilities is further compounded by a high attrition rate of trained P&O personnel (2).

Recommendation: Further investment into developing sufficient training places with an appropriate professional structure for P&O practitioners is required to underpin both LEAD and COMPASS.

7.1.10 Ongoing development of outcome measures

Development and psychometric testing of outcome measures is an evolving process. In time new outcome measures or improved forms of existing outcome measures will be available. If outcome

measures included in the COMPASS are updated, it will be important to ensure that a consistent form of the outcome measure is used to allow comparison over time and comparison across populations. Crosswalks to older versions of outcome measures can be developed to facilitate some degree of comparison. Regular review of available outcome measures and the associated evidence to support their use will be needed as new psychometric reporting emerges.

Recommendation: The COMPASS should be reviewed and revised periodically as new outcome measures and evidence becomes available.

7.1.11 Limitations of balancing ICF categories

Consensus participants were asked to rank ICF chapters corresponding to the concepts or constructs included in included outcome measures. This was done to ensure significant duplication of constructs or unconscious omittance of a construct was avoided. Ultimately all constructs were included. The absence of an agreed ICF Core Set to map the outcome measures in the COMPASS against leads to uncertainty that all relevant constructs are indeed covered by the COMPASS. A weighting of importance might have allowed exclusion of some of these constructs.

The ICF chapter on environmental factors includes a small number of items relevant to assistive technology and covers wearable devices such as lower limb prosthetics but it lacks granularity and detail for some aspects relevant to people with LLA.

7.1.12 Additional categories beyond the ICF relevant to people with LLA

Three additional unranked categories not included in the ICF were added to the consideration of constructs explored by outcome measures. They were socket comfort, satisfaction with prosthesis and ability to don the prosthesis. All were felt to be important constructs and outcomes by the consensus process participants. As no ranking of these constructs was sought, they were both unranked and unweighted. There may be other non ICF categories of importance not revealed by this process leading to overall uncertainty about the constructs covered by the COMPASS.

7.1.13 Changing technology

The evolution of technology will affect the data collection for both the LEAD and the COMPASS and presents some challenges and opportunities.

Data collected in the COMPASS and LEAD could benefit greatly from smart phone or tablet data collection methods. For the COMPASS an application (app) that prompts clinicians with instructions for PerfOMs and offers in-app timing would be an advantage. Collection of PROMs from people with LLA in a waiting area on a tablet directly, or remotely on their own device would be time efficient and reduce the time burden for clinicians. For the LEAD a similar application could ask questions directly and have in built logic checks (preventing obvious mistakes such as amputation before date of birth etc), to improve data quality.

Such applications could also be used for benchmarking results against deidentified and aggregate data. After data entry, users may be able to compare their outcomes to others in similar circumstances, clinicians would be able to compare the outcomes they obtain with their clients to other clinicians, managers would be able to view centre level data to understand key aspects of the operation that require the focus of quality improvement and policy makers would be able to view data to inform policy decisions. Flexibility of design of such a platform can allow for other locally relevant data to be collected in parallel to the LEAD.

Data elements for prosthetic components and rehabilitation interventions in the LEAD reflect current state of technology available in routine clinical practice. While categories, rather than an exhaustive list of products which is liable to change have been preferred, it is possible whole new categories of prosthetic components may become available in the future. As such, an update of the categories such as prosthetic components and assistive devices, may be needed in time.

Recommendation: The LEAD should be reviewed and revised as new categories of prosthetic components and rehabilitation interventions become available.

Recommendation: As detailed in Section 7.1.1 a data collection platform should be developed to allow collection of LEAD and COMPASS data that allows smart phone, tablet or computer input of data items directly.

Recommendation: A data collection platform should allow real time benchmarking information to be provided to users and clinicians.

7.1.14 Identification of unmet need

Indications of unmet need can be examined within the LEAD by comparing existing data elements such as sex distributions against existing census data to establish under-representation as discussed in section 6.5.6. The LEAD can also be augmented with local level data on geography, ethnicity, or

other local demographic information to establish the presence of barriers to access. The possible addition of locally relevant data items in encouraged, with the LEAD representing a minimum uniform core dataset.

Recommendation: That locally relevant data items such as ethnicity, location (zip codes or region) and others are collected in parallel with the LEAD.

Recommendation: A dataset to comprehensively capture all unmet need for lower limb prostheses would need different data items to the LEAD.

7.1.15 Need for ongoing collaboration to ensure change and improvement in service delivery

Data collected by the LEAD and COMPASS can be a positive driver of change and improvement. This data needs to be shared, analysed and discussed for conclusions to be drawn and improvements to be realised. While data privacy is a primary concern, deidentified and aggregate reporting is critical to be able to answer real world questions about resource allocation to drive improved outcomes. Data contained in the LEAD will in time be able to highlight issues relevant to policy makers, such as which additional members of a multidisciplinary team are likely to bring about improved outcomes, what effect does alternative prosthetic components have on outcomes and are there signs of significant unmet need. Comparison and collaboration will allow local knowledge and experience to be quantified and shared, as well as to be used to establish best practice and make key recommendations.

Recommendation: To support decision making at all levels, analysis of deidentified data reported in aggregated form should be regularly performed.

7.1.16 Need for future momentum and adaption to ensure the relevance of LEAD and COMPASS

Development of the LEAD and COMPASS are the primary steps in improving routine data collection worldwide. Commitment from stakeholders to implement and evolve the LEAD and COMPASS will be needed overtime. As has been discussed in section 7.1.10, development of new or improved outcome measures may require changes to the COMPASS to be made in time. New categories of componentry or evolution of rehabilitation practices may need to be incorporated into the LEAD. Most registries go through a period of evolution, based on feedback from users, clinicians, researchers, and other data consumers. A data field might be found to have so much missing data,

or be found to add little to analysis that it requires deletion or adaption. The inability to perform certain analyses may prompt the call for inclusion of a previously omitted data item. Evolution through such critique and consideration should be encouraged including deletion of underutilised data items as well as adding of new ones.

Recommendation: As detailed in Section 7.1.10 The COMPASS should be reviewed and revised periodically as new outcome measures and evidence becomes available.

Recommendation: The LEAD is a dataset that requires evolution and adaption overtime.

Recommendation: Funding should be provided to an international organisation to develop, pilot and implement the LEAD and the COMPASS.

7.1.17 Implications for broader AT sector

Implications for assistive technology (AT) beyond rehabilitation for people with LLA include learning from the demonstrated methodology used to construct a recommended list of outcome measures (COMPASS) and a CDS (LEAD). Of particular interest should be the expert panel, which was used for this project to review the results of the systematic review to ascertain which outcome measures have adequate psychometric properties so as to avoid measures that lack evidence of validity or reliability being routinely recommended. Application of the ICF chapters has some flexibility of future application in other fields in that various health conditions (with or without ICF core sets) could be mapped and reviewed to determine outcome measures which explore important constructs.

LEAD data will be able to provide valuable information about assistive technology and its effect on outcomes of people with LLA and this may be able to be generalisable to other health conditions.

Recommendation: The broader AT sector should review this process to support future efforts and adapt it as needed.

7.1.18 Implications for rehabilitation sector

The development of the LEAD and COMPASS offer an opportunity for standardised reporting on outcomes as opposed to just the quantity of services offered. In the past the number of devices provided was a key metric and the reporting of outcomes done in an ad hoc manner. A move towards standard reporting of outcomes alongside services enables reflection on what is limiting the

outcomes of people with LLA and what can be done to improve it. Collection of demographics and surgical/limb absence data as well as other potential confounders within the LEAD will allow for nuanced explanations within project or programme reporting and for specific corrective action to be determined.

Recommendation: That the rehabilitation sector should move further towards reporting outcomes, including patient reported outcomes, in addition to the quantity of services provided.

7.1.19 Implications for an ageing population

Many countries around the world are confronting changing age demographic profiles with an anticipation of a larger percentage of the population being elderly. The chronic diseases which are increasingly prevalent in the elderly predispose to disability. The prevalence of diabetes is increasing rapidly in LMIC and in turn the medium- and long-term complications are likely to increase as well, resulting in increasing generalised debility and specifically micro and macrovascular disease as well as neuropathy predisposing to lower limb amputation. The increased need for services for people with LLA for the elderly will require accurate information about AT and rehabilitation treatments that are effective in improving outcomes, to guide scarce resource allocation.

7.1.20 Need for funding of rehabilitation services

There is a need to establish the case for funding rehabilitation services worldwide in order to improve and maintain access to services. Policy and decision makers generally wish to pay for services that have been demonstrated to improve outcomes, rather than services that do not. LEAD and COMPASS allow for real world determination of improved outcomes as a result of rehabilitation interventions and can thus be used effectively to build the case for funding. Where no services are available, examination of similar contexts can give insights about the future possibilities. Where services need improvement, examination of similar contexts where different models of care and priority setting has occurred can give insight into the changes and improvements that should be sought.

It is important to build a strong case for all aspects of rehabilitation and assistive technology provision to guide investment towards improved outcomes.

7.2 COMPASS Recommendations

Six key recommendations regarding the use of outcome measures for individuals with LLAs arose from the consensus process and these were presented in Section 4.2 and discussed in Section 4.3.1. These recommendations are as follows:

1. Recommendation 1: That the AMP, TUG, 2MWT, PEQ – Residual Limb Health, PEQ – Utility, and TAPES-R make up the ISPO lower-limb COMPASS.
2. Recommendation 2: That the CHAMP and 6MWT, two additional PerfOMs recommended for high functioning individuals with LLA make up the COMPASS+.
3. Recommendation 3: That the PSFS make up the COMPASS Adjunct.
4. Recommendation 4: That a generic HRQoL outcome measure, such as the EQ-5D-5L or PROMIS© Brief Profile (PROMIS-29) be used to supplement the COMPASS.
5. Recommendation 5: That outcome measures suited to low- and middle-income countries (LMICs) are developed with a focus on activities such as sitting cross-legged, kneeling, squatting, and other culturally important mobility related activities.
6. Recommendation 6: That translation, validation, and open sharing of translated outcome measures included in the COMPASS, COMPASS+, and COMPASS Adjunct occurs.

7.3 LEAD Recommendations

The consensus process to define the LEAD (Section 6.5) has resulted in a CDS for people with LLA that spans demographic, surgical/limb absence, rehabilitation and outcome data elements. Collection of the LEAD will result in an ability to analyse outcomes of people with LLA and can inform health policy, centre management while enhancing project or programme reporting.

Recommendation: That the LEAD is used in routine practice to collect data for registries and databases.

Agreement was reached on many themes including objectives, target population, time burden, recruitment, the compulsory or voluntary participation of users, data collection methods, provider information, funding, stakeholders, registry design, data access and multinational registry solutions. These have been detailed in section 6.2.2 and discussed in section 6.3. Development of each registry or database will face unique constraints such as existing data collection practices and norms that

must be accommodated as well as local laws and regulations. The diversity of circumstances that a registry custodian might face cannot be fully anticipated which makes it impossible to provide definitive recommendations on each theme. This report should act as useful guidance and support in registry or database development.

Recommendation: That those responsible for developing or maintaining existing registries and databases consider the discussion in this report on themes such as; objectives, target population, time burden, recruitment, the compulsory or voluntary participation of users, data collection methods, provider information, funding, stakeholders, registry design, data access and multinational registry solutions.

7.4 Multinational registry

Although most registries to date have been national in design, there is growing interest in working across countries to harmonize standards, share data, and contribute to research and development of leading practices at one level while simultaneously facilitating demonstration of quality and safe service delivery at all levels. The motivation for this work was to assist in the promotion of increased service delivery and improved service provision around the globe. Data pooled across countries will enable comparison and lead to more generalizable evidence in terms of device and service performance and patient outcomes. A broader range of practice settings can capture higher variation in performance and provide more opportunity for discovering and analysing best clinical practices.

International registry networks are not without their challenges. The data harmonization required for pooling data must build on consensus from different stakeholders that bridges health systems, reimbursement structures, and cultures. Importantly, consensus has been reached in this work. A pragmatic approach was used with the most interested countries to achieve harmonization regarding the data elements and data dictionary. Beyond harmonization, actual sharing of data is made more difficult by divergent national rules. Privacy laws, for example, vary by country and sometimes by region, allowing greater or lesser ease for global aggregation of de-identified patient-level data. An approach of selecting those countries with data privacy laws that will facilitative international collaboration is one suggested strategy. Operationally, developing an international registry network is challenging. Communication, for example, must rely more heavily on video teleconferences that must span a broad range of global time zones and require a single common

language or use of simultaneous translation. Notably, this project achieved this international cooperation and managed to organize regular, virtual meetings to facilitate convergence of data definitions.

There was unanimous consensus that an international registry or platform that allowed for compliance with local data laws and was available in participants country or territory would be a good solution for improved data collection. While many general principles are similar for participant enrolment and retention, there are different customs or regulations regarding contract language, requirements for ethics committee or other submissions, informed consent, and allowable approaches to patient retention in different countries. An international registry will need to reference national and local regulations that extends across countries.

Registries may be subject to specific procedures depending on the data collected, the authorization obtained, and applicable governmental regulations. A solution for an international registry is to store only de-identified data in a cloud-based platform under the auspices of the data custodian. This de-identified centrally-stored data would be available in aggregate form, while local health services as data owners would maintain access to their own data in a fully identifiable format. This would allow users/patients to be contacted, specific analysis to be performed and comparisons made using other data collected in the registry. The data custodian for the central repository could allow access to data via annual reports or web-based statistics which would have numerous advantages for global policy making.

For any international effort there are some specific issues related to data protection and ensuring all information shared centrally is not identifiable. Most registries have reporting guidelines in place so that if only a small number of people meet a specific category their results will not be reported since information related to their care may allow those viewing the data to discern an individual's protected health information. In the case of an international central repository, using deidentified data would preclude sharing of protected health information to this central location where there are small numbers of people in one jurisdiction with specific circumstances in the LEAD. Exact timings of events (in terms of month and year) can also be replaced with intervals between events to prevent any potential identification.

Access to or publication of data which can be associated with any jurisdiction would require authorisation from that jurisdiction. Reporting of global and regional averages for various other categories within the LEAD and COMPASS would allow comparisons of models of funding and care outcomes to be made and various other determinations of effectiveness in a variety of contexts.

The data custodian of an international repository should be a trusted international organisation with the resources on an ongoing basis to undertake the various functions required of custodianship.

Recommendation: An international registry should be developed that allows for data comparison and aggregated reporting under the custodianship of a trusted international organisation.

References

1. Clarke L, Puli L, Ridgewell E, Dillon MP, Anderson S. Regulation of the global orthotist/prosthetist workforce, and what we might learn from allied health professions with international-level regulatory support: a narrative review. Hum Resour Health. 2021 Dec;19(1):83.
2. Pryor W, Smith F. Impacts of the Nippon Foundation's investment in Exceed Worldwide programs. ISPO, Brussels; 2019.

8.0 LEAD and COMPASS Conclusion

The LEAD and COMPASS are a significant development for provision of service for people with LLA. They align with the CRPD obligations on state parties to collect information and data to inform policy related to people with disabilities by establishing recommendations on what data should be collected for people with LLA. Widespread implementation of the LEAD and COMPASS will contribute to achievement of the SDGs, which focus on good health and well-being and reduced inequality and underpin the project goals of informing the investment case for lower limb prosthetic and rehabilitation services.

The LEAD and COMPASS represent a coordinated, integrated, multidisciplinary international effort to standardise routine data collection. Many challenges exist in providing rehabilitation services worldwide especially in LMICs. Data from the LEAD and COMPASS can inform decision making and be a powerful driver for access to high quality, cost effective and sustainable rehabilitation interventions for people with LLA.

Consistent collection, storage and use of such data has many challenges, but is worthwhile and will have far reaching benefits for people with LLA. Collection and analysis of these data elements can educate, motivate and empower individuals, inform and underpin the treatment plans jointly negotiated with clinicians, and be invaluable for managers planning, monitoring and improving service delivery and quality as well as giving a solid foundation to inform policy and funding decisions.

There is no single prescribed method to implement the LEAD and COMPASS at either the local or country level. Regionally or internationally the development of a common platform that allows ownership and possibly storage of data within the local jurisdiction, but shares aggregated and deidentified information centrally would present many opportunities for comparison and learning and is recommended to allow maximal benefits to be obtained from the data.

The LEAD is a CDS and other locally relevant data or datasets can be collected in parallel or developed as additional modules to complement the LEAD. Not all items can be incorporated into a CDS, and additional data items or data banks could be developed and collected in parallel for specific projects or research protocols. While some may consider a CDS as lacking some details, the strength of the LEAD is that the items can be collected in a manageable time frame even in busy clinical practice, which will enhance compliance with data collection.

A transformation offered by LEAD and COMPASS, the routine collection of core data items of demographics, surgical and limb absence interventions, rehabilitation interventions and outcomes offers us the ability to answer questions of interest at the local level with local data. The advantage of gathering and analysing local data is that it accounts for all the local variability that might affect outcomes. This is in contrast to many published studies about effectiveness of rehabilitation interventions for people with LLA which are not generalisable to other contexts and are often limited by small sample sizes.

When widely adopted the LEAD and COMPASS will allow comparisons to be made to other jurisdictions with different service delivery models, prescribing patterns, component choices or composition of the multi-disciplinary team. Establishment of best practice can be determined through such comparison within the prevailing resource constraints. The deficiencies of small sample size can be addressed with aggregation of data from several jurisdictions to establish datasets for less common clinical presentations allowing for clearer recommendations to be made.

We recommend that the LEAD and COMPASS be promoted, adopted, implemented and interrogated widely in order to assist the aims of ATScale to markedly increase accessibility to and benefit from prosthetic services.

9.0 Acknowledgements

ISPO would like to acknowledge the following individuals and organisations:

The project was funded by USAID in support of ATscale and managed through UNOPS.

LEAD (CDS): Bengt Söderberg, Barbara Goedde, Ben Mortenson, Bernd Schwien, Dave Kruppa, Fabian Pedraza, Joanne Hebenton, Jonathan Naber, Manasi Joshi, Kenton Kaufman, Lisa Gramnäs, Puni Mamdoo, Brendan McCormack, Saeed Zahedi, Samual Hale, Soikat Ghosh, Stuart Campbell, Andreas Hahn, Carlotte Kiekens, Chris Likando, Nachiappan Chockalingam. Benje Theron, Kerstin Hagberg, Giulia Oggero, Almah Kuambu, Didier Cooreman, Lee Brentnall, Trevor Binedell, Mohammad Sobuh, Mahpekay Sidiqy, Gulapar Srisawasdi, John Ondiege, Zulfiqaraki Abbas, Fiona Davie-Smith, Roberto Araujo.

COMPASS (Outcome Measures): Abmir Mitra, Andreas Hahn, Barbara Goedde, Breda Clancy, Christopher Hutchison, Carson Harte, Chantel Ostler, Christian Gasy, Cristina Garces, David Boone, Ghassan Ali, Gavin Jenkins, Julia Engkasan, Katie O'Halloran, Kajsa Lindberg, Katherine Cook, Luphiwo Mduzana, Maria Paleo, Michael McGrath, Nicolás Muñoz, Niles Jonkergouw, Oscar Portillo, Grazielle Oliveira Da silva, Phil Stevens, Robert Gailey, Roxanne Maritz, Rob Verwaard, Oinam Sanjoy, Sisary Kheng, Surona Visagie, Yudith Chivata, Sarah Hughes, Carolina Schiappacasse, Bernd Schwien, Brendan McCormack, Lisa Gramnäs.

LEAD Data Dictionary definitions: Puni Mamdoo, Stuart Campbell and Kenton Kaufman.

The expert panel which appraised the psychometric strength of outcome measures: David Boone, Kerstin Hagberg, Linda Resnik, Deirdre Desmond, Brian Hafner, Robert Gailey Matthew Major, Franco Franchignoni and Friedbert Kohler.

Members of the International Society of Prosthetics and Orthotics (ISPO) Industry Advisory Group (IAG): Andreas Hahn, David Moser Kenton Kaufman, Alan Hutchison, Saeed Zahedi, Brian Hafner, Bengt Soderberg, Christopher Hutchison, Samuel Hale, Berit Hamer and Friedbert Kohler.

Project team members: Principal Investigator: Friedbert Kohler, Project Manager: Gregory Halford, Scientific Lead: Jade Tan, Technical Officers: Martina Lukin and Brodwen McBain