Evaluation Studies of Robotic Rollators by the User Perspective: A Systematic Review

Christian Werner[a], Phoebe Ullrich[a*], Milad Geravand[b], Angelika Peer[c], Klaus Hauer[a]

[a] Department of Geriatric Research, AGAPLESION Bethanien-Hospital/Geriatric Center at the University of Heidelberg, Germany.
[b] Department Robot and Assistant Systems, Fraunhofer Institute for Manufacturing Engineering and Automation IPA, Stuttgart, Germany.
[c] Bristol Robotics Laboratory, University of the West of England, Bristol, United Kingdom.

*CW and PU contributed equally to this work

Corresponding author:

Klaus Hauer (KH)
Department of Geriatric Research, AGAPLESION Bethanien-Hospital/Geriatric Center at the University of Heidelberg, Rohrbacher Str. 149, 69126 Heidelberg, Germany
Phone: +49 6221 319 2150, Fax: +49 6221 319 1505
E-Mail: khauer@bethanien-heidelberg.de (KH)

E-mail of other authors:

christian.werner@bethanien-heidelberg.de (CW)
phoebe.koepp@bethanien-heidelberg.de (PU)
milad.geravand@ipa.fraunhofer.de
angelika.peer@brl.ac.uk

Running head:

Evaluation of Robotic Rollators

Word count abstract:	299
Word count text manuscript (without abstract, acknowledgments, disclosures, and references):	5130
Number of references:	48
Number of figures:	1

Abstract

Background: Robotic rollators enhance the basic functions of established devices by technically advanced physical, cognitive, or sensory support to increase autonomy in persons with severe impairment. In the evaluation of such Ambient Assisted Living solutions, both the technical and user perspectives are important to prove usability, effectiveness, and safety, and to ensure adequate device application.

***Objective*:** The aim of this systematic review is to summarize the methodology of studies evaluating robotic rollators with focus on the user perspective and to give recommendations for future evaluation studies.

Methods: A systematic literature search up to December 31, 2014 was conducted based on the Cochrane Review methodology using the electronic databases PubMed and IEEE Xplore. Articles were selected according to the following inclusion criteria: Evaluation studies of robotic rollators documenting human-robot interaction, no case reports, published in English language.

Results: Twenty-eight studies were identified that met the predefined inclusion criteria. Large heterogeneity in the definitions of the target user group, study populations, study designs, and assessment methods was found across the included studies. No generic methodology to evaluate robotic rollators could be identified. We found major methodological shortcomings related to insufficient sample descriptions and sample sizes, and lack of appropriate, standardized and validated assessment methods. Long-term use in habitual environment was also not evaluated.

Conclusions: Apart from the heterogeneity, methodological deficits in most of the identified studies became apparent. Recommendations for future evaluation studies include: clear definition of target user group, adequate selection of subjects, inclusion of other assistive mobility devices for comparison, evaluation of the habitual use of advanced prototypes, adequate assessment strategy with established, standardized and validated methods, and statistical analysis of study results. Assessment strategies may additionally focus on specific functionalities of the robotic rollators allowing an individually tailored assessment of innovative features to document their added value.

Key words

Systematic review, Evaluation studies, Ambient assisted living, Robotics, Rollator, Walker, Self-help devices, Human-robot interaction, Mobility, User experience

Introduction

In older persons, the ability to move independently represents a hallmark of autonomous living [1] and quality of life [2], while being physically active is associated with numerous positive health outcomes [3, 4]. However, sensory, motor or cognitive impairments restrict mobility in frail, older persons [5]. Motor key functions such as standing, walking, or transfers are substantial challenges for their daily activities leading to high risk exposure of falls as documented in residents of senior homes [6]. Effects of motor impairment are augmented by sensory deficits such as visual impairment, leading to restricted functional independence [7], or by cognitive impairment, leading to spatio-temporal disorientation or executive dysfunction [8]. To overcome or compensate for such impairments and to improve the quality of life of affected persons, assistive devices as in walking aids (e.g. canes, walkers, rollators) have been developed with an early focus on mobility support. They provide support of postural stability and mobility [9], reduce risk of falling [10], and improve activity and participation [11]. However, such conventional mobility devices may not cover the needs of persons suffering from major functional or cognitive impairments.

In the context of Ambient Assisted Living (AAL), robotically augmented rollators with various high-tech functionalities have been developed to provide physical, sensory and cognitive assistance, and/or health monitoring for further support [12]. The development and evaluation of such a robotic rollator (RR) is still a new, emerging research field mainly driven by technical engineering goals. However, as technical functionalities translate into assistive devices for use of the target population, for which these have been developed, the human-robot interaction and user perspective shifts in the development focus. Apart from the sheer technical evaluation of concepts and functionalities, needs, requirements, and preferences of potential users will have to guide the development and evaluation of assistive technology devices [13, 14]. In addition to technical testing, which verifies the functional capability of devices, an evaluation with focus on user performance, physical demands, and subjective experiences of the RR is essential to prove the usability, ensure safety, and demonstrate the added value for the intended user group. The change from technical to user perspective may, however, lead to specific methodological challenges including the study design and assessment strategy. To our knowledge, no systematic review on the evaluation of RRs with focus on the user perspective has been published. Therefore, the aim of this systematic review was to summarize the methodology of studies evaluating the human-robot interaction from a user perspective and to give recommendations for future evaluation studies.

Methods

Initial search terms were compiled and iteratively refined by team members with expertise in the clinical and in the technical research field. The literature search was conducted using the electronic databases PubMed and IEEE Xplore. Search terms included both controlled vocabulary (i.e. MeSH Terms, IEEE Terms) and keywords of relevance identified during searches. The detailed search

strategy used in PubMed, which was modified for IEEE Xplore, is presented in the online supplementary table 1.

Manual searches were performed to identify additional studies by scanning reference lists of relevant articles and by reviewing key authors' own databases. Studies were searched with focus on the evaluation of a RR (or robotic wheeled walker) by experiments, trials, or interventions in human beings independent of the type of outcome measurement. No restrictions regarding age or health status of the subjects were made. Single case reports were excluded. For the purpose of this review the term 'robotic' includes the normal function of a rollator enhanced by additional physical, sensory, or cognitive robotic support while walking, also including sit-to-stand transfers. Studies evaluating solely monitoring functionalities without taking into account any user supporting functionalities or the subjective user experience were excluded. The search was limited to articles in the English language published up to December 31, 2014.

The selection process was conducted following the methodology as described in the method guidelines of the Cochrane Collaboration [15]. Titles and abstracts were identified by the standardized search strategy. For abstracts which met the inclusion criteria or for those with unclear status, full-text articles were analyzed for inclusion. Each step of study selection, based on predefined eligibility criteria, was performed independently by two reviewers (PU, CW). Any disagreements were resolved by consensus or third-party adjudication (KH). After inclusion, data on the user group, sample characteristics, and the methodological approach were extracted by one researcher (CW) and confirmed by two other researchers (PU, DS). If an article described more than one study, the results for each study were extracted separately.

Results

A total of 8989 articles were identified through database searching, and another 79 were added through manual searches. After removing duplicates, the initial search resulted in 8876 articles. Of these, 235 were found to be related to the search topic based on title and abstract. After reviewing full texts, 148 articles were excluded as they did not meet the predefined inclusion criteria (Fig. 1). Another 63 were discarded, as these articles described either identical experiments with the same RR, or various stages of development of a certain RR. In both cases, the article providing the most comprehensive information with focus on the user perspective was included. If different articles contained similar information, the one with the most recent development stage was included. Twenty-four articles published between 2001 and 2015 were identified for inclusion in the review. As two articles reported on two [16, 17] and one article on three independent studies [18], the final data extraction was based on 28 studies[†]. The detailed review results extracted for each study are presented

[†] When necessary, the individual studies of these articles are distinguished with numeric coding (i.e. [16[1,2]], [17[1,2]], [18[1,2,3]])

in the online supplementary table 2, containing information on the names of devices, the definition of user groups, study sample, study object, study design, and selected assessment methods.

(Please insert figure 1 about here)

Fig. 1: Flow chart of the study selection process and extraction methodology

User Group Definitions

Apart from two articles [19, 20], all mentioned a target user group for the RR; however, their definition differed substantially in accuracy and explicitness. Five articles provided a generic description in broad terms such as 'elderly (disabled) people'' [21-25], two defined users by setting-specific characteristics such as 'persons in nursing and assisted living homes', partly amended by disease-related criteria (e.g. Alzheimer's disease, stroke) [26, 27], and ten provided brief information on users' motor-functional (e.g. 'with mobility problems'), cognitive (e.g. 'with cognitive impairment') and/or visual status (e.g. 'visually impaired') [17, 18, 28-35], but without staging impairment levels based on any screening or assessment instrument. Three articles described users by disease categories (e.g. Parkinson's disease, hemiplegia) [16, 36, 37] without detailed information on the patients' functional impairment level. Specific impairment-related definitions based on established, validated assessment methods (i.e. Walking Index for Spinal Cord Injury [WISCI II], Functional Ambulation Classification) were documented in only two articles [12, 38].

Study Samples

The mean sample size of studies was 7.2 (standard deviation [SD] ± 4.3). The exact number of subjects was not reported in five studies [18[1,2,3], 35, 37]. No study presented a sample size calculation.

Samples differed considerably regarding age, impairments, or diseases. The age of subjects ranged from 14 [22] to 97 years [31] with age information lacking in half of the studies (14 of 28) [16[1], 17[1], 18[1,2,3], 20, 23, 25, 27-29, 35, 37].

Thirteen studies included subjects with motor, functional, cognitive, visual and/or neurological impairments [12, 16[1,2], 17[1,2], 26, 27, 30-32, 34, 36, 38], whereas a convenient (e.g. 'ordinary adult males') [19, 20, 23, 24, 33], mixed (e.g. 'healthy subjects and subjects with motor and cognitive impairment') [18[1,2,3], 21, 22, 29, 35, 37] or setting-specific sample (e.g. 'residents of retirement facility') [28] was used in 14 studies. In studies including impaired subjects, definitions and staging of the severity level of impairment were mostly absent (15 of 20) [17[1,2], 18[2,3], 22, 26, 29-32, 34, 35, 37, 38]. In only six studies, motor-functional or cognitive impairment levels were defined by established and validated screening or assessment instruments (e.g. Timed up and Go [TUG], Mini-Mental State Examination) [12, 16[1,2], 21, 27, 36].

In ten studies, subjects did not match with the predefined user group [18[1,2,3], 22-24, 27, 28, 33, 37]. However, due to the unspecific and wide-ranging user group definitions given in a number of articles, most studies (15 of 28) were carried out with subjects who were covered by these broad definitions

[12, 16[1,2], 17[1,2], 21, 26, 29-32, 34-36, 38]. In three studies, a user group definition and/or a description of the study sample was completely missing [19, 20, 25].

Design of Studies

Depending on study objectives, three different types of studies were performed: (1) observational studies; (2) comparative studies, or (3) interventional studies.

Observational Studies

Fourteen articles reported on observational studies [12, 18, 20, 22, 24, 29, 35, 37] or single observational experiments as part of their studies [16, 17, 23, 26, 28, 33], focusing predominantly on the verification of technical capability and/or the subjective user evaluation of RR. User performance was used as the study object in only one of these studies [26]. In observational studies/experiments, outcomes were only descriptively presented, without providing any reference values.

Comparative Studies

Fourteen articles included comparative studies [19, 21, 25, 27, 28, 30-32, 34, 38] or single comparative experiments in addition to observations [16, 17, 26, 33]. Comparisons were further distinguished into four categories: (1) 'inter-device comparisons' in which RR and conventional devices (e.g. cane, folding/wheeled walker) or fully unassisted walking/sit-to-stand transfers were compared [19, 21, 26, 27, 30, 32, 34, 38]; (2) 'intra-device comparisons' in which different assistance levels (e.g. activated vs. non-activated obstacle avoidance), interface designs, or development stages of the same RR were compared [17[2], 19, 25-28, 30, 31, 33, 34]; (3) comparisons in a pre/post-test study design with focus on the user experience [34] or the technical functionality [23], assessed before and after/over a series of trials; and (4) comparisons between outcomes of a newly developed robotic monitoring functionality and those of an external criterion measure as a reference measurement [16[2]].

Interventional Studies

Two articles described studies that used an interventional approach, providing training opportunities with the RR [16, 36]. In one study, the subjects' gait performance with the robotic gait assistance system was assessed on six consecutive days [16[1]]. However, subjects seemed to use the RR only during test procedures and not in their daily routine. Although the ultimate research hypothesis for this 'interventional' approach was lacking, we assumed that the repeated use represented a type of training intervention in order for the subjects to get used to using the RR. In the other study, a four-week randomized controlled trial was conducted to evaluate the effects of ambulation training with a RR compared to a traditional rehabilitation therapy method using parallel bars [36]. In this study, assessment methods were used to evaluate the subjects' motor-functional performance after the robot-assisted training intervention.

Statistical Analysis

An inferential statistical analysis of outcomes was included in only three studies [19, 34, 36]. In 25 studies, outcomes were presented using solely descriptive or qualitative data (e.g. frequencies, means, SDs, and user comments) [12, 16[1,2], 17[1,2], 18[1,2,3], 20-33, 35, 37, 38].

Assessment Methods

Assessment measures used in identified studies can be classified into five categories:

(1) established clinical performance-based measures assessing subjects' functional ability to perform a requested task by simple quantitative time-, range-, or rating-based outcomes (e.g. gait speed, walking distance, rating score) or by more detailed, qualitative outcomes captured by external technical measures (e.g. step time, double support time); (2) tailored assessment methods in terms of self-designed performance-based measures specifically tailored to specific functionalities of the RR (e.g. guidance system, obstacle avoidance). In addition to simple quantifiable time- or count-based outcomes (e.g. walking time, number of collisions), these assessment methods predominantly used more technique-based and qualitative outcomes (e.g. path deviation, distance to obstacle); (3) assessment methods used to evaluate the subject's physical and physiological demands during the use of the RR; (4) subjective evaluation measures to assess a user's experience with the RR; and (5) technical evaluation measures to assess the technical capability of the RR.

As technical evaluation measures used in nine studies [12, 16[2], 18[1,2], 20, 22-24, 33], exclusively focused on the technical verification of the RR with limited relevance for the user perspective, we do not further address and discuss these measures in this review.

Clinical Performance-Based Measures

Established clinical performance-based measures were used in three studies [21, 32, 36]. In one of these, the subjects' gait and functional performance with the RR were assessed by the 4-meter walk test (4MWT), a modified version of the TUG, and spatio-temporal gait parameters (i.e. step time, double support time) captured by video camera during both tests [21]. Other studies documented the subjects' motor performance by the 6-minute walk test (6mWT), 10-meter walk test (10MWT), and Performance Oriented Mobility Assessment (POMA) [36] or only by the 10MWT [32]. The most frequently used outcomes were gait speed [21, 32, 36], completion time [21], or walking distance and rating scores for functional performance (POMA) [36].

In one study, an established screening test for assessing the functional ability of subjects to perform activities of daily living (ADL) was used (Barthel ADL Index) [36].

Tailored Assessment Measures

In ten studies, assessment strategies included self-designed performance-based measures specifically tailored to specific robotic functionalities [16[1,2], 17[2], 19, 25-28, 31, 34]. Obstacle avoidance and guidance systems were evaluated while subjects completed walking paths [25, 28] or obstacle courses [17[2], 31, 34], navigation and localization systems while performing navigational tasks [26, 27], and gait assistance systems by analyzing the subject's gait during robot-assisted walking [16[1,2], 19]. Simple quantifiable outcomes of these tests included number of collisions [26, 31, 34], reorientations [34], navigational mistakes [27] or abnormal gait patterns [16[1,2]], walking time [34], or achievement of task [26]. More specifically tailored, technique-based outcomes, as used in eight studies, comprised of deviations from an optimal path [17[2], 25, 28, 31], distance to obstacles [17, 26], maximum speed and walking distance [26], mean and SD of robot's velocity [19], and gait variability (i.e. SD of gait speed/step length) [16[1,2]]. To obtain such technically advanced outcomes, five studies used the data flow created by the technical systems installed on the RR, including laser rangefinders (LRF) [16[1,2], 28], a video camera and sonar sensors [17[2]], or a web camera [31]. In the other three studies, information on the technical measure to capture these outcomes was nonexistent [19, 25, 26]. Out of the studies that determined outcomes with the robot-integrated technical systems, only one seemed to process raw data (LRF data) into outcome variables (i.e. path deviation) by using an already established method for robust position estimation of mobile robots in indoor environments ('Monte Carlo localization') [28]. In the other four studies, it remained unclear whether raw data was analyzed by self-designed or potentially established methods [16[1,2], 17[2], 31].

In two inter-device comparative studies, a bicycle speedometer attached to the conventional device [16] or a LRF placed in the test environment [26] was used to assess technically advanced outcomes such as walking distance or gait variability also when not using the RR. However, a reference, or any information on the psychometric quality of these methods, was missing in both studies.

In four studies including tailored assessment measures, test procedures appear to be non-standardized [16[2], 26, 34] or have been insufficiently described [28].

Evaluation of Physical and Physiological Demands

Four studies assessed subjects' physical and physiological demands with motorized RR during time-based performance-based measures (i.e. navigational trail, 10MWT) [26, 32] or during walking with standardized gait speed [19, 33]. In two studies, the exertion of force applied to steer the RR was measured using the force/torque sensors integrated on the robot's handles [19, 26]. One also reported on forces required to operate a conventional walker, but did not mention the method to capture these forces [26]. The other study additionally evaluated the oxygen consumption (VO_2) and metabolic cost of transport (metabolic cost per unit of mass and distance travelled) during robot-assisted gait using open-circuit respirometry [19]. In the remaining two studies, the muscle activity in the lower extremities was recorded by electromyography (EMG) [32, 33], and one also measured torso kinematics by a tri-axial accelerometer attached to the subject's back [32].

Subjective Evaluation Measures

Nineteen studies included measures to evaluate the subjects' experience with the RR [12, 16[1],17[1,2],18[1,3],19, 22-24, 26-30, 34, 35, 37, 38]. However, assessment instruments to perform such subjective evaluations varied widely in methodological quality. Nine studies documented solely non-specific comments of non-standardized surveys [16[1], 17[2], 18[1], 22, 24, 28, 29, 35, 37], three used standardized (dichotomous) questions [27, 30, 38], four used self-designed structured questionnaires, each with different multi-stage rating scales (e.g. 1 to 5, 0 to 100) [12, 17[1], 19, 34], two mentioned the use of questionnaires but did not provide detailed information on contents or a reference [18[3], 26], and one presented results of the subjective evaluation by response categories referring to different items but without mentioning the assessment instrument used for this purpose [23]. Most frequently used outcomes of standardized surveys included maneuverability [12, 17[1], 38], safety [12, 30, 38], and comfort [12, 19, 34].

Discussion

The aim of this systematic review was to summarize the methodology of evaluation studies of RRs with focus on the user perspective. Identified studies showed large heterogeneity in definitions of potential users, study population, study design, and assessment methods. We found major methodological shortcomings related to insufficient sample descriptions and sample sizes, lack of appropriate, standardized and validated assessment instruments, and lack of statistical analysis of study results. No generic methodology to evaluate RRs could be identified.

User Group Definitions

The majority of user group definitions seemed inadequate to guide a technical development of an AAL system. Generic, setting-specific, non-specific impairment-based or disease-oriented definitions do not relate to specific functional impairments of potential users, but cover users with a wide range of different functional abilities and requirements. The effective design of AAL systems in such heterogeneous user groups may be not feasible. The main goal of an AAL system should rather be to overcome or compensate for specific impaired functions. Clear impairment-related definitions are therefore mandatory to specifically tailor AAL developments for specific impairments of users and to ensure that innovative functionalities effectively address a user's needs. When such specific impairment-related definitions are additionally based on standardized and validated assessment methods with established cut-off values, a general comparability of developments and evaluations will be feasible.

Definitions according to impairment levels will in turn allow specifications such as risk stratification of potential users. With this, the user group will be further classified opening up the option to exclude persons with no or minor impairment, with no need for assistive devices, or with

advanced impairment or unacceptable risk exposure when using the device (triage). Another specification may focus on the main function of the specific device. For example, when an AAL system such as a RR basically supports gait performance, a specific definition based on standardized and validated gait assessment (e.g. 10MWT) will be superior compared to less specific definitions such as general functional scores (e.g. Barthel ADL Index).

As the user group of RRs may be old and multi-morbid persons, also highly prevalent age-associated impairments might be included in the definitions, depending on the specific functionalities or complexity of devices (e.g. inclusion of cognitive impairment with respect to navigation functions in disoriented persons).

Study Samples

Overall, sample sizes seemed rather limited to give a consistent picture of the user perspective. Surprisingly, the statistical analysis of documented data was not in the focus of studies as only a very limited number included such analyses (3 of 28) and none of these presented a sample size calculation as a prerequisite of statistical analysis.

A remarkable number of studies (10 of 28) evaluated RRs in persons who were not covered by the predefined user group, considerably limiting the user perspective of these studies. Study results with inadequate, convenient, or insufficiently described samples may not suffice to allow conclusions for persons with specific impairments which may represent the potential users of the RR. To ensure that RRs meet a user's needs and requirements and become successful in the market, it seems mandatory to involve the intended users at all stages of the design and evaluation process of such assistive robotic technologies [39-41].

Design of Studies

Observational Studies

The most heterogeneous group of studies covered observational studies that used solely descriptive data presentations without providing any reference or comparative values. Findings and conclusions of these studies were thus mainly based on the authors' subjective perception and appraisal. However, when using standardized and validated outcome measures with well-established cut-off values or other assistive mobility devices for comparison, such observations lose their merely subjective and study-specific nature and enable the objective appraisal of outcomes related to other studies or the documentation of an added value of the RR compared to other devices. From a user as well as a technical perspective, observational studies that descriptively presented non-classifiable or non-comparable outcomes therefore seem to have limited value.

Comparative Studies

The documentation and perception of an added value of the RR is of utmost importance for potential users. Innovative high-tech developments may be fascinating and mandatory for engineering research; however, they may also lead to rather complicated devices for everyday use, not easy to maneuver, too complex to operate, or too expensive to afford. A comparison of RRs with established, low-tech devices ('inter-device comparative study design') may therefore be useful to demonstrate to users the benefit of RR usage.

Comparisons may also be used for the evaluation of single functionalities to document the effect of a specified functionality (e.g. activated guidance system) or the progress of a new development stage. Such an 'intra-device comparative' study design allows a tailored assessment of the subjects' functional performances, physical and physiological demands, and user experience in specific assistance levels or development stages of the RR.

Frail, older persons may initially be intimidated by the robot's appearance in early stages of development (e.g. without casing, exposed hardware) which may in turn result in a more negative user perception before actually having used the RR. Subjective user evaluations, in a pre/ post-test study design, provide the opportunity to assess the subjects' initial impressions of the RR and whether there are potentially negative prejudices, which may, however, be overcome after actual use of the RR.

Independent of different types of comparative studies, such a study design should definitely include a statistical analysis to compare results which was however seldom used in the identified studies.

Interventional Studies

An interventional study design represents a new aspect in evaluation studies with strong focus on the user perspective. Newly developed RRs may not necessarily meet a user's acceptance or provide usability and efficiency when using them for the first time. Insufficient training opportunities or instruction prior to assessment measures may jeopardize study outcomes [42]. An adequate practice time therefore seems mandatory to prevent initial problems in operating the RR, and may further increase the impact on outcomes. Particularly when comparing RRs with a subject's own conventional assistive devices, brief instructions may not be sufficient, as subjects are already much more familiar and better trained with their own devices.

Overall, we identified a lack of studies investigating usability of RRs in natural environments with adequate long-term evaluation of habitual use. The development and evaluation of RRs seemed to occur rather in engineering laboratories than in clinical settings, as already reported for other robotic assistance systems (e.g. service robots, robotic exoskeleton) [43]. This may be explained by the fact that most of the identified studies evaluated research prototypes in rather early development stages, not yet ready for market launch. In such stages, it is important to manipulate specific variables of a prototype in order to investigate their effects precisely and to optimize technical functionalities accordingly [41]. Since laboratory evaluations also require less time and provide highly standardized conditions, a restricted experimental study design may have been favored. However, for the ultimate

goal of RRs to assist mobility of impaired persons in daily life, tests for habitual use seem to be mandatory documenting risk, experience-based perception of use, and quality of life with high relevance for users as well as caregivers.

Assessment Methods

Clinical Performance-Based Measures

Internationally well-established, clinical performance-based measures allow a worldwide comparability of results, but may be insufficient to cover the particular added value of specific robotic functionalities (e.g. obstacle avoidance, navigation assistance) as the outcome variables do not necessarily refer to the subjects' abilities potentially affected by the RR [42]. In addition, clinical assessment methods may be limited by subjective rating (POMA) or limited with respect to less detailed, unidimensional outcomes such as gait speed (4MWT, 10MWT) or task completion time (TUG). Augmenting such measures with technical assessment systems (e.g. video analysis system) allows a multidimensional analysis of the subjects' gait, including outcomes related to insecure gait or postural (in-)stability (e.g. width of base of support, double vs. single limb support) and reduction of falling risk as a main target of RRs.

Even established and validated assessment methods may have their limitations when inadequately used. Outcomes such as gait speed (4MWT) and task completion time (TUG) may be inappropriate when comparing a non-motorized, conventional device with a motorized RR with limited maximum speed. In such comparisons, a superior outcome for the low-tech device seems almost mandatory and may indicate an insufficient selection of a study outcome. The use of ADL scales (e.g. Barthel ADL Index) to evaluate effects of a robot-assisted ambulation training appears also inappropriate, since they include, if any, only very few sub-items targeting the subject's walking ability.

Another potential methodological pitfall may be related to performance-based outcome variables with ambiguous consequences: a motorized RR will improve gait speed in less impaired persons without substantial risk. However, improved performance may be traded off by a substantially higher risk of falling in more impaired persons.

Tailored Assessment Methods

The quality of an assessment strategy substantially depends on the appropriateness of methods with focus on the newly developed functionalities to document the added value of RRs. Clinical performance-based measures may be attractive because of their well-established psychometric properties; however, they have been developed for clinical purpose and may not cover new functionalities in innovative assistive technologies [42]. An assessment strategy specifically tailored to the specific functionality to be evaluated may help to achieve this goal. In RR, depending on the functionalities installed, a huge data flow created by the robot-integrated sensing technique already exists to control motoror cognitive assistance systems. Using this data flow for assessment purposes

may allow highly qualitative and quantitative tailored assessments exactly tuned to the newly developed functionality in order to document the added value of the RR. For example, when focusing on functionalities providing navigational assistance, the data flow from laser sensors, which is used to feed back the position of the RR, could be processed into a superior assessment of walking trajectories during a navigational task. When using such data for assessment purpose, it seems mandatory to examine or to provide sufficient information on the psychometric qualities of the robot-integrated sensor technique and the analysis method used to process raw data into the outcome variables. However, it appeared that only one study used an already established method for this approach [28]. Furthermore, to ensure reliable, reproducible, and comparable outcomes, the test procedure of tailored assessment measures has to be also clearly standardized.

Evaluation of Physical and Physiological Demands

Measures such as EMG, respirometry, accelerometry, or measurements of applied steering forces to the RR allow a detailed insight into relevant physical and physiological effects on objective parameters, which may be indicators for the subject's individual physical exertion (e.g. VO_2, muscle activity). However, some of these rather laborious measures (e.g. EMG, respirometry) seem less amenable for old and multi-morbid persons and may have therefore been used predominantly in studies including only young, healthy adults [19, 33]. To prevent overtaxing by test conditions, alternative methods to evaluate physical exertion are available which may increase amenability by standardized and validated subjective rating (e.g. [44]).

Subjective Evaluation Measures

In studies including subjective evaluation measures, a wide range of methods (e.g. non-specific comments, self-designed questionnaires) related to a variety of different aspects of the subject's experience with the RR was used which may considerably limit the comparability of outcomes. The overall lack of already established, validated questionnaires for the subjective evaluation of assistive technology (e.g. [45-47]) might be due to two reasons: (1) established questionnaires have been developed for a generic evaluation of a wide range of assistive technology devices but may be limited for evaluating specific functionalities of individual devices [45]; (2) some questionnaire items may also be inappropriate to evaluate prototypes after a short-term experiment in a restricted test scenario, covering aspects such as quality of life, usability in daily routine, durability, or services [45-47] whose assessment may only be feasible after habitual use of the devices over an extended period of time. However, the subjective evaluation measures used in the identified studies rather targeted the subject's actual experience directly after using the RR. This may explain the use of self-developed questionnaires including items already assessable after short-term use in an artificial setting (e.g. maneuverability, safety, ease of use). However, only when these questionnaires have been validated

before application and internationally established cut-off values are available, such assessment instruments guarantee high psychometric quality and allow comparability of study results [48].

Limitations

Only information available in the articles was evaluated in this review, although the authors may have used additional or more detailed methodology, not stated in articles. The fact that the evaluation of AAL prototypes may require elaborate and costly ethical application and study procedures ('Medical Product Act') may have prevented RRs to be tested in comprehensive studies with adequate sample sizes and the target user group as well as in natural environments with adequate long-term evaluation of habitual use. The role of clinical partners in AAL research projects may offer opportunities to solve such problems. Clinical partners may be able to provide specific impairment-based user group definitions, to recruit a satisfactory number of potentially adequate subjects, and to investigate the habitual use of AAL systems in natural environments.

Conclusions

Apart from the heterogeneity, methodological deficits in most of the identified studies became apparent. Recommendations for future evaluation studies include: (1) clear definition of target user group by valid, specific impairment-based criteria; (2) adequate selection of subjects with predefined inclusion criteria representative for potential users; (3) inclusion of other assistive mobility devices for comparison; (4) inclusion of the habitual use of advanced prototypes in evaluation rather than mere short-term, restricted, experimental test scenarios for single functionalities of prototypes not finalized for use in the target user group; (5) selection of established, standardized, and validated assessment methods; (6) implementation of a specifically tailored assessment strategy, focusing on specific functionalities of the RR, and (7) statistical analysis of study results. These recommendations, given for RRs, may also apply in general for the development and evaluation of AAL systems with focus on the user perspective.

Acknowledgments

This systematic review was carried out within the 7th Framework Program of the European Union, ICT Challenge 2, Cognitive Systems and Robotics, contract "EU-FP7-ICT-2011-9 2.1 - 600769 - MOBOT: Intelligent Active MObility Assistance RoBOT Integrating Multimodal Sensory Processing, Proactive Autonomy and Adaptive Interaction". The authors are solely responsible for the content of this manuscript which does not necessarily represent the views of the European Union. We thank Denis Schrank (DS) for assistance with data extraction and Alex Smith for English editing of the manuscript.

Conflict of Interest

The authors have no conflict of interests to declare.

References

Asterisks indicated articles included in the synthesis

1 Hirvensalo M, Rantanen T, Heikkinen E: Mobility difficulties and physical activity as predictors of mortality and loss of independence in the community-living older population. J Am Geriatr Soc 2000;48:493-498.

2 Groessl EJ, Kaplan RM, Rejeski WJ, Katula JA, King AC, Frierson G, Glynn NW, Hsu FC, Walkup M, Pahor M: Health-related quality of life in older adults at risk for disability. Am J Prev Med 2007;33:214-218.

3 Chodzko-Zajko WJ, Proctor DN, Fiatarone Singh MA, Minson CT, Nigg CR, Salem GJ, Skinner JS: American college of sports medicine position stand. Exercise and physical activity for older adults. Med Sci Sports Exerc 2009;41:1510-1530.

4 McAuley E, Konopack JF, Motl RW, Morris KS, Doerksen SE, Rosengren KR: Physical activity and quality of life in older adults: Influence of health status and self-efficacy. Ann Behav Med 2006;31:99-103.

5 Chaudhry SI, McAvay G, Ning Y, Allore HG, Newman AB, Gill TM: Geriatric impairments and disability: The cardiovascular health study. J Am Geriatr Soc 2010;58:1686-1692.

6 Rapp K, Becker C, Cameron ID, Konig HH, Buchele G: Epidemiology of falls in residential aged care: Analysis of more than 70,000 falls from residents of bavarian nursing homes. J Am Med Dir Assoc 2012;13:187 e181-186.

7 Raina P, Wong M, Massfeller H: The relationship between sensory impairment and functional independence among elderly. BMC Geriatr 2004;4:3.

8 Pai MC, Jacobs WJ: Topographical disorientation in community-residing patients with alzheimer's disease. Int J Geriatr Psychiatry 2004;19:250-255.

9 Bateni H, Maki BE: Assistive devices for balance and mobility: Benefits, demands, and adverse consequences. Arch Phys Med Rehabil 2005;86:134-145.

10 Graafmans WC, Lips P, Wijlhuizen GJ, Pluijm SM, Bouter LM: Daily physical activity and the use of a walking aid in relation to falls in elderly people in a residential care setting. Z Gerontol Geriatr 2003;36:23-28.

11 Salminen AL, Brandt A, Samuelsson K, Toytari O, Malmivaara A: Mobility devices to promote activity and participation: A systematic review. J Rehabil Med 2009;41:697-706.

12* Frizera-Neto A, Ceres R, Rocon E, Pons J: Empowering and assisting natural human mobility: The simbiosis walker. Int J Adv Robot Syst 2011;8:34-50.

13 European Commission: Guidelines on medical devices. Clinical evaluation: A guide for manufacturers and notified bodies. http://ec.europa.eu/health/medical-devices/files/meddev/2_7_1rev_3_en.pdf, 2009.

14 Schulz R, Wahl HW, Matthews JT, De Vito Dabbs A, Beach SR, Czaja SJ: Advancing the aging and technology agenda in gerontology. The Gerontologist 2015;55:724-734.

15 Higgins J, Green S (eds): Cochrane handbook for systematic reviews of interventions 5.1.0, Chichester, UK, John Wiley & Sons, 2011.

16* Mou W-H, Ming-Fang C, Chien-Ke L, Yuan-Han H, Shih-Huan T, Li-Chen F: Context-aware assisted interactive robotic walker for parkinson's disease patients: IEEE/RSJ Int Conf Intell Robot Syst. Vilamoura, Portugal, 2012, pp 329-334.

17* Yu H, Spenko M, Dubowsky S: An adaptive shared control system for an intelligent mobility aid for the elderly. Autonomous Robots 2003;15:53-66.

18* Bühler C, Heck H, Nedza J, Wallbruch R: Evaluation of the mobil walking & lifting aid; in Marincek C, Bühler C, Knops H, Andrich R (eds): Assistive technology added value to the quality of life. Amsterdam, ISO Press, 2001, vol 10, pp 210-215.

19* Grondin SL, Li Q: Intelligent control of a smart walker and its performance evaluation: IEEE Int Conf Rehabil Robot. Seattle, USA, 2013, pp 6650346.

20* Taghvaei S, Hirata Y, Kosuge K: Vision-based human state estimation to control an intelligent passive walker: IEEE/SICE Int Symp Syst Integr. Sendai, Japan, 2010, pp 146-151.

21* Rumeau P, Pasqui V, Vigourou N: A generic method for the assessment of smart walkers *[full paper]*. Gerontechnology 2012;11. http://gerontechnology.info/index.php/journal/article/download-SuppFile/1833/gt.2012.11.02.480.422.

22* Wasson G, Sheth P, Huang C, Alwan M: Intelligent mobility aids for the elderly; in Alwan M, Felder R (eds): Eldercare technology for clinical practitioners. Totowa, NJ, Humana Press, 2008, pp 53-76.

23* Xu W, Jian H, YongJi W, Hong C: Study of reinforcement learning based shared control of walking-aid robot: IEEE/SICE Int Symp Syst Integr. Kobe, Japan, 2013, pp 282-287.

24* Ko C-H, Yi-Hung H, Yao-Tse C, Agrawal SK, Kuu-Young Y: Guidance and obstacle avoidance of passive robot walking helper based on receding horizon control: IEEE Int Conf Autom Sci Eng. Taipei, Taiwan, 2014, pp 1032-1037.

25* Hirata Y, Oscar C, Jr., Asami H, Kosuge K: Human-adaptive motion control of active and passive type walking support system: IEEE Int Work Adv Robot Soc Imp. Nagoya, Japan, 2005, pp 139-144.
26* Graf B: An adaptive guidance system for robotic walking aids. J Comput Inf Technol 2009;17:109-120.
27* Kulyukin V, Kutiyanawala A, LoPresti E, Matthews J, Simpson R: Iwalker: Toward a rollator-mounted wayfinding system for the elderly: IEEE Int Conf Radio Freq Identif. Las Vegas, USA, 2008, pp 303-311.
28* Morris A, Donamukkala R, Kapuria A, Steinfeld A, Matthews JT, Dunbar-Jacob J, Thrun S: A robotic walker that provides guidance: IEEE Int Conf Robot Autom. Pittsburgh, USA, 2003, 1, pp 25-30.
29* Glover J, Holstius D, Manojlovich M, Montgomery K, Powers A, Wu J, Kiesler S, Matthews J, Thrun S: A robotically-augmented walker for older adults. Technical report cmu-cs03-170. Computer Science Department, Carnegie Mellon University, Pittsburgh, US, 2003.
30* Chugo D, Asawa T, Kitamura T, Jia S, Takase K: A motion control of a robotic walker for continuous assistance during standing, walking and seating operation: IEEE/RSJ Int Conf Intell Robot Syst. St. Louis, USA, 2009, pp 4487-4492.
31* Kikuchi T, Tanaka T, Tanida S, Kobayashi K, Mitobe K: Basic study on gait rehabilitation system with intelligently controllable walker (i-walker): IEEE Int Conf Robot Biomim. Tianjin, China, 2010, pp 277-282.
32* Tamura T, Sekine M, Kuno H, Fujie M, Mori A, Andoh K: Evaluation of walkers for elderly people: IEEE Int Conf Eng Med Biol Soc. Istanbul, Turkey, 2001, 2, pp 1391-1392.
33* Jang J, Yu S, Han J, Han C: Development of a walking assistive service robot for rehabilitation of elderly people; in Takahashi Y (ed): Service robot applications, InTech, 2008, pp 139-158. http://www.intechopen.com/books/service_robot_applications/development_of_a_walking_assistive_service_robot_for_rehabilitation_of_elderly_people.
34* Rentschler AJ, Simpson R, Cooper RA, Boninger ML: Clinical evaluation of guido robotic walker. J Rehabil Res Dev 2008;45:1281-1293.
35* Mori H, Kotani S, Saneyoshi K, Sanada H, Kobayashi Y, Mototsune A: Research and development project for practical use of robotic travel aid for the visually impaired: Int Conf Disab Virt Real Ass Tech. Veszprém, Hungary, 2002, pp 123-130.
36* Annicchiarico R: Enhancing service delivering, improving quality of life, preserving independence through assistive technology. Stud Health Technol Inform 2012;180:14-18.
37* Lee J, Kyung K, Jongbae K, Won-Kyung S: Essential feedback on first prototypes of smart mobile walker and upper extremity assistive robot: Int Conf Ubiq Robot Amb Intell. Daejeon, Korea, 2012, pp 65-66.
38* Lee G, Ohnuma T, Nak Young C, Soon-Geul L: Walking intent-based movement control for jaist active robotic walker. IEEE Trans Syst Man Cyber 2014;44:665-672.
39 Tsui K, Feil-Seifer D, Matarić M, Yanco H: Performance evaluation methods for assistive robotic technology; in Madhavan R, Tunstel E, Messina E (eds): Performance evaluation and benchmarking of intelligent systems. New York, US, Springer, 2009, pp 41-66.
40 Scherer M: The change in emphasis from people to person: Introduction to the special issue on assistive technology. Disabil Rehabil 2002;24:1-4.
41 Salminen A, Petrie H: Evaluating assistive technology prototypes: Laboratory or real life contexts?; in Placencia Porrero I, Ballabio E (eds): Improving the quality of life for the european citizen: Technology for inclusive design and equality. Amsterdam, NL, IOS Press, 1998, pp 414-419.
42 Fuhrer M: Assistive technology outcomes challenges met and yet unmet. Am J Phys Med Rehabil 2001;80:529-535.
43 Pearce A, Adair B, Miller K, Ozanne E, Said C, Santamaria N, Morris M: Robotics to enable older adults to remain living at home. J Aging Res 2012;2012:538169.
44 Borg G: Psychophysical bases of perceived exertion. Med Sci Sports Exerc 1982;14:377-381.
45 Demers L, Weiss-Lambrou R, Ska B: Development of the quebec user evaluation of satisfaction with assistive technology (quest). Assist Technol 1996;8:3-13.
46 Day H, Jutai J: Measuring the psychosocial impact of assistive devices: The piads. Can J Psychol 1996;9:159-168.
47 Scherer MJ, Cushman L: Measuring subjective quality of life following spinal cord injury: A validation study of the assistive technology device predisposition assessment. Disabil Rehabil 2001;23:387-393.
48 Bartneck C, Kulić D, Croft E, Zoghbi S: Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots. Int J of Soc Robotics 2009;1:71-81.

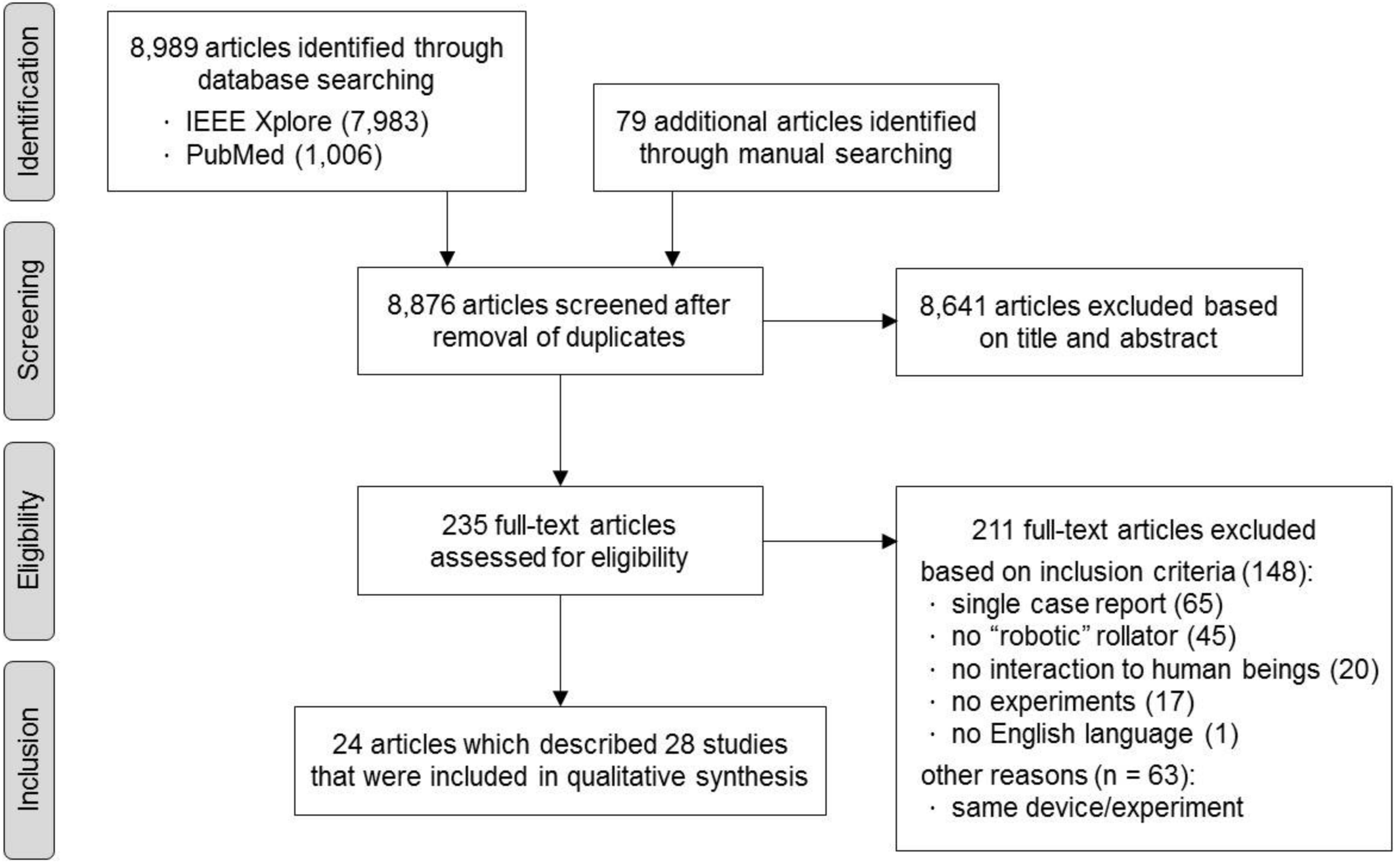
Identification
Screening
Eligibility
Inclusion
8,989 articles identified through database searching
· IEEE Xplore (7,983)
· PubMed (1,006)
79 additional articles identified through manual searching
8,876 articles screened after removal of duplicates
8,641 articles excluded based on title and abstract
235 full-text articles assessed for eligibility
211 full-text articles excluded
based on inclusion criteria (148):
· single case report (65)
· no "robotic" rollator (45)
· no interaction to human beings (20)
· no experiments (17)
· no English language (1)
other reasons (n = 63):
· same device/experiment
24 articles which described 28 studies that were included in qualitative synthesis

Table 1. Overview of the search term used in PubMed

Assistive mobility device	Robotic functionality	Gait/mobility support	Evaluation measure
#1 'robotics'[Mesh]	#14 'electric power supplies'[Mesh]	#23 'gait'[Mesh]	#32 'evaluation studies as topic'[Mesh]
#2 'walkers'[Mesh]	#15 robot*[tiab]	#24 'Walking'[Mesh]	#33 'Technology Assessment, Biomedical'[Mesh]
#3 'self-help devices'[Mesh]	#16 smart[tiab]	#25 'Dependent Ambulation'[Mesh]	#34 evaluat*[tiab]
#4 'biomedical technology'[Mesh]	#17 intelligent[tiab]	#26 gait[tiab]	#35 assess*[tiab]
#5 robot*[tiab]	#18 power*[tiab]	#27 walk*[tiab]	#36 measur*[tiab]
#6 rollator*[tiab]	#19 electric[tiab]	#28 ambulant*[tiab]	#37 trial*[tiab]
#7 mobile platform*[tiab]	#20 motorized[tiab]	#29 mobility[tiab]	#38 experiment*[tiab]
#8 mobility aid*[tiab]	#21 motorised[tiab]	#30 OR (#23-#29)	#39 test*[tiab]
#9 mobility device*[tiab]	#22 OR (#14-#22)	#31 (#13 AND #22 AND #30)	#40 clinical[tiab]
#10 assistive device*[tiab]	#23 (#13 AND #22)		#41 OR (#32-#40)
#11 assistive system*[tiab]			#42 (#13 AND #22 AND #30 AND 41)
#12 walking aid*[tiab]			
#13 OR (#1-#12)			

Table 2. Study characteristics and assessment methods of the 28 studies included in this systematic review

Name of device Authors [Ref. No.]	User group definition	Study sample	Study object	Study design	Assessment methods Type: outcome measurement: outcome variable
Context-aware Assisted Interactive RObotic Walker (CAIROW) Mou et al. 2012 [16[1,2]]	PD patients	Study 1 n = 6 (F = n/a) Age: n/a PD patients of senior care unit; mHY, stage range 1.5-3	UP	IV: repeated assessment on six consecutive days	TAM: gait analysis on straight walking path[b]; CAIROW gait analysis system (based on LRF)[b]: SD of gait speed/step length; expert rating of gait[d]: number of abnormal gait patterns (festinating gait, freezing of gait)
			UE	OB	SEM: user comments[c] after gait analysis
		Study 2 n = 7 (F = n/a) Mean age: 86 yrs PD patients of senior care unit; mHY, stage range 1-3	UP	Inter-DC: walking with CAIROW vs. normal walking (with own/ with-out assistive device)	TAM: gait analysis on walking path with obstacles, people randomly passing by, up- and down-going slopes, short section for backward walking[d]; CAIROW gait analysis system[b] or LRF[c] placed in test environment when normal walking: SD of gait speed/step length; expert rating of gait[d]: number of abnormal gait patterns
			TC (gait analysis system)	EC: gait analysis system vs. expert rating	TEM (see original article for details)
Care-O-bot II Graf 2009 [26]	Elderly people in home environment	n = 6 (F = 5) Age range: 86-92 yrs Inhabitants of an old people's residence using mobility aids in daily life	UP, PD	Inter-/intra-DC: target mode (robot-determined motion control) vs. direct control mode (user-deter-mined motion control) vs. conventional walker	TAM: navigation trail in old people's residence with a ramp, tables, and people randomly passing by[d]; robot's guidance system[c], bicycle speedometer[c] mounted on conventional walker: walking time, number of collisions, maximum speed, walking distance, distance to obstacle PHY: force/torque sensors[c] in robot's handles, force measurement when using conventional walker not reported[d]: pushing force
			UP	OB	TAM: navigation trail in old people's residence with transition between ground floor and 1st floor, a ramp, tables, people randomly passing by[d]: achievement of target
			UE	OB	SEM: questionnaire[b] after navigation trail: n/a
Chugo group walker Chugo et al. 2009 [30]	Elderly people in need for nursing in daily routine	n = 7 (F = n/a) Age: ≥ 67 yrs People in need of long-term care at level I or II in Japanese Long-term Insurance System	UE	Inter-/intra-DC: STS transfer without assistance vs. with previous/novel STS assistance system	SEM: questionnaire[b] after STS transfer : ease of standing up, fear of falling (1= inferior, 3 = same, 5 = better feeling compared to STS transfer without assistance)
CO-Operative Locomotion Aide (COOL-Aide) Wasson et al. 2008 [22]	Elderly people	n = 12 (F = 2) Mean age (SD): 36.8 (18.1) yrs Healthy subjects (n = 8), subjects with disorders affecting mobility (cerebral palsy, familial torsion dystonia) (n = 8) note: (1) total sample, (2) - (5) subsample: only healthy subjects	TC (guidance, user intent detection and obstacle avoidance system)	OB	TEM (see original article for details)
			TC (obstacle avoidance system with vs. without stability preservation)	Intra-DC: standard vs. stability-preserved obstacle avoidance	TEM (see original article for details)
			UE	OB	SEM: user comments[d] after performing a set of short obstacle courses
Gait Rehabilitation Service Robot (GRSR) Jang et al. 2008 [33]	Disabled or elderly with mobility problems or paralysis; weighing up to 75 kg	n = 2 (F = 0) Mean age (SD): 28.5 (2.1) yrs Ordinary adult males	TC (guidance system)	OB	TEM (see original article for details)
			PD	Intra-DC: 40/20 % body weight support vs. full body weight	PHY: EMG[a] during straight walking with standardized gait speed of 0.2 m/s: muscle activity of lower extremities (EMG signal) (quadriceps, hamstrings, gastrocnemius, tibialis anterior)

Table 2. (continued)

Name of device Authors [Ref. No.]	User group definition	Study sample	Study object	Study design	Assessment methods Type: outcome measurement: outcome variable
Guido Rentschler et al. 2008 [34]	Frail elderly people with visual impairment	n = 17 (F = n/a) Mean age (SD): 85.3 (7.0) yrs Residents of a supportive living facility/nursing home with visual impairment due to macular degeneration, cataract, glaucoma or other reasons; mean time (SD) since onset of visual impairment: 20.4 (13.0) yrs; ambulatory (≥ 20 min within 90 min period) with limited assistance	UP	Inter-/intra-DC: Guido vs. conventional assistive mobility device or normal walking (with own/ no assistive device); automatic (user-determined motion control) vs. manual mode (shared user-robot motion control)	TAM: obstacle course with randomly placed obstacles before each trial[d]: walking time, number of obstacle/wall collisions, number of reorientations
			UE	PPC: before and after 3 trials	SEM: Subjective Mobility Questionnaire[b] after obstacle course: appearance, ease of use, usefulness in living enviroment, embarrassment (1 = best score; 5 = worst score)
Hitachi walker Tamura et al. 2001 [32]	Elderly people who have difficulty walking	n = 6 (F = n/a) Mean age (SD): 82 (7.9) yrs Subjects ambulatory with supervision (n = 4), subjects in need for walking assistance (n = 2)	UP PD	Inter-DC: Hitachi vs. caster vs. conventional walker; robot vs. parallel bars	CPM:10MWT[a]: gait speed PHY: EMG[c], tri-axial accelerometer[c] during non-standardized gait speed (10MWT): muscle activity (EMG signal), trunk acceleration
HUST walking-aid robot Xu et al. 2013 [23]	Elderly or disabled people	n = 3 (F = n/a) Age: n/a Volunteering subjects with/ without experience using robot; one subject with restricted knee joint to imitate lower limb disorders	TC (motion control system)	PPC: autonomous learning process of HUST in motion behavior over a series of trials	TEM (see original article for details)
			UE	OB	SEM: subjective evaluation after completing a series of obstacle courses, assessment measure not reported[d]: flexibility, comfort, maneuverability, obstacle avoidance
i-Go Ko et al. 2014 [24]	Elderly people	n = 3 (F = n/a) Age: "in their twenties"	TC (guidance system) UE	OB	TEM (see original article for details) SEM: user comments[d] after completing an S-shaped walking path
Intelligent Mobility Platform (IMP) Glover 2003 [29]	Older adults (primarily without major visual or cognitive impairment)	n = 6 (F = n/a) Age: n/a Residents of a care facility with/ without need for walker	UE	OB	SEM: user comments[d] after presentation and informal testing of the robot
iWalker Kulyukin et al. 2008 [27]	Persons with stroke, early- to mid-stage AD, traumatic brain injury, macular degeneration, cataracts, visual impairment; primarily in nursing and assisted living homes	n = 4 (F = n/a) age: n/a Clients of in-home supportive service currently using cane, walker or bot, with history of way finding problems; MMSE, mean score (SD): 26 (3.6)	UP	Inter-DC: iWalker vs. conventional device (cane/walker) accompanied by researcher	TAM: several navigation trails[b] : walking time, number of navigational mistakes
			UE	Intra-DC: map-based (+ auditory cues) vs. text-and-arrow-based (+ auditory cues) user interface design	SEM: dichotomous question[b]: choice of user interface; user comments[d]

Table 2. (continued)

Name of device Authors [Ref. No.]	User group definition	Study sample	Study object	Study design	Assessment methods Type: outcome measurement: outcome variable
i-Walker (EU) Annicchiarico 2012 [36]	Post-stroke patients with hemiparesis	n = 20 (F = 11) Mean age: 59.9 yrs Acute hemiparetic stroke patients (event < 1 yrs) receiving rehabilitation treatment; MMSE score ≥ 20; CNS upper & lower limb > 0	UP	IV (RCT): robot-assisted ambulatory training (EG) vs. in parallel bars (CG) (4 weeks, 5x a week)	CPM: POMA[a]: total score; 6mWT[a]: walking distance; 10MWT[a]: gait speed ADL screening: Barthel ADL Index[a]: score
i-Walker (Japan) Kikuchi et al. 2010 [31]	Patients with imbalanced motor/sensory functions (e.g. hemiplegic patients), difficulties in smooth walking	n = 6 (F = 2) Mean age (SD): 88.7 (6.1) yrs Residents of elder care facility with wheelchair due to loss of vision/muscle strength which occasionally train walking with forearm caster walker; chronic disease: stroke, dementia, muscle atrophy, high blood pressure, heart failure, AD, cataract, PD	UP	Intra-DC: passive vs. active robot motion control system	TAM: walking path with obstacles[b], robot-integrated web camera[c]: deviations from a path marked on the floor, number of collisions
JAIST Active Robotic Walker (JARoW) Lee et al. 2014 [38]	Elderly people with certain level of ambulatory capability (FAC score 4-5)	n = 5 (F = 4) Age range: 75-84 yrs Subjects using traditional walkers in daily routine	UE	Inter-DC: JARoW vs. conventional walker	SEM: questionnaire[b] after walking around for 10 min: ease of walking, safety, maneuverability, suggestions for improvements
MOBIL walking & lifting aid Bühler et al. 2001 [18[1]]	Frail, elderly and walking disabled people	<u>Study 1</u> n ≥ 2 (F = n/a) Age: n/a Selected users, technical and rehabilitation experts	TC (overall system functionality) UE	OB	TEM (see original article for details) SEM: user/expert ratings, comments and interviews[d]
MOBIL test bed [18[2]]	Frail, elderly and walking disabled people	<u>Study 2</u> n ≥ 2 (F = n/a) Age: n/a Rehabilitation engineers, walking impaired persons	TC (overall system functionality)	OB	TEM (see original article for details)
MOBIL walking & lifting aid, MOBIL test bed [18[3]]	Frail, elderly and walking disabled people	<u>Study 3</u> n ≥ 2 (F = n/a) Age: n/a Community-dwelling people, institutionalized elderly disabled people, care staff	UE	OB	SEM: questionnaire[b] after demonstration, video presentations, practical trials: n/a

Table 2. (continued)

Name of device Authors [Ref. No.]	User group definition	Study sample	Study object	Study design	Assessment methods Type: outcome measurement: outcome variable
Nomad XR 4000 Morris et al. 2003 [28]	Frail older people with cognitive impairment	n = 4 (F = n/a) Age: n/a Residents of a retirement facility	UP	Intra-DC: passive (no navigational assistance) vs. active (with navigational assistance) vs. forced mode (full robot motion control)	TAM: navigational trail[d]; robot's navigation system (based on LRF, 'Monte Carlo localization')[a]: deviation from optimal path
			UE	OB	SEM: user comments after navigational trails[d]
Personal Aid for Mobility and Monitoring (PAMM SmartWalker) Yu et al. 2003 [17[1,2]]	Independently living or institutionalized elderly people with mobility difficulties due to physical frailty and/or disorientation due to age and sickness	Study 1 n = 8 (F = n/a) Age: n/a Elderly residents of assisted living facility with mobility aid	UE	OB	SEM: questionnaire[b] after free driving at facility: ease of control, going straight, turning, heaviness, support, satisfaction (1 = worst score, 5 = best score)
		Study 2 n = 8 (F = 5) Age range: 84-95 yrs Elderly residents of assisted living facility with need for walkers	UP	Intra-DC: full robot motion control vs. adaptive shared user-robot motion control vs. without any motion control	TAM: wall-limited walking path through assisted living facility[d]; robot's vision-based localization system (based on charged-coupled device camera)[b]: deviations from robot-generated, pre-planned path, distance to wall
			UE	OB	SEM: user comments[d]
Robotic Mobility Platform (RMP) Grondin & Qinggou 2013 [19]	n/a	n = 10 (F = 5) Mean age (SD): 24.6 (3.0) Subjects without previous/current gait-related injuries and without experience in using rollators or robotic walkers	UP, PD	Intra-DC: novel vs. previous motion control system	TAM: walking with targeted velocity of 1 m/s through a circular path in low-traffic hallways[b]; technical outcome measurement not reported[d]: mean and SD of robot velocity; PHY: force/torque sensor[a] under robot's left handle: pushing force
			PD	Inter-/intra-DC: novel vs. previous motion control system vs. conventional rollator vs. no assistive device	PHY: walking with targeted velocity of 1 m/s through the circular path[c] (use of a Hall effect sensor mounted on the conventional rollator to display target velocity); respirometry[a]: metabolic cost of transport, oxygen consumption
			UE	Intra-DC: novel vs. previous motion control system	SEM: questionnaire[b]: comfort, intuition, speed control, exertion, overall experience (0 = worst score, 5 = best score)
robuWALKER Rumeau et al. 2012 [21]	elderly people	n = 8 (F = 5) Mean age (SD): 82.6 (8.7) yrs Healthy elderly (n = 4): 4MWT < 4s, TUG < 13s, MMSE score ≥ 26; elderly patients with motor & cognitive impairment (n = 4): 4MWT > 4s, TUG > 13s, MMSE mean score (SD): 20 (3.5); all subjects without experience in using walking frames	UP	Inter-DC: robuWalker vs. conventional walker	CPM: 4MWT[a]: gait speed, modified TUG[a]: completion time; gait analysis by video recordings[c] during 4MWT and TUG: step time, double support time

Table 2. (continued)

Name of device Authors [Ref. No.]	User group definition	Study sample	Study object	Study design	Assessment methods Type: outcome measurement: outcome variable
Robotic Travel Aid (RoTA) Mori et al. [35]	visually impaired community-dwelling people, hospital patients, or residents of senior homes loss of ability to walk with mobility aids for the blinds	n > 60 (F = n/a) Age: n/a Blind and weak-sighted elderly people	UE	OB	SEM: user comments[d] after walking course
RT Walker Taghvaei et al. 2010 [20]	n/a	n = 2 (F = n/a) Age: n/a	TC (motion control system)	OB	TEM (see original article for details)
SIMBIOSIS Walker Frizera-Neto et al. 2011 [12]	SCI patients mainly using wheelchair, but usually able to walk for short periods of time with assistance of device, WISCI II = 16	n = 8 (F = n/a) Age: n/a Subjects with preserved cognitive functions; ability to (1) maintain standing position, (2) walk 10 m without assistance of another person and with or without support of a mobility aid, and (3) to grasp; WISCI II, mean score (SD): 15.9 (2.9)	TC (user intent detection system) UE	OB OB	TEM (see original article for details) SEM: questionnaire[b] after completing U-shaped walking path: maneuverability, safety, posture & comfort (0 = worst score, 100 = best score)
Smart Mobile Walker (SMW) Lee et al. 2012 [37]	elderly people, people with hemiplegia, people with incomplete SCI	n ≥ 2 (F = n/a) Age: n/a Stroke patients, SCI patients, clinical experts	UE	OB	SEM: user comments/interviews[d] after demonstrations
Walking Helper Hirata et al. 2005 [25]	elderly people, disabled people	n = 8 (F = n/a) Age: n/a	UP	Intra-DC: novel vs. traditional motion control system	TAM: following S-shaped walking path[b] (marked on the floor); technical outcome measurement not reported[d]: deviation from path marked on the floor

Abbreviations: PD = Parkinson's disease; F = females; n/a = not available; mHY = modified Hoehn and Yahr Scale; UP = User performance; UE = User experience; IV = interventional; OB = observational; TAM = tailored assessment measure; LRF = laser rangefinder; SD = standard deviation; SEM = subjective evaluation measure; TC = technical capability; inter-DC = inter-device comparative; EC = comparison with external criterion measure; TEM = technical evaluation measure; PD = physical/physiological demands; intra-DC = intra-device comparative; PHY = evaluation of physical or physiological demands; STS = sit-to-stand; EMG = electromyography; PPC = pretest-posttest comparative; CPM = clinical performance-based measure; 10MWT = 10-meter walk test; MMSE = Mini-Mental State Examination; CNS = Canadian Neurological Scale; RCT = randomized controlled intervention trial; EG = experimental group; CG = control group; POMA = Performance Oriented Mobility Assessment; 6mWT = 6-minute walk test; ADL = activities of daily living; AD = Alzheimer's disease; FAC = Functional Ambulation Classification; TUG = Timed Up and Go; 4MWT = 4-meter walk test; WISCI = Walking Index for Spinal Cord Injury; SCI = Spinal Cord Injury.

[a] established, standardized and validated assessment test or outcome measurement.

[b] standardized, but not validated test procedure or outcome measurement.

[c] potentially an established outcome measurement, but no reference given.

[d] non-standardized or unclear test procedure or outcome measurement.