

RESEARCH

Open Access

A comprehensive systematic review of the development process of 104 patient-reported outcomes (PROs) for physical activity in chronically ill and elderly people

Anja Frei^{1,2*}, Kate Williams³, Anders Vetsch^{1,2}, Fabienne Dobbels⁴, Laura Jacobs⁵, Katja Rüdell³ and Milo A Puhan^{1,6}, for the PROactive consortium

Abstract

Background: Capturing dimensions of physical activity relevant to patients may provide a unique perspective for clinical studies of chronically ill patients. However, the quality of the development of existing instruments is uncertain. The aim of this systematic review was to assess the development process of patient-reported outcome (PRO) instruments including their initial validation to measure physical activity in chronically ill or elderly patient populations.

Methods: We conducted a systematic literature search of electronic databases (Medline, Embase, Psycinfo, Cinahl) and hand searches. We included studies describing the original development of fully structured instruments measuring dimensions of physical activity or related constructs in chronically ill or elderly. We broadened the population to elderly because they are likely to share physical activity limitations. At least two reviewers independently conducted title and abstract screening and full text assessment. We evaluated instruments in terms of their aim, items identification and selection, domain development, test-retest reliability, internal consistency, validity and responsiveness.

Results: Of the 2542 references from the database search and 89 from the hand search, 103 full texts which covered 104 instruments met our inclusion criteria. For almost half of the instruments the authors clearly described the aim of the instruments before the scales were developed. For item identification, patient input was used in 38% of the instruments and in 32% adaptation of existing scales and/or unsystematic literature searches were the only sources for the generation of items. For item reduction, in 56% of the instruments patient input was used and in 33% the item reduction process was not clearly described. Test-retest reliability was assessed for 61%, validity for 85% and responsiveness to change for 19% of the instruments.

Conclusions: Many PRO instruments exist to measure dimensions of physical activity in chronically ill and elderly patient populations, which reflects the relevance of this outcome. However, the development processes often lacked definitions of the instruments' aims and patient input. If PROs for physical activity were to be used in clinical trials more attention needs to be paid to the establishment of content validity through patient input and to the assessment of their evaluative measurement properties.

* Correspondence: anja.frei@usz.ch

¹Horten Centre for Patient-oriented Research, University Hospital of Zurich, Switzerland

Full list of author information is available at the end of the article

Background

Physical activity is crucial to chronically ill patients' functioning in daily life. The evidence of the protective role of physical activity for the prevention and management of chronic diseases has been well established over recent decades [1,2]. Physical activity is a multidimensional construct and defined as "any bodily movement produced by the contraction of skeletal muscle that increases energy expenditure above a basal level" [3].

The assessment of physical activity as an outcome measure provides a unique perspective in chronic disease research not only for observational studies, but also for drug and nondrug clinical trials. Furthermore, evidence from trials regarding physical activity as a patient-reported outcome (PRO) could inform patients about treatment options that address relevant aspects of their daily life. Investigators who are interested in measuring physical activity face the challenge of not only choosing an instrument that serves their study aim, but that has also been carefully developed and validated. These instruments should have strong psychometric properties such as stability over time (test-retest reliability) and the capacity to detect even small effects (responsiveness to change). In addition, investigators need to be certain that the instruments reflect the dimensions of physical activity that are relevant to patients.

It is currently unclear whether available instruments to measure physical activity fulfil these requirements. Therefore, the aim of this systematic review, which is part of the Innovative Medicines Initiative PROactive project (<http://www.proactivecopd.com> a project jointly funded by the European Commission and the European Federation of Pharmaceutical Industries and Associations 'EFPIA'), was to identify existing fully structured PROs (questionnaires, scales) measuring physical activity (frequency, intensity and total amount), and/or symptoms (physical and mental) or complaints/concerns associated with physical activity in chronically ill or elderly patient populations. We broadened the population to elderly because they are likely to share some characteristics regarding physical activity with chronically ill patients. Furthermore, the systematic review aimed to evaluate the methodological rigour with which the retrieved instruments were developed and initially validated as a part of the development process. Therefore, we restricted our review to the first validations of the instruments as part of the development process. In this paper we focused on the methods used in the development of the physical activity instruments. The content and the format of the included instruments are reviewed in another paper.

Methods

A study protocol (unregistered) guided the entire review process. We followed standard systematic review methodology as outlined in the handbooks of the Centre for Reviews and Dissemination [4] and the Cochrane Collaboration. The reporting follows the PRISMA statement that recently replaced the former guidelines of reporting of systematic reviews and meta-analyses (QUOROM statement) [5].

Eligibility criteria

We considered the following criteria for inclusion and exclusion:

Population

We included PRO instruments developed for patients with chronic disease or elderly people. Elderly people were included because chronic illnesses usually affect people in later stages of life. In addition, we supplemented the electronic database search with explicit search terms for COPD patients. This is because this systematic review is part of the PROactive project, which aims to develop and validate two PRO instruments for COPD patients [6].

Types of instrument

We included fully structured instruments (scales or questionnaires) with standardised questions and answer options which were reported by the patient (self-reported). We only included interviewer administered instruments if the information was self-reported by the patient and we excluded instruments that required a rating by an interviewer.

Content of instrument/assessment of physical activity

We included instruments measuring dimensions of physical activity or related constructs. We considered the following definition for physical activity according to the U.S. Department of Health and Human Services [3]: "Physical activity is defined as any bodily movement produced by the contraction of skeletal muscle that increases energy expenditure above a basal level". This definition of physical activity is broad and encompasses activities of daily living, sports and activities for personal fulfilment. We did not restrict the search to instruments measuring the frequency, intensity and total amount of physical activity, but also considered instruments assessing "related constructs" and/or subscales that focused on symptoms (physical and mental) or complaints/concerns associated with physical activity. All of the instruments we included contained at least one physical activity subscale. We only included instruments whose items we could access from the publication or from the developers. We did not have any language or publication date restrictions.

Study design

We included both cross-sectional and longitudinal studies which described the development (including item generation, piloting etc) or modifications of the original instrument and the initial validation (psychometric properties, cross-sectional or longitudinal) of the original instrument. Since we focused on the methods used for the development process of the instruments, the article had to describe a minimum of the development or first validation process, for example, a description about item identification or selection and/or at least one assessment of test-retest, responsiveness or validity in a publication that was clearly the original. We excluded studies that used an eligible instrument as an outcome measure and were not designed to initially validate this instrument. We also excluded studies that reported the validation of instruments in additional languages and/or populations.

Information sources

Electronic database searches

We searched the electronic databases Medline, Embase, PsycINFO and CINAHL on September 18th 2009.

Hand searches

We conducted the following hand searches to complement the electronic database search results: We searched for original development studies of instruments from articles which were excluded for the reason "validation only" or "used as outcome measures"; we scanned the reference lists of the full texts; we searched the Patient-Reported Outcome and Quality of Life Instruments Database (PROQolid) on March 10 2010, search term: "physical functioning" questionnaires; and we contacted experts in the field and asked them to check if our list of included instruments was complete or if we missed any instruments.

Search

For the electronic database search, we used the following search terms: (physical activity OR functioning OR function OR motor activity OR activities of daily living OR walking OR activity OR exercise) AND (questionnaire* OR scale OR instrument OR tool OR diary OR assessment OR self-report OR measure*) AND (valid*) AND (chronic disease OR elderly OR COPD OR chronic lung disease OR chronic obstructive lung disease) NOT (athletic performance OR sports OR children OR adolescent).

Study selection

Title and abstract screening

Two pairs of two reviewers each used a title and abstract screening document to independently review the title and abstract of every article retrieved by the

database search. Decisions to include or exclude were recorded in the RefWorks-COS file (0 = exclude, 1 = order for full text assessment, 2 = only validation study of existing instrument, 3 = related study (e.g. reviews), do not order but may be useful reference). We ordered all articles that were deemed potentially eligible by at least one reviewer.

Full text screening

Two pairs of two reviewers each independently evaluated the full texts and made a decision on inclusion or exclusion according to the predefined selection criteria. They recorded their decision on a paper form together with the reason for exclusion (not relevant patient group; instrument does not measure dimensions of physical activity or related constructs; instrument is not self-reported (e.g. functional or exercise test like time to stand up from a chair or 6 minutes walking test); instrument with all its items is not available from the publication or from the developers; instrument is used as an outcome measure/study is not designed to validate this instrument, respectively; validation study only (e.g. additional languages, populations etc.); other). If the two reviewers could not agree, a third reviewer decided whether to include or exclude. Studies that did not fulfil all of the predefined criteria were excluded and their bibliographic details were listed with the specific reason for exclusion.

Piloting the study selection process

Initially, all reviewers piloted the selection process by applying the inclusion and exclusion criteria to the 50 first references for titles and abstracts screening and the first 30 papers for full text assessment. Inclusion and exclusion criteria were refined and clarified based on this piloting process.

Dealing with lack of information

We made three attempts to contact authors by e-mail in the following conditions: 1) If it was unclear from the full text article whether the study fulfilled the inclusion and exclusion criteria; 2) If the included development study lacked information on how the instrument was developed in order to complete data extraction; 3) If the included development study lacked information on the instrument's content (items, introduction question, recall period etc.). If we failed to retrieve the relevant information from the author, this was reported on the data extraction form.

Dealing with duplicate publications

In cases where multiple papers were published (e.g. translations, reporting on different outcomes etc.), we treated the study with multiple reports as a single study but made reference to all publications.

Data extraction process

We created standardised data extraction forms based on a form used in a previous review [7] to record the

relevant information from the articles. The data extraction forms were piloted twice by four reviewers including 8 instruments for the first and 6 instruments for the second pilot. The forms and categories were then adapted and refined where necessary. The first reviewers extracted the data and stored it in a MS Word file. The second reviewers then independently extracted the data and compared their results with that of the first reviewers. These changes were made using the 'track changes' mode. The file was sent back to the first reviewer in order to come to an agreement. When an agreement could not be reached a third reviewer was consulted.

Data extraction

We extracted data from the development studies regarding the instruments' development and initial validation process. We used pre-defined categories and answer options including numerical indications, fixed texts such as "yes/no", multiple choice and free text. We extracted data for the following categories:

Development of instruments

Aim of instrument We distinguished between 3 categories: First, if the aim of the instrument was clearly described by the authors before the instrument was developed, the classification was "described". We differentiated between the four aims "evaluative" (detection of changes over time, typically for evaluation of treatments), "discriminative" (detection of differences between patients, e.g. for phenotyping), "predictive" (prediction of future health outcomes, e.g. hospital admissions or death) and "planning" (planning of treatment, e.g. detection of areas with low scorings to target patient education accordingly). Second, if the aim was not explicitly described by the authors before development but could be identified from the context, the classification was "not clearly described, but presumably (e.g. evaluative)". Third, if the purpose of the instrument was not reported and could not be identified we used the classification "not described".

Identification of items To describe the identification of the items, we differentiated between five categories of sources of item generation (several answer options possible): patients and elderly (target population); experts (e.g. clinical experts, health professionals, care givers etc., also includes supplementation or modification of existing items through experts); significant others (e.g. family members, care givers); literature; and adaptation of existing instruments. We also described the method of item identification in brackets, for example, interviews or focus groups, systematic or unsystematic searches.

Selection of items We reported the approach used by the authors to select items for the final instrument by

differentiating between the following four sources: patients quantitative; patients qualitative; experts quantitative; experts qualitative. We provided specific details in brackets, for example, "Patients: quantitative (e.g. factor analysis)", "Patients: qualitative (e.g. focus group)", "Experts: quantitative (e.g. relevance)" or "Experts: qualitative (e.g. interviews)".

Development of domains We recorded the method of how the domains were defined, i.e. if they were defined a priori (the authors predefined domains and items which belong to these domains without statistical analyses but based on their clinical/research experience or opinion) or if domains were statistically defined by factor analysis.

Initial validation of instruments

Test-retest We recorded if test-retest reliability (reproducibility) was examined and described the statistical method used, for example, intra-class correlation coefficients, coefficient of variation, Pearson or Spearman correlation coefficients or t-tests.

Internal consistency We recorded if internal consistency reliability was assessed and described the statistical method used, for example, Cronbach's alpha, corrected item total correlation or Cronbach's alpha excluding item analysis.

Validity We recorded if validity was assessed and if so, the type of validity that the authors described to assess (in quotation marks) and the statistical methods used (in brackets).

Responsiveness We recorded any approaches to assess responsiveness (i.e. the ability of an instrument to detect changes over time) and we reported the statistical methods used.

Minimal important difference (MID) We reported if the MID was examined and the statistical methods (e.g. anchor- or distribution-based approaches) used.

Summary of conducted initial validation assessments according to aim of instrument The aim of the instrument determines the measurement properties, which should be assessed in the validation process. The assessment of test-retest reliability and internal consistency is important for each instrument development, regardless of whether the instrument's aim is evaluative, discriminative, predictive or planning. For instruments with an evaluative aim, the longitudinal testing of the validity is of special interest whereas for instruments with discriminative or planning aims, cross-sectional testing of the validity is sufficient. For instruments with evaluative aims, the assessment of responsiveness and the MID is crucial because they aim to detect changes over time.

We summarised the assessed psychometric properties of the instruments for which the authors clearly described an aim before the instruments was developed.

Synthesis of results

We described the results of the data extraction in structured tables according to the categories described above (see Additional file 1). We synthesised the data on the instruments' development and initial validation in a narrative way and in integrated tables. We used numbers and proportions to describe the results quantitatively. These frequencies were calculated using SPSS (Version 18.0).

Results

Study selection

Figure 1 shows the flow diagram of the identification of the studies. The electronic database search produced 2542 references. After title and abstract screening, 2268 of these were excluded resulting in 274 articles for full text assessment. This included 5 Japanese and one Chinese language article which were provisionally included due to their English abstract but were not included in the current analysis as we were unable to translate them [8-13]. Hand searches of reference sections and of excluded articles revealed an additional 70 instruments/development studies for full text assessment. The search of the PROQolid database produced a further 58 instruments, 19 of which were included for full text assessment after title and abstract screening. One additional instrument was retrieved from the consultation with experts. Therefore, a total of 364 papers were included for full text assessment.

Following full text assessment, a further 255 were excluded resulting in 104 instruments from 103 full texts (the article of Mannerkorpi & Hernelid (2005) [14] provided information for the development process of two instruments) included in the review [14-117]. The most frequent reasons for exclusion were instrument is not self-reported ($n = 71$), followed by instrument does not measure physical activity ($n = 66$), validation study only ($n = 35$) and instrument used as an outcome measure ($n = 29$). The references of all excluded articles after full text assessment are summarised in Additional file 2.

Study characteristics

Additional file 1 summarises the extracted data for the development and initial validation process of the reviewed instruments.

Aim of instrument

For almost half of the instruments ($n = 49$, 47.1%), the authors clearly described the aim of the instruments before the scales were developed. One aim was described for 26 instruments (53.1%) and more than one for 23 instruments (46.9%). The most frequently described aim was evaluative ($n = 33$), followed by discriminative ($n = 26$), planning ($n = 13$) and predictive (n

= 5). For 43.3% of the instruments ($n = 45$), the authors did not clearly describe one or several aims but they could be presumed from the context (presumably discriminative: $n = 32$, presumably evaluative: $n = 24$, presumably planning: $n = 9$, presumably predictive: $n = 9$). For 10 instruments (9.6%), the authors did not describe an aim.

Identification of items

For 39 instruments (37.5%) items were identified with patient input, either with patient input only or with patient input together with other sources (adaptation of existing instruments, experts and/or literature). Adaptation of existing instruments and/or unsystematic literature searches only were the source for item identification of 33 instruments (31.7%), and expert input only or expert input additionally to literature and adaptation was the source for item identification of 14 instruments (13.5%). For the development of 18 instruments (17.3%), item identification was not reported or not clearly described. Table 1 describes the sources which were used to identify the items of the included instruments, ordered by frequency.

The most frequently used method to generate patient input was "interviews with patients" only (for 24 of 39 instruments). Focus groups were less frequently conducted (for 5 of 39 instruments) and for only 1 instrument both interviews and focus groups were conducted. For 7 instruments, the method of generating patient input was not reported and for 2 instruments, patient input was described as "clinical interactions" or "open ended survey". The methods used to obtain expert input were more diverse and varied from interviews with experts to workshops, ratings of relevance, unspecified discussions and undefined consideration of clinical opinion. Literature searches were always conducted unsystematically.

Selection of items

For 58 instruments (55.8%), patient input was used for item reduction, and for 12 instruments (11.5%) the items were selected by expert input only. For 34 instruments (32.7%), item reduction was not clearly described (see Table 2). Where patient input was used for item selection ($n = 58$), the methods were predominantly quantitative ($n = 31$, 53.4%) and conducted by factor analysis (17 of 31 instruments). Less frequently used methods included item-total correlations, Rasch analyses and consideration of response rates and floor/ceiling effects. Qualitative methods, either alone or in addition to quantitative methods, were used in the selection of items for 46.6% ($n = 27$) of the instruments. Most frequently, qualitative patient input for item selection was generated by patient interviews (10 of 27 instruments). Less frequently focus groups and cognitive interviews/debriefings were used.

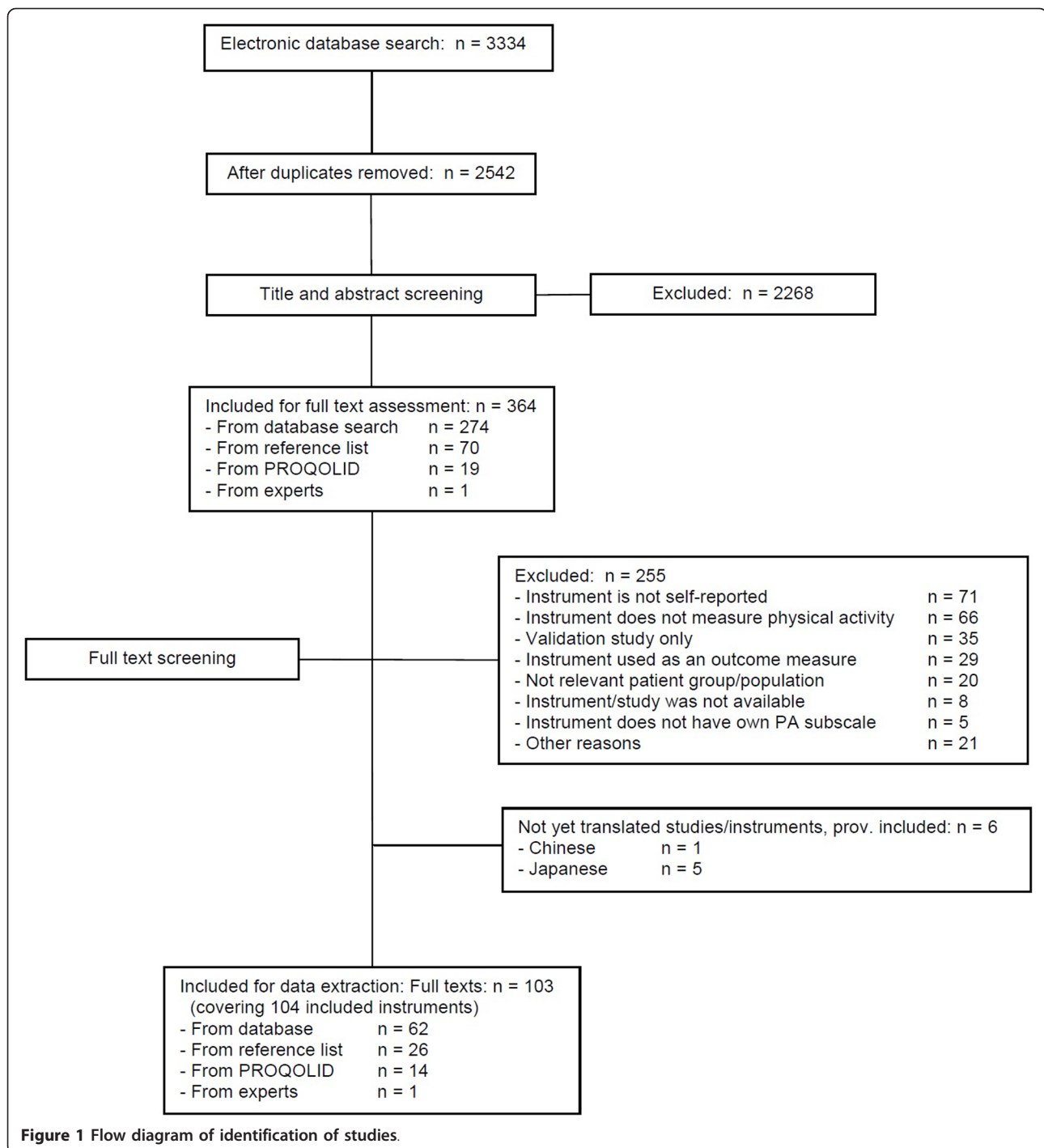


Figure 1 Flow diagram of identification of studies.

Development of domains

The domains were more often developed by factor analysis ($n = 36$, 34.6%) than by a priori specifications ($n = 16$, 15.4%). For half of the instruments, the development of the domains was not reported ($n = 42$, 40.4%) or was not applicable ($n = 8$, 7.7%). The domains of two instruments were developed by Rasch analysis.

Test-retest

Test-retest reliability was assessed for 63 instruments (60.6%). The most frequently used statistical methods were intraclass correlation coefficients either alone ($n = 18$) or together with other methods ($n = 5$). This was followed by Pearson correlation coefficient ($n = 10$), unspecified correlations ($n = 9$), various types of t-tests

Table 1 Sources of item identification of the included instruments (n = 104)¹⁾

Sources of item identification	n	%
Adaptation of existing scales only	18	17.3%
Patients & experts & literature (unsystematic search) ¹⁾	13	12.5%
Patients only	12	11.5%
Literature only (unsystematic search)	10	9.6%
Experts and literature (unsystematic search)	7	6.7%
Patients and literature (unsystematic search)	6	5.8%
Adaptation and literature (unsystematic search)	4	3.8%
Patients and experts	4	3.8%
Experts only	3	2.9%
Experts and adaptation and literature (unsystematic search)	2	1.9%
Patients and adaptation	2	1.9%
Adaptation and literature (systematic search)	1	1%
Adaptation and experts	1	1%
Patients and adaptation and literature (unsystematic search)	1	1%
Patients and experts and adaptation	1	1%
Significant others and literature (unsystematic search) and adaptation	1	1%
Not reported/not clearly described	18	17.3%

¹⁾ For data extraction details, please see Additional file 1

(either alone or together with other methods, n = 6) and various other methods (n = 15). 41 development studies (39.4%) did not report on assessing test-retest reliability.

Internal consistency

Internal consistency was assessed in 62 development studies (59.6%). Most frequently internal consistency was assessed by Cronbach's alpha alone (n = 46) or Cronbach's alpha together with other methods (n = 10).

Validity

Eighty-eight studies reported on the assessment of validity (84.6%). The most frequently assessed type of validity

Table 2 Source and method for item selection of the included instruments (n = 104)¹⁾

Source and method for item selection	n	%	n	%
Selection with patient input			58	55.8%
Patients quantitative	21	20.2%		
Patients qualitative	13	12.5%		
Patients qualitative and quantitative	6	5.8%		
Patients quantitative and experts qualitative	6	5.8%		
Patients and experts qualitative	5	4.8%		
Patients and experts quantitative	4	3.8%		
Patients and experts qualitative, patients quantitative	3	2.9%		
Selection with expert input only			12	11.5%
Qualitative	5	4.8%		
Quantitative	5	4.8%		
Quantitative and qualitative	2	1.9%		
Not reported (n = 33) and N/A (n = 1)			34	32.7%

¹⁾ For data extraction details, please see Additional file 1

that the authors described was construct validity (n = 43), followed by convergent/convergence validity (n = 19), discriminant validity (n = 18), concurrent validity (n = 16), content validity (n = 12), criterion validity (n = 11), predictive validity (n = 6), divergent validity (n = 4) and face validity (n = 4). For 25 instruments, the authors did not specify or name the type of validity tested. Most authors reported several types of validity. Validity was most frequently assessed with a correlational approach.

Responsiveness

The assessment of responsiveness was reported for 20 instruments only (19.2%). Several methods were used.

MID

Only 3 development studies reported on the MID (2.9%).

Summary of initial validation assessments according to aim of instrument

Table 3 refers to the instruments for which an aim was clearly described before the instrument was developed (n = 49, some studies described more than one aim). The table shows the number and percentage of instruments which assessed each psychometric property. The majority of instruments with a defined aim assessed validity in the initial validation process, regardless of the kind of aim, whereas test-retest was assessed for fewer instruments. For 40.6% of the instruments with an evaluative aim, responsiveness was assessed and the MID for 6.3%.

Discussion

Our systematic review showed that there are many existing PRO instruments measuring various dimensions of physical activity, highlighting the importance of this concept as an outcome measure. The methodological quality of the development process varied considerably across the 104 included instruments. For the majority of the instruments, the aim either was not clearly described or not described at all before the instruments were developed. In addition, patients were often not involved in the item identification process of new instruments, making the adaptation of existing scales, unsystematic literature searches and/or expert input the only sources of item generation. Several instruments used quantitative patient input for item selection, but a surprisingly high number of studies did not describe or report on how items were selected. Also, the quality of the initial validation varied widely between the instruments. Internal consistency and test-retest reliability were assessed more frequently than responsiveness to change. The MID was estimated for only 3 instruments. Some instruments defined an evaluative aim; however, responsiveness was assessed in less than half of these. Many studies assessed construct validity while content validity was assessed for only a minority of the instruments.

Table 3 Conducted initial validation assessments according to described aims of instruments¹⁾

Described aim of instrument	Test-retest		Internal consistency		Validity		Responsiveness		MID ²⁾	
	n	% ³⁾	n	% ³⁾	n	% ³⁾	n	% ³⁾	n	% ³⁾
Evaluative (n = 33)	23	69.7%	21	63.6%	32	96.9%	13	39.4%	2	6.1%
Discriminative (n = 26)	15	57.7%	18	69.2%	24	92.3%	6	23.1%	1	3.8%
Planning (n = 13)	11	84.6%	7	53.8%	12	92.3%	3	23.1%	1	7.7%
Predictive (n = 5)	1	20.0%	0	0%	3	60.0%	0	0%	0	0%

¹⁾ For data extraction details, please see Additional file 1

²⁾ MID = Minimal important difference

³⁾ % in relation to the corresponding aim

Over the last decades, physical activity instruments were traditionally used predominantly in epidemiological research to measure physical activity as a potential determinant of health outcomes [1,2]. This requires that the instruments are able to discriminate between people in order to identify different levels of physical activity that might be associated with different health outcomes. In recent years, there has been growing interest in physical activity as a PRO measure. For example in obesity research, studies examine the effect of interventions on physical activity [118-120]. The use of physical activity instruments as outcome measures has implications for the development and initial validation process of these scales. Since PROs should be able to detect changes over time, their evaluative power is essential. Consequently, development and initial validation studies should go beyond cross-sectional studies and assess responsiveness to change and the MID in prospective follow-up studies [7].

PROs for symptoms, health-related quality of life but also for physical activity have become a prevalent outcome in clinical trials. Over the last ten years many new PROs have been developed and validated and it can be expected that in the near future an increasing number of claims on the effectiveness of drugs will be made based on PROs. As a consequence, both the U.S. Food and Drug Administration (FDA) and the European Medicines Agency (EMA) have developed guidance documents on the requirements for PRO instruments that would allow making drug claims. A key evaluation point for the FDA is the evidence on content validity. Content validity describes the extent of how the instrument measures the concept of interest, which is specific to the population, condition and treatments to be studied. The FDA explicitly asks for patient input for item generation through qualitative research to ensure content validity in the development process of a new instrument [121-123].

Although all of the PRO instruments included in this systematic review were developed before the finalisation of the FDA guidance document in December 2009, it is still surprising that in less than one third of the included

studies authors reported on qualitative research for item generation such as patient interviews or focus groups, and a minority declared explicitly to have tested content validity of the newly developed instruments. These findings, along with the fact of poor reporting on item selection methods, indicate that only few physical activity PRO instruments would currently fulfil the FDA and EMA requirements for outcome measures. While the need to establish content validity has been recognised for many years, there has been little pressure to conduct qualitative research as illustrated in our systematic review. This is likely to change; at least in the field of clinical trials as investigators developing new instruments can now follow the FDA and EMA guidance to establish content validity more formally through qualitative research. Existing instruments are in a more difficult position, although they may in retrospect support their relevance to patients through additional qualitative research. For example, one may examine whether the constructs measured by existing instruments align with what patients perceive to be important, or if important aspects are missing.

One strength of this systematic review was the adherence to rigorous systematic review methodology along with the broad search strategy to identify existing physical activity instruments and subscales/domains. We supplemented the systematic database searches by a comprehensive hand search as well as by a PROQolid database search. As we aimed to identify any relevant instruments, we kept the inclusion criteria broad by using the definition for physical activity as described in the "2008 Physical activity guideline for Americans" [3]. Such a broad perspective could also be perceived as a limitation. Although we paid great attention to carefully defining the inclusion criteria, we cannot exclude the possibility of having missed questionnaires. Also, the decision about inclusion or exclusion of the instruments was sometimes ambiguous as for example for instruments assessing specific types of physical activity for chronic illnesses such as multiple sclerosis or chronic pain. In such cases we tried to adopt systematically and scientifically defensible decision criteria for inclusion or

exclusion. For multiple sclerosis patients, for example, we did not consider physical activity instruments aiming at impaired hand motor activity but we included those assessing physical activity limitations which are more general and which could also be relevant for other chronic illnesses like “Walking ability” [54] or “Physical functioning” [93]. Another example includes activity limitations due to pain, where we excluded some instruments such as those targeting specialised pain coping activities, but included instruments such as the Activities of Daily Living Scale [71]. We focused solely on publications of the development and initial validation, which to some extent may underestimate the rigour of the overall development process. Undoubtedly some instruments might have had additional validation studies which we have not included in this review. However, we suspect that many instruments were introduced into research and practice rather rapidly without further validation, and, if validations were conducted during the development process, it is likely that the authors would have published these results as part of the development paper.

Conclusion

Our systematic review showed that there are many existing PRO instruments measuring physical activity in chronically ill and elderly patient populations, highlighting the importance of this concept as an outcome measure. However, the development processes often lacked definitions of the instruments’ aims and patient input. If PROs for physical activity are to be used in clinical trials, there needs to be more focus on establishing content validity through patient input, and assessing their evaluative measurement properties.

Additional material

Additional file 1: Data extraction results: Development and initial validation process of the reviewed instruments. Summary of the extracted data for the development and initial validation process of the reviewed instruments according to the categories aim of instruments, identification of items, selection of items (item reduction), development of domains, test-retest, internal consistency, validity, responsiveness and MID.

Additional file 2: References list of excluded articles after full text assessment. List of all references of articles which have been excluded after full text assessment.

Acknowledgements

The study was conducted within the PROactive project which is funded by the Innovative Medicines Initiative Joint Undertaking (IMI JU # 115011). The authors would also like to thank Laura Jacobs for her support in the early stages of the project as well as the PROactive group: Caterina Brindicci and Tim Higenbottam (Chiesi Farmaceutici S.A.), Thierry Trooster and Fabienne Dobbels (Katholieke Universiteit Leuven), Margaret X. Tabberer (Glaxo Smith Kline), Roberto Rabinovitch and Bill McNee (University of Edinburgh, Old

College South Bridge), Ioannis Vogiatzis (Thorax Research Foundation, Athens), Michael Polkey and Nick Hopkinson (Royal Brompton and Harefield NHS Foundation Trust), Judith Garcia-Aymerich (Municipal Institute of Medical Research, Barcelona), Milo Puhán and Anja Frei (Universität of Zürich, Zürich), Thys van der Molen and Corina De Jong (University Medical Center, Groningen), Pim de Boer (Netherlands Asthma Foundation, Leusden), Ian Jarrod (British Lung Foundation, UK), Paul McBride (Choice Healthcare Solution, UK), Nadia Kamel (European Respiratory Society, Lausanne), Katja Rudell and Frederick J. Wilson (Pfizer Ltd), Nathalie Ivanoff (Almirall), Karoly Kulich and Alistair Glendenning (Novartis), Niklas X. Karlsson and Solange Corriol-Rohou (AstraZeneca AB), Enkeleida Nikai (UCB) and Damijen Erzen (Boehringer Ingelheim).

Author details

¹Horten Centre for Patient-oriented Research, University Hospital of Zurich, Switzerland. ²Institute of General Practice and Health Services Research, University Hospital of Zurich, Switzerland. ³Patient Reported Outcomes Centre of Excellence, Global Market Access, Primary Care Business Unit, Pfizer Ltd, Walton Oaks, Surrey, United Kingdom. ⁴Centre for Health Services and Nursing Research, post-doctoral researcher FWO Vlaanderen, Katholieke Universiteit Leuven, Leuven, Belgium. ⁵Respiratory Sciences, Katholieke Universiteit Leuven, Leuven, Belgium. ⁶Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Johns Hopkins University, Baltimore (MD), USA.

Authors’ contributions

MP and KR led the systematic review. MP, KR, AF, FD and LJ developed the conceptual framework and the study protocol. MP, KR and AF coordinated the review. MP, FD, KR and AF conducted the electronic database searches; KW, AV, AF, KR and MP conducted the additional searches. AF coordinated the references in RefWorks. KR and FD (1st reviewers), KW and LJ (2nd reviewers) and MP and AF (3rd reviewers) screened titles and abstracts. KW and KR (1st reviewers), AF and AV (2nd reviewers) and MP (3rd reviewer) assessed full texts of the identified studies and extracted the relevant data. AF conducted the statistical analysis. AF and MP drafted the manuscript. All authors contributed to revising the manuscript and approved the final version.

Authors’ information

Katja Rüdell is an honorary lecturer of health psychology at the University of Kent, UK

Competing interests

Anja Frei, Anders Vetsch, Fabienne Dobbels, Laura Jacobs and Milo Puhán have no competing interests. Katja Rüdell is an employee of Pfizer Ltd and Kate Williams is contracted by Pfizer Ltd. The current work is methodological and provides no competitive advantage or disadvantage. No medical writers were used to support the writing of this manuscript.

Received: 21 December 2010 Accepted: 20 December 2011

Published: 20 December 2011

References

1. Lagerros YT, Lagiou P: **Assessment of physical activity and energy expenditure in epidemiological research of chronic diseases.** *European Journal of Epidemiology* 2007, **22**(6):353-362.
2. Valanou EM, Bamia C, Trichopoulos A: **Methodology of physical-activity and energy-expenditure assessment: A review.** *Journal of Public Health* 2006, **14**(2):58-65.
3. U.S. Department of Health and Human Services: **2008 Physical Activity Guidelines for Americans.** [http://www.health.gov/paguidelines/].
4. **Systematic Reviews. CRD’s guidance for undertaking reviews in health care.** York: Centre for Reviews and Dissemination, University of York; 2009.
5. Moher D, Liberati A, Tetzlaff J, Altman DG: **Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement.** *BMJ* 2009, **339**:332-336.
6. PROactive:[http://www.proactivecopd.com/].
7. Frei A, Svarin A, Steurer-Stey C, Puhán MA: **Self-efficacy instruments for patients with chronic diseases suffer from methodological limitations—a systematic review.** *Health Qual Life Outcomes* 2009, **7**:86.

8. ChongHua W, Li G, XiaoMei L: **Development of the General Module for the System of Quality of Life Instruments for Patients with Chronic Disease: Items selection and structure of the general module.** *Chinese Mental Health Journal* 2005, **19**(11):723-726.
9. Eto F, Tanaka M, Chishima M, Igarashi M, Mizoguchi T, Wada H, Iijima S: **[Comprehensive activities of daily living (ADL) index for the elderly].** *Nippon Ronen Igakkai Zasshi - Japanese Journal of Geriatrics* 1992, **29**(11):841-848.
10. Hashimoto S, Aoki R, Tamakoshi A, Shibazaki S, Nagai M, Kawakami N, Ikari A, Ojima T, Ohno Y: **[Development of index of social activities for the elderly].** *Nippon Koshu Eisei Zasshi - Japanese Journal of Public Health* 1997, **44**(10):760-768.
11. Horiuchi T, Kobayashi Y, Hosoi T, Ishibashi H, Yamamoto S, Yatomi N: **[The assessment of the reliability and the validity of the EQOOL questionnaire of osteoporotics-QOL assessment of elderly osteoporotics by EQOOL].** *Nippon Ronen Igakkai Zasshi - Japanese Journal of Geriatrics* 2005, **42**(2):229-234.
12. Inaba Y, Obuchi S, Oka K, Arai T, Nagasawa H, Shiba Y, Kojima M: **[Development of a rating scale for self-efficacy of physical activity in frail elderly people].** *Nippon Ronen Igakkai Zasshi - Japanese Journal of Geriatrics* 2006, **43**(6):761-768.
13. Kozaki K, Murata H, Kikuchi R, Sugiyama Y, Hasegawa H, Igata A, Toba K: **["Activity scale for the elderly" as a measurement for the QOL of local elderly individuals and the assessment of the influence of age and exercise].** *Nippon Ronen Igakkai Zasshi - Japanese Journal of Geriatrics* 2008, **45**(2):188-195.
14. Mannerkorpi K, Hernelid C: **Leisure Time Physical Activity Instrument and Physical Activity at Home and Work Instrument. Development, face validity, construct validity and test-retest reliability for subjects with fibromyalgia.** *Disability & Rehabilitation* 2005, **27**(12):695-701.
15. Alvarez-Gutierrez FJ, Miravittles M, Calle M, Gobartt E, Lopez F, Martin A, Grupo de Estudio EIME: **[Impact of chronic obstructive pulmonary disease on activities of daily living: results of the EIME multicenter study].** *Archivos de Bronconeumologia* 2007, **43**(2):64-72.
16. Anagnostis C, Gatchel RJ, Mayer TG: **The pain disability questionnaire: a new psychometrically sound measure for chronic musculoskeletal disorders.** *Spine* 2004, **29**(20):2290-2302.
17. Anderson KO, Dowds BN, Pelletz RE, Edwards WT, Peeters-Asdourian C: **Development and initial validation of a scale to measure self-efficacy beliefs in patients with chronic pain.** *Pain* 1995, **63**(1):77-84.
18. Arbuckle TY, Gold DP, Chaikelson JS, Lapidus S: **Measurement of activity in the elderly: The Activities Checklist.** *Canadian Journal on Aging* 1994, **13**(4):550-565.
19. Avlund K, Kreiner S, Schultz-Larsen K: **Functional ability scales for the elderly: a validation study.** *European journal of public health* 1996, **6**(1):35-42.
20. Avlund K, Schultz-Larsen K, Kreiner S: **The measurement of instrumental ADL: content validity and construct validity.** *Aging-Clinical & Experimental Research* 1993, **5**(5):371-383.
21. Baiardini I, Braido F, Fassio O, Tarantini F, Pasquali M, Tarchino F, Berlendis A, Canonica GW: **A new tool to assess and monitor the burden of chronic cough on quality of life: Chronic Cough Impact Questionnaire.** *Allergy: European Journal of Allergy and Clinical Immunology* 2005, **60**(4):482.
22. Bergner MPD, Bobbitt RAPD, Carter WBPD, Gilson BSMD: **The Sickness Impact Profile: Development and Final Revision of a Health Status Measure.** *Medical care* 1981, **19**(8):787-805.
23. Bula CJ, Martin E, Rochat S, Piot-Ziegler C: **Validation of an adapted falls efficacy scale in older rehabilitation patients.** *Archives of Physical Medicine & Rehabilitation* 2008, **89**(2):291-296.
24. Burckhardt CS, Clark SR, Bennett RM: **The fibromyalgia impact questionnaire: development and validation.** *The Journal of rheumatology* 1991, **18**(5):728-733.
25. Calin A, Garrett S, Whitelock H, Kennedy LG, O'Hea J, Mallorie P, Jenkinson T: **A new approach to defining functional ability in ankylosing spondylitis: the development of the Bath Ankylosing Spondylitis Functional Index.** *The Journal of rheumatology* 1994, **21**(12):2281-2285.
26. Cardol M, de Haan RJ, de Jong BA, van den Bos GA, de Groot IJ: **Psychometric properties of the Impact on Participation and Autonomy Questionnaire.** *Archives of Physical Medicine & Rehabilitation* 2001, **82**(2):210-216.
27. Cardol M, de Haan RJ, van den Bos GA, de Jong BA, de Groot IJ: **The development of a handicap assessment questionnaire: the Impact on Participation and Autonomy (IPA).** *Clinical rehabilitation* 1999, **13**(5):411-419.
28. Carone M, Bertolotti G, Anchi F, Zotti AM, Donner CF, Jones PW: **Analysis of factors that characterize health impairment in patients with chronic respiratory failure. Quality of Life in Chronic Respiratory Failure Group.** *European Respiratory Journal* 1999, **13**(6):1293-1300.
29. Caspersen CJ, Bloemberg BPM, Saris WHM, Merritt RK, Kromhout D: **The Prevalence of Selected Physical Activities and Their Relation with Coronary Heart Disease Risk Factors in Elderly Men: The Zutphen Study, 1985.** *American Journal of Epidemiology* 1991, **133**(11):1078-1092.
30. Chou K: **Hong Kong Chinese Everyday Competence Scale: a validation study.** *Clinical gerontologist* 2003, **26**(1):43-51.
31. Clark DO, Callahan CM, Counsell SR: **Reliability and validity of a steadiness score.** *Journal of the American Geriatrics Society* 2005, **53**(9):1582-1586.
32. Clark MS, Bond MJ: **The Adelaide Activities Profile: a measure of the life-style activities of elderly people.** *Aging-Clinical & Experimental Research* 1995, **7**(4):174-184.
33. Dallosso HM, Morgan K, Bassey EJ, Ebrahim SB, Fentem PH, Arie TH: **Levels of customary physical activity among the old and the very old living at home.** *Journal of Epidemiology & Community Health* 1988, **42**(2):121-127.
34. Davis AH, Figueredo AJ, Fahy BF, Rawiworrakul T: **Reliability and validity of the Exercise Self-Regulatory Efficacy Scale for individuals with chronic obstructive pulmonary disease.** *Heart & Lung* 2007, **36**(3):205-216.
35. Di Pietro L, Caspersen CJ, Ostfeld AM, Nadel ER: **A survey for assessing physical activity among older adults.** *Medicine & Science in Sports & Exercise* 1993, **25**(5):628-642.
36. Dorevitch MI, Cossar RM, Bailey FJ, Bisset T, Lewis SJ, Wise LA, MacLennan WJ: **The accuracy of self and informant ratings of physical functional capacity in the elderly.** *Journal of clinical epidemiology* 1992, **45**(7):791-798.
37. Dunderdale K, Thompson DR, Beer SF, Furze G, Miles JNV: **Development and validation of a patient-centered health-related quality-of-life measure: the Chronic Heart Failure Assessment Tool.** *Journal of Cardiovascular Nursing* 2008, **23**(4):364-370.
38. Eakin EG, Resnikoff PM, Prewitt LM, Ries AL, Kaplan RM: **Validation of a new dyspnea measure: the UCSD Shortness of Breath Questionnaire.** *Chest* 1998, **113**(3):619-624.
39. Eakman AM: **A reliability and validity study of the Meaningful Activity Participation Assessment.** University of Southern California; 2007.
40. Fillenbaum GG: **Screening the elderly. A brief instrumental activities of daily living measure.** *Journal of the American Geriatrics Society* 1985, **33**(10):698-706.
41. Fillenbaum GG, Smyer MA: **The development, validity, and reliability of the OARS multidimensional functional assessment questionnaire.** *Journal of gerontology* 1981, **36**(4):428-434.
42. Finch M, Kane RL, Philp I: **Developing a new metric for ADLs.** *Journal of the American Geriatrics Society* 1995, **43**(8):877-884.
43. Follick MJ, Ahern DK, Laser-Wolston N: **Evaluation of a daily activity diary for chronic pain patients.** *Pain* 1984, **19**(4):373-382.
44. Frederiks CM, te Wierik MJ, Visser AP, Sturmans F: **The functional status and utilization of care of elderly people living at home.** *Journal of community health* 1990, **15**(5):307-317.
45. Friedenreich CM, Courneya KS, Bryant HE: **The lifetime total physical activity questionnaire: development and reliability.** *Medicine & Science in Sports & Exercise* 1998, **30**(2):266-274.
46. Fries JF, Spitz PW, Young DY: **The dimensions of health outcomes: the health assessment questionnaire, disability and pain scales.** *The Journal of rheumatology* 1982, **9**(5):789-793.
47. Garrad J, Bennett AE: **A validated interview schedule for use in population surveys of chronic disease and disability.** *British journal of preventive & social medicine* 1971, **25**(2):97-104.
48. Garrod R, Bestall JC, Paul EA, Wedzicha JA, Jones PW: **Development and validation of a standardized measure of activity of daily living in patients with severe COPD: the London Chest Activity of Daily Living scale (LCADL).** *Respiratory medicine* 2000, **94**(6):589-596.
49. Guyatt GH, Berman LB, Townsend M, Pugsley SO, Chambers LW: **A measure of quality of life for clinical trials in chronic lung disease.** *Thorax* 1987, **42**(10):773-778.

50. Guyatt GH, Eagle DJ, Sackett B, Willan A, Griffith L, McLroy W, Patterson CJ, Turpie I: **Measuring quality of life in the frail elderly.** *Journal of clinical epidemiology* 1993, **46**(12):1433-1444.
51. Harwood RH, Rogers A, Dickinson E, Ebrahim S: **Measuring handicap: the London Handicap Scale, a new outcome measure for chronic disease.** *Quality in Health Care* 1994, **3**(1):11-16.
52. Helmes E, Hodsman A, Lazowski D, Bhardwaj A, Crilly R, Nichol P, Drost D, Vanderburgh L, Pederson L: **A Questionnaire To Evaluate Disability in Osteoporotic Patients With Vertebral Compression Fractures.** *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences* 1995, **50A**(2):M91-M98.
53. Hlatky MA, Boineau RE, Higginbotham MB, Lee KL, Mark DB, Califf RM, Cobb FR, Pryor DB: **A brief self-administered questionnaire to determine functional capacity (the Duke Activity Status Index).** *American Journal of Cardiology* 1989, **64**(10):651-654.
54. Hobart JC, Riazi A, Lamping DL, Fitzpatrick R, Thompson AJ: **Measuring the impact of MS on walking ability: The 12-Item MS Walking Scale (MSWS-12).** *Neurology* 2003, **60**(1):31-36.
55. Holbrook M, Skilbeck CE: **An activities index for use with stroke patients.** *Age & Ageing* 1983, **12**(2):166-170.
56. Hyland ME: **The Living with Asthma Questionnaire.** *Respiratory medicine* 1991, **85**(Supplement 2):13-16.
57. Jacobs JE, Maillé AR, Akkermans RP, van Weel C, Grol RP: **Assessing the quality of life of adults with chronic respiratory diseases in routine primary care: construction and first validation of the 10-Item Respiratory Illness Questionnaire-monitoring 10 (RIQ-MON10).** *Quality of Life Research* 2004, **13**(6):1117-1127.
58. Jette AM, Deniston OL: **Inter-observer reliability of a functional status assessment instrument.** *Journal of chronic diseases* 1978, **31**(9-10):573-580.
59. Jones PW, Quirk FH, Baveystock CM, Littlejohns P: **A self-complete measure of health status for chronic airflow limitation. The St. George's Respiratory Questionnaire.** *The American Review of Respiratory Disease* 1992, **145**(6):1321.
60. Kaplan RM, Sieber WJ, Ganiats TG: **The quality of well-being scale: comparison of the interviewer-administered version with a self-administered questionnaire.** *Psychology and Health* 1997, **12**(Journal Article):783-791.
61. Keith RA, Granger CV, Hamilton BB, Sherwin FS: **The functional independence measure: a new tool for rehabilitation.** *Advances in Clinical Rehabilitation* 1987, **1**(Journal Article):6-18.
62. Kempen GI, Suurmeijer TP: **The development of a hierarchical polytomous ADL-IADL scale for noninstitutionalized elders.** *Gerontologist* 1990, **30**(4):497-502.
63. Kuhl K, Schurmann W, Rief W: **COPD disability index (CDI) - a new instrument to assess COPD-related disability.** *COPD-Disability-Index (CDI) - ein neues Verfahren zur Erfassung der COPD-bedingten Beeinträchtigung* 2009, **63**(3):136.
64. Lareau SC, Carrieri-Kohlman V, Janson-Bjerklie S, Roos PJ: **Development and testing of the Pulmonary Functional Status and Dyspnea Questionnaire (PFSDQ).** *Heart & Lung* 1994, **23**(3):242-250.
65. Lareau SC, Meek PM, Roos PJ: **Development and testing of the modified version of the Pulmonary Functional Status and Dyspnea Questionnaire (PFSDQ-M).** *Heart & Lung* 1998, **27**(3):159-168.
66. Lee L, Friesen M, Lambert IR, Loudon RG: **Evaluation of Dyspnea During Physical and Speech Activities in Patients With Pulmonary Diseases.** *Chest* 1998, **113**(3):625-632.
67. Leidy NK: **Psychometric properties of the functional performance inventory in patients with chronic obstructive pulmonary disease.** *Nursing research* 1999, **48**(1):20-28.
68. Lerner D, Amick BC III, Rogers WH, Malspeis S, Bungay K, Cynn D: **The Work Limitations Questionnaire.** *Medical care* 2001, **39**(1):72-85.
69. Letrait M, Lurie A, Bean K, Mesbah M, Venot A, Strauch G, Grandordy BM, Chwalow J: **The Asthma Impact Record (AIR) index: a rating scale to evaluate the quality of life of asthmatic patients in France.** *European Respiratory Journal* 1996, **9**(6):1167-1173.
70. Lewin RJ, Thompson DR, Martin CR, Stuckey N, Devlen J, Michaelson S, Maguire P: **Validation of the Cardiovascular Limitations and Symptoms Profile (CLASP) in chronic stable angina.** *Journal of cardiopulmonary rehabilitation* 2002, **22**(3):184-191.
71. Linton SJ: **Activities of daily living scale for patients with chronic pain.** *Perceptual & Motor Skills* 1990, **71**(3 Pt 1):722.
72. Liu B, Woo J, Tang N, Ng K, Ip R, Yu A: **Assessment of total energy expenditure in a Chinese population by a physical activity questionnaire: examination of validity.** *International Journal of Food Sciences & Nutrition* 2001, **52**(3):269-282.
73. Maillé AR, Koning CJM, Zwinderman AH, Willems LNA, Dijkman JH, Kaptein AA: **The development of the []Quality-of-Life for Respiratory Illness Questionnaire (QOL-RIQ): a disease-specific quality-of-life questionnaire for patients with mild to moderate chronic non-specific lung disease.** *Respiratory Medicine* 1997, **91**(5):297-309.
74. Mathias SD, Bussel JB, George JN, McMillan R, Okano GJ, Nichol JL: **A disease-specific measure of health-related quality of life in adults with chronic immune thrombocytopenic purpura: psychometric testing in an open-label clinical trial.** *Clinical therapeutics* 2007, **29**(5):950-962.
75. Mathuranath PS, George A, Cherian PJ, Mathew R, Sarma PS: **Instrumental activities of daily living scale for dementia screening in elderly people.** *International Psychogeriatrics* 2005, **17**(3):461-474.
76. Mayer J, Mooney V, Matheson L, Leggett S, Verna J, Balourdas G, DeFilippo G: **Reliability and validity of a new computer-administered pictorial activity and task sort.** *Journal of Occupational Rehabilitation* 2005, **15**(2):203-213.
77. McHorney CA, Ware JE Jr, Lu JF, Sherbourne CD: **The MOS 36-Item Short-Form Health Survey (SF-36): III. Tests of data quality, scaling assumptions, and reliability across diverse patient groups.** *Medical care* 1994, **32**(1):40-66.
78. Migliore Norweg A, Whiteson J, Demetis S, Rey M: **A new functional status outcome measure of dyspnea and anxiety for adults with lung disease: the dyspnea management questionnaire.** *Journal of cardiopulmonary rehabilitation* 2006, **26**(6):395-404.
79. Moriarty D, Zack M, Kobau R: **The Centers for Disease Control and Prevention's Healthy Days Measures - Population tracking of perceived physical and mental health over time.** *Health and Quality of Life Outcomes* 2003, **1**(1):37.
80. Morimoto M, Takai K, Nakajima K, Kagawa K: **Development of the chronic obstructive pulmonary disease activity rating scale: reliability, validity and factorial structure.** *Nursing & health sciences* 2003, **5**(1):23-30.
81. Morris WW, Buckwalter KC, Cleary TA, Gilmer JS: **Issues related to the validation of the Iowa Self-Assessment Inventory.** *Educational and Psychological Measurement* 1989, **49**(4):853-861.
82. Myers J, Do D, Herbert W, Ribisl P, Froelicher VF: **A nomogram to predict exercise capacity from a specific activity questionnaire and clinical data.** *American Journal of Cardiology* 1994, **73**(8):591-596.
83. Nijs J, Vaes P, McGregor N, Van Hoof E, De Meirleir K: **Psychometric properties of the Dutch Chronic Fatigue Syndrome-Activities and Participation Questionnaire (CFS-APQ).** *Physical Therapy* 2003, **83**(5):444-454.
84. Nouri FM, Lincoln NB: **An extended activities of daily living scale for stroke patients.** *Clinical rehabilitation* 1987, **1**(4):301-305.
85. Parkerson GRJMDMPH, Gehlbach SHMDMPH, Wagner EHMDMPH, James SAPD, Clapp NERNMPH, Muhlbaier LHMS: **The Duke-UNC Health Profile: An Adult Health Status Instrument for Primary Care.** *Medical care* 1981, **19**(8):806-828.
86. Pluijm SM, Bardage C, Nikula S, Blumstein T, Jylha M, Minicuci N, Zunzunegui MV, Pedersen NL, Deeg DJ: **A harmonized measure of activities of daily living was a reliable and valid instrument for comparing disability in older people across countries.** *Journal of clinical epidemiology* 2005, **58**(10):1015-1023.
87. Rankin SL, Briffa TG, Morton AR, Hung J: **A specific activity questionnaire to measure the functional capacity of cardiac patients.** *American Journal of Cardiology* 1996, **77**(14):1220-1223.
88. Regensteiner JG, Steiner JF, Panzer RJ, Hiatt WR: **Evaluation of walking impairment by questionnaire in patients with peripheral arterial disease.** *J Vasc Med Biol* 1990, **2**:142-152.
89. Rejeski WJ, Ettinger JWH, Schumaker S, James P, Burns R, Elam JT: **Assessing performance-related disability in patients with knee osteoarthritis.** *Osteoarthritis and Cartilage* 1995, **3**(3):157-167.
90. Resnick B, Jenkins LS: **Testing the reliability and validity of the Self-Efficacy for Exercise Scale.** *Nursing research* 2000, **49**(3):154-159.
91. Rimmer JH, Riley BB, Rubin SS: **A new measure for assessing the physical activity behaviors of persons with disabilities and chronic health conditions: the Physical Activity and Disability Survey.** *American Journal of Health Promotion* 2001, **16**(1):34-42.

92. Roland M, Morris R: A study of the natural history of back pain. Part I: development of a reliable and sensitive measure of disability in low-back pain. *Spine* 1983, **8**(2):141-144.
93. Rotstein Z, Barak Y, Noy S, Achiron A: Quality of life in multiple sclerosis: development and validation of the 'RAYS' scale and comparison with the SF-36. *International Journal for Quality in Health Care* 2000, **12**(6):511-517.
94. Schag AC, Heinrich RL, Aadland RL, Ganz PA: Assessing Problems of Cancer Patients: Psychometric Properties of the Cancer Inventory of Problem Situations. *Health Psychology* 1990, **9**(1):83-102.
95. Schultz-Larsen K, Avlund K, Kreiner S: Functional ability of community dwelling elderly. Criterion-related validity of a new measure of functional ability. *Journal of clinical epidemiology* 1992, **45**(11):1315-1326.
96. Shah S, Vanclay F, Cooper B: Improving the sensitivity of the Barthel Index for stroke rehabilitation. *Journal of clinical epidemiology* 1989, **42**(8):703-709.
97. Sintonen H: The 15-D Measure of Health Related Quality of Life: Reliability, Validity and Sensitivity of its Health State Descriptive System. 1994.
98. Sintonen H: The 15D instrument of health-related quality of life: properties and applications. *Annals of Medicine* 2001, **33**(5):328-336.
99. So CT, Man DWK: Development and validation of an activities of daily living inventory for the rehabilitation of patients with chronic obstructive pulmonary disease. *OTJR: Occupation, Participation & Health* 2008, **28**(4):149-159.
100. Stel VS, Smit JH, Pluijij SM, Visser M, Deeg DJ, Lips P: Comparison of the LASA Physical Activity Questionnaire with a 7-day diary and pedometer. *Journal of clinical epidemiology* 2004, **57**(3):252-258.
101. Stewart AL, Mills KM, King AC, Haskell WL, Gillis D, Ritter PL: CHAMPS physical activity questionnaire for older adults: outcomes for interventions. *Medicine & Science in Sports & Exercise* 2001, **33**(7):1126-1141.
102. Tinetti ME, Richman D, Powell L: Falls efficacy as a measure of fear of falling. *Journal of gerontology* 1990, **45**(6):P239-243.
103. Topolski TD, LoGerfo J, Patrick DL, Williams B, Walwick J, Patrick MB: The Rapid Assessment of Physical Activity (RAPA) among older adults. *Prev Chronic Dis* 2006, **3**(4):A118.
104. Tu SP, McDonnell MB, Spertus JA, Steele BG, Fihn SD: A new self-administered questionnaire to monitor health-related quality of life in patients with COPD. Ambulatory Care Quality Improvement Project (ACQUIP) Investigators. *Chest* 1997, **112**(3):614-622.
105. Tugwell P, Bombardier C, Buchanan WW, Goldsmith CH, Grace E, Hanna B: The MACTAR Patient Preference Disability Questionnaire—an individualized functional priority approach for assessing improvement in physical disability in clinical trials in rheumatoid arthritis. *The Journal of rheumatology* 1987, **14**(3):446-451.
106. van der Molen T, Willemse BW, Schokker S, ten Hacken NH, Postma DS, Juniper EF: Development, validity and responsiveness of the Clinical COPD Questionnaire. *Health & Quality of Life Outcomes* 2003, **1**(Journal Article):13.
107. Verbunt JA: Reliability and validity of the PAD questionnaire: a measure to assess pain-related decline in physical activity. *Journal of Rehabilitation Medicine* 2008, **40**(1):9-14.
108. Voorrips LE, Ravelli AC, Dongelmans PC, Deurenberg P, Van Staveren WA: A physical activity questionnaire for the elderly. *Medicine & Science in Sports & Exercise* 1991, **23**(8):974-979.
109. Washburn RA, Smith KW, Jette AM, Janney CA: The Physical Activity Scale for the Elderly (PASE): development and evaluation. *Journal of clinical epidemiology* 1993, **46**(2):153-162.
110. Weaver TE, Narsavage GL, Guilfoyle MJ: The development and psychometric evaluation of the Pulmonary Functional Status Scale: an instrument to assess functional status in pulmonary disease. *Journal of cardiopulmonary rehabilitation* 1998, **18**(2):105-111.
111. Wigal JK, Creer TL, Kotses H: The COPD Self-Efficacy Scale. *Chest* 1991, **99**(5):1193-1196.
112. Windisch W, Freidel K, Schucher B, Baumann H, Wiebel M, Matthys H, Petermann F: The Severe Respiratory Insufficiency (SRI) Questionnaire A specific measure of health-related quality of life in patients receiving home mechanical ventilation. *Journal of clinical epidemiology* 2003, **56**(8):752-759.
113. Yohannes AM, Roomi J, Winn S, Connolly MJ: The Manchester Respiratory Activities of Daily Living questionnaire: development, reliability, validity, and responsiveness to pulmonary rehabilitation. *Journal of the American Geriatrics Society* 2000, **48**(11):1496-1500.
114. Yoza Y, Ariyoshi K, Honda S, Taniguchi H, Senjyu H: Development of an activity of daily living scale for patients with COPD: the Activity of Daily Living Dyspnoea scale. *Respirology* 2009, **14**(3):429-435.
115. Zaragoza J, Lugli-Rivero Z: Development and Validation of a Quality of Life Questionnaire for Patients with Chronic Respiratory Disease (CV-PERC): Preliminary Results. *Construcción y validación del instrumento Calidad de Vida en Pacientes con Enfermedades Respiratorias Crónicas (CV-PERC) Resultados preliminares* 2009, **45**(2):81.
116. Zhou YQ, Chen SY, Jiang LD, Guo CY, Shen ZY, Huang PX, Wang JY: Development and evaluation of the quality of life instrument in chronic liver disease patients with minimal hepatic encephalopathy. *Journal of Gastroenterology & Hepatology* 2009, **24**(3):408-415.
117. Zisberg A: Influence of routine on functional status in elderly: development and validation of an instrument to measure routine. University of Washington; 2005.
118. Lombard C, Deeks A, Jolley D, Ball K, Teede H: A low intensity, community based lifestyle programme to prevent weight gain in women with young children: cluster randomised controlled trial. *BMJ (Clinical research ed)* 2010, **341**:c3215.
119. Neumark-Sztainer DR, Friend SE, Flattum CF, Hannan PJ, Story MT, Bauer KW, Feldman SB, Petrich CA: New moves-preventing weight-related problems in adolescent girls: A group-randomized study. *American Journal of Preventive Medicine* 2010, **39**(5):421-432.
120. Pekmezci DW, Neighbors CJ, Lee CS, Gans KM, Bock BC, Morrow KM, Marquez B, Dunsiger S, Marcus BH: A Culturally Adapted Physical Activity Intervention for Latinas. A Randomized Controlled Trial. *American Journal of Preventive Medicine* 2009, **37**(6):495-500.
121. U.S. Department of Health and Human Services: Guidance for industry - Patient-reported outcome measures: Use in medical product development to support labeling claims. Edited by: Administration FaD. Silver Spring, MD; 2009.
122. Bottomley A, Jones D, Claassens L: Patient-reported outcomes: Assessment and current perspectives of the guidelines of the Food and Drug Administration and the reflection paper of the European Medicines Agency. *European Journal of Cancer* 2009, **45**(3):347-353.
123. European Medicines Agency (EMA): CHMP Reflection Paper on the Regulatory Guidance for the Use of Health Related Quality of Life Measures in the Evaluation of Medicinal Products. 2006 [http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2009/09/WC500003637.pdf].

doi:10.1186/1477-7525-9-116

Cite this article as: Frei et al.: A comprehensive systematic review of the development process of 104 patient-reported outcomes (PROs) for physical activity in chronically ill and elderly people. *Health and Quality of Life Outcomes* 2011 **9**:116.

Submit your next manuscript to BioMed Central and take full advantage of:

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit

