Accepted Manuscript

What adherence measures should be used in trials of home-based rehabilitation interventions? A systematic review of the validity, reliability and acceptability of measures

Rachael Frost, PhD, Sara Levati, MSc, Doreen McClurg, PhD, Marian Brady, PhD, Brian Williams, PhD

PII: S0003-9993(16)31135-2

DOI: 10.1016/j.apmr.2016.08.482

Reference: YAPMR 56685

To appear in: *ARCHIVES OF PHYSICAL MEDICINE AND REHABILITATION*

Received Date: 29 April 2016

Revised Date: 26 August 2016

Accepted Date: 31 August 2016

Please cite this article as: Frost R, Levati S, McClurg D, Brady M, Williams B, What adherence measures should be used in trials of home-based rehabilitation interventions? A systematic review of the validity, reliability and acceptability of measures, *ARCHIVES OF PHYSICAL MEDICINE AND REHABILITATION* (2016), doi: 10.1016/j.apmr.2016.08.482.

What adherence measures should be used in trials of home-based rehabilitation interventions? A systematic review of the validity, reliability and acceptability of measures

Rachael Frost, PhD, NMAHP Research Unit, Glasgow Caledonian University

Sara Levati, MSc, NMAHP Research Unit, Glasgow Caledonian University

Doreen McClurg, PhD NMAHP Research Unit, Glasgow Caledonian University

Marian Brady, PhD, NMAHP Research Unit, Glasgow Caledonian University

Brian Williams, PhD, School of Health & Social Care, Edinburgh Napier University.

Corresponding author: Rachael Frost, University College London, Research Department of Primary Care and Population Health, Royal Free Campus, Rowland Hill Street, London, NW3 2PF, Tel: 0207 8302881, email: rachael.frost@ucl.ac.uk

Keywords: Patient compliance, reliability and validity, rehabilitation

Conflict of interest: The authors declare they have no conflicts of interest.

Acknowledgements: The authors would like to thank Dr Pauline Campbell, Heather Strachan, Dr Katie Thomson and Dr Barbara Farquharson for undertaking secondary data extraction and quality assessment for this review.

Acknowledgement of financial support: Glasgow Caledonian University PhD studentship

PROSPERO ID: CRD42013004084.

What adherence measures should be used in trials of home-based rehabilitation interventions? A systematic review of the validity, reliability and acceptability of measures

Keywords: Patient compliance, reliability validity, rehabilitation

Supplementary files: Medline search strategy, table of included studies, reference list of included studies

Abbreviations:

COSMIN	COnsensus-based Standards for the selection of health Measurement INstruments.
MPA	measurement property assessment
OT	occupational therapist
PT	physiotherapist
RCT	randomised controlled trial
SLT	speech and language therapist

Abstract

Objective: To systematically review methods for measuring adherence used in home-based rehabilitation trials, and evaluate their validity, reliability and acceptability.

Data sources: Phase 1: We searched CENTRAL, EED and HTA (Jan 2000-April 2013) to identify adherence measures used in randomised controlled trials of allied health professional home-based rehabilitation interventions. Phase 2: We searched Medline, Embase, CINAHL, AMED, PsycINFO, CENTRAL, ProQuest and Web of Science (inception-April 2015) for measurement property assessments (MPAs) for each measure.

Study selection: Studies assessing the validity, reliability or acceptability of adherence measures

Data extraction: Two reviewers independently extracted data on participant and measure characteristics, measurement properties evaluated, evaluation methods and outcome statistics and assessed study quality using the COSMIN checklist.

Data synthesis: Phase 1: We included 8 adherence measures (n=56 trials). Phase 2: From 222 MPAs identified in 109 studies, 22 high quality MPAs were narratively synthesised. Low quality studies were used as supporting data. StepWatch Activity Monitor validly and acceptably measured short term step count adherence. The Problematic Experiences of Therapy Scale validly and reliably assessed adherence to vestibular rehabilitation exercises. Adherence diaries had moderately-high validity and acceptability across limited populations. The Borg 6-20 scale, Bassett & Prapavessis' scale and the Yamax CW series had insufficient validity. Low quality evidence supported use of the Joint Protection Behaviour Assessment Polar A1 series heart monitors were considered acceptable by one study.

Conclusions: Current rehabilitation adherence measures are limited. Some possess promising validity and acceptability for certain parameters of adherence, situations and populations and should be used in these situations. Rigorous evaluation of adherence measures in a broader range of populations is needed.

Keywords: Patient compliance, reliability and validity, rehabilitation

PROSPERO ID: CRD42013004084.

Adherence is the extent to which a person's behaviour coincides with agreed clinical recommendations.[1] Documenting participant adherence in clinical practice is necessary to monitor the patient's progress and help determine whether improvements (or lack of) is to be attributed to non/adherence or ineffectiveness of the prescribed therapy. Similarly, within clinical trials it is essential to measure adherence to answer the same question of attribution at a larger level, assess the impact of the intervention dose upon effectiveness, and to assist in identifying non-adherent patient subgroups.[2] This is particularly vital within home-based rehabilitation interventions, where therapists expect greater independent patient engagement to prescribed therapeutic activities between formal therapy sessions. Prescribed home activities, e.g. home exercises, are an essential component within many allied health professional rehabilitation therapies, such as physiotherapy or occupational therapy. This reflects the increasing focus on functionally relevant rehabilitation, early supported discharge,[3] maximising patient engagement with rehabilitation[4] and self-management.[5] Documenting adherence within clinical trials and practice can also provide an indication of the acceptability of an intervention to patients.

Given its vital role, the choice of adherence measurement method(s) should be guided by rigorous evidence of their respective measurement properties. Three prior systematic reviews have been undertaken in this area, focussing on: self-report adherence measures in home-based rehabilitation;[6] patient or provider adherence questionnaires in physiotherapy[7] and measures assessing adherence to non-pharmacological self-management in musculoskeletal conditions.[8] All concluded that the available trials included largely self-developed questionnaires that lacked sufficient evidence of measurement properties.[6,8,9] A broader perspective was therefore required to encompass other methods in addition to questionnaires, based on methods currently used in clinical trials. Consequently, this review aimed 1) to

identify adherence measurement methods used in rehabilitation clinical trials since 2000 and 2) to evaluate their validity, reliability and acceptability.

Methods

To address both review aims, we used a two-phase approach. In Phase 1 we identified recently used adherence measurement methods, and in Phase 2 we evaluated these methods according to the level of evidence for their measurement properties. The review protocol was registered in PROSPERO (ID CRD42013004084) and is reported according to PRISMA guidelines.[10]

Defining adherence

Adherence is commonly defined in general terms, such as the World Health Organisation definition: "*the extent to which a person's behaviour – taking medication, following a diet, and/or executing lifestyle changes, corresponds with agreed recommendations from a health care provider*" (p.3).[1] Whilst the breadth of this definition allows it to apply widely across many therapy types, it lacks the detail required to inform a useful operational definition for use in clinical practice or trials. Rehabilitation interventions are typically complex in nature and combine a number of parameters, to which patients may differentially adhere. Rehabilitation prescriptions, similar to exercise or physical activity prescriptions, appear often to be characterised by four parameters in reviews or trials: frequency, duration, intensity and accuracy.[11–14] For example, stroke patients seeking to improve mobility may be asked to carry out three balance exercises for five minutes each seven times a week. Despite adherence to the frequency of seven times per week, the patient may exercise for a shorter

duration than recommended, may carry out just one of the three exercises or may carry out an exercise incorrectly.

Adherence was therefore operationalised within this review as the extent to which individuals undertake a prescribed behaviour accurately and at the agreed frequency, intensity and duration (see Figure 1). Measures assessing adherence to one or more of these parameters were included, in order to make recommendations across specific parameters and types of rehabilitation. "General adherence" was also included to identify any questionnaires based on the broader concept only.

Phase 1 – Identifying currently used adherence measures

Phase 1 aimed to collate a sample of adherence measurement methods used in home-based rehabilitation randomised controlled trials (RCTs). Rehabilitation is defined as the health strategy applied by professionals "that aims to enable people with health conditions experiencing or likely to experience disability achieve and maintain optimal functioning in interaction with the environment." (p.282).[15] Physiotherapy (PT), occupational therapy (OT) and speech and language therapy (SLT) rehabilitation interventions were selected as allied health professionals whose therapies most commonly contain home-based components. We searched the Cochrane Central Register of Controlled Trials (CENTRAL), the NHS Economic Evaluation Database and the Health Technology Assessment database in April 2013 as a comprehensive source of rehabilitation clinical trials. We used the keywords adherence, compliance and rehabilitation (see Supplementary File 1). We limited the review to post-2000 as it was anticipated that relevant adherence measures developed before 2000 would carry forward into more recent usage. Hand searching was not used as adherence

research is reported across multiple disciplines and research areas rather than within specific journals.

Inclusion criteria:

- **Study design**: RCTs, including protocols of RCTs
- **Participants**: adults with a health condition of any duration and severity
- **Interventions**: rehabilitation interventions including at least one of the following as part of a prescribed therapeutic regimen: modifications to the home environment or strategies to improve activities of daily living, home-based physical or language exercises or home-based interventions led by PTs, OTs or SLTs or an unspecified professional but the intervention met all other inclusion criteria; interventions to increase adherence to one of the above interventions.
- **Comparators**: any
- **Outcomes**: any method of measuring adherence to the concepts outlined above, including proxy measures, to the home-based component of the intervention.
- Studies carried out in countries where English is the primary language to ensure applicability to English-speaking populations.

Exclusion criteria:

Studies were excluded if they assessed the following: healthcare professional adherence to guidelines or study protocols; clinic- or hospital-based adherence only; group- or class-based adherence only; nutritional or pharmacological interventions only; primary prevention or screening initiatives; increasing physical activity in general rather than prescribed therapy.

One reviewer screened titles, abstracts and full texts for relevant clinical trials, taking an inclusive approach and checking with a second reviewer (SL) in cases of uncertainty. Both reviewers (RF and SL) extracted data from included studies using a standardised data extraction form, regarding intervention characteristics; sample demographics; adherence measurement method used and component of adherence measured; adherence definition and outcome used; assessment location; completion rates; and references to relevant measurement property studies. In cases of disagreement, consensus was reached through discussion or consultation with a third reviewer (BW). Risk of bias was not assessed as we aimed to compile measurement methods rather than utilise the trials' findings. Titles of adherence measurement methods identified in Phase 1 contributed to Phase 2.

Phase 2 – Evaluating the measurement properties of each method

Within Phase 2 we aimed to evaluate the validity, reliability and acceptability of each named measurement method located in Phase 1, defined as:

i. Validity: whether an instrument measures what it intends to[16], including:
    a. Criterion validity: the closeness of a measure with the recognised gold standard or how well it predicts future outcomes.[16]
    b. Construct validity: testing a hypothesised network of relationships and inferring the validity of the instrument from the results of these tests.[16,17]
    c. Structural validity: the degree to which questionnaire scores reflect the dimensionality of the constructs measured.[18]
    d. Face validity: the relevance and clarity of the measure at face value according to respondents or investigators' assessments.[19,20]

e. Content validity: systematic examination of the extent to which the instrument covers all elements requiring measurement in sufficient detail.[19]
f. Responsiveness to change: a measure's ability to detect change, ideally those that are clinically important.[21]

ii. Reliability: the extent to which a measure is free from random error.[17]
a. Test-retest reliability: reproducibility of a measure over a short period of time where the variable is not expected to change.[20]
b. Measurement error: the discrepancy between the observable concept measured and the actual underlying variable.[17]
c. Inter-rater reliability: the agreement between two or more raters assessing the same population.[20]
d. Intra-rater reliability: the agreement between the same rater on the same subject on the same occasion [22]
e. Internal consistency: the homogeneity of scale items [20]

iii. Acceptability: the patient's willingness or ability to complete a measure,[23] including data from any study type regarding wear time or rates (devices), completion rates, qualitative interviews, focus groups or think aloud studies and survey opinions or rating scales.[24,25]

Measurement properties were based on Classical Test Theory concepts with Item Response Theory (a questionnaire-specific theory that models the relationship between questionnaire items and the person's level of the construct[26]) MPAs include where relevant e.g. internal consistency, structural validity . Acceptability was considered a third key characteristic as adherence measures often require participants to wear or complete instruments more frequently than other outcome measures.

Medline, CENTRAL, ProQuest Nursing & Allied Health, EMBASE, CINAHL, AMED and Web of Science Core Collection were searched initially from inception to April 2015 (see Appendix 1 for Medline example of search terms). An earlier version of this review can be found as a conference abstract.[27] For each measure the title, with synonyms where applicable, was combined with acceptability search terms and Terwee et al's[28] MPA study precise filter. Subject headings were adapted for each database. Hand searching and consultation of topic experts were infeasible in such a diverse topic area and searching for ongoing MPAs was not possible as clinical trials registries are focussed on trials only.

Inclusion criteria:

- **Participants:** adults (healthy or clinical populations).
- **Study types:** studies assessing one or more MPAs outlined above in relation to the frequency, intensity, duration, accuracy or general adherence of an exercise or activity.
- **Setting:** laboratory and 'real-world' assessments.
- **Adherence measure:** the specific model or questionnaire type listed in Phase 1 only.
- **Comparator**: any comparator that could be classed or was described as a gold standard (criterion only) or measured a related aspect to the adherence component measured (construct only)

Exclusion Criteria:

We excluded papers: not written in English; cross-cultural validity assessments, and therefore studies where the measure was used or administered in a language other than English; where the relevant measure was used to validate another measure; where the measure assessed symptoms, functional limitations or total energy expenditure rather than an adherence parameter; water-based activity; articles focussed on sports science applications rather than health science; conference abstracts (limited information) or reviews (relevant systematic review reference lists were screened).

Study screening was undertaken as per Phase 1. Two reviewers (RF and either HS, BF, KT or PC) independently extracted data regarding: population, MPA type, sample size, activity, comparator(s) used, statistical methods, results and conclusions. Both independently assessed study quality using the COSMIN 4-point checklist [29] and resolved disagreements through discussion. COSMIN scores measurement property studies as Poor, Fair, Good or Excellent based on their methodological features according to a least-score-counts system. Though this checklist has limited applicability to electronic measures as it was developed for patient-reported outcome measures, it is the only comprehensive, well-developed checklist currently available for MPAs. We intended to synthesise studies of all quality; however, due to a large number of small, lower quality studies, the protocol was refined to include only Excellent or Good studies in the main narrative synthesis. This ensured that conclusions were based on high quality evidence, whilst Poor or Fair rated studies were used in a sensitivity analysis to see if they confirmed, refuted or extended the higher quality study findings. Study authors were contacted where possible in the event of missing data.

Studies were tabulated according to measurement method, MPA type and parameters of adherence the method was validated for. We aggregated studies using the Centre for Reviews and Dissemination's narrative synthesis approach.[30] Whilst statistics such as limits of agreement are in the original units and so have a more straightforward interpretation [31], there

is little consensus as to the interpretation of statistics which give a value between 0 and 1 (e.g. correlations, kappa, alpha). As we did not plan to conduct meta-analyses, we grouped values to assist comparisons. A minimum acceptable value was not used as we accepted that this would differ according to the measurement needs of different situations. High values are generally considered to be >0.70, preferably >0.80 [32–35], therefore we used the following cut offs, based on commonly used rules of thumb, to classify correlations, alpha, kappa and percentage wear/completion rates:[36]

- Poor: 0.00-0.19, 0-19%
- Fair: 0.20-0.39, 20-39%
- Moderate: 0.40-0.59, 40-59%
- Good: 0.60-0.79, 60-79%
- Excellent: >0.80, 80-100%

Other acceptability results were descriptively summarised due to the heterogeneity of the methods used (e.g. qualitative interviews, completion rates).

Results

Figure 2 shows the flow of studies throughout Phase 1 and 2.

[Figure 2 about here]

Phase 1- Identifying currently used adherence measures

Within Phase 1, 56 datasets of 59 full texts were included out of 1174 initial references and 209 full texts. Twenty eight were checked with a second reviewer (SL). Interventions were classified as discipline-specific as per the professional described in the text, and were largely

physiotherapy-based (n=36). Musculoskeletal conditions (n=27) were most commonly treated in the included trials. Thirty five single and 21 combinations of adherence measurement methods were identified (see Table 1). Frequency adherence was most commonly measured (n=44), followed by duration (n=15), intensity (n=14) and general adherence (n=12). Accuracy was only measured in four RCTs. Adherence diaries were assumed to measure frequency only if no further details were given. Common adherence outcomes used were average percentage sessions (n=17), average number of sessions (n=14) and percentage achieving minimum adherence levels (n=10).

Seven named methods were identified. One questionnaire used in two studies[37,38] was not named but the RCT reports contained measurement property information. This scale, termed Bassett & Prapavessis' scale after the study authors, was included in Phase 2 but as Phase 2 search strategies incorporated measure titles further measurement property searches were not feasible for this scale. "Cited by" functions did not reveal further studies. We therefore evaluated the following eight methods in Phase 2, which are summarised in Table 2 along with their MPAs.

[Table 2 about here]

Phase 2 – Evaluating the measurement properties of each method

The initial and updated results were combined, de-duplicated and rescreened as necessary. Out of 6926 hits across both reviews, 869 full texts were screened and 109 studies including 222 MPAs were included (18 articles checked by a second reviewer). After applying COSMIN criteria [29], 22 Excellent or Good MPAs were included in the synthesis, 153 low quality studies were used as supporting data and 47 acceptability studies were evaluated.

These are summarised alongside a description of each measure in Table 2, with details of each study tabulated in Supplementary File 2. Three MPAs are awaiting further information.[39–41]

To summarise, the evidence for most measures was limited. The StepWatch Activity Monitor appeared to be the most valid measure of adhering to a daily step count, but the evidence base consisted largely of short-term laboratory studies, was inconsistent across populations and lacked predictive validity (see Table 2). It appeared to be reliable and acceptable to wear for one week and up to 28 days. Adherence diaries had good to excellent criterion validity in the limited populations they were validated in, but lacked predictive validity of functional outcomes. Evidence for their reliability was scarce, but acceptability ranged from moderate to excellent (50-100% return rates). Regarding questionnaires, the Problematic Experiences of Therapy Scale had greater validity, reliability and acceptability for assessing general adherence than Bassett & Prapavessis' scale, though both had limited MPAs in single populations. The Borg 6-20 scale and CW series pedometers had inadequate validity, though these measures appeared to be reliable. The Joint Protection Behaviour Assessment had low quality supporting data for validity and reliability, whilst the Polar A1 heart rate monitor series had good acceptability in healthy adults but no other validity or reliability assessments. Sensitivity analyses largely confirmed the findings in broader patient populations and contributed reliability data.

Discussion

In this systematic review we found that adherence diaries were the most commonly used measures, usually for assessing adherence to how frequently a home-based behaviour was carried out. Self-developed questionnaires were also common, whilst most named methods

were sparsely used. The eight named methods identified had limited evidence, with suggestions that the StepWatch Activity Monitor and adherence diaries may be valid and acceptable within certain populations. Other methods lacked measurement properties or were assessed only in limited populations.

Strengths

As found in previous reviews of adherence to physiotherapy, rehabilitation and self-management adherence systematic reviews, we found an abundance of self-developed questionnaires and diaries.[6,8,9] However, these reviews found little evidence of measurement properties for any of the included measures. The larger volume found in this review is likely to arise from including electronic measures and aggregating diaries (often considered as a single type of measure).

In order to confirm the relevance of the measures considered in Phase 2 above we updated our Phase 1 search in August 2016. Out of the 41 new studies identified in the update adherence diaries (34 studies), Step Watch Activity Monitors (2 studies) Yamax CW-701, Borg 6-20 scale and the Problematic Experiences of Therapy Scale (each 1 study) continued to be reported. Additional non-named methods were also reported (as in our Phase 1 review) including sensors in hardware or software (n=5), self-developed questionnaires (n=6), telephone interviews (n=6), carer reports (n=1) and an accuracy checklist developed for the study (n=1). Some newly emerging measures were also reported within isolated studies including the Exercise Adherence Rating Scale, the Omron HJ-720ITC Pocket pedometer, the Borg CR-10 scale, the Accusplit pedometer, and the Adherence Assessment. These new methods remain avenues for further review alongside measures developed in non-English languages, in trials not indexed in CENTRAL or not yet employed in a rehabilitation clinical trial.

To our best knowledge this review is the first to provide a rigorous assessment and summary of multiple types of adherence measures across a broad range of interventions, participants and professionals. In particular, previous reviews have neglected to evaluate the acceptability of each measure, which remains a vital part of adherence measurement, particularly when measures are worn or completed on a daily basis. Comparison across electronic, provider report and self-report methods, whilst complex, is vital for decision making and so this review has greater utility than one of a single measure or type of measure. Further strengths include the two-phase approach which ensured that relevant measures were assessed and the use of an explicit conceptual underpinning often absent in adherence measurement. We searched for a wide variety of measurement properties and two reviewers independently assessed study quality using the COSMIN checklist. Only one main protocol refinement occurred, which was to include only high quality studies, but this was deemed reasonable as it allowed recommendations to be made on the basis of the most rigorous evidence.

Limitations

Within Phase 2, some relevant measurement property assessments may not have been located due to inadequate definition, classification and reporting of these studies. Common limitations in the evidence base located included small sample sizes and suboptimal statistics in validity and reliability assessments. Only a small number of included studies were of high quality. Most were of Fair or Poor quality and used only small sample sizes. A large majority of the StepWatch studies were carried out in a lab, which limits generalisability to use in a home-based situation where a wider range of activity is likely to be recorded. Whilst laboratory environments lessen the clinical applicability of these studies, they were included as they provided some validity information and for some tools (e.g. the StepWatch Activity Monitor) assessing criterion validity outside of a laboratory is challenging.

Other methods were tested in only limited populations e.g. the Borg scale was usually validated for activities in healthy adults, despite its increasingly common usage in rehabilitation. Diaries lacked reliability assessments, whilst all measures had a paucity of reliability, responsiveness to change and predictive validity studies. Acceptability was rarely formally assessed, despite wear and completion being important components of electronic devices such as activity monitors or diaries. Defining adequate comparators was also problematic as some included methods were used to validate others.[42] Gold standards were unavailable for some types of rehabilitation activity or for assessing adherence to behaviour accuracy.

Implications for clinical practice

When selecting adherence measures for use in clinical trials or clinical practice, conceptual adherence definitions need to be utilised. This permits a measure to be selected according to the level of rigorous evidence of measurement properties available for the relevant components. The main recommendations for using adherence measures in clinical trials and practice are summarised in Table 3. Most measures were validated in specific participant populations and prior to using a measure, clinicians should check it is validated for that population. Consequently our findings are likely to have the greatest relevance to physiotherapy and exercise-based interventions, as this was where most measures were used and evaluated, though some measures (e.g. adherence diaries) were used across all intervention types.

[Table 3 about here]

Implications for future research

Further well-designed, adequately powered studies, particularly reliability studies, evaluating a measure in therapeutic situations are required to inform future adherence measure selection. Formal qualitative evaluations by service users are required to further assess acceptability studies and better reporting of quantitative acceptability data. Identifying the most suitable measures for different populations will optimise their use in trials and clinical practice. Furthermore, this review showed that reviewing existing electronic measures (e.g. pedometers) warrants further investigation to determine their validity and acceptability for measuring adherence. The development of new questionnaires based upon a thorough adherence conceptualisation that takes accuracy or intensity into account may also be valuable. However current methods also offer potential for development and testing. This should be prioritised to avoid the multitude of self-developed questionnaires that are not comparable, as identified in the first phase of this review. Utilising adherence measures in RCTs presents further opportunities to collect feasibility, acceptability and MPA data regarding adherence measures. These should be reported clearly or separately to enable location of this data in future reviews.

Conclusion

Currently, there is no gold standard of adherence measurement for home-based therapies. Methods included in this review are limited by the quality of evidence of their measurement properties or their limited applicability across interventions. However, in light of the available evidence, StepWatch Activity Monitors are likely to be valid and acceptable to assess adherence to walking interventions, adherence diaries can approximate adherence to intervention frequency and duration and the Problematic Experiences of Therapy Scale can validly and reliably assess general adherence across vestibular rehabilitation populations.

Further study into which measures are most suitable for intervention parameters and patient populations and clearer reporting is required.

Tables

Figures

418 **References**

419 1. World Health Organization. Adherence to long-term therapies: evidence for action.
420 Geneva: 2003.

421 2. Kehoe SH, Chheda PS, Sahariah S, Baird J, Fall CHD. Reporting of participant
422 compliance in randomized controlled trials of nutrition supplements during pregnancy.
423 Matern. Child Nutr. 2009;5:97–103.

424 3. National Institute for Health and Care Excellence. Stroke rehabilitation: Long-term
425 rehabilitation after stroke. Manchester:

426 4. National Institute for Health and Care Excellence. MI – secondary prevention:
427 Secondary prevention in primary and secondary care for patients following a
428 myocardial infarction.

429 5. Long-term Conditions Alliance Scotland. “Gaun Yersel!”: The Self Management
430 Strategy for Long Term Conditions in Scotland. Glasgow: 2008.

431 6. Bollen JC, Dean SG, Siegert RJ, Howe TE, Goodwin V. A systematic review of
432 measures of self-reported adherence to unsupervised home-based rehabilitation
433 exercise programmes, and their psychometric properties. BMJ Open. 2014;4:e005044.

434 7. Holden M, Haywood K, Potia T, Gee M, McLean S. Recommendations for exercise
435 adherence measures in musculoskeletal settings: a systematic review and consensus
436 meeting (protocol). Syst. Rev. 2014;3:10.

437 8. Hall AM, Kamper SJ, Hernon M, Lonsdale C, Hurley DA, Hughes K, et al.
438 Measurement tools for adherence to non-pharmacological self-management treatment
439 for chronic musculoskeletal conditions: a systematic review. Arch. Phys. Med.
440 Rehabil. 2015;96:552–62.

441 9. Mclean S, Holden M, Haywood K, Potia T, Gee M, Mallett R, et al. Exercise

adherence measures - why we need to start again. Findings of a systematic review and consensus workshop. Physiotherapy. 2015;101:e981–2.

10. Moher D, Liberati A, Tetzlaff J, Altman DG, Grp P. Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement (Reprinted from Annals of Internal Medicine). Phys. Ther. 2009;89:873–80.

11. Page SJ, Schmid A, Harris JE. Optimizing terminology for stroke motor rehabilitation: recommendations from the American Congress of Rehabilitation Medicine Stroke Movement Interventions Subcommittee. Arch. Phys. Med. Rehabil. 2012;93:1395–9.

12. Pollock M. Prescribing Exercise for Fitness and Adherence. In: Exercise Adherence: Its Impact on Public Health. Champaign: Human Kinetics; 1988. p. 259–77.

13. Simek EM, McPhate L, Haines TP. Adherence to and efficacy of home exercise programs to prevent falls: A systematic review and meta-analysis of the impact of exercise program characteristics. Prev. Med. (Baltim). 2012;55:262–75.

14. Smith J, Lewis J, Prichard D. Physiotherapy exercise programmes: Are instructional exercise sheets effective? Physiother. Theory Pract. 2005;21:93–102.

15. Stucki G, Cieza A, Melvin J. The International Classification of Functioning, Disability and Health (ICF): a unifying model for the conceptual description of the rehabilitation strategy. J. Rehabil. Med. 2007;39:279–85.

16. Streiner DL, Norman G. Health Measurement Scales: A practical guide to their development and use. 3rd ed. Oxford: Oxford University Press; 2003.

17. Carmines EG, Zeller RA. Reliability and Validity Assessment. London: Sage Publications; 1979.

18. Mokkink L, Terwee C, Patrick D, Alonso J, Stratford P, Knol D, et al. COSMIN checklist manual. 2012;

19. Vitolins M, Rand C, Rapp S, Ribisl P, Sevick M. Measuring adherence to behavioral and medical interventions. Control. Clin. Trials. 2000;21:188S – 94S.

20. Bowling A. Techniques of questionnaire design. In: Bowling A, Ebrahim S, editors. Handbook of Health Research Methods. Maidenhead: Open University Press; 2005. p. 394–427.

21. Guyatt G, Walter S, Norman G. Measuring Change Over Time: Assessing the usefulness of evaluative instruments. J. Chronic Dis. 1987;40:171–8.

22. Mokkink L, Terwee C, Patrick D, Alonso J, Stratford P, Knol D, et al. The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. Qual. Life Res. 2010;19:539–49.

23. Haywood KL, Hargreaves J, White R, Lamb SE. Reviewing measures of outcome: reliability of data extraction. J. Eval. Clin. Pract. 2004;10:329–37.

24. De Bleser L, De Geest S, Vincke B, Ruppar T, Vanhaecke J, Dobbels F. How to test electronic adherence monitoring devices for use in daily life: a conceptual framework. Comput. Inform. Nurs. 2011;29:489–95.

25. Haywood KL, Staniszewska S, Chapman S. Quality and acceptability of patient-reported outcome measures used in chronic fatigue syndrome/myalgic encephalomyelitis (CFS/ME): a systematic review. Qual. Life Res. 2012;21:35–52.

26. Edelen MO, Reeve BB. Applying item response theory (IRT) modeling to questionnaire development, evaluation, and refinement. Qual. Life Res. 2007;16 Suppl 1:5–18.

27. Frost R, Brady M, McClurg D, Williams B. A systematic review of adherence measurement methods currently used in randomised controlled trials of home-based

rehabilitation interventions. Clin. Rehabil. 2015;29:396–7.

28. Terwee CB, Jansma EP, Riphagen II, De Vet HCW. Development of a methodological PubMed search filter for finding studies on measurement properties of measurement instruments. Qual. Life Res. 2009;18:1115–23.

29. COSMIN. COSMIN checklist with 4-point scale [Internet]. 2011;Available from: http://www.cosmin.nl/images/upload/files/COSMIN checklist with 4-point scale 22 juni 2011.pdf

30. Centre for Reviews and Dissemination. Systematic reviews: CRD's guidance for undertaking reviews in health care. York: University of York; 2008.

31. de Vet HCW, Terwee CB, Knol DL, Bouter LM. When to use agreement versus reliability measures. J. Clin. Epidemiol. 2006;59:1033–9.

32. Kottner J, Audigé L, Brorson S, Donner A, Gajewski BJ, Hróbjartsson A, et al. Guidelines for Reporting Reliability and Agreement Studies (GRRAS) were proposed. J. Clin. Epidemiol. 2011;64:96–106.

33. Terwee C, Mokkink L. Qualitative Attributes and Measurement Properties of Physical Activity Questionnaires. Sport. Med. 2010;40:525–37.

34. Fitzpatrick R, Davey C, Buxton M, Jones D. Evaluating patient-based outcome measures for use in clinical trials. Health Technol. Assess. (Rockv). 1998;2:14.

35. Cicchetti D V. The precision of reliability and validity estimates re-visited: distinguishing between clinical and statistical significance of sample size requirements. J. Clin. Exp. Neuropsychol. 2001;23:695–700.

36. Viera AJ, Garrett JM. Understanding interobserver agreement: The kappa statistic. Fam. Med. 2005;37:360–3.

513 37. Bassett SF, Prapavessis H. A test of an adherence-enhancing adjunct to physiotherapy
514 steeped in the protection motivation theory. Physiother. Theory Pract. 2011;27:360–
515 72.

516 38. Bassett SF, Prapavessis H. Home-based physical therapy intervention with adherence-
517 enhancing strategies versus clinic-based management for patients with ankle sprains.
518 Phys. Ther. 2007;87:1132–43.

519 39. Barak S. Habitual ambulatory activity measurement post-stroke. Diss. Abstr. Int. Sect.
520 B Sci. Eng. 2009;

521 40. Bergman RJ, Bassett Jr DR, Muthukrishnan S, Klein DA. Validity of 2 devices for
522 measuring steps taken by older adults in assisted-living facilities. J. Phys. Act. Health.
523 2008;5 Suppl 1:S166–75.

524 41. Wajciechowski J, Gayle R, Andrew R, Dintiman G. The accuracy of radio telemetry
525 heart rate monitoring during exercise. Clin. Kinesiol. 1991;45:9–12.

526 42. Franklin PD, McLaughlin J, Boisvert CB, Li W, Ayers DC. Pilot study of methods to
527 document quantity and variation of independent patient exercise and activity after total
528 knee arthroplasty. J. Arthroplasty. 2006;21:157–63.

Table 1. Number of measures found in Phase 1, by type and adherence parameter measured

Measurement Type(s)	n	Freq	Dur	In	Accu	Gen	Unclear	Ref	Named Methods
Questionnaires (patient)	11	3	2	0	0	7	0	1–11	Problematic Experiences of Therapy Scale
Questionnaire (provider)	2	0	0	0	1	1	0	12,13	Joint Protection Behaviour Assessment
Diary	20	18	2	2	0	1	1	14–33	Adherence diary
Electronic method	1	1	0	0	0	0	0	34	-
Telephone interview	1	1	0	0	0	0	0	35	-
Questionnaire (patient and provider)	4	3	2	0	3	1	0	36–39	-
Diary + provider questionnaire	2	2	0	1	0	1	0	40,41	Adherence diary
Diary + patient questionnaire	1	1	0	1	0	0	0	42	Adherence diary
Diary + telephone interview	3	3	0	1	0	1	0	43–45	Adherence diary
Diary + heart rate (self-assessed)	2	2	1	2	0	0	0	46,47	Adherence diary
Diary + electronic method	3	3	2	1	0	0	0	48–50	StepWatch Activity Monitor Yamax Digiwalker CW-701 Adherence diary
Diary + telephone interview + questionnaire	1	1	1	1	0	0	0	51	Adherence diary
Diary + telephone interview + electronic method	1	1	1	1	0	0	0	52	A1 & FS1 heart rate monitors Adherence diary
Diary + Borg rating of perceived exertion	1	1	1	1	0	0	0	53	Borg 6-20 RPE Adherence diary

Diary + Borg rating of perceived exertion + self-assessed heart rate	2	2	2	2	0	0	0	[54,55]	Borg 6-20 RPE Adherence diary
Diary + telephone interview + Borg rating of perceived exertion + self-assessed heart rate	1	1	1	1	0	0	0	[56]	Borg 6-20 RPE Adherence diary

Key: n=number of trials containing this measure; Freq=frequency, Dur=duration, In=Intensity, Accu=accuracy, Gen=general adherence, Ref=reference. The reference list for included studies can be found in Supplementary File 3.

Table 2. Summary of each included measure and its measurement properties

Measure	Description	Validity	Reliability	Acceptability
StepWatch Activity Monitor (SAM)	Research-grade ankle-worn activity monitor.[57] Described as a pedometer, accelerometer or activity monitor as the internal mechanisms have not been disclosed.[58]	*High quality studies (n=5):* Small percentage error and mean bias and high percentage accuracy compared to direct observation for measuring step counts in healthy populations, individuals with COPD and individuals with MS in laboratory settings.[59–61] Fair predictive validity in persons with intermittent claudication for changes in Peak Walking Time.[48]	*High quality studies (n=0)*	N=24. Highly acceptable across populations for 1-28 days' wear (most commonly worn for 1 week), including persons with MS, TKA, neurological conditions, sarcoma, lower limb prosthesis, knee OA and older, sedentary and obese adults. In most studies >90% patients complied with SAM wear, but this was variable in stroke survivors and lower in persons with dementia, persons with intermittent claudication and healthy adults.[48,62–77] Most wore the SAM for >6 out of 7 days per week[63,66,78–80] and >11 hours per day.[69,80,81]
		Low quality studies (n=20): Small mean bias and percentage error and high percentage accuracy were confirmed in older adults, healthy volunteers and individuals with COPD, neurological conditions and mobility limitations).[82–98] Lower validity in persons with dementia,[77] cycling activity,[86] outdoor walking on a paretic limb[95] and when attached to a cane.[98] Moderate construct validity for activity intensity compared to a diary.[85]	*Low quality studies (n=14):* Excellent test-retest reliability for step counts same day to 3 weeks apart in the lab or home/ community in persons who are healthy or with neurological conditions (wider LOA in community).[62,72,75,77,85,94,97,99–101] Excellent inter-rater reliability in healthy adults.[87]	

Problematic Experiences of Therapy Scale (PETS)	12-item scale measuring general non-adherence - the degree to which socially acceptable reasons prevented patients adhering e.g. symptom severity/aggravation, efficacy doubts, practical challenges.[102]	*High quality studies (n=2):* Excellent structural validity in two populations with chronic dizziness from vestibular conditions.[103] *Low quality studies (n=3):* The PETS could differentiate between self-identified rehabilitation adherers or maintainers in Meniere's disease and dizziness patients.[11,103]	*High quality studies (n=2):* Excellent internal consistency in two populations with dizziness.[103] *Low quality studies (n=0)*	N=1. High completion rates in a Meniere's disease rehabilitation study (225/240)[11].
Adherence diaries (AD)	ADs were defined by their function of regular (usually daily) patient self-report of an activity. All AD types were aggregated.	*High quality studies (n=6):* Moderate-excellent criterion validity for measuring adherence to exercise frequency and duration compared to a heart rate monitor, pedometers and radiofrequency identification card system in sedentary women, older adults, cancer patients and pregnant women.[104–107] Fair to no predictive validity for walking adherence and changes in fitness in sedentary women[106]. Good construct validity was found compared to the Physical Activity Questionnaire in cancer patients.[105]	*High quality studies (n=0).*	N=19. Ranged evenly from moderate to high (50-100% return rates) across a variety of patient populations recording adherence from 2 weeks to 12 months.[14,21,24,25,29,32,33,54,65,105,108–116] Higher return rates were found in persons with TKA, systematic sclerosis, heart failure, coronary heart disease, diabetes, Crohn's,

		Low quality studies (n=19): Good to excellent criterion validity compared to a range of objective comparators in varied populations (older adults, knee arthroplasty patients, individuals with pain conditions, brain injury and SLE).[65,117–122] Low to moderate predictive validity for functional outcome measures in individuals with COPD, sedentary women, individuals with radial fracture, total knee arthroplasty patients and patients with implantable cardiac defibrillators.[65,109,123–125]. Moderate to good construct validity for exercise-related constructs in sedentary women and healthy adults, but lower validity for behavioural constructs in persons with Huntingdon's disease and sedentary women,.[125–128] Adherence predicted maintenance in sedentary women.[129] Diaries were responsive to short-term adherence changes in pulmonary rehabilitation.[130]	*Low quality studies (n=1):* Good test-retest reliability in pregnant women.[104]	elbow pain and osteoarthritis. Lower (50-75%) return rates were found in stroke survivors and patients with rotator cuff tears, risk factors for diabetes and after stem cell transplant. Mixed return rates were found in persons with COPD and back pain. Strategies that appeared to have a higher return rate included remuneration,[25] weekly collection[65] and weekly review[115]. Monthly collection did not engender particularly high return rates[108] and studies using reminders had mixed return rates.[29,33,105,112]
Bassett & Prapavessis'	Self-report scale measuring general adherence (rated 1-5) to 5 dimensions of home-based	*High quality studies (n=2):* Poor predictive validity for adherence and functional outcomes in patients with ankle sprains.[2] Fair construct validity compared	*High quality studies (n=1):* Good internal consistency between scale items in patients with ankle	N=0.

scale	physiotherapy: exercises, ice, rest, strapping and elevation.[3]	to intentions to adhere in patients with ankle sprains.[2]	sprains.[2]	
		Low quality studies (n=0)	*Low quality studies (n=1):* Good internal consistency in patients with ankle sprains.[3]	
Borg 6-20 rating of perceived exertion scale	Simple 15-grade scale of self-reported exertion commonly used in rehabilitation, exertion testing and training.[131] Only single estimates of intensity for one activity were included as the most relevant to rehabilitation adherence recording.	*High quality studies (n=3):* Fair criterion validity compared to a heart rate monitor in older adults in two activities.[132] Fair construct validity compared to other walk parameters (e.g. gait speed).[133]	*High quality studies* N=0.	N=0.
		Low quality studies (n=26): Poor to excellent criterion validity compared to objective intensity measures in healthy adults and pregnant women.[134–143] Low construct validity with walking distance travelled in patients with MS and stroke survivors[18,144] but good with speed and function in healthy adults with a foot orthosis and patients with MS.[18,145] Responsive to changes in walking, exercise and ADL intensity in healthy adults.[137,143,146–150] Content validity in patients with brain injury and low back pain and	*Low quality studies (n=5):* Good test-retest reliability in ADLs, walking, resistance training and cycling in healthy adults and individuals with MS.[18,141,142,146,149,152]	

		healthy students.[151]		
Yamax Digiwalker CW series	Yamax pedometer which records and displays the number of steps taken. It has a two week memory and a three year battery life.[153] All CW series contain the same internal mechanisms and so all were included.	*High quality studies (n=1):* Limited criterion validity compared to a GT1M ActiGraph accelerometer in pregnant women (overcounted at high step rates and undercounted at low step rates). Moderate to good 'active' and 'inactive' classifications.[154]	*High quality studies (n=0)*	N=1. CW-701 had data for 58/61 pregnant women for four days' wear.[154]
		Low quality studies (n=1): Poor criterion validity (high percentage error) in older adults.[155]	*Low quality studies (n=1)* Good inter-rater reliability in older adults.[155]	
Joint Protection Behaviour Assessment (JPBA)	20-task observational scale assessing performance accuracy of arthritis joint protection behaviours when making a hot drink and snack in a kitchen.[156] Behaviours are graded as correct, partially correct or	*High quality studies (n=0)*	*High quality studies (n=0)*	N=1. 83/127 individuals with rheumatoid arthritis agreed to be recorded performing the JPBA.[12]
		Low quality studies (n=6): Fair construct validity with hand impairment,[156,157] but higher with pain, perceived helplessness and reduced grip strength in persons with rheumatoid arthritis.[156,158] Responsive to changes in joint protection training in healthy adults.[159] Good face and	*Low quality studies:* Excellent test-retest, inter- and intra-rater reliability and internal consistency in healthy adults and individuals with	

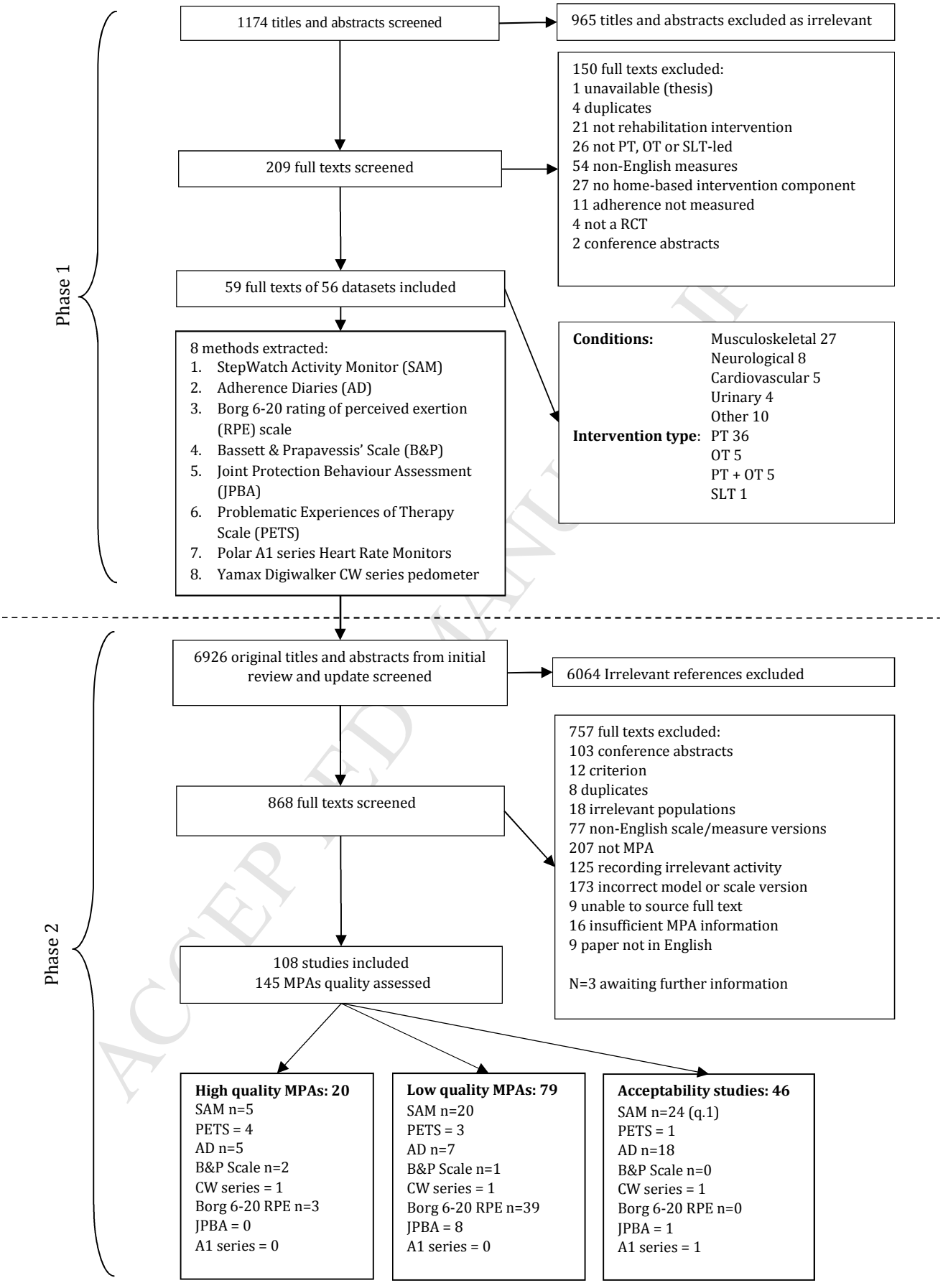
	incorrect and converted to a percentage score.[156]	content validity to researchers and OTs.[156]	rheumatoid arthritis.[156,159]	
Polar A1 series heart rate monitors	A family of Polar heart rate monitors. The models from this family with the same T31 transmitter include the FS1, A1, FT1, FT4, FT60, FT7 and RCX5. All these models were included in this review, though the A1 and FS1 may no longer be in production.[160]	*High quality studies (n=0)* *Low quality studies (n=0)*	*High quality studies (n=0).* *Low quality studies (n=0)*	N=1. The Polar FT60 was used in 76% of exercise session by healthy adults. Interviews showed that adults found the polar monitor motivational, fun and increased understanding of exercise. However, it was unsuitable for certain sports, could be forgotten and the guidance was not always applicable for people.[161]

The reference list for included studies can be found in Supplementary File 3.

Table 3. Implications for adherence measures identified in this review

Measure	Implications from this review
StepWatch Activity Monitor	➢ Valid for assessing step frequency in persons with COPD and multiple sclerosis, but lower predictive validity and in persons with dementia and irregular walking activity (e.g. outdoor walking on a paretic limb) ➢ Likely to be reliable for use in the community in persons with neurological conditions (e.g. stroke survivors, persons with Parkinson's), but this is low quality evidence ➢ Acceptable for 7 days wear in persons with neurological conditions and knee osteoarthritis and older, sedentary or obese adults
Problematic Experiences of Therapy Scale	➢ Can be recommended in chronic dizziness populations arising from vestibular conditions and where barriers and facilitators require assessment ➢ Requires testing in a wider variety of populations
Adherence diaries	➢ Can be used with high validity for recording activity frequency in sedentary women, older adults, cancer patients and pregnant women and potentially individuals with pain conditions, brain injury, SLE or after total knee arthroplasty ➢ Lacks predictive validity of functional outcomes ➢ Requires further reliability testing ➢ Mixed, moderate to excellent return rates across a wide variety of populations. Remuneration, weekly collection and weekly review appeared to increase completion rates; monthly collection and reminders had mixed results
Bassett & Prapavessis' scale	➢ Not currently recommended to assess general adherence: some reliability in ankle sprain populations but low construct validity

Borg 6-20 rating of perceived exertion scale	➢ Not currently recommended to assess intensity adherence: only fair validity in older populations, though may be reliable and responsive to change
Yamax Digiwalker CW series	➢ May be acceptable but cannot be recommended above other measures as it lacks evidence of good validity ➢ Pedometer models with good supporting evidence should be selected
Joint Protection Behaviour Assessment	➢ Recommended for assessing accuracy adherence of joint protection behaviour in patients with rheumatoid arthritis, though evidence is limited
Polar A1 heart rate monitor series	➢ May be acceptable to healthy adults but not currently recommended due to a lack of evidence ➢ Heart monitor models with good supporting evidence should be selected



Supplementary File 2. Table of included validity, reliability and acceptability studies.

Reference	MPA	Para	Population (n, descriptor, mean (SD) age, %female)	Activity	Comparator	Statistic and outcome	Qual
StepWatch Activity Monitor high quality studies (n=5)							
Feito 2012a[162]	Crit val	F, I	n=65, healthy volunteers, full sample (n=71): normal=27.8yrs (8.0), overweight=34.6yrs (14.2), obese=31.5yrs (11.1), 55%f	Lab; 5 minute walk on motorized treadmill at 3 different speeds (40, 67 and 94 m/min).	Trained observer with hand tally counter	Mean bias close to zero with 95% prediction interval: ±8 steps/min 95-102% steps recorded across different speeds. Pearson correlations: slow speed r=0.635, moderate speed r=0.500, fast speed r=0.558 (all p<0.001)	G
Feito 2012b[163]	Crit val	F, I	n=56, healthy individuals with a range of BMI values, normal=28.3yrs (10.5), overweight=31.2yrs (9.9), obese=29.0yrs (7.9), 50%f	Lab; 5 x 100 step walks on a treadmill at different speeds (40, 54, 67, 80 and 94 m/min)	Trained observer with hand tally counter	100±1% accuracy at slowest speed, >97% accuracy at faster speeds. No effect of BMI.	G
Hiatt 2011[48]	Crit val	D	n=62, intermittent claudication patients randomised to take propionyl-L-carnitine or placebo, G1 n=30, 66.6yrs (8.8), 17%f, G2 n-32 67.4yrs (8.7), 38%f	Home/community; 30-50 min walking 2-3 times per week; daily activities for 7 days at screening, 3 mo and 6 mo	Change in Peak Walking Time between baseline and 6mo	Changes in SAM ambulatory activity r=0.34 (p=0.013) Changes in SAM dose (mins of exercise) r=0.259 (p=0.048)	G
Moy 2012[164]	Crit val	F	n=127, stable COPD patients >40,	Lab; 244m walking course	Observer	Mean bias (95% LOA): +3 steps (-13.53 to	G

Reference	MPA	Para	Population (n, descriptor, mean (SD) age, %female)	Activity	Comparator	Statistic and outcome	Qual
			71.0yrs (8.0), 2%f	at usual speed		20.11 steps). >90% accuracy in 133/134 participants. No effect of BMI.	
Sandroff 2014[61]	Crit val	F	n=63, ambulatory individuals with multiple sclerosis, 50.7yrs (9.2), 76%f	Lab; 3 x 6min walk test around a rectangular hallway at comfortable, fast and slow walking speeds	Direct observation by research assistant using hand tally counter	Comfortable walking speed = 99.8% accuracy, fast 99.9%, slow 99.0%. High disability and low speed were less accurate.	G
StepWatch Activity Monitor low quality studies (n=35)							
Algase 2003[77]	Crit val	D	N=40, individuals with dementia, (all subject n=178) 85.3 (6.3), 75%f	Nursing home; duration of wandering in two 4hr periods	Trained observers recording using a bar code reader	Multiple regression controlled for age, sex and mini-mental state examination score: SAM predicted 63.6% of the variance in time spent wandering ($p<0.001$). Time in motion 16.8% SAM vs 15.4% observation.	P
Bergman 2008[165]	Crit val	F	N=21, older adults living in assisted living facilities, 78.6 (13.1), 76%f	Assisted living facility; walking course, 161m walk at a self-selected pace	Observer with hand tally counter	Mean bias = -11.3 (SE 2.56) (overestimation) ($p<0.001$), 95% prediction interval = -18.01 to -4.65. Correlations $r^2=0.99$ ($p<0.001$)	P
Bowden 2007[101]	Crit val	F	n=11, individuals with incomplete spinal cord injury with no more than minimal assistance required for walking, 45.5 (range 21-63), 18%f	Lab; 1 x 6 minute walk test at usual pace over series of hallways, 2 x 10 minute walk tests at self-selected pace, completed at 2 different times	Observer with manual handheld counter	Percentage accuracy (smaller quantity as percentage of larger quantity) 6 minute walk test = 97%, 10 minute walk test = 97%	P

Reference	MPA	Para	Population (n, descriptor, mean (SD) age, %female)	Activity	Comparator	Statistic and outcome	Qual
				in randomised order (4 hrs - 1 week later)			
Busse 2009[85]	Crit val	F	n=18, healthy volunteers, 26.1 (range 22-39), gender NR	Lab; Walking an indoor circuit for ~10mins (200m), including sit-to-stand transitions, completion of kitchen tasks and shoe removal; and an outdoor circuit for ~20mins (1100m), including uneven ground, lifts, ramps	Observer of videotaped walk using handheld step counter by one researcher with excellent intra-rater reliability. Overall ICC=0.99 for intra-rater reliability, but poor inter-rater reliability (ICC 0.26)	Indoor: mean (SD) dif = 5.76% (5.18). LOA = -4.6 to 16.2 steps. Outdoor: mean (SD) dif = 2.82% (7.47). LOA -12.2 to 17.8 steps. Percentage accuracy: Indoor = 96.1% (3.5), outdoor = 99.6% (1.1). Percentage error: indoor = 3.9% (3.5), outdoor = 0.4% (1.1)	P
Carr 2012[86]	Crit val	D, I	N=36, healthy adults, 23 (3.7), 55%f	Lab; 60min testing session including 6 sedentary and light activity activities for 8 min each (middle 6min compared)	Observer watching activities for fidelity	Percentage accuracy for light intensity: Walking 1.0mph: 86.1%. Pedalling 7.0mph 54.4%. Pedalling 15.0mph 23.5%. Root mean square error for minutes correctly coded = 3.33min	F
Ford 2010[92]	Crit val	F	n=12, individuals with Parkinson's disease, 67.2 (SD NR), 8.3%f	Lab; 1 min walk around the lab	Single observer	Percentage accuracy 98%	P
Foster 2005[87]	Crit val	F	n=20, healthy adults 50% lean 30 (13), 50% obese 32 (7), age range = 21-51yrs, 50%f	Lab; 15min walks at 1, 2 and 3 mph each. Level ground walking at 1 and 1.85mph each for 25min.	Single observer using electronic counter	Percentage accuracy 99.7% ±0.67 ICC=0.9995	P
Fulk 2014[93]	Crit val	F	n=26, diagnosis of stroke or traumatic	Lab; 2-minute Walk Test at	Observed step count of videoed	Mean difference = 4.7 steps (1.11-8.35). No	P

Reference	MPA	Para	Population (n, descriptor, mean (SD) age, %female)	Activity	Comparator	Statistic and outcome	Qual
			brain injury, able to walk with minimal assistance, able to follow study commands and give informed consent. (full sample n=50) 52.9 (15.1), 32%f	normal, comfortable pace with SAM on less affected side	walk on two separate occasions (ICC=0.99),with first count used in analysis	relationship between SAM error and gait speed, Berg balance scale or Fugl-Meyer score ICC=0.97 (0.92-0.99)	
Hartsell 2002[88]	Crit val	F	N=10, healthy adults, 43.2 (14.1), mixed (NR).	Lab; walking course; 4x530m walk in athletic shoes or fibreglass cast (TCC) on one leg, over flat ground and stairs each at self-selected pace	Mean of 2 observers (r=0.9923-0.9999)	Percentage error: flat surface: athletic shoe 0.136%, TCC 0.206%. Stairs: -3.648% athletic shoe and -5.697% TCC (undercounting). ANOVA: significant effects for walking surfaces.	P
Karabulut 2005[89]	Crit val	F, I, A	N=20, healthy adults, 28 (3.7), 50%f	Lab; treadmill; 3min walks at a variety of speeds, 3min each of heel tapping, leg swinging, cycle ergometer and (n=10) driving.	Observer with hand tally counter (2nd min only)	Mean bias = 0.9 steps min^{-1}, prediction interval = -2.3 to +4.1 steps min^{-1}. Mean step counts within 1% at all speeds. SAM responsive to heel tapping, leg swinging and cycling but not driving.	P
Macko 2002[94]	Crit val	F	n=16, >55 yrs of age with remote ischemic stroke (>6 months), with residual hemiparetic gait deficits and some preserved capacity for ambulation, 67 (7), mixed (NR)	Rehabilitation centre; walking course; 2x6min floor walk at self-selected pace, 2x1min floor at self-selected comfortable and fastest pace using normal adaptive device/orthosis	Observer with hand tally	Percentage accuracy: self-selected pace 98.5%±-1.0 (P<0.01), fast walking pace 97.7±-2.0* (p<0.01) First 6min walk 98.8±1.1 2nd 6min walk 98.7±1.2	P
Mudge 2007[166]	Crit val	F	n=25 chronic stroke patients, median age = 69 (range 42-79), 32%f	Lab; 6 trials on a 6m walkway without shoes at a self-selected	3-dimensional gait analysis	Pearson's r=0.959 (non-paretic limb) and r=0.896 (paretic limb)	P

Reference	MPA	Para	Population (n, descriptor, mean (SD) age, %female)	Activity	Comparator	Statistic and outcome	Qual
				pace			
	Crit val	F	n=21, chronic stroke patients, full sample median age = 69 (range 42-79), 32%f	Lab; indoor: 8m at self-selected pace and 8m at fast pace, outdoor: 200m course including steps, inclines and declines wearing usual footwear at a self-selected pace (with rest if required).	On/off event footswitches taped to the foot	95% LOA: non-paretic limb ±9 steps, paretic limb ±57 steps Percentage error: non-paretic limb -1.3% (range, -4.5% to 2.5%), paretic limb -4.2% (range, -42% to 16%). Pearson correlations: non-paretic limb r=0.999, paretic limb r=0.963	P
Ng 2012[167]	Crit val	F, I	N=20, chronic obstructive pulmonary disease patients with functional limitation, 73 (8.5), 60%f	Lab; walking course; self-selected slow and normal paces with and without a rollator for 5min each.	Observer (average of 30s interval at start of 2nd, 3rd and 4th minute)	Mean bias = +2 steps/min, 95% LOA 6 steps/min (-4 to 8 steps/min). No effect of rollator or walking speed on validity of step rate.	P
Resnick 2001[97]	Crit val	F	N=30, older adults (65+) with Parkinson's (n=3), previous hip fracture (n=10) or evidence of degenerative joint disease and/or osteoporosis (n=17), 86 (6.1), 73%f	Lab; 1min walk at self-selected speed over carpet, repeated after a 2min rest	Mean of two observers (experienced nurses). Inter-rater reliability = 0.98	Correlations (type not stated) r=0.95 (p<0.05) % accuracy = 96% % error = 4.0±3.1% (range 0-12%)	F
Schmidt 2011[168]	Crit val	F	N=20, individuals with Parkinson's disease (n=11, 66.8 (SD NR)) or multiple sclerosis (n=9, 55.8 (SD NR)),	Lab; walking course; 3 walks at usual speed over the GaitMat II	GaitMat II	Pearson correlations: Multiple sclerosis r=0.99, Parkinson's disease r=1.0. Mean strides: 15.55 (SAM) and 15.85 (GM).	P

Reference	MPA	Para	Population (n, descriptor, mean (SD) age, %female)	Activity	Comparator	Statistic and outcome	Qual
			65%f				
Shepherd 1999[90]	Crit val	F	n=29, healthy individuals able to comfortably walk a mile and two flights of stairs. 42.3 (15.3), 72%f	Lab; 2 trials of: 1) brisk walking around a 400m track 2) slow walking for 10m (household pace) 3) ascend 11 steps, 4) descend 11 steps	Single observer with handheld counter	Percentage accuracy mean (SD) (positive=over-counting). overall = 0.54% (0.7), 1) 0.31 (0.7) 2) 5.25% (5.7) 3) 3.58% (5.2), 4) 7.25% (11.6). Not affected by BMI, gender or lower leg surgery.	P
Storti 2008[83]	Crit val	F, I	N=34, 65+ and able to walk independently without an assistive device, 79.2 (6), 71%f	Lab; walking course; walked 100 steps on level surface at self-selected pace	Observer with handheld step counter	Percentage error: total +6.9, slow gait = +6.5, middle-speed gait = +6.6, fast gait = +2.8 (SAM over-counting) Absolute percentage error total 5.7 (5.0), slow 6.6 (5.7), medium 6.6 (5.5), fast 3.6 (2.9)	F
Wendland 2012[98]	Crit val	F	N=15, healthy adults using an assistive device, able to ambulate >10m without rest, (full sample n=16) 75.6 (SD NR), mixed (NR).	Lab; walking course; 2x10m each over linoleum, pavement, grass, up and down a ramp, and up and down stairs. SAM attached to cane and right leg.	Observer for leg strides and observer for cane strides with handheld tally counters	Percentage accuracy: leg = 93.4%, cane = 84.7%. Stairs less accurate ($p<0.001$).	P
Busse 2009[85]	Cons val	F, D, I	n=22, healthy volunteers, 26.9 (22-45), gender NR	Home/community; Everyday activities for 4 days	4-day activity diary (main activity recorded in 15-min blocks) and classified into inactive, low, moderate and	Spearman's for counts of 15min blocks in activity level Inactive $\rho=0.47$ ($p<0.05$), low $\rho=0.42$ ($p<0.05$), medium $\rho=0.48$ ($p<0.05$), high $\rho=0.59$ ($p<0.01$)	P

Reference	MPA	Para	Population (n, descriptor, mean (SD) age, %female)	Activity	Comparator	Statistic and outcome	Qual
					vigorous based on METs		
Bowden 2007[91101]	TRR	F	n=11, diagnosis of incomplete spinal cord injury with no more than minimal assistance required for walking, 45.5yrs (range 21-63), 18.2%f	Lab; 1 x 6MWT at usual pace over series of hallways, 2 x 10m WTs at self-selected pace, completed at 2 different times in randomised order	4hrs to 1 week later both tests repeated	ICC (2,1) 6MWT = 0.99, 10mWT = 0.97	P
Busse 2009[85]	TRR	F	n=20, healthy volunteers, 26.15 (range 17-38), gender NR	Lab; 3 outdoor 20min circuit walks with ramps, lifts etc using metronome to standardise cadence	Three walks of same circuit	ICC=0.96	P
Busse 2004[62]	TRR	F, D, I	n=10 healthy adults, 43.3 (18.9), 40%f; n=10 ambulant neurological patients with impairments from different pathologies with restricted walking mobility but able to walk >10m without assistance, 59.4 (13.4), 50%f	Home/community; Everyday activities for two 7-day monitoring periods (SAM worn for 24hr/day ad removed for bathing)	1-3 weeks apart	Healthy step count: ICC=0.89 day to day CV=28%, week to week CV 8.8%. Peak activity index ICC=0.98. 20min sustained activity ICC=0.75. 30min sustained activity ICC=0.71, 60min sustained activity = 0.57. Neurological patients step count ICC=0.86 day to day CV=30%, week to week 12%. Peak activity index ICC=0.82, 20min sustained activity ICC=0.94. 30min sustained activity ICC=0.90, 60min sustained activity = 0.95.	P
Haeuber	TRR	F	n=17, >50, remote ischaemic stroke	Home/community; total strides	Average per day of two 48 hr	ICC=0.96 (p<0.001)	P

Reference	MPA	Para	Population (n, descriptor, mean (SD) age, %female)	Activity	Comparator	Statistic and outcome	Qual
2004[169]			over 6mth ago, residual hemiparetic gait defects but capacity for ambulation with assistive device. 65 (6), gender NR	over 48 hours	periods, up to 3 weeks' apart		
Algase 2003[77]	TRR	D	Sample size not clear, individuals with dementia, (full sample n=178) 85.3 (6.3), 75%f	Nursing home; free-living wandering in 1-4 four-hour periods	Time interval: 3 days	Pearson's correlations r=0.71 (p<0.001)	F
Mudge 2008[75]	TRR	F, I	N=40, >6 months post-stroke, able to walk independently but with some residual difficulty, 69.2 (12.6), 43%f	Home/community; mean steps in free-living three day period	Time interval: 1 week (same 3-day period)	Total step count ICC = 0.989; CV = 10.7%; Medium rate steps: ICC = 0.964; CV = 17.8%; High rate steps: ICC=0.926; CV=37.6%; Low rate steps: ICC=0.953; CV=11.1%	F
Macko 2002[94]	TRR	F	N=16, patients >55 yrs with remote ischemic stroke (>6 months), with residual hemiparetic gait deficits and some preserved capacity for ambulation, 67 (7), mixed (NR).	Rehabilitation centre; walking course; 2x6min walks at self-selected pace using their normal adaptive device/orthosis	Time interval: >=1 day (NR)	ICC r= 0.975, P < 0.0001	P
Mudge 2010[170]	TRR	F, I	N=15, healthy adults, (full sample n=30) 27.7 (8.9), 50%f	Home/community; 3 days free-living activity	Time interval: 1 week (same 3-day period)	Mean steps/day: ICC=0.895; CV=11.8%; Medium rate steps: ICC=0.854; CV=13.0%; High rate steps: ICC=0.744; CV=36.9%	P
Resnick 2001[97]	TRR	F	N=30, older adults (65+) with Parkinson's (n=3), previous hip fracture	Lab; 1min walk at self-selected speed over carpet, repeated	Time interval: 2min	ICC r=0.84	F

Reference	MPA	Para	Population (n, descriptor, mean (SD) age, %female)	Activity	Comparator	Statistic and outcome	Qual
			(n=10) or evidence of degenerative joint disease and/or osteoporosis (n=17), 86 (6.1), 73%f	after a 2min rest			
Subramony 2012[72]	TRR	F, D, I	N=19, ambulatory (with/without an assistive device) individuals with different spinocerebellar ataxias, 56 (10.7), 79%f	Home/ community; free-living wear for 8 days	Time interval: days 1-3 compared to days 5-7	Percentage time in activity: low speed ICC=0.872, moderate speed ICC=0.886, high speed ICC=0.606. Percentage steps: low speed ICC=0.912, moderate speed ICC=0.893, high speed ICC=0.793. Average daily step count ICC=0.900. Steps/min ICC=0.864.	P
Foster 2005[87]	Inter-rater rel	F	n=20, healthy adults 50% lean 30 (13), 50% obese 32 (7), age range = 21-51yrs, 50%f	Lab; 15min walks at 1, 2 and 3 mph each. Level ground walking at 1 and 1.85mph each for 25min.	SAM worn on inside of left ankle and outside of right ankle during same trials.	Mean bias 0.18±0.28 steps/min at 1ph, 0.18±0.31 steps/min at 2 mph, and 0.04±0.06 steps/min at 3 mph. Hall walking = 0.02 steps/min compared to treadmill measures	P
Bowden 2007[101]	ME	F	n=11, diagnosis of incomplete spinal cord injury with no more than minimal assistance required for walking, range 21-63, 18.2%f	Lab; 1 x 6 minute walk test at usual pace over series of hallways, 2 x 10minute walk test at self-selected pace, completed at 2 different times in randomised order (4 hrs - 1 week later)	4hrs to 1 week later both tests repeated	Standard error of measurement 6 minute walk test = 6.0 steps 10 minute walk test = 0.76 steps	P
Mudge 2008[75]	ME	F, I	N=40, >6 months post-stroke, able to	Home/community; free-living	Time interval: 1 week (same 3-	95% LOA (absolute,%):	F

Reference	MPA	Para	Population (n, descriptor, mean (SD) age, %female)	Activity	Comparator	Statistic and outcome	Qual
			walk independently but with some residual difficulty, 69.2 (12.6), 43%f	mean steps over 3 days;	day period)	Total 3 day step count = ±1801 (37.8) Medium rate steps = ±836 (87.1%) High rate steps = ±1750 (153%) Low rate steps =±1643 (63.6%)	
Mudge 2010[170]	ME	F, I	N=15, healthy adults, (full sample n=30) 27.7 (8.9), 50%f	Home/community; 3 days free-living activity	Time interval: 1 week (same 3-day period)	95% LOA (absolute, %) Mean steps/day: 3341 (39.1%) Medium rate steps: 2111 (53.5%) High rate steps: 2521 (122%)	P
Ng 2012[167]	Resp to change	F, I	N=20, chronic obstructive pulmonary disease patients with functional limitation, 73 (8.5), 60%f	Lab; walking course; self-selected slow and normal paces with and without a rollator for 5min each	Step rate at slow and normal paces and with/out rollator (speed regulated by audio signals)	ANOVA: significant effect of walking speed ($F_{1,19}$ = 88.69; $p < 0.01$) on step rate as measured by SAM. Significant effect of rollator ($F_{1,19}$=12.39, p=0.02). No interactions between speed and rollator.	P
StepWatch Activity Monitor acceptability studies (n=24)							
Algase 2003[77]	Acc	D, O	n=72, ambulatory nursing home residents with dementia, full sample (n=178) 85.3 (6.31), 75.3%	Wandering activity for 4x4hr periods	Wear rates	29.2% wore SAM for all 4 periods. 83.3% accepted a device for any period. MMSE and age did not predict device acceptance. N=288 periods: 57.98% periods had available data, 0.69% periods had equipment failure, 0% project/staff problems, 1.48% setting issues, 28.80% subject issues, other = 11.0% . SAM was added later in study	-
Algase 2003[77]	Acc	D, O	n=17, nursing home staff, age NR,	Patients with dementia wearing	Rating scale	0-5 scale (unacceptable to highly acceptable):	-

Reference	MPA	Para	Population (n, descriptor, mean (SD) age, %female)	Activity	Comparator	Statistic and outcome	Qual
			gender NR	SAM for 4x4hr period		Appearance: 3.50 (0.76), Comfort: 3.47 (0.80), Concealment: 3.74 (0.61), Easy application: 3.19 (0.98), Ease of cleaning: 3.25 (0.92), Location: 3.76 (0.55), Safety: 3.71 (0.59), Size: 3.53 (0.79), Weight 3.76 (0.49). Rated second most highly on six scores out of four devices	
Barak 2014[171]	Acc	F	n=408, >18, 5-30 days after stroke without contraindications to exercise, 62.02 (12.74), 45.1%f	Everyday activities for 2 days, removed for bathing, showering, swimming or sleeping	Wear time	Inferred adherence per day = activity (>2 steps) within each six hour time period (6am-12pm, 12-6 and 6-12am) for each day. Day 1 = 68.1% adherence, Day 2 = 60.8%, Both = 52.9%, Either day = 76.0%. Logistic regression indicated that older individuals with better balance self-efficacy and walking endurance were more likely to adhere to the SAM protocol. Written information and reminders given.	-
Bergman 2005[81]	Acc	F	n=37, >65 living in independent living (n=17), assisted living (n=8) and nursing home facilities (n=12) in Knoxville, 85.81 (4.16), 70.3%f	Everyday activities for 1 full weekday, removed only for bathing	Wear time	Average wear time = 13.66 (1.26) hours. Retirement homes (n=17) = 12.63 (1.43) Assisted living (n=8) = 13.82 (1.26) Nursing home (n=12) = 14.13 (0.84). Reminders and instructions were provided to participants and staff.	-
Busse 2004[62]	Acc	F, D,	n=10 healthy adults, 43.3 (18.9), 40%f;	Everyday activity for two 7-day	Wear rates	"All subjects were compliant in continuous	-

Reference	MPA	Para	Population (n, descriptor, mean (SD) age, %female)	Activity	Comparator	Statistic and outcome	Qual
		I	n=10 ambulant neurological patients with impairments from different pathologies and restricted walking mobility 59.4 (13.4), 50%f.	periods. SAM worn for 24hr/day		wearing of the monitor throughout the monitoring period. This was confirmed by visual inspection of the data" (No clear definition of non-compliance or visual inspection)	
Cavanaugh 2011[63]	Acc	F, D, I	n=21, ambulatory, community-dwelling, multiple sclerosis patients, 57.6 (12.7), 57.1%f	Everyday activities for 7 days (waking hours only except bathing, sleeping or swimming)	Wear rates	19 (90%) completed >=6 days of recording. Range = 3-7 days	-
Cavanaugh 2012[78]	Acc	F, D, I	n=57, Parkinson's disease patients. Of 33 complete data 67.06 (8.75), 33.3%f	All activity for 7 days, except bathing, swimming or showering, worn on least affected leg. SAM worn at baseline and 12 month follow-up	Wear rates	57 wore monitors at baseline, 37 at following year. Data recording problems (incorrect wear and computer docking issues) = 4 (10.8%). Mean (SD) days of wear = 6.7 (1.1) at baseline (n=57), 6.4 (1.0) at 1 year (n=33). In a few cases, participants decided to wear the monitor 1-2 additional days. In a few cases, activity data from a day were excluded due to minimal activity compared to all other days	-
Danks 2014[64]	Acc	F, D, I	n=23, stroke survivors (>6 months post stroke), walking without assistance (devices allowed), n=16 completers, 66 (range 40-78), 19%f	SAM on non-paretic leg, worn for all waking hours (except bathing and swimming) for 4 weeks	Wear rates	2/23 (8.7%) withdrew due to difficulty attaching the SAM or consistently wearing it. 2/19 (10.5%) did not return SAM with minimum 3 full days/week captured activity as per protocol and one admitted to inflating her baseline step activity.	-

Reference	MPA	Para	Population (n, descriptor, mean (SD) age, %female)	Activity	Comparator	Statistic and outcome	Qual
						2 withdrew for reasons unrelated to study	
Franklin 2006[65]	Acc	F, I	n=8, primary total knee arthroplasty patients with varying characteristics, age and gender NR	6 knee exercises during week 3-12 after surgery. SAM worn for 4 continuous days before surgery and during postoperative week 6.	Wear rates	All patients successfully wore the SAM before and after surgery. No complaints or problems. SAM returned by post	-
Gundle 2014[79]	Acc	F	n=29, lower extremity sarcoma (primary or recurrent) patients treated with limb salvage, 55yrs (range 22-76), 62%f	Everyday activities for 7 days (waking hours only except bathing, sleeping or swimming)	Wear rates	Patients wearing SAM upside down, incorrectly positing or non-wear for >3hr a day were excluded to give n=29. Mean days of data collection in included patients = 12 (3), range 6-16. Non-wear was not defined and n excluded not reported.	-
Hiatt 2011[48]	Acc	F, D, I	n=69 randomised, 62 analysed; intermittent claudication >=1 yr, 67 (SD NR), 37.8%f	Home-based walking exercise 2-3 times per week initially for 30-50min per session; everyday activities for >=10hr/day for seven days at screening, 3 mo and 6 mo	Wear rates	Baseline: 83.5% recorded >10hr ambulatory activity (unclear how detected), 6mo 63% recorded >10hr, 3mo NR	
Kong 2014[66]	Acc	F, D, I	n=46 (37 completed), inactive obese or overweight pregnant women, 26.95 (SD NR), 100%f	Walking programme, increasing from 50-150min/week or no intervention. SAM worn for 4 x 1 week periods	Wear rates	>3 days of valid data: timepoint 1 n=31/37 (84%), timepoint 2 n=36/37 (97%), timepoint 3 n=35/37 (95%), timepoint 4 n=35/37 (95%). Exclusions were mainly due to missing data and SAM	-

Reference	MPA	Para	Population (n, descriptor, mean (SD) age, %female)	Activity	Comparator	Statistic and outcome	Qual
						misplacement. Mean=6 days for each timepoint per ppt. No difference in compliance between intervention or control.	
Moy 2014a[172], Moy 2014b[173], Danilack 2014[174]	Acc	F	n=173, chronic obstructive pulmonary disease patients, >40yrs, stable clinical condition, 71 (8), 1.2%f	Everyday activities for 14 days	Wear rates	5/173 (2.9%) had >=8 no-wear days (defined as <200 steps and <8hrs wear time). 81/2338 days (3.5%, 167x14) met no-wear criteria in final sample. 98 participants wore the monitor twice - 122/3766 days (3%) were no-wear days.. Subsample in separate study had 48/1428 (3%) no-wear days. Unclear how many participants wore the SAM for the entire 14 days	-
Mudge 2008[75]	Acc	F, I	n=54, >6mo post-stroke, able to walk independently but with some residual difficulty, completers (n=40) 69.2 (12.6) , 40%f	Everyday activities for 3 days one week and same 3 days following week, removing for sleeping and showering.	Wear rates	13/54 (24.1%) did not wear SAM for six full days. 40/54 (74.1%) had full six days, n=1 withdrew. Written instructions were provided for SAM.	-
Mudge 2009[68]	Acc	F, I	n=50, >6mo post-stroke, able to walk independently but with some residual difficulty, 67.4 (12.5), 40.8%f	Everyday activities for 3 days, SAM attached to non-paretic leg	Wear rates	49/50 (98%) had 3 complete days of data (not defined)	-
Mudge 2010[170]	Acc	F, I	n=30, healthy adults, 27.7 (8.9), 50%f	Everyday activities for 3 days one week and same 3 days following week, removing for	Wear rates	2 x 3 days = 50%, 2x 2 days = 50%. High attrition. Written instructions were provided for SAM. No significant differences between completers and	-

Reference	MPA	Para	Population (n, descriptor, mean (SD) age, %female)	Activity	Comparator	Statistic and outcome	Qual
				sleeping and showering.		non-completers	
Nguyen 2011[69]	Acc	F, D, I	n=17, COPD patients who had completed pulmonary rehabilitation, 68 (11), 64.7%f	Everyday activity (waking hours only) for 14 days at baseline, 3 mo and 6mo (42 days total)	Wear rates	564 person-days of free-living ambulatory activity were recorded. 33.2 (9.9) valid days (>=10hrs of monitor wear) per person. Mean = 13.9 (0.3) waking hours recorded. 39% had 14 days at each timepoint. 89% had 11 days for each timepoint.	-
Nguyen 2011[69]	Acc	F, D, I	n=60 healthy older adults aged 60-80yrs, 70 (6), 51.7%f	Everyday activity for 7 days	Wear time	Average monitoring days per person = 7.0 (1.5), 442 person days total	-
Nguyen 2013[80]	Acc	F	n=148, chronic obstructive pulmonary disease patients, 66,5 (8.8), 22%f	Everyday activities for 7 days (waking hours only)	Wear time	Median wear time = 7 days. Valid day >=10hr (600min) monitor wear. High anxiety symptoms mean = 874 mins/day wear, low anxiety symptoms mean = 899min/day wear (p=0.29)	-
Parker 2010[70]	Acc	F, D, I	n=27, >18, lower limb prosthesis for >1yr. Full sample (n=52, SAM subsample) were age 55.2 (15.8), 21.2%f.	Everyday activities for 7 days	Wear time	Non-wear (not defined): 4 days = 2 participants (7.4%), 6 days = 3 participants (11.1%)	-
Roos 2012[71]	Acc	F, D, I	n=54 stroke survivors able to walk without assistance from another person, 63.7 (10.4), gender NR; n=18 retired or semi-retired older adults living in the community without walking deficits,	Everyday activity for 3 days (waking hours only except bathing and swimming)	Wear rates	7/72 (9.7%) did not have 3 days of ambulation activity (3/54 (5.6%) stroke survivors, 4/18 (22.2%) older adults)	-

Reference	MPA	Para	Population (n, descriptor, mean (SD) age, %female)	Activity	Comparator	Statistic and outcome	Qual
			68.9 (6.2), gender NR				
Subramony 2012[72]	Acc	F, D, I	n=19, ambulatory individuals with different spinocerebellar ataxias, 56 (10.66), 78.9%f	Everyday activities for 7 x 24hr periods (8 days), day and night	Wear rates	"patients...wore it faithfully through all activities with no interruptions in the recordings."	-
Varma 2014[73]	Acc	F, D, I	n=195, >=60yrs, 66.8 (5.6), 76.5%f. Obesity (8.3%), hypertension (71.2%), osteoarthritis (61.8%) and diabetes (32.6%) fairly prevalent.	Everyday activity for 3-7 days	Wear time	Average data = 4.9 days. Average of 0.8 days (16.4%) removed from analysis. 8/195 non-compliant (defined as a) <201 total steps/day, b) days with <6hr of any activity between wake and sleep c) days with 6hr of consecutive inactivity (<1 step) between wake and sleep and d) subjects self-reported in diary that they hadn't complied). Final 187 participants provided 4.3 days' data (range 1-9) each. Unclear how this subset were chosen from larger RCT. Vague number of days' wear prescribed.	-
White 2012[175]	Acc	F	n=1343, community dwelling adults >50 with a previous knee injury or operation, body weight >median value for age and sex-specific group, knee OA confirmed radiographically, 63.1 (7.8), 60%f	Daily walking for 7 days, waking hours only	Wear rates	Out 1343 eligible participants, 1116 (83%) received a SAM and 1018 (93%) wore it for 3+ days. Of the 229 who did not receive a SAM, 72% refused, 16% had impairments preventing use, 7% had no device available to them, and 5% had other reasons. Unclear why high refusal rate or why no	-

Reference	MPA	Para	Population (n, descriptor, mean (SD) age, %female)	Activity	Comparator	Statistic and outcome	Qual
						devices were available for some. Times were omitted when no steps for >180mins consecutively.	
Problematic Experiences of Therapy Scale high quality studies (n=4)							
Kirby 2014[103]	Int cons	Gen	n=128, patients with chronic dizziness, original sample (n=170) G1 63.9yrs (15.2), 71%f, G2 61.0yrs (14.4), 71%f	Home; up to 12 weeks' dizziness rehabilitation exercises. PETS completed at 12 weeks post-treatment assessment.	Subscale items	Symptoms: α=0.91 Uncertainty: α=0.96 Doubts about efficacy: α=0.94 Practical problems: α=0.84	E
	Int cons	Gen	n=225, Meniere's disease patients with dizziness symptoms, original sample (n=227) G1: 58.0yrs (11.4), 73%f, G2: 60.0yrs (13.6), 63%f	Home; up to 12 weeks' dizziness rehabilitation exercises. PETS completed at 12 weeks post-treatment assessment.	Subscale items	Symptoms: α=0.91 Uncertainty: α=0.93 Doubts about efficacy: α=0.84 Practical problems: α=0.87	E
	Struct val	Gen	n=128, patients with chronic dizziness (labyrinthine cause), original sample (n=170) G1 63.93yrs (15.21), 71%F, G2 61.01yrs (14.42), 71%F	Home; up to 12 weeks' dizziness rehabilitation exercises. PETS completed at 12 weeks post-treatment assessment.	Scale items	PCA: Four factor solution corresponding to 4 hypothesised subscales, accounting for 84% of the variance. All items loaded onto one factor for >=0.67 and <0.10 on others. All factor eigenvalues >0.9. Subscale	E

Reference	MPA	Para	Population (n, descriptor, mean (SD) age, %female)	Activity	Comparator	Statistic and outcome	Qual
						correlations range from -0.22 to -0.53.	
	Struct val	Gen	n=225, Meniere's disease patients with dizziness symptoms, original sample (n=227) G1: 58.0yrs (11.4), 72.5%f, G2: 60.0yrs (13.6), 62.5%f	Home; up to 12 weeks' dizziness rehabilitation exercises. PETS completed at 12 weeks post-treatment assessment.	Scale items	PCA: Four factor solution corresponding to the 4 hypothesised subscales, accounting for 81% of the variance. All factor loadings >=0.60 and <0.11 on other factors. All factor eigenvalues >1. Subscale correlations ranged from 0.12 to 0.36.	E
Problematic Experiences of Therapy Scale low quality studies (n=3)							
Yardley 2006[11]	Crit val	Gen	N=223, Meniere disease, dizziness or imbalance symptoms over 12 months, VR group (full sample n=120) 58 (11.4), 73%f, SC group (full sample n=120) 60.0 (13.6), 62.5%f	Home/ community; vestibular rehabilitation or symptom control 3 months	Self-reported adherence for 9-12 weeks or until asymptomatic (2 questions)	T-test: PETS subscales scores significantly higher in non-adherent group (all p<0.01)	P
Kirby 2014[103]	Cons val	Gen	n=128, patients with chronic dizziness (labyrinthine cause), original sample (n=170) G1 n=83 63.93 (15.21), 71%F, g (N=87) 61.01 (14.42), 71%F	Up to 12 weeks' dizziness rehabilitation exercises. PETS completed at 12 weeks post-treatment assessment and coded into "no barriers" or "some barriers" for each subscale.	12 weeks: Participant self-report adhering for >9 weeks or until asymptomatic	Chi-squared between low adherers. Symptoms: some barriers 47%, no barriers 14.6% (p<0.001). Uncertainty: some barriers 51%, no barriers 26.4% (p<0.01). Doubts: some barriers 50%, no barriers 20% (p<0.001). Practical problems: some barriers 42.9%, no barriers 25% (p<0.05).	P

Reference	MPA	Para	Population (n, descriptor, mean (SD) age, %female)	Activity	Comparator	Statistic and outcome	Qual
Kirby 2014[103]	Cons val	Gen	n=227, Meniere's disease patients with dizziness symptoms in the last 12mo, Original sample G1: n=120 58.0 (11.4), 72.5%f, G2: n=120 60.0 (13.6), 62.5%f	Up to 12 weeks' dizziness rehabilitation exercises. PETS completed at 12 weeks post-treatment assessment and coded into "no barriers" or "some barriers" for each subscale.	6 mo: self-report adhering after 12 weeks (any duration)	Chi-squared for maintenance. Symptoms: no barriers 47.5% ($p<0.01$), doubts 47.5% ($p<0.01$).	P
Problematic Experiences of Therapy Scale acceptability studies (n=1)							
Yardley 2006[11]		G	n=240 (2 intervention groups), individuals with Meniere's disease experiencing dizziness or imbalance in last 12 mo; VR group=58.0 (11.4), 72.5%f; SC group= 60.0 (13.6), 62.5%f	Home-based booklet vestibular rehabilitation (VR; daily balance training exercises and how to tailor to symptoms) or symptom control (SC; relaxation and breathing exercises) for 3 mo	Return rates	225/240 (93.8%) completed PETS at 3mo. No information on individual item rates. PETS was packaged with other questionnaires.	-
Adherence diaries high quality studies (n=6)							
Wilbur 2001[106]	Crit val	F, D	n=156, sedentary African American and Caucasian women, mean (SD) age NR (range 45-65yrs), 100%f	Home/community; moderate intensity 24-week walking program.	Polar Vantage XL heart rate monitors	Frequency = +4.33 (SD 7.09) walks on log ($r=0.962$, $p<0.01$), duration = +5.0 (SD 8.08) min on monitor ($r=0.536$, $p<0.001$)	E
	Crit val	F, D	n=139, sedentary African American and Caucasian women, mean (SD) age NR	Home/community; moderate intensity 24-week walking	Change in fitness (VO_{2max} in a treadmill test) between	Frequency of walks and change in VO_{2max} = 0.270 ($p<0.01$); average duration and change in VO_{2max} =	G

Reference	MPA	Para	Population (n, descriptor, mean (SD) age, %female)	Activity	Comparator	Statistic and outcome	Qual
			(range 45-65yrs), 100%f	program.	baseline and post-intervention.	-0.088 (NS). Similar correlations found between monitor variables and change in VO_{2max}.	
Jeffrey 2012[107]	Crit val	F	n=135, adults over 60 with osteopenia of the hip or spine, age 82.3yrs (7.1), 67%f	Home/community; use of active or sham vibrating platform for 10min per day for up to 3yrs	Radio frequency identification card system	Mean bias close 0.02 and narrow LOA (graph only), ICC=0.96	G
Lindseth 2005[104]	Crit val	D	n=94, women within the first 12 weeks of pregnancy, 27yrs (4.6), 100%f	Home/community; exercise for 3 days at 14 and 28 weeks' gestation	Mean Accu-split Power Stride pedometer counts per day	r=0.49 (p<0.02)	G
Shang 2009[105]	Crit val	F, D, I	n=126, newly diagnosed cancer patients, 60.2yrs (10.6), 61%f	Home/community; walking programme 5 days per week for 5-35 weeks	Pedometer steps (brand NR) worn for whole study (intervention) or first and last two weeks of study (control)	ρ=0.42 (p<0.001)	G
	Cons val	F, D, I	n=126, newly diagnosed cancer patients, 60.2yrs (10.6), 61%f	Home/community; walking programme 5 days per week for 5-35 weeks	Physical Activity Questionnaire METs of previous 4 weeks (administered at end of study)	ρ=0.67 (p<0.001)	G
Adherence diary low quality studies (n=30)							
Castro 2002[111]	Crit val	F, D, O	n=9, sedentary and healthy post-menopausal women providing unpaid care to a relative with dementia, total sample (n=51) 62.2 (9.3), 100%f	Hone; exercise programme of increasing intensity - 4x 30-40mins per week for 12 mo.	Solid-state two-channel portable microprocessor recording heart and body movement for one 3 day period	87.5% agreement between continuous bouts of physical activity at moderate intensity heart rate as recorded by the monitor and logs.	P
Dougherty	Crit val	D	n=77, single or dual chamber	Home; 8 week home-based	Fitness (peak VO2 measured by	Test unclear, participants achieving >=80%	F

Reference	MPA	Para	Population (n, descriptor, mean (SD) age, %female)	Activity	Comparator	Statistic and outcome	Qual
2015[123]			Implantable Cardioverter Defibrillator patients taking beta-blockers and willing to complete the exercise program, of n=84 at start 56.1(12.1), 20.2%f	aerobic training followed by 16 week maintenance with increases in heart rate targets	CPET using symptom-limited treadmill test)	adherence during aerobic conditioning achieved significantly higher peak VO2 (27.7 (7.0) vs 24.3 (6.7), p=0.03) and associated exercise outcomes (data NR)	
Franklin 2006[65]	Crit val	F	n=8, total knee arthroplasty patients, total sample (n=21) 69 (SD NR), 67%f	Home; daily leg exercises for 9 weeks	SAM-recorded periods of sustained step activity for 4 days during week 6	All diary-reported exercise sets were recorded as high activity peaks on the SAM, plus extra	P
	Crit val	F, O	n=21, total knee arthroplasty patients, 69 (SD NR), 67%f	Home; daily leg exercises for 9 weeks	Physical composite score of SF-12	Regression: daily repeats in leg exercise and PCS changes: slope = 0.34 (p=0.10), knee reflex repeats and PCS changes: slope = 0.31 (p=0.09)	P
Jakicic 1998[118]	Crit val	F, D	N=50, overweight women, mean & SD NR (range 25-50), 100%f	Home/ community; part of 20 week trial comparing long (1x20-40min per day) and short (2-4x10min per day) bouts of exercise	Tri-Trac accelerometers (6 days between randomly allocated between weeks 5 and 10)	29 (58%) under/accurately reported session frequency, mean difference -1.5±2.4 sessions, 88.5±24.2% sessions matched 21 (42%) over-reported, mean difference 2.9±2.3 sessions, 44.0±28.1% sessions matched 26 (52%) under/accurately reported mins per week: mean difference in duration -42.8±45.5 mins 24 (48%) over-reported: mean difference in duration 71.5±78.4 mins	P

Reference	MPA	Para	Population (n, descriptor, mean (SD) age, %female)	Activity	Comparator	Statistic and outcome	Qual
Lyngcoln 2005[176]	Crit val	F	n=15, individuals =<18 with a distal radial fracture managed conservatively in a cast, 65.1 (11.1), 93%f	Home; home-based hand therapy exercises for six weeks	Functional status scale (modified Levine questionnaire)	Pearson's r=0.63 (p<0.05)	P
					Jebsen test of hand function	Spearman's Item 1: ρ=0.40, item 2: ρ=0.51, item 3: ρ=0.25, item 4: ρ=0.26, item 5: ρ=0.54 (p<0.05), item 6: ρ=0.39, item 7: ρ=0.32	P
					Active wrist extension	Pearson's r=0.46	P
					Hand dynamometer (grip strength)	Pearson's r=0.41	P
					Pain (VAS)	Pearson's r=0.54 (p<0.05)	P
					all of the above	Pearson's r=0.44 Number of exercises performed and all outcome measures r =0.29	P
McAuley 1991[125]	Crit val	F, D	n=48, sedentary healthy female university employees, 39 (SD NR), 100%f	Home/community; twice weekly supervised 1hr exercise classes for eight weeks, plus home aerobic exercise of >15min	Body weight (calibrated balance)	MANOVA by participants >median overall adherence: p<0.1	P
					Body fat (three site method of skinfold thickness)	MANOVA by participants >median overall adherence: NS	P
Moseley 2006[120]	Crit val	F	N=51, complex regional pain syndrome type 1 of one limb diagnosis from their treating practitioner, full sample n=67:	Home; RCT of overt vs covert adherence monitoring of computer-based motor imagery	In-house software recording performance time and duration.	Overt monitoring (n=24): 5% (95% CI 0.51–9.48) underestimation. Covert monitoring (n=27): 10% (95% CI 3.0–16.9)	P

Reference	MPA	Para	Population (n, descriptor, mean (SD) age, %female)	Activity	Comparator	Statistic and outcome	Qual
			32 (10), 48%f	home exercise programme		overestimation. Longer symptom duration correlated with greater inaccuracy.	
Sassi-Dambron 1994[177]	Crit val	F, D	n=42, patients with chronic obstructive pulmonary disease, 61.4 (7.6), 24%f	Home/community; initially 3 short walks increasing to one long walk up to a goal of at >30 min per day for 8 weeks	Maximum exercise tolerance (Maximum METs, symptom-limited treadmill test collecting blood gases)	Pearson's: Total minutes walked r=0.32 (p<0.05), total days walked r=0.05	F
					Maximum exercise tolerance (peak VO2, symptom-limited treadmill test collecting blood gases)	Total minutes walked r=0.18, total days walked r=0.05	F
					Endurance time (constant work treadmill test)	Total minutes walked r=0.37 (p<0.05), total days walked r=0.22	F
Shaw 2005[119]	Crit val	D	n=4, chronic traumatic brain injury >1 yr prior to participation with relative hemiparesis, full sample (n=22) 39.3 (14.4), 35%f	Home; constraint-induced movement therapy (mitt wear) and other behavioural techniques for 2 weeks	Sensor and timing device sewn into mitt	Median ICC = 0.97	P
	Crit val	F, D	n=22, chronic traumatic brain injury >1 yr prior to participation with relative hemiparesis, 39.3 (14.4), 35%f	Home; constraint-induced movement therapy (mitt wear) and other behavioural techniques for 2 weeks	Motor Activity Log Quality of Movement	T-test adherent (>57%) vs non-adherent = 1.8 vs 1.3 (p=0.065), correlation in less adherent participants r = 0.68, none among more adherent participants	P

Reference	MPA	Para	Population (n, descriptor, mean (SD) age, %female)	Activity	Comparator	Statistic and outcome	Qual
Wilcox 2004[178]	Crit val	F, D	n=18, >=65yrs, no cardiovascular disease, stroke or musculoskeletal problems, active =<2 twice a week, full sample (n=103) 70.2 (4.1), 65%f	Home; either Fit & Firm - brisk walking and weights or resistance bands or Stretch & Flex - home stretches twice per week, both 12 mo	Solid-state portable microprocessor (Vitalog Corp) recording heart and body movement for 3 days	10/12 (83%) in Fit & Firm condition showed evidence of an exercise bout of >=20min on days they reported engaging in a home exercise session. Only 1/6 in Stretch & Flex condition.	P
Yuen 2013[122]	Crit val	F, D	N=11, sedentary African-American women with systemic lupus erythematosus experiencing fatigue, 48.8 (14), 100%f	Home; WiiFit exercises for 30min, 3 times a week for 10wk	WiiFit records	Mean difference (95% LOA) session duration: 3.8min (35 to -27 mins), 12.7% difference. ICC=0.40 (95% CI 0.27-0.51). 72% sessions matched between methods.	P
Aurilio 2000[126]	Cons val	F, D, O	n=30, sedentary healthy women aged 30-50yrs, 41 (6.3) 100%f	Home/community; 12 week walking programme	Behavioral Risk Factor Surveillance System Exercise questionnaire (telephone interview)	Days walked per week ICC= 0.77 (p=0.01), Spearman's ρ=0.62 (p<0.01) Mins walked per week ICC=0.08 (p=0.34), ρ=0.54 (p<0.01) Miles walked per week ICC= 0.04 (P=0.43), ρ=0.63 (p<0.01)	F
Henry 1999[179]	Cons val	F	N=15, healthy adults over 65, 72.8 (SD NR, range 67-82), 73%f	Home; 2, 5 or 8 general strengthening exercises 10 times a day	Performance accuracy assessment tool developed for study (scored by PTs), inter-rater reliability 0.87 for first exercise and 0.93 for second	Correlations: r=0.54	P
Khalil 2012[180]	Cons val	F	n=15, mid-stage Huntingdon's disease	Home; PT-prescribed exercises	Intrinsic Motivation Inventory	Spearman's correlations	P

Reference	MPA	Para	Population (n, descriptor, mean (SD) age, %female)	Activity	Comparator	Statistic and outcome	Qual
			with difficulties with walking or balance and stable medical regimen, 53.6 (range 25-78), 47%f	on a DVD at least 3 times/week for 8 weeks	(multidimensional questionnaire about perceived interest, enjoyment, competence, effort, value and usefulness while performing a given activity)	Subscales: Interest/enjoyment: 0.09, perceived competence: 0.39, effort/importance: 0.37, pressure/tension -0.63 ($p<0.05$), value/usefulness: -0.24	
McAuley 1991[125]	Cons val	F, D	n=48, sedentary healthy female university employees, 39 (SD NR), 100%f	Home/community; twice weekly supervised 1hr exercise classes for eight weeks, plus home aerobic exercise of >15min	Self-motivation (self-motivation inventory)	MANOVA by participants >median overall adherence: NS	F
					Self-efficacy (questionnaire of barriers)	MANOVA by participants >median overall adherence: $F(3.43) = 3.37$, $p<0.05$	F
					Post-program perceptions (self-developed questionnaire of program success, goal achievement, improvements in conditioning and class enjoyment)	MANOVA by participants >median overall adherence: NS	P
Wilbur 2005[181]	Cons val	F, D	n=72, sedentary healthy, employed Black and White women 45-65, full sample (n=90) 49.9 (4.8), 100%f	Home/community; home-based moderately intense walking programme 4 times per week in a target heart rate range, progressing from 20 to 30 min	Exercise recorded by AD in maintenance phase	Multiple regression: in a model of exercise self-efficacy, physiological measures, background characteristics and adherence during intervention phase, adherence during intervention ($p<0.01$) and self-efficacy ($p=0.02$) were significant	F

Reference	MPA	Para	Population (n, descriptor, mean (SD) age, %female)	Activity	Comparator	Statistic and outcome	Qual
				over 24 weeks, followed by 24 weeks maintenance stage		predictors of walking during maintenance (40% variance explained overall)	
Steele 2008[130]	Resp to change	D	n=106, adults >45 with chronic lung disease and shortness of breath with diminished functioning due to a pulmonary problem who completed pulmonary rehabilitation, 67 (SD NR), 8%f	Home/community; adherence to exercise after pulmonary rehabilitation	Exercise adherence intervention with weekly phone calls, 1 home visit, pedometer and exercise handbook.	T-test self-reported minutes of activity increased in adherence intervention group short term: intervention group 3(39), control -13 (26) (p=0.015). Long term: intervention 1(45), control -8 (31) (p=0.335)	F
Lindseth 2005[104]	TRR	D	N=94, women within the first 12 weeks of pregnancy, 27 (4.6), 100%f	Home/community; activity recorded for three days at 14 and 28 weeks' gestation	Time interval: 14 weeks	Pearson's correlations: r=0.61 (p<0.01)	F
Adherence diaries acceptability studies (n=19)							
Ada 2003[14]	Acc	U	n=14 (control group), stroke survivors 6mo-5yrs previously, 66 (11), 28.6%f	Home exercise programme (strength, balance, coordination) 3 times per week for 4 weeks	Return rate	Return rate = 8/14 (57.1%) Two subjects had lost their logs and 1 subject was lost to follow-up. Logs returned independently of sessions	-
Bauldoff 2001[113]	Acc	F, D	n=408, >18, 5-30 days after stroke without contraindications to exercise, 62.02 (12.74), 45.1%f	Home-based 8 week walking programme with or without music, , 2-5 days per week for >20mins	Return rates	100% complied with log recording. Logs were returned every 4 weeks at appointments. pedometer-recorded distance walked also recorded (also completed by all)	-
Bodrie	Acc	F, D,	n=40, discharged from phase II cardiac	12 weeks home-based	Return rates	Mailed exercise logs to the investigator every 2	-

Reference	MPA	Para	Population (n, descriptor, mean (SD) age, %female)	Activity	Comparator	Statistic and outcome	Qual
1999[112]		I, O	rehabilitation, aged >=45 (male) or 55 (female), prior or current coronary heart disease and risk factors, 69 (11), 33.3%f	prescribed cardiac rehabilitation exercise programme		weeks. n=3 (15%) stopped mailing log and so ended study. 7 others dropped out for other reasons. Up to 3 telephone and mail reminders per diary: additional calls or letters required: 0: n=19, 1: n=4, 2: n=4, 3: n=3	
Castro 2002[111]	Acc	F, D, I, O	n=51 (exercise group), sedentary and healthy post-menopausal women >=50 living with and providing unpaid care to a relative with dementia; 62.16 (9.33), 100%f	Exercise programme of increasing intensity - 4x 30-40mins per week for 12 mo	Return rates	Mean return rate = 8.81/12 (SD 4.39). Returned monthly by mail, with phone contacts to obtain info if not returned	-
Dyson 1997[182]	Acc	D, O	n=93 , participants with increased fasting plasma glucose (range of 5.5 to 7.7 mmol * L -1 on two consecutive tests 2 weeks) and >=1 risk factor for diabetes, Full sample (n=227) 50 (9), 59%f	20-30mins exercise 2-3 times per week, increasing to 5-6 times per week over 12mo	Return rates	Return rate: 51/93 (55%) returned 3/4 diaries, 15 (16%) returned none. Diaries were collected at each 3mo visit, unclear how many returned diaries in control group	-
Franklin 2006[65]	Acc	F	n=31, primary total knee arthroplasty patients, 69yrs (SD NR), %f NR	6 knee exercises during weeks 3-12 post-TKA	Completion rates	3/31 (10%) returned blank logs. 21 remaining participants recorded >=3 days exercise per week. Log completion consistent over weeks 3-12. Weekly collection by study coordinator, high attrition (n=7).	-

Reference	MPA	Para	Population (n, descriptor, mean (SD) age, %female)	Activity	Comparator	Statistic and outcome	Qual
Frost 2004[183]	Acc	F, D, O	n=26, 55-75yrs, OA or degenerative joint disease of the hip diagnosis, unilateral total arthroplasty and completion a course of outpatient or home-based physical therapy, experimental 66.2 (5.2), 84.6%f; control 65.9 (6.8), 63.5%f	Specialized Motivational Exercise Counselling intervention or to a control group that received usual care recorded for 2 months in diary	Return rates	1 participant did not return diaries at weeks 7 and 8 due to a life-threatening injury to a family member. 1 participant did not return diaries for weeks 5-8 as they developed a pressure ulcer on their heel.	-
Koumantakis 2005[21]	Acc	F	n=45, low back pain, SEE group: n=29, 39.2 (11.4) %f NR, GE group: n=26, 35.2 (9.7), %f NR	General exercise with (SEE) or without (GE) trunk muscle stabilisation exercises for up to 30mins three times per week for 8 weeks	Completion rates	35/45 (77.8%) completed a diary (not defined)	-
Long 2004[24]	Acc	F	n=312, low back pain patients, n=206 completing study 42.2 (SD NR), 45%f	Lumbar exercises (three different types) for 2 weeks	Return rates	68% (137/201) returned diaries	-
Loudon 1999[115]	Acc	F, D, I, O	n=12 (completed trial), adults with mildly active or remitted Crohn's disease not involved in any exercise in the previous year, 38.3(7.5), 83.3%f	Three sessions per week of structured walking, either indoor as a group or individually, progressing from 20 to 35 mins per session and increasing distance and intensity over 12 weeks	Completion rates	"all logbooks were kept in order and were found to be well documented after the 12 week program. All data in the logbooks were complete for all 12 subjects." Logbooks were reviewed weekly by one investigator, including those who missed group sessions	-

Reference	MPA	Para	Population (n, descriptor, mean (SD) age, %female)	Activity	Comparator	Statistic and outcome	Qual
Martinez-Silvestrini 2005[25]	Acc	F	n=94 (three groups), chronic (>3mo) lateral elbow pain, 45.5 (7.7), 46.8%f	Stretching, concentric or eccentric exercises (3x10 sets per day) for six weeks	Return rates	100% return rate in all subjects completing analysis (81/94) Failure to enter daily data for >10 days was considered non-compliant and resulted in exclusion. N=1 didn't comply, but unclear if this with question naire, diary, or low adherence. Subjects remunerated for completed log books	-
Roddey 2002[29]	Acc	G	n=108, full-thickness rotator cuff tear patients undergoing arthroscopic repair, G1 n=54 58.7, (10.6), 35%f, G2 n=54 57.2 (9.1), 39%f	Home PT exercises, either through a videotape or 4 PT sessions for 6 mo	Completion rates	n=73/106 (68.9%) returned all four logs. Compliance criteria determined a priori: "fully compliant" = all 4 logs and 70% adherence, n=61; "partially compliant" = 3-4 logs and 50-69% adherence, n=12; "noncompliant" = <3 logs or <50% adherence, n=33 (31.1%). 2 subjects lost to follow up. Telephone reminders used. Logs returned every 6 weeks with SAE	-
Sassi-Dambron 1994[177]	Acc	F, D	n=57, symptomatic chronic obstructive pulmonary disease patients, n=42 sample responding 61.4(7.6), 24%f	Daily walks, initially 3 short walks increasing to one long walk of up to 30 mins (also supervised exercise: treadmill and an upper-body ergometer, upper-body weight training	Completion rates	42/57 had completed diaries "the others were either not collected or not completed" (numbers for each unclear)	-

Reference	MPA	Para	Population (n, descriptor, mean (SD) age, %female)	Activity	Comparator	Statistic and outcome	Qual
				over 8 week block - not recorded in diary)			
Schachter 2003[54]	Acc	F, D	n=143, sedentary women with fibromyalgia, groups (1) long bout exercise (LBE) (n=51) 41.3 (8.67) 100%f; (2) short bouts exercise (SBE) (n=56) 41.9 (8.57), 100%f and (3) control (no exercise) (n=36) 42.5 (6.69), 100%f	16 week progressive low-impact aerobics programme using a videotape, in long or short bouts	Return rates	45/56 (80.4%) in short bout group, 42/51 (82.4%) in long bout group completed the study and submitted logs (81.3% overall) (completion not defined).=	-
Shang 2009[105]	Acc	F, D, I	n=126, newly diagnosed cancer patients aged 21+ with no evidence of metastatic disease, scheduled to receive chemotherapy or radiotherapy, 60.2 (10.6), 39%f	Individualised home-based walking and muscle-strengthening exercise program 5 days per week, throughout cancer treatment (5-35 weeks). Control participants continued their usual physical activity."	Completion rates	17 participants (13.49%, 4 (5.9%) in intervention, 13 (22.4%) in control) had "significant" missing data for certain weeks. Missing data were imputed from telephone logs (correlated highly with exercise log on other weeks). Logs were mailed back at the end of each week. Research nurses would call if logs were not returned on time.	-
Webb-Peploe 2000[32]	Acc	F, D, I	n=24, patients with ischaemic and idiopathic dilated cardiomyopathy, 53 (SD NR), 4.2%f	Progressive exercises and 20min bicycle ergometry at least 5 days a week for 8 weeks	Completion rates	18/24 (75%) completed diaries, 16/24 (66.7%) correctly filled out revolutions pedalled per day. Unclear if 18 participants completing diaries were same as included in final analysis	-
Williams	Acc	F, D,	n=46, non-insulin dependent diabetes	Usual exercise over 2 weeks	Return rates	Return rate = 100% (also fully completed, not	-

Reference	MPA	Para	Population (n, descriptor, mean (SD) age, %female)	Activity	Comparator	Statistic and outcome	Qual
1996[114]		I, O	mellitus patients, 60.3 (SD NR), 45.7%f			defined)	
Wilson 2005[110]	Acc	F, D, I	n=13, aged 18-65 who received blood stem cell or bone marrow transplant >6mo prior to participation and low leisure time physical activity, n=17 full sample, 48.9 (10.4), 64.7%f	Exercise 3+ times/week for 20mins continuously in their HR training zone for 12 weeks	Return rates	9/13 (69.2%) returned completed exercise diaries. High number of withdrawals and refusals (76% intervention acceptability)	-
Yuen 2012[33]	Acc	F	n=26 (intervention group), adults with systemic sclerosis, 51.9 (14.3), 80.8%f	Daily orofacial exercises, teeth brushing and flossing for 6 mo	Return rates	Return rate = 11/13 (84.6%) 2 did not return monthly charts; unclear if all charts were returned for others. Diaries were posted (SAE provided) with telephone reminders	-
Bassett & Prapavessis' scale high quality studies (n=3)							
Bassett 2011[2]	Int cons	Gen	n=70, patients with an ankle sprain undergoing PT, G1 35.9yrs (13.4), G2 34.9yrs (12.2), G3 34.9yrs (13.1), 57%f	Home; PT program with PMT, attention control or no information	Cronbach's alpha was calculated from the means of participants' mean scores for each subscale	α=0.63	G
	Cons val	Gen	n=69, patients with an ankle sprain undergoing PT, (full sample n=70) G1 35.9yrs (13.4), G2 34.9yrs (12.2), G3 34.9yrs (13.1), 57%f	Home; PT program with PMT, attention control or no information	Intentions to attend clinic appointments and to adhere to home-based therapy (2 items based on Theory of Planned Behaviour, 7 point Likert scale)	Home exercise subscale & intentions to attend clinic appointments r=0.24 (p=0.05) Home exercise subscale & intentions to adhere to home therapy r=0.25 (p=0.05) Ankle elevation adherence and intentions to adhere to home therapy r=0.38 (p=0.01)	G

Reference	MPA	Para	Population (n, descriptor, mean (SD) age, %female)	Activity	Comparator	Statistic and outcome	Qual
						Other variables not correlated	
	Crit val	Gen	n=69, patients with an ankle sprain undergoing PT, (full sample n=70) G1 35.9yrs (13.4), G2 34.9yrs (12.2), G3 34.9yrs (13.1), 57%f	Home; PT program with PMT, attention control or no information	Ankle function (Lower Limb Task Questionnaire and Motor Activity Scale) at the end of the PT programme.	No significant correlations	G
Bassett & Prapavessis' scale low quality studies (n=1)							
Bassett 2007[3]	Int cons	Gen	N=47, diagnosis of acute ankle sprain, 30 (12.4), 40%f	Clinic; participants randomised to home-based or clinic-based three-phase physical therapy programme	Patients rated their adherence at the beginning of each clinic appointment. Means were calculated for each participant and subscale	Cronbach's α=0.78	F
Borg 6-20 Rating of Perceived Exertion scale high quality studies (n=3)							
Miller 1985[132]	Crit val	I	n=113, healthy adults, f=64.8yrs (SD NR) m=64.3yrs (SD NR), 52%f	Lab; walking on the spot for 2min at brisk, comfortable pace	Heart rate (Exersentry heart rate monitor)	r=0.34 p=0.0002	E
	Crit val	I	n=89, healthy adults, f=64.8yrs (SD NR) m=64.3yrs (SD NR), 52%f	Lab; 600m walk at brisk, comfortable pace	Heart rate (Exersentry heart rate monitor).	R=0.33 p=0.002	G
Julius 2012[133]	Cons val	I	n=50, 65+ adults with mobility limitations, 76.8yrs (5.5), 66%f	Lab; ~15m walk at self-selected, comfortable pace	Gait speed (GaitMatII)	ρ=-0.16 (p=0.27)	G
					Modified Gait Abnormality Rating Scale	ρ=0.21 (p=0.15)	G
					Energy cost of 3min treadmill walk at self-selected pace	ρ=0.01 (p=0.95)	G

Reference	MPA	Para	Population (n, descriptor, mean (SD) age, %female)	Activity	Comparator	Statistic and outcome	Qual
					(oxygen consumption by open-circuit spirometry)		
					Late Life Function and Disability Questionnaire	Function subscale ρ=-0.17 (p=0.24) basic lower extremity subscale ρ=-0.20 (p=0.17), advanced lower extremity subscale ρ=-0.11 (p=0.47), disability subscale ρ=-0.07 (p=0.61)	G
					Survey of Activities and Fear of Falling in the Elderly	Fear subscale ρ=0.26 (p=0.07), activity subscale ρ=0.13 (p=0.35), restriction subscale ρ=0.02 (p=0.88)	G
					Physical activity during daily activities (Actigraph accelerometer)	ρ=0.30 (p=0.04)	G
					Gait Efficacy Scale	ρ=-0.33 (p=0.02)	G
Borg 6-20 scale low quality studies (n=57)							
Gamberale 1972[138]	Crit val	I	N=12, adult healthy men, 26.5 (SD NR, range 20-35), 0%f	Lab; randomly assigned 6min exercise tasks including lifting weights, pushing a wheelbarrow and cycling	Heart rate (telemetry, Medenik, Honeywell) at randomly chosen values for each workload	Pearson's: Wheelbarrow activity r=0.42, lifting weights r=0.64, cycle ergometer r=0.94	P
Goslin 1986[139]	Crit val	I	N=10, healthy Caucasian males, 24.3 (2.8), 0%f	Lab; treadmill tests with varying backpack loads and speeds	Heart rate (Hewlett-Packard telemetry)	Correlations: r=0.47	P
					Oxygen uptake (VO_2),	r=0.75	P

Reference	MPA	Para	Population (n, descriptor, mean (SD) age, %female)	Activity	Comparator	Statistic and outcome	Qual
					Ventilation index (VI) (open circuit chamber with Beckman OM-14 and LB-2 oxygen and carbon dioxide analysers).	r=0.58	P
Goss 2003[184]	Crit val	I	n=24, healthy adults, F=22.9 (5.1), M=22.4 (1.6), 50%f	Lab; 12 6min exercise trials on a Nordic Track Total Body System, with different combinations of arm and leg exercises. Three six min exercise trials were completed per session, separate by >24hr.	Oxygen consumption (ml/kg/min) - open circuit spirometry	Pearson's r=0.52	P
					Oxygen consumption (%VO2 peak) - open circuit spirometry	r=0.54	P
					Respiratory exchange ratio - open circuit spirometry	r=0.52	P
					Heart rate - Eaton Care Telemetry	r=0.42	P
Lagally 2004[185]	Crit val	I	n=20, 10 novice and 10 recreationally trained women, full sample (n=28) novice 21.6 (1.5) 100%f, recreational 21.9 (2.2) 100%f	Lab; 8 repetitions at 60% 1RM, 6 repetitions at 80% 1RM of a bench press exercise	Muscle activity using electromyography (MP100 EMG system)	No significant correlations (statistics NR) between RPE and EMG	P
O'Neill 1992[134]	Crit val	I	N=48, healthy women with uncomplicated singleton pregnancies, 30(5), 100%f	Location NR; aerobics class at 13-28 weeks gestation, RPE estimated at the end of each exercise track	Heart rate (Polar Sports Tester PE3000)	Pearson's r=0.27 (p>0.05)	F

Reference	MPA	Para	Population (n, descriptor, mean (SD) age, %female)	Activity	Comparator	Statistic and outcome	Qual
O'Neill 1992[134]	Crit val	I	N=11, healthy women with uncomplicated singleton pregnancies, 30(3), 100%f	Location NR; 26min treadmill exercise at 23-28 and 34-37 weeks gestation and again at 8+ weeks after delivery.	ECG (Hewlett Packard 1405A)	Pearson's: r=0.83 p<0.01	P
O'Neill 1992[134]	Crit val	I	N=12, healthy women with uncomplicated singleton pregnancies, 32(4), 100%f	Location NR; 12min exercise on a bicycle ergometer at 34-38 weeks gestation and at 8+ weeks postpartum	ECG (Hewlett Packard 1405A)	Pearson's: r=0.74, p<0.015	P
O'Neill 1992[134]	Crit val	I	N=24, healthy women with uncomplicated singleton pregnancies, 30(3), 100%f	Location NR; 30min circuit training between 20-28 weeks gestation	Heart rate (Polar Sports Tester PE3000)	Pearson's r=0.39 p>0.05	P
O'Neill 1992[134]	Crit val	I	N=29, healthy women with uncomplicated singleton pregnancies, 31(4), 100%f	Location NR; aerobics class at 29-39 weeks gestation	Heart rate (Polar Sports Tester PE3000)	Pearson's r=0.35 p>0.05	P
Pandolf 1978[186]	Crit val	I	n=15, highly fit males, 20.2 (SD NR, range 18-22), 0%f	Lab; climbing and descending a laddermill and stool stepping at three different rates, using foot over foot climbing and both feet to same rung climbing, for five mins each	Heart rate (Sanborn model 100 Viso Recorder)	Regression: foot over foot climb descent r=0.56, ascent r=0.74, both feet to same rung descent r=0.23, ascent r=0.53, stool stepping r=0.74	P
					Oxygen consumption (expired air and spirometer)	Foot over foot climb descent r=0.60, ascent r=0.72, both feet to same rung descent r=0.45, ascent r=0.63, stool stepping r=0.82.	P
Pollock	Crit val	I	n=13 healthy adults, 53.5 (5.4), 85%f	Lab; WiiFit session, including	Heart rate (30s left radial pulse	Pearson's r=0.32 (p value not calculated due to	P

Reference	MPA	Para	Population (n, descriptor, mean (SD) age, %female)	Activity	Comparator	Statistic and outcome	Qual
2013[135]				5min warm up, exercise from two WiiFit categories for 15min each, 5min cool down, performed on two days with different exercises. RPE assessed during final 30s of each exercise category.	palpation by experienced study coordinators)	repeated measures, mixed effects model analysis found significant association p<0.001)	
Row 2012[187]	Crit val	I	n=21, healthy older adults, 76.6 (5.5), 43%f	Fitness centre; concentric and eccentric resistance training using a seated leg press with 50% to 150% body weight loads (4-5 reps), administered in a random order	%1RM lifted in a second session that equated to the loads lifted in the first session. Lowest load in each 10% range and corresponding RPE were used.	Regression: average RPE for each load strongly predicted average %1RM for each load (R^2 = 99.5%, p<0.001)	P
Schaeffer 1995[136]	Crit val	I	N=16, healthy women with previous instructional experience in aerobic dance, 23.0 (3.7), 100%f	Lab; 8 trials - 1min each for 8 minutes x 3 (T1, T2, T3) including 4 steps (jumping jack, power jack, jog and march) at 2 cadences (124 or 138 bpm) along with a leader	Heart rate (CIC Polar heart monitor)	T1: r=-0.18, T2 r=0.01, T3 r=0.26. Partial correlations controlling for absolute oxygen consumption: T1: r=-0.16, T2 r=0.02, T3 r=0.25	P
					Absolute VO2 consumption (Sensormedics, 2900 measurement cart)	T1: r= -0.13, T2 r= -0.01, T3 r=0.14	P
					Relative VO2 consumption (Sensormedics, 2900	T1: r=0.25, T2 r=0.20, T3 r= -0.02	P

Reference	MPA	Para	Population (n, descriptor, mean (SD) age, %female)	Activity	Comparator	Statistic and outcome	Qual
					measurement cart)		
					%VO2 max(Sensormedics, 2900 measurement cart)	T1: r=0.30, T2 r=0.08, T3 r=0.01	P
					%max HR (CIC Polar monitor)	T1: r=-0.02, T2 r=-0.02, T3 r=0.33. Partial correlations controlling for absolute oxygen consumption T1 r= -0.03, T2 r= -0.02, T3 r=0.34	P
					Volume of carbon dioxide expired	T1: r=-0.02, T2 r=0.04, T3 r=0.21. Partial correlational analyses controlling for absolute oxygen consumption T1 r=0.33, T2 r=0.25, T3 r=0.35	P
					Ventilation rate (Sensormedics, 2900 measurement cart)	T1: r=0.23, T2 r=0.20, T3 r=0.32. Partial correlational analyses controlling for absolute oxygen consumption T1 r=0.51, (P<0.05) T2 r=0.48, T3 r=0.33	P
					O2 pulse	T1: r=-0.01, T2 r= -0.05, T3 r=0.01	P
					Gross energy cost (kcal/min)	T1: r= -0.11, T2 r=0.00, T3 r=0.22. Partial correlational analyses controlling for absolute oxygen consumption T1 r=0.35, T2 r=0.27, T3 r=0.20	P
					Net energy cost (kcal/min)	T1: r= 0.15, T2 r= -0.13, T3 r=0.02. Partial correlational analyses controlling for absolute	P

Reference	MPA	Para	Population (n, descriptor, mean (SD) age, %female)	Activity	Comparator	Statistic and outcome	Qual
						oxygen consumption T1 r=0.08, T2 r= -0.39, T3 r= -0.39	
					Respiratory exchange ratio (Sensormedics, 2900 measurement cart)	T1: r=0.37, T2 r=0.34, T3 r=0.43. Partial correlational analyses controlling for absolute oxygen consumption T1 r=0.40, T2 r=0.34, T3 r=0.47	P
Schaeffer-Gerschutz 2000[137]	Crit val	I	N=25, aerobically trained women, 21.0 (1.0), 100%f	Lab; 4 combinations of 3min aerobic steps include dynamic (D) and static (S) high and low impact arm exercises in a random order, following videotaped directions and with a 3min break in between	Heart rate (Quinton 4000 ECG)	Pearson's correlations High impact: D r=0.23, S r=0.34, Low impact D r=0.20, S r= -0.14	P
					Percentage maximum heart rate (Quinton 4000 ECG)	High impact: D r=0.27, S r=0.43 (sig p<0.03), Low impact D r=0.19, S r= -0.18	P
					Relative oxygen consumption (Sensormedics, 2900 measurement cart)	High impact: D r= -0.07, S r=0.00, Low impact D r=0.06, S r= 0.15	P
					Percentage of maximum oxygen consumption (Sensormedics, 2900 measurement cart)	High impact: D r=0.12, S r=0.16, Low impact D r=0.14, S r= 0.15	P
					Absolute oxygen consumption (Sensormedics, 2900 measurement cart)	High impact: D r= -0.12, S r= -0.20, Low impact D r= -0.05, S r= 0.13	P
					Ventilation (Sensormedics, 2900 measurement cart)	High impact: D r=0.36 S r=0.09, Low impact D r=0.25, S r= 0.42 (sig p<0.03)	P

Reference	MPA	Para	Population (n, descriptor, mean (SD) age, %female)	Activity	Comparator	Statistic and outcome	Qual
					Ventilatory equivalent per oxygen consumption (Sensormedics, 2900 measurement cart)	High impact: D r=0.62, S r=0.40, Low impact D r=0.39, S r= 0.45 (all p<0.03)	P
Stamford 1976[142]	Crit val	I	n=14, female undergraduate students, 18.7 (SD NR), 100%f	Lab; 6 cycling, treadmill walking, treadmill jogging and stool stepping tasks performed at a variety of intensities and for differing lengths of time in a randomised order (including interval tasks).	Identical tasks were performed with interval RPE rated every minute of exercise	Pearson's correlations ranged between 0.71 to 0.90 for all activities (p<0.01)	P
Eng 2002[188]	Cons val	I	n=25, individuals >1 year post-stroke, 62.6 (8.5), 32%f	Lab; 6 minute walk test and 12 minute walk test, estimation of exertion at end of each test	Distance walked (m), measured by amount undertaken on 42m path	Pearson's 6MWT r= -0.10, 12MWT r= -0.06	P
Fry 2005[18]	Cons val	I	n=12, adults with MS able to ambulate for >6min, 47.3 (10.6), 75%f	Lab; static standing balance test	Test scores (best out of 3)	Spearman's ρ= -0.72 (p=0.01)	P
				Lab; functional stair test	Test scores (best out of 3)	ρ=0.70 (p=0.01)	P
				Lab; sit-to-stand test	Test scores (best out of 3)	ρ=0.51 (p=0.09)	P
				Lab; 6-minute walk test (metres)	Test score	ρ= -0.31 (p=0.33)	P
Okhovatian 1997[145]	Cons val	I	n=10, able-bodied subjects wearing a knee-ankle-foot orthosis and using	Location unclear; 5 min of walking around looped track at	Speed (calculated by simultaneously recording time	Correlations r=0.733 (p<0.01)	P

Reference	MPA	Para	Population (n, descriptor, mean (SD) age, %female)	Activity	Comparator	Statistic and outcome	Qual
			crutches, 26.7 (SE 1.3), gender NR	preferred speed, slow speed and fast speed	and distance)		
Hills 2006[146]	Resp to change	I	n=50, obese (n=30, age=47.8 (10.8), gender NR) or non-obese (n=20, 36.9 (12.4), gender NR) sedentary non-smokers	Grass track; walking on a level 2km once each day for three days, at "walking for pleasure" speed for first two days and maximum pace manageable on last day	Walking for pleasure speed vs maximum pace	Mean RPE values significantly higher for both groups (F=133,1, p<0.01)	P
Kravitz 2003[147]	Resp to change	I	n=18, men and women aged 20-32 from boxing exercise classes, 22.0 (2.8), 33%f	Lab; 2min boxing bouts at varying tempos	60, 72, 84, 96, 108 and 120 punches per min tempos, established by a metronome	Friedman non-parametric ANOVA. Significant differences (p<0.05) between RPE ranks (2.3, 2.4, 2.9, 3.2, 4.2 for each respective tempo)	P
Lagally 2002[148]	Resp to change	I	n=19, healthy adults, F=21.8 (2.7), M=23.2 (3.6), 47%f	Lab; 7 resistance exercises	15 repetitions of 30%1RM, 5 repetitions of 90%1RM	ANOVA: All seven exercises showed significantly higher RPE at higher intensity (p<0.01)	P
Lagally 2004[185]	Resp to change	I	n=28, 14 novice and 14 recreationally trained women, novice 21.6 (1.5) 100%f, recreational 21.9 (2.2) 100%f	Lab; 8 repetitions at 60% 1RM, 6 repetitions at 80% 1RM of a bench press exercise	Increase from 60% 1Rm to 80%1RM	ANOVA - RPE significantly higher (11.29 vs 13.39, p<0.01) at 80% 1RM	P
Leidy 1997[149]	Resp to change	I	n=20, healthy adults, 35.8 (12.4), 80%f	Lab; 2mins of: Light activities: conversing, writing, reading, playing cards, standing and waiting; moderate: polishing, sweeping, dressing, folding	Light, moderate and heavy activities	Friedman non-parametric ANOVA: RPE varied by activity intensity in the order hypothesised (p<0.001). Post-hoc tests found significant differences between light and heavy and heavy and moderate activity.	P

Reference	MPA	Para	Population (n, descriptor, mean (SD) age, %female)	Activity	Comparator	Statistic and outcome	Qual
				clothes, level walking; heavy: stair climbing, hustle walking, pushing and pulling a vacuum, carrying groceries, lifting and moving objects			
Schaeffer-Gerschutz 2000[137]	Resp to change	I	N=25, aerobically trained women, 21.0 (1.0), 100%f	Lab; 4 combinations of 3min aerobic steps include dynamic and static high and low impact arm exercises in a random order, following videotaped directions and with a 3min break in between	High vs low impact exercises	ANOVA: 10.92 and 12.16 (high impact) vs 9.00 and 9.36 (low impact). F=34.72 (p<0.03)	P
Vasquez 2013[150]	Resp to change	I	n=12, healthy males with >2yrs experience of back squats, 21.9 (1.3), 0%f	Lab; back squats: 3 repetitions of 50%1RM and to volitional failure, repeated with 70%1RM and 90%1RM in a randomised order with 10min rest in between	Hypothesised differences between 3 repetitions at each intensity but not reps to volitional failure	ANOVA and one within-subjects factor - significant main effect for condition (F=42.8, p<0.001) and significant differences between 3 repetition intensities (50=9.5, 70=11.7, 90=15.3, p<0.001). No sig differences between those to volitional failure (50=16.7, 50=16.5, 90=17.4).	P
Dawes 2005[189]	FCV	I	n=19, individuals with acquired brain injury (age range 30-60, 37%f), n=16, individuals with chronic low back pain	Lab; participants asked to imagine they are cycling up a progressively steeper hill to a	VAS and percentage ratings	All groups followed an S-shaped curve increase from nothing to maximum compared to the mean VAS. Confidence intervals were larger in the centre	P

Reference	MPA	Para	Population (n, descriptor, mean (SD) age, %female)	Activity	Comparator	Statistic and outcome	Qual
			(age range 23-55, 50%f), n=20 healthy students (age range 19-25, 50%f)	point where they are unable to continue. Each verbal anchor, administered in a random order, from the 6-20 was rated on a 20-cm VAS (limits nothing at all and maximum) and given a percentage rating. New VAS and blank cards were given for each anchor and the previous one hidden. Participants rated both breathlessness and leg fatigue, though as there were no significant differences only breathlessness and the VAS were used in the comparison.		of the scale and significant differences were found between some anchors but not others, though this varied between groups.	
Fry 2005[18]	TRR	I	n=12, adults with MS able to ambulate for >6min, 47.3 (10.6), 75%f	Lab; Static standing balance test (best trial out of 3)	1 week	Spearman's: ρ=0.77 (p=0.00)	P
				Lab; functional stair test (best trial out of 3)	1 week	ρ=0.86 (p=0.00)	P
				Lab; sit-to-stand test (best trial out of 3)	1 week	ρ=0.70 (p=0.01)	P

Reference	MPA	Para	Population (n, descriptor, mean (SD) age, %female)	Activity	Comparator	Statistic and outcome	Qual
				6-minute walk test (rated at end of each individual test)	1 week	ρ=0.96 (p=0.00)	P
Leidy 1997[149]	TRR	I	n=18, healthy adults, full sample (n=20): 35.80 (12.37), 80%f	Lab; 2 mins of each activity. Light activities: conversing, writing, reading, playing cards, standing and waiting; moderate: polishing, sweeping, dressing, folding clothes, level walking; heavy: stair climbing, hustle walking, pushing and pulling a vacuum, carrying groceries, lifting and moving objects	Within 1 week (mean = 2.8 (1.7))	Unclear; no significant differences in RPE (data NR)	P
Row 2012[187]	TRR	I	n=21, healthy older adults, 76.6 (5.5), 43%f	Fitness centre; concentric and eccentric resistance training using a seated leg press with 50% to 150% body weight loads (4-5 repetitions), administered in a random order	Second presentation of the same five loads at the end of the session	ICC=0.729	P
Skatrud-Mickelson	TRR	I	n=21, healthy adults aged 18-74 of all BMI classes, full sample n=117 (61.2%	Lab; 0.29 mile indoor lap 1) very slow walk, 2) normal	Mail survey 6-8 weeks later asking participants to recall the	Wilcoxon signed ranks test: significant difference between median ranks (p=0.02) between times	P

Reference	MPA	Para	Population (n, descriptor, mean (SD) age, %female)	Activity	Comparator	Statistic and outcome	Qual
2011[152]			aged 18-49, 38.8% aged 50-74) 57%f	paced walk, 3) brisk walk or jog	RPE for the three laps		
Stamford 1976[142]	TRR	I	n=14, female undergraduate students, 18.7 (SD NR), 100%f	Lab; 2 x one set of cycle ergometer tasks including a range of intensities	Part of 12 work tasks in a randomised order over 4 sessions - could be in the same session or 2-8 days later	Pearson's r=0.90	P
Hills 2006[146]	ME	I	n=50, obese (n=30, age=47.8 (10.8), gender NR) or non-obese (n=20, 36.9 (12.4), gender NR) sedentary non-smokers	Grass track; walking on a level 2km track once each day for two days, at "walking for pleasure" speed	1 day	Mean bias = -0.1, LOA = 2.1	F
Yamax Digiwalker CW series high quality studies (n=1)							

Reference	MPA	Para	Population (n, descriptor, mean (SD) age, %female)	Activity	Comparator	Statistic and outcome	Qual
Kinunnen 2011[154]	Crit val	F, O	n=58, overweight and obese pregnant women (BMI>25), median age 32 yrs (IQR 27-36), 100%f	Home/community; everyday activity for 4 days	GT1M Actigraph (time in sedentary, light, moderate and vigorous activity and step count)	LOA for mean value (6026 steps) = -2690 to 2656 steps/day. Lowest step count 906 steps (LOA -297 to 4897) Highest 12018 steps (LOA -4753 to 33) No effect of BMI and gestational age. Steps/day Spearman's ρ=0.78 (0.59-0.90) p<0.001 >=8000 steps/day (CW) and >=30min moderate-vigorous physical activity per day (GT1M) k=0.45 (0.24-0.67), >8000 or <8000 steps/day (CW & GT1M) k=0.63 (0.43 to 0.83) Wilcoxon signed-rank test for absolute step count between devices = medians 5961 vs 5687 (p=0.37)	G
Yamax Digiwalker CW series low quality studies (n=2)							
Martin 2012[155]	Crit val	F	n=18, community dwelling older adults, BMI<30, able to ambulate without assistance for >100m, 63.6 (SD NR), 67%f	Lab; walks at 50, 66 and 80 steps/min (in time with metronome) and self-selected speed on a 40m indoor track (8 total walks)	Average of 2 observers with handheld counters (if within 5% steps). 100% agreement for 88% trials, no discrepancies >5%.	Percentage error 50 steps/min = 66.8%, 66= 40.8%, 80=22.7% and SS= 4.8%	P
Martin 2012[155]	Inter-rater rel	F	n=18, community dwelling older adults, BMI<30, able to ambulate without assistance for >100m, 63.6 (SD NR),	Lab; walks at self-selected speed on a 40m indoor track	Three pedometers of the same brand randomly assigned to a participant and compared	ICC=0.70 (0.20-0.89)	P

Reference	MPA	Para	Population (n, descriptor, mean (SD) age, %female)	Activity	Comparator	Statistic and outcome	Qual
			67%f				
Yamax Digiwalker CW series acceptability studies (n=1)							
Kinunnen 2011[154]	Acc	F, O	n=93, overweight and obese pregnant women, median 13 weeks' gestation, median age 32 (IQR 27-36), 100% f	Steps per day averaged over 4 days	Missing data	n=3 women did not have pedometer data (unclear if out of 61 with complete accelerometer data (worn simultaneously) or out of 93 original sample). Accelerometer had much higher missing data. Authors discuss that in some cases step counts were much lower on pedometer than accelerometer, suggesting non-wear or a tilt angle that did not properly detect steps	-
Joint Protection Behaviour Assessment low quality studies (N=22)							
Hammond 1999b[157]	Cons val	A	n=35, RA patients with wrist or metacarophalangeal involvement and some restriction in ability to perform ADLs; 55.17 (9.39), 83%f	Home; use of joint protection in ADLs after a group education programme	Grip strength (digital dynamometer)	Spearman's ρ=-0.11 (NS)	F
					Hand Joint Alignment and Motion Sale (ROM and deformity)	ρ=0.06 (NS)	F
					Frequency of joint protection practice (7pt scale 1=once a week, 7=daily)	ρ=0.47 (significant). Also predicted in regression model (β=5.35, p=0.02)	P
					Hand pain (VAS)	ρ=-0.02	F

Reference	MPA	Para	Population (n, descriptor, mean (SD) age, %female)	Activity	Comparator	Statistic and outcome	Qual
Hammond 2002[190]	Cons val	A	n=30; RA diagnosis with wrist or hand involvement, >18, able to perform household tasks but hand pain on activity; 52.3 (12.08), 90%f	Home; use of joint protection in ADLs after an education programme run by OTs	Perceived helplessness (Rheumatology Attitudes Index Part 1)	ρ= -0.43 (p=0.03)	F
					Perceived control of arthritis (Rheumatology Attitudes Index Part 2)	ρ= -0.38 (p=0.05)	F
					Attending more sessions	ρ=0.39 (p=0.04)	F
					Change in overall pain (VAS)	ρ= -0.36 (p=0.07)	P
					Change in hand pain (VAS)	ρ= -0.35 (p=0.08)	P
Hammond 1999a[156]	Cons val	A	N=24, Group A: "Normal" - no RA or history of hand dysfunction, 40.5 (7.9), 83%f	Use of joint protection in ADLs	Extreme groups; Group A and B JPBA scores compared	Mann-Whitney U: Group A median = 0%, IQR = 0%. Group B median = 23.01%, IQR 6.48-31.88% U=175, p<0.0001	F
		A	N=20, Group B: RA diagnosis by consultant rheumatologist, history of hand dysfunction, difficulty with kitchen activities, 57.2 (9.9), 65%f	Use of joint protection in ADLs			
Hammond 1999a[156]	Cons val	A	N=35, rheumatoid arthritis patients, 55.2 (9.4), 83%f	Use of joint protection in ADLs	Hand pain (VAS, HAQ pain scale)	Spearman's VAS ρ=0.51 (p<0.001), functional pain score ρ=0.38 (p<0.05)	F
					Hand impairment (Joint Alignment and Motion Scale)	ρ=0.22 (NS)	F
					Grip strength (Digital	ρ=-0.54 (p<0.001)	F

Reference	MPA	Para	Population (n, descriptor, mean (SD) age, %female)	Activity	Comparator	Statistic and outcome	Qual
					Dynamometer)		
					Number of painful joints (ACR criteria)	ρ=0.41 (p<0.01)	F
					Functional disability (HAQ)	ρ=0.33 (p<0.05)	F
Hammond 1999a[156]	FCV	A	Face validity: NR Content validity: n=7 experienced rheumatology OTs	Use of joint protection in ADLs	Face validity: 20 JPBA tasks rated according to whether they involved the 5 joint principles. Content validity: 124 codes of behaviour definitions (normal, joint protection and functional adaptations) were developed from literature and video observations of RA. Seven rheumatology OTs reviewed each behaviour code and scored it as correct, partially correct or incorrect.	Each task was rated as being appropriate for assessing 2-5 joint protection principles Kappa: κ=0.6 overall (range for individual tasks 0.46-1.00 (all p<0.01)). 41.13% (51/124) codes had total agreement.	F
Klompenhouwer 2000[159]	Resp to change	A	N = 6 participants: healthy adults (junior OT students), age NR, gender NR	Lab; 3x3 groups of observers each rated a videotape of 6	JPBA tasks performed with 1) no joint protection knowledge,	Mean scores 1) 0.06	P

Reference	MPA	Para	Population (n, descriptor, mean (SD) age, %female)	Activity	Comparator	Statistic and outcome	Qual
			N=9 observers (different junior OT students), age NR, gender NR	unique performances and 2 duplicates of JPBA performances	2) after 1 hr joint protection instructions and 3) with verbal guidance. Unique rating scores compared at each manipulated level. >0.20 considered clinically significant difference.	2) 0.38 3) 0.82 All >0.20 difference	
Hammond 1999a[156]	TRR	A	N=20, Rheumatoid arthritis diagnosis, history of hand dysfunction, difficulty with kitchen activities, 57.2 (9.9), 65%f	Use of joint protection in ADLs	Time interval: approx. 8 weeks	Spearman's ρ=0.91 p<0.0001	P
Hammond 1999a[156]	Inter-rater rel	A	4 OTs with no recent rheumatology experience, 1 researcher, age and gender of OTs and sample NR	Use of joint protection in ADLs	10 videotaped JPBAs of people with RA were scored by each OT with regular consultation of the manual	Kappa: OT 1 κ=0.88, 94.1%; OT 2 κ=0.80, 92.1%; OT 3 κ=0.71, 87.5%; OT 4 κ=0.68, 81.6%	P
Klompenhouwer 2000[159]	Intra-rater rel	A	N = 6 participants: healthy adults (junior OT students), age NR, gender NR N=9 observers (different junior OT students), age NR, gender NR	Lab; JPBA tasks performed with 1) no joint protection knowledge, 2) after 1 hr joint protection instructions and 3) with verbal guidance	3x3 groups of observers each rated a videotape of 6 unique performances and 2 duplicates; duplicates assessed	ICC = 0.97 (0.92-0.99)	P
Klompenhouwer 2000[159]	Inter-rater rel	A	N = 6 participants: healthy adults (junior OT students), age NR, gender NR N=9 observers (different junior OT	Lab; JPBA tasks performed with 1) no joint protection knowledge, 2) after 1 hr joint	3x3 groups of observers each rated a videotape of 6 unique performances and 2 duplicates;	ICC = 0.93 (0.83-0.97)	P

Reference	MPA	Para	Population (n, descriptor, mean (SD) age, %female)	Activity	Comparator	Statistic and outcome	Qual
			students), age NR, gender NR	protection instructions and 3) with verbal guidance	unique performances assessed		
Klompenhouwer 2000[159]	Int cons	A	N = 6 participants: healthy adults (junior OT students), age NR, gender NR N=9 observers (different junior OT students), age NR, gender NR	Lab; JPBA tasks performed with 1) no joint protection knowledge, 2) after 1 hr joint protection instructions and 3) with verbal guidance	3x3 groups of observers each rated a videotape of 6 unique performances and 2 duplicates; unique rating scores divided into S-JPBA and A-JPBA	Cronbach's alpha = 0.95	P
Joint Protection Behaviour Assessment acceptability studies (N=1)							
Hammond 2004[12]	Acc	A	n=127, rheumatoid arthritis patients experiencing hand pain on activity; mean age for control group: 51 years range: (45-59.25); mean age for joint protection programme: 52 years range: (44-59), 76%f	Using joint protection strategies in ADLs for 48mo after a standard arthritis education programme, including 2.5hrs of joint protection	Performance rates	83/127 agreed to be recorded performing JPBA (44/49 at 48mo in intervention, 39/58 in control) Unclear if others were assessed but not videoed. High refusal in standard group may relate to non-intervention	-
Polar A1 series HRMs acceptability studies (n=1)							
Segerstahl 2011[161]	Acc	F, D, I	n=30, healthy adults sampled on exercise background and motivation, 30.0 (6.3), 50%f	Structured and non-structured exercise, including swimming, running, cycling, strength training, climbing, horseback riding, walking, soccer,	Wear rates, experiences of using the HRM	HRM used in 291/383 (76.0%) of sessions reported in a diary. 28/30 (93.0%) chose to use it regularly. 92 (24.0%) reported sessions were carried out without the HRM	-

Reference	MPA	Para	Population (n, descriptor, mean (SD) age, %female)	Activity	Comparator	Statistic and outcome	Qual
				basketball and gardening, over 3 weeks		Semi-structured interviews and diaries: Common reasons for non-use were: inconvenience/awkwardness associated with the chest strap, perceived unsuitability of heart rate monitoring for specific sports such as rock climbing or windsurfing, lack of time or forgetting to bring it along when exercising Benefits to HRM: monitors helped understand cause and effect in exercise behaviour, challenge or validate subjective feelings, optimise performance, highlight training patterns, was motivational and fun and offered a sense of accomplishment. Limitations: lack of surety about the appropriateness of the monitor's guidance and whether it was specific enough, further detail needed in manuals about target behaviours, unsuitability for certain situations, data	

Reference	MPA	Para	Population (n, descriptor, mean (SD) age, %female)	Activity	Comparator	Statistic and outcome	Qual
						incompleteness and privacy concerns. Participants were highly motivated, young, fit, healthy and computer literate. 66.7% had prior experience of using a HRM.	

Abbreviations: MPA=measurement property assessment, Crit val=criterion validity, Cons val=construct validity, FCV=face and content validity, Resp to change=responsiveness to change, TRR=test-retest reliability; rel=reliability; Int cons=internal consistency, ME=measurement error, NR=not reported, Para=parameter assessed, F=frequency, D=duration, I=intensity, A=accuracy, Gen=general adherence, O=other, n=number of participants, SD=standard deviation,G1=group 1, %f=percentage female, Qual=COSMIN quality rating (F=Fair, P=Poor), ICC=intra-class correlation coefficient, LOA=limits of agreement, SAM=StepWatch Activity Monitor, PETS=Problematic Experiences of Therapy Scale, JPBA=Joint Protection Behaviour Assessment, RPE=rating of perceived exertion, ADL=activities of daily living, HAQ=health assessment questionnaire, VAS=visual analogue scale, NS=non-significant, m=metre, s=seconds, min=minutes, hrs=hours, mo=months, mph=miles per hour, RCT=randomised controlled trial, 1RM=1 repetition maximum

Supplementary File 1: List of search terms

Phase 1 Search

CENTRAL, EED and HTA (2000-April 2013)

(Title, abstract, keywords: “patient compliance” OR Title, abstract, keywords: compliance OR Title, abstract, keywords: adherence) AND (All text: “rehabilitation” OR All text: rehabilitation)

Phase 2 Searches (Medline example)

Search strategies were adapted with headings relevant to each database.

Publication type

1. Validation studies

MeSh

1. Reproducibility of results
2. Psychometrics
3. Observer variation
4. Discriminant analysis

Ti+ab

5. Reproducib*
6. Psychometr*

7. Clinimetri*
8. Clinometr*
9. Observer variation
10. Reliab*
11. Valid*
12. Coefficient
13. “internal consistency”
14. (Cronbach* AND (alpha OR alphas))
15. “item correlation”
16. “item correlations”
17. “item selection”
18. “item selections”
19. “item reduction”
20. “item reductions”
21. Test-retest
22. (test AND retest)
23. (reliab* AND (test OR retest))
24. Stability
25. Interrater
26. Inter-rater
27. Intrarater
28. Intra-rater
29. Intertester
30. Inter-tester
31. Intratester
32. Intra-tester
33. Interobserver
34. Inter-observer

35. Intraobserver
36. Intra-observer
37. Intertechnician
38. Inter-technician
39. Intratechnician
40. Intra-technician
41. Interexaminer
42. Inter-examiner
43. Intraexaminer
44. Intra-examiner
45. Interassay
46. Inter-assay
47. Intraassay
48. Intra-assay
49. Interindividual
50. Inter-individual
51. Intraindividual
52. Intra-individual
53. Interparticipant
54. Inter-participant
55. Intraparticipant
56. Intra-participant
57. Kappa
58. Kappa’s
59. Kappas
60. “coefficient of variation”
61. Generaliza*
62. Generalisa*

63. Concordance
64. (intraclass AND correlation*)
65. Discriminative
66. “known group”
67. “Factor analysis”
68. “Factor analyses”
69. “factor structure”
70. “factor structures”
71. Dimensionality
72. Subscale*
73. “multitrait scaling analysis”
74. “multitrait scaling analyses”
75. “Item discriminant”
76. “Interscale correlation”
77. “Interscale correlations”
78. ((Error OR errors) AND (measure* OR correlat* OR evaluat* OR accuracy* OR accurate OR precision OR mean))
79. “individual variability”
80. “interval variability”
81. “rate variability”
82. “variability analysis”
83. (uncertainty AND (measurement OR measuring))
84. “standard error of measurement”
85. Sensitiv*
86. Responsive*
87. (limit AND detection)
88. “minimum detectable concentration”
89. Interpretab*

90. (small* AND (real OR detectable) AND (change OR difference))
91. “Meaningful change”
92. “minimal important change”
93. “minimal important difference”
94. “minimally important change”
95. “minimally important difference”
96. “minimal detectable change”
97. “minimal detectable difference”
98. “minimally detectable change”
99. “minimally detectable difference”
100. “minimal real change”
101. “minimal real difference”
102. “minimally real change”
103. “minimally real difference”
104. “ceiling effect”
105. “floor effect”
106. “item response model”
107. IRT
108. Rasch
109. “Differential item functioning”
110. DIF
111. “computer adaptive testing”
112. “Item bank”
113. “cross-cultural equivalence”
114. qualitative
115. interpret*
116. rating*
117. attach*

118. meaning*
119. impact*
120. burden
121. feasib*
122. “missing data”
123. “missing values”
124. “data loss”
125. (response OR non-response OR nonresponse)
126. “refusal rate”
127. understand*
128. completion
129. comprehens*
130. wear
131. non-wear
132. nonwear
133. comfort*
134. discomfort
135. eas*
136. appearance
137. safe*
138. (location OR placement)
139. size
140. conceal*
141. usab*
142. utility
143. satisf*
144. accepta*
145. willing*

146. ability
147. benefit
148. performance
149. obtrusive*
150. pilot*
151. workload

Text word (TX)

1. Agreement
2. Precision
3. Imprecision
4. “precise values”
5. Repeatab*
6. ((replica* OR repeated) AND (measure OR measures OR findings OR result OR results OR test OR tests))

All the above terms were searched using OR, and the exclusion filter was applied using NOT.

Exclusion filter (All terms combined using OR)

Publication type

1. “addresses”
2. “biography”
3. “case reports”
4. “comment”
5. “directory”
6. “editorial”
7. “festschrift”

8. “interview”
9. “lectures”
10. “legal cases”
11. “legislation”
12. “letter”
13. “news”
14. “newspaper article”
15. “patient education handout”
16. “popular works”
17. “congresses”
18. “consensus development conference”
19. “consensus development conference, nih”
20. “practice guideline”

MeSH

21. NOT (“animals” NOT “humans”)

Measure search terms

Problematic experiences of therapy scale	TI, AB “problematic experiences of therapy scale”
StepWatch Activity Monitor	TI, AB “step activity monitor” OR stepwatch OR (monitor AND orthocare) OR (monitor AND cyma) OR (monitor AND modus) OR (SAM AND monitor* AND step)

Adherence diary	TI, AB "Exercise diary" OR "Exercise diaries" OR "Home diary" OR "Home diaries" OR ((Logbook OR logbooks) AND (adherence OR compliance OR activity OR exercise)) OR "Activity diary" OR "Activity diaries" OR "Activity log" OR "Activity logs" OR ("Treatment log" AND (home OR adherence OR exercise OR compliance)) OR ("Treatment logs" AND (home OR adherence OR exercise OR compliance)) OR ("Treatment diary" AND (home OR adherence OR exercise OR compliance)) OR ("Treatment diaries" AND (home OR adherence OR exercise OR compliance)) OR "Compliance diary" OR "Compliance diaries" OR "Adherence diary" OR "Adherence diaries" OR "Adherence log" OR "Adherence logs" OR "Compliance log" OR "Compliance logs" OR "Exercise log" OR "Exercise logs" ("Training diaries" OR "training diary") AND (home OR adherence OR exercise OR compliance) OR ("Training log" OR "training logs") AND (home OR adherence OR exercise OR compliance)
Borg scale	TI "perceived exertion" OR TI "Borg" OR (TI "RPE" AND AB (Borg OR "perceived exertion"))
JPBA	TI, AB "joint protection behaviour assessment" OR "joint protection behavior assessment" OR JPBA
Yamax Digiwalker CW-701	(Yamax AND (Digiwalker* OR Digi-walker)) OR (yamax AND pedometer*) OR ((digiwalker OR digi-walker) AND pedometer*) OR ((digiwalker OR digi-walker) AND CW*) OR (Yamax AND CW*)

Polar A1 & FS1 heart rate monitors	TI, AB (Polar AND heart AND monitor*)