

# The impact of measuring patient-reported outcomes in clinical practice: a systematic review of the literature

J. M. Valderas · A. Kotzeva · M. Espallargues · G. Guyatt · C. E. Ferrans ·  
M. Y. Halyard · D. A. Revicki · T. Symonds · A. Parada · J. Alonso

Accepted: 23 November 2007 / Published online: 4 January 2008  
© Springer Science+Business Media B.V. 2007

## Abstract

**Objective** The purpose of this paper is to summarize the best evidence regarding the impact of providing patient-reported outcomes (PRO) information to health care professionals in daily clinical practice.

**Methods** Systematic review of randomized clinical trials (Medline, Cochrane Library; reference lists of previous systematic reviews; and requests to authors and experts in the field).

**Results** Out of 1,861 identified references published between 1978 and 2007, 34 articles corresponding to 28 original studies proved eligible. Most trials (19) were conducted in primary care settings performed in the USA (21) and assessed adult patients (25). Information provided to professionals included generic health status (10), mental health (14), and other (6). Most studies suffered from

methodologic limitations, including analysis that did not correspond with the unit of allocation. In most trials, the impact of PRO was limited. Fifteen of 23 studies (65%) measuring process of care observed at least one significant result favoring the intervention, as did eight of 17 (47%) that measured outcomes of care.

**Conclusions** Methodological concerns limit the strength of inference regarding the impact of providing PRO information to clinicians. Results suggest great heterogeneity of impact; contexts and interventions that will yield important benefits remain to be clearly defined.

**Keywords** Patient-reported outcomes · Quality of life · Health status indicators · Outcome assessment · Clinical practice

---

J. M. Valderas  
National Primary Care Research and Development Center,  
The University of Manchester,  
Manchester, UK  
e-mail: jose.valderas@manchester.ac.uk

J. M. Valderas · J. Alonso (✉)  
Health Services Research Unit, Institut Municipal d'Investigació  
Mèdica (IMIM-Hospital del Mar), PRBB. Doctor Aiguader, 88,  
08003 Barcelona, Spain  
e-mail: jalonso@imim.es

A. Kotzeva · M. Espallargues · A. Parada  
Catalan Agency for Health Technology Assessment and  
Research (CAHTA), Roc Boronat, 81-95, 08005 Barcelona,  
Spain

A. Kotzeva  
e-mail: akotzeva@aatrm.catsalut.net

M. Espallargues  
e-mail: mespallargues@aatrm.catsalut.net

G. Guyatt  
McMaster University, Hamilton, Canada  
e-mail: guyatt@mcmaster.ca

C. E. Ferrans  
College of Nursing (MC 802), University of Illinois at Chicago,  
Chicago, IL, USA

M. Y. Halyard  
Mayo Clinic, Scottsdale, AZ, USA

D. A. Revicki  
Center for Health Outcomes Research, United Biosource  
Corporation, Bethesda, MD, USA

T. Symonds  
Worldwide Outcomes Research (IPC160), Pfizer Global  
Research and Development, Kent, UK

J. Alonso  
Department of Experimental and Life Sciences,  
Universitat Pompeu Fabra, Barcelona, Spain

## Abbreviations

PRO Patient-reported outcomes  
RCTs Randomized control trials

## Introduction

Patient-reported outcomes (PRO) are measurements of any aspect of a patient's health status that come directly from the patient [1]. There are two ways clinicians and their patients may benefit from these measures, which are usually the result of a standardized instrument or questionnaire. First, using instruments measuring PRO in clinical research, investigators can provide important evidence to inform clinicians' and patients' decisions about treatment alternatives. PRO measures may also provide added value from their use in daily clinical practice. This article explores this second potential use via a systematic review of randomized clinical trials.

### Potential benefits in daily clinical practice

For over a decade, investigators have been expressing interest in the use of PRO assessments in daily clinical practice, although it must be acknowledged that PRO researchers have been much more interested than have practicing clinicians. PRO assessments may have several potential benefits in daily clinical practice. First, patient information collected using standardized questionnaires may facilitate detection of physical or psychological problems that might otherwise be overlooked. PRO instruments can also be applied as standardized measures to monitor disease progression and to provide information about the impact of prescribed treatment [2–5]. Another benefit of using PRO measures in routine clinical care may be in facilitating patient–clinician communication and thus promoting the model of shared decision making. Patients and clinicians often need to establish common priorities and expectations regarding the outcomes of treatment and illness [6]. Establishing a common understanding may be important for meeting patients' disparate needs and for improving patients' satisfaction with health care and their adherence to prescribed treatment [7]. PRO measurement in clinical care may be also used to monitor outcomes as a strategy for quality improvement or to reward presumed superior care.

### Evidence of impact of patient-reported outcomes measures in clinical care

Many practical and attitudinal barriers exist to the effective use of PRO instruments for application in clinical practice [8, 9]. Most questionnaires are lengthy, and both patients and

providers may perceive them as burdensome. Clinicians must receive data and the associated interpretation promptly and in an understandable manner, and this may involve appreciable resources [10, 11]. Although available evidence supports that validity and reliability of these measures are comparable with routinely used clinical measures [12], skepticism about their clinical meaning also inhibits their use in practice [13]. Not least, the use of these measures might cause unintended harm, even if from a theoretical perspective only. Physical or psychological problems that might otherwise be overlooked may make them more of a concern for the patient. Rather than facilitating doctor–patient communication, this information might possibly interfere with it, and it could also force the discussion into areas that the clinician has little control over. Considering all these issues, before the use of PRO measures in clinical practice can be recommended at all, there is a need for rigorous evaluation of the impact of the use of PRO measures in clinical practice, ideally by conducting randomized control trials (RCTs).

Previous systematic reviews of RCTs on the use of PRO measures in clinical practice have varied in their conclusions [2, 3, 14, 15]. Greenhalgh et al. [2] suggested that feedback on overall patient assessment increases the detection of psychological and, to a lesser extent, functional problems but found little evidence of changes in management or outcomes. Later, Gilbody et al. [14] concluded that PRO benefit in improving psychosocial outcomes of patients with psychiatric disorders managed in nonpsychiatric setting was insufficient to mandate their use. Espallargues et al. [3] identified 23 RCTs or quasi-RCTs with considerable heterogeneity of results precluding definitive recommendations concerning their use. Because a number of clinical trials have been published since the most recent of these reviews, we undertook a systematic review to assess the best available evidence regarding the impact of routine use of PRO measures in daily primary and secondary health care on process of care, satisfaction with care, and health outcomes.

## Methods

We developed a detailed protocol describing the following consecutive stages of the process: (1) definition of eligibility criteria, (2) search of relevant published articles, (3) screening of titles and abstracts for eligibility, (4) full-text eligibility evaluation of potentially eligible studies, (5) validity assessment and data extraction, and (6) data analysis.

### Eligibility criteria

Studies were eligible if they met all of the following inclusion criteria:

1. They were RCTs in which individual physicians, groups of physicians (e.g., hospitals, practices), or patients were randomly allocated to one or more intervention groups and to a control group.

2. Participating patients were attending a health practitioner's office, an outpatient clinic, an emergency room, or a hospital.

3. Studies compared replicable interventions consisting of administration of standardized PRO questionnaire(s) and subsequent feedback to health care professionals versus routine clinical practice without any administration of PRO measures. Questionnaire results were disclosed only to the clinicians in the intervention group, with or without additional education concerning the optimal application of this information.

4. At least one of the following outcomes was reported: mortality, morbidity, health-related quality of life and related measures, clinician behavior, clinician impressions, patient satisfaction, or costs (health services use).

5. Language of publication was English, French, German, Italian, Russian, or Spanish, expanding the language restrictions of previous reviews.

We excluded studies in which patients in the intervention group had received clinical care from providers with a different skill set than that of those providing care to patients in the control group, and also studies in which PROs were endpoints of the trial but not included in feedback to the providers.

Before starting the review process, a pilot test of eligibility criteria was performed on a sample of articles, and the result was discussed in a team meeting in order to refine the criteria and increase concordance within the team.

#### Search strategy and data sources

A previous systematic review [3] published in 2000 provided a starting point for our work. This review used a comprehensive search strategy, and we therefore assumed that it had identified all possible eligible studies for the years 1966–1997. All 23 papers identified in that previous search were included in the full-text eligibility evaluation (stage 4 in “Methods”). We therefore searched for new studies only conducted from January 1998 onward.

A professional librarian (AP) formulated the search strategy (an updated version of that used in the 2000 review) and performed the search, including modification of the strategy on the basis of initial results. To overcome heterogeneity and the absence of pattern in key words and terms of the original search strategy, the search strategy used for this report was organized in three blocks, each capturing: (1) potential RCTs, (2) selected questionnaires and provision of feedback, and (3) PRO (details available

from the authors). The search was conveniently devised to be performed in Medline and the Cochrane Library, including the Cochrane Database of Systematic Reviews (CDSR), the Cochrane Controlled Trials Register (CCTR), and the Database of Abstracts of Reviews of Effectiveness (DARE). The Medline search was updated during the editorial process of the manuscript, with the last update being performed on 7 September 2007.

Other sources of potentially eligible articles included reference lists of all prior reviews on the subject, as well as (later in the process) references of studies that had been selected as eligible for our review. Authors and experts in the field (e.g., members of the research team and other colleagues involved in this area of research) also provided information about other published or unpublished studies of which they were aware.

#### Screening and eligibility evaluation

Six teams of two reviewers participated in all stages of the study selection process, and the number of publications reviewed was distributed equally among these pairs. During the screening stage, each reviewer in a pair evaluated all titles and abstracts of the primary studies identified in the bibliographic search to determine whether the study met our predetermined eligibility criteria. If either reviewer felt there was any possibility that an article would fulfill our eligibility criteria, they selected the reference for full-text evaluation (“low threshold” strategy).

In the subsequent stage of the study selection process, again, each investigator reviewed independently the full text of all papers assigned to his or her pair to determine eligibility and then completed the specially designed Eligibility Evaluation Form. A reviewer was never assigned an article that he or she had selected in the previous (screening) stage. Existing disagreements in each pair were resolved by discussion until consensus was reached. In the cases when this did not happen, an additional reviewer made the final decision on eligibility of the particular article.

Several publications would eventually result from the same study. Duplicate publications of the same study (i.e., studies that were conducted on the same population and used the same intervention, although they may have reported different analyses or may have been published in different formats) were classified according to the Decision Tree for Identification of Patterns of Duplicate Publications proposed by von Elm et al. [16] and were based on comparison of similarity of samples and outcomes of pairs of duplicates. As a result, we produced a final list of all eligible articles (both duplicate and not) corresponding to all the relevant studies.

## Data extraction and validity assessment

Data extraction included the following variables: (1) characteristics of participants in the study (both patients and health care professionals), (2) clinical area of practice, setting, and country, (3) number of participants (or cases and controls) randomized, not included, excluded, partially followed up and lost, (4) unit of randomization and unit of analysis (i.e., patient, physician, or other provider, practice, or hospital), (5) time period of observation (6) characteristics of the intervention (including content, format, source, recipient and timing), (7) type of PRO used to provide the feedback, and (8) all reported results on process and outcomes of care and on satisfaction with care.

Two reviewers independently extracted data and assessed each study's quality using a specifically designed Data Abstraction Form. Disagreements were resolved by discussion. Individual validity of the studies was assessed using a modified version of the Jadad scale [17]. The characteristics being assessed were randomization (up to 2 points), statistical analysis consistent with unit of randomization and with clustering (when needed) (1.5), blinding (1.5), and loss to follow-up (2). Theoretically, scores ranged from 0 to 7, with higher scores indicating better study quality.

To organize and systematically present the possible outcome variables, we identified those most frequently observed in previous reviews, and a list of the most common definitions of outcomes was developed. This list was provided to the reviewers, together with the Data Abstraction Form and a procedure manual containing the main guidelines and working definitions.

## Analysis

Characteristics of individual studies were reported, including setting, participants, methodology, instruments used, and design of implemented interventions. All study outcomes were identified and classified on the basis of their consistency (that is, the extent to which the outcome measures were defined similarly and monitored similarly). According to our conceptual framework, we classified the study outcomes in three main groups, starting from the most proximal to the moment of intervention, with other outcomes expected to occur later [18]: (1) process of care, with subgroups referring to the conceptual stages of the care as a whole—patient–provider communication, provider behavior (diagnosis, treatment, and use of health services) and patient behavior (compliance with treatment and change of attitude), (2) outcomes of care, with subgroups on patients' general health and self-perceived health status, and (3) satisfaction with care for both patients and

clinicians. In this model, changes in process of care would mediate further changes in either outcomes of care or satisfaction with care.

Given the high variability of endpoints of impact of routine provision of PRO feedback to health care professionals that was observed and the high number of endpoints being assessed in only two studies or fewer (limiting the generalizability of results), we selected indicators for further evaluation on the basis of two criteria: (1) their position in the conceptual framework (i.e., stage of care continuum as described above), and (2) frequency, in order to maximize comparability between studies.

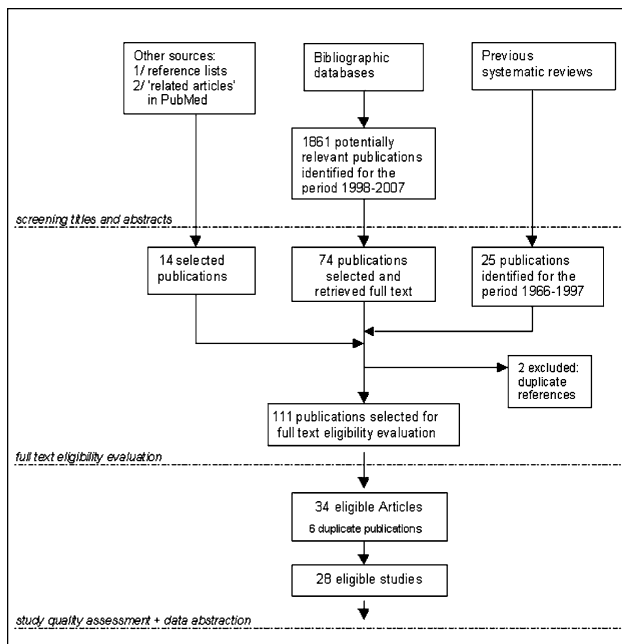
After the stage of study eligibility evaluation, the level of interreviewer agreement within each team and the median for each response item was calculated (Cohen's kappa [19]) to evaluate quantitatively the reliability of gathered evidence [20]. Kappa has a range of 0–1.00, with larger values indicating better reliability. For each eligibility criterion (presented as a separate response item in Form A), we calculated the median kappa for all reviewers. We used the statistical package STATA 8.0 for all data analyses.

## Results

The bibliographic search identified 1,861 potentially relevant publications, which we screened on the basis of their titles and abstracts (Fig. 1). All articles were identified by means of either the electronic search or by hand searching reference lists. After excluding all that were clearly non-randomized clinical trials or did not fulfill the selection criteria for our review, we selected 74 publications (71 in English, one in Spanish, one in German, and one in French) and obtained the full text of each article for more detailed eligibility evaluation. These references were added to the 25 publications reported by the previous systematic reviews [2, 3, 14, 15], together with another 14 publications identified as potentially relevant from other sources, including an additional systematic review retrieved during the updates [21] (Fig. 1).

In total, reviewers independently evaluated 111 full-text reports for eligibility, with substantial overall agreement, presented by a median kappa of 0.90 (range 0.53–1.00). We identified six pairs of duplicate publications, with four reporting identical samples and different outcomes (references 28–29, 36–37, 38–39, and 63–64), and two reporting both identical samples and identical outcomes (references 24–34 and 25–30). Thus, 34 publications corresponding to 28 eligible RCTs were included and reported in this systematic review (Table 1).

The majority of studies were performed in the USA (75%) and in primary care settings (67.9%). More than half



**Fig. 1** Flow chart of study selection

of the studies (57.1%) included only fully trained physicians (general practitioners, internists, and other specialists). In the remainder, a variety of health professionals participated, including residents and trainees and (in five studies) nurses and physician assistants. Patients, mainly adults, often came from a variety of restricted population groups: geriatric, mental health, or with a specific diagnosis.

Units of randomization in 14 of the trials were patients (in six) or physicians and groups of physicians (or practices, clinic modules, etc.) (in eight). The design of 12 RCTs included screening before randomization. The mean quality score was 3.41, with individual scores ranging from 2.0 to 5.0 points (Table 1). The most common quality limitations included analysis of data not keeping to the design (that is, analyzing data as if patients had been randomized when the unit of allocation was clinicians or groups of clinicians) and the inclusion of large numbers of outcomes, making it difficult to establish the significance of one or two positive results in a particular trial.

Studies varied in the ways in which the intervention was designed and implemented in the clinical care setting. Single feedback to clinicians was performed in eight studies and consisted of measurement and provision of feedback on patient's self-reported health status to the intervention-group clinicians occurring just once. The rest of the trials (20) performed some kind of complex intervention: feedback supplemented with either other intervention(s) to clinicians and/or patients or multiple feedback (PRO results provided to clinicians more than once) (Table 2).

The type and content of the information provided to clinicians differed in its characteristics across studies. In 14 studies, only the individual scores of the intervention questionnaire (e.g., each patient's results) were fed back, and in 13, the scores were accompanied by ranges of scores and/or explanations of individual scores. In one study only, a note was attached to patients' visit forms indicating to the physician that this person scored as "mildly depressed" or "severely depressed" [53]. In four trials, clinicians received diagnosis and/or treatment recommendations and guidelines on patient management. Other information, such as longitudinal information (previous scores for that patient), special notifications, summaries, etc., was provided together with the feedback in ten studies. In 15 publications, the presentation form of feedback was described: narrative in nine, graphic in two, and narrative and graphic in three. For the rest, such information was not provided.

The interval between questionnaire administration and provision of feedback to practitioners varied from immediate to 6 months. In most studies (19), provision of feedback was thoroughly described and occurred on the same day of the patient's visit and PRO measurement. The scored results reached the clinician before or during the consultation and were given personally by the research assistant, by the patient, or attached to the visit chart. In two of these cases, feedback was provided as a combination of "just before" and "just after" the visit [27, 33, 36]. In another five studies, clinicians received feedback on patients' scores after more than 1 month. The remaining four trials left this information unclear.

Intervention(s) other than feedback were part of the design and targeted practitioners in 20 of the trials and patients in another 5 studies. Oral or written presentation of the study (e.g., protocol, objectives, hypotheses tested) had been given to the participating clinical staff in 4 of the trials. Educational sessions were conducted before feeding back any information in 12 trials. These sessions focused on explanations of the instruments being used, building skills to use them, and the provision of interpretation aids (mailed or handed out). In 2 of these studies, the intervention group clinicians participated in a pilot study or in a training session. Another common additional intervention, observed in 7 of the trials, was the provision of guidelines, algorithms and/or tailored recommendations on how to manage specific patients or conditions. In two studies, the investigative team provided assistance with arranging follow-up visits and/or referrals, and phone calls to the patients.

The interventions other than feedback aimed at patients included: (1) written and oral presentation of the study and its goals by the research staff [43], (2) series of educational group sessions (six sessions plus a booster session

**Table 1** Studies on impact of routine patient-reported outcome measurement and feedback in clinical practice ( $n = 28$ )

Reference	Design	Unit of randomization	Setting	Clinicians	Patients ( $n$ )	Sample size calculations <sup>a</sup>	Intervention	Intervention instruments	Quality score
Brody et al. [22]	RCT (2 intervention and 1 control group) Screening before randomization	Practices ( $n = 4$ )	Outpatient (hospital-based), USA	Residents in internal medicine ( $n = 60$ )	Mental health problems ( $n = 103$ )	Not reported	Feedback plus additional intervention (clinicians)	GHQ (12-item, specific, psychiatric)	2.0
Calkins et al. [23]	RCT (2 groups) Screening before randomization	Practice teams ( $n = 4$ )	Outpatient (hospital-based), USA	Internists and residents ( $n = 60$ )	Adults, known, disabled ( $n = 497$ )	Not reported	Feedback plus additional intervention (clinicians)	FSQ (34-item, generic)	3.5
Callahan et al. [63] Callahan et al. [64]	RCT (2 groups)	Practices ( $n = 32$ )	Primary care, USA	General practitioners ( $n = 111$ )	Aged ( $\geq 60$ years) Depressed ( $n = 222$ )	Not reported	Feedback plus additional interventions (patients and clinicians)	HAMD (specific, psychiatric)	2.5
Dailey et al. [26]	RCT (2 groups) Screening before randomization	Patients	Dental practices, UK	General dental practitioners ( $n = 8$ )	Adults, new ( $n = 123$ )	Based on a difference of 1.2 on the Spielberger state anxiety scale	Feedback plus additional intervention (clinicians)	MDAS	3.5
Detmar et al. [27]	RCT (crossover) Screening before randomization	Physicians	Outpatient cancer clinic, the Netherlands	Oncologists ( $n = 10$ )	Oncology patients, known adults, known ( $n = 273$ )	Based on a moderate between-group difference (effect size = 0.40) in frequency of problems discussion	Feedback plus additional interventions (patients and clinicians)	EORTC QLQ-C30 (specific, neoplasm)	3.5
Dowrick et al. [28] Dowrick et al. [29]	RCT (2 groups plus detected control group) Screening before randomization	Patients	Primary care, UK	General practitioners ( $n = 9$ )	Adults (16–64 years), depressed, known ( $n = 116$ )	Based on a 15% difference in depression score	Feedback plus additional intervention (clinicians)	BDI (specific, psychiatric)	4.5
German et al. [25] Shapiro et al. [30]	RCT (2 intervention and 3 control groups) Screening before randomization	Patients	Primary care, USA	General practitioners and residents ( $n = 45$ ), physician assistants	Adults ( $n = 809$ )	Not reported	Feedback	GHQ (28-item, specific, psychiatric); DIS, incl. Mini Mental)	3.0

Table 1 continued

Reference	Design	Unit of randomization	Setting	Clinicians	Patients (n)	Sample size calculations <sup>a</sup>	Intervention	Intervention instruments	Quality score
Goldsmith et al. [31]	RCT (1 intervention and 3 control groups)	Physicians	Primary care, USA	General practitioners and residents (n = 27)	Adults (≥55 years) With ≥1 chronic conditions (n = 62)	Not reported	Feedback plus additional intervention (clinicians)	SIP (136-item, generic)	4.5
Hooper et al. [32]	RCT (2 groups)	Patients	Primary care, USA	General practitioners and internists (n = 14)	Adults, new and known (n = 1,469)	Not reported	Feedback	GHQ (28-item, specific, psychiatric)	3.5
Kazis et al. [33]	RCT (1 intervention and 2 control groups)	Patients	Outpatient, USA	Arthritis specialists (n = 27)	Adults, arthritic (n = 1,920)	Based on mild to moderate effects (ES = 0.20–0.50)	Feedback	AIMS (specific, arthritis) MHAQ health status report (generic)	2.5
Linn et al. [34]	RCT (5 intervention and 1 control group)	Patients	Outpatient, USA	Internists and residents (n = 78)	Adults, new General medical (n = 150)	Not reported	Feedback	Zung-SDS; GDI (specific, psychiatric)	3.0
Magruder-Habib et al. [35]	RCT (1 intervention and 3 control groups) Screening before randomization	Patients	Primary care, USA	Internists, residents, and physician assistants (n = 12)	Adults, new and known Depressed (n = 100)	Not reported	Feedback	Zung-SDS (specific, psychiatric)	4.5
Mathias et al. [36]	RCT (2 groups)	Practices (n = 23)	Primary care (HMO), USA	General practitioners (n = 75)	Adult (21–65 years) not new Depressed, anxious (n = 573)	Not reported	Feedback plus additional intervention (clinicians)	SCL-90-R (specific, psychiatric) SF-36 (generic)	2.5
Mazonson et al. [37]	Screening before randomization	Patients (by day of visit)	Emergency department, Canada	Emergency department physicians, geriatric nurses	Geriatric (≥65 years) (n = 388)	Not reported	Feedback plus additional interventions (patients and clinicians)	ISAR; standardized geriatric nursing assessment	3.5
McCusker et al. [38]	RCT (2 groups)	Patients (by day of visit)	Emergency department, Canada	Emergency department physicians, geriatric nurses	Geriatric (≥65 years) (n = 388)	Not reported	Feedback plus additional interventions (patients and clinicians)	ISAR; standardized geriatric nursing assessment	3.5
McCusker et al. [39]	Screening before randomization	Patients (by day of visit)	Emergency department, Canada	Emergency department physicians, geriatric nurses	Geriatric (≥65 years) (n = 388)	Not reported	Feedback plus additional interventions (patients and clinicians)	ISAR; standardized geriatric nursing assessment	3.5

Table 1 continued

Reference	Design	Unit of randomization	Setting	Clinicians	Patients (n)	Sample size calculations <sup>a</sup>	Intervention	Intervention instruments	Quality score
Moore et al. [40]	RCT (2 groups) Screening before randomization	Patients	Primary care, USA	Residents	Adults (20–60 years), new and known (n = 96)	Not reported	Feedback	SDS (20-item, specific, psychiatric)	3.0
Rand et al. [41]	RCT (2 groups)	Physicians	Primary care, USA	Family practice residents (n = 23)	Adults, new and known (n = 1,040)	Not reported	Feedback plus additional intervention (clinicians)	GHQ (28-item, specific, psychiatric)	3.0
Rubenstein et al. [58]	RCT (2 groups) Screening before randomization	Physicians	Primary care, USA	Internists and general practitioners (n = 76)	Adults, known Disabled (n = 649)	Not reported	Feedback plus additional intervention (clinicians)	Beth Israel/UCLA FSQ (generic)	3.0
Rubenstein et al. [27]	RCT (2 groups)	Clinic modules (n = 2)	Primary care, USA	General internist residents (n = 73)	Adults New (n = 557)	Not reported	Feedback plus additional interventions (patients and clinicians)	FSQ (34-item, generic)	4.0
Saitz et al. [44]	RCT (2 groups)	Physicians	Primary care, USA	General practitioners and faculty physicians (n = 41)	Adults (n = 312)	Based on a 50% increase in counseling from 33% + decrease of 1.3 drinks per drinking day	Feedback plus additional intervention (clinicians)	CAGE (specific, alcohol screening); additional questions	4.5
Smith [45]	RCT (2 groups)	Patients (by day of week)	Primary care, UK	General practitioners (n = 8)	Mixed pediatric and adult (n = 1,382)	Not reported	Feedback plus additional intervention (clinicians)	30-item GHQ (specific, psychiatric)	3.0
Velikova et al. [46]	RCT (1 intervention and 2 control groups)	Patients	Outpatient oncology clinic, UK	Oncology consultants and physicians in training (n = 28)	Oncology patients, adults, known (n = 286)	Not reported	Feedback plus additional interventions (clinicians)	EORTC QLQ - C30 (specific, neoplasm) HADS (specific psychiatry)	4.5
Wagner et al. [47]	RCT (2 groups)	Patients	Outpatient, USA	Neurologists (epileptologists) (n = 2)	Adults (≥18 years) not new Epileptic (n = 163)	Not reported	Feedback plus additional intervention (clinicians)	SF-36 (generic)	3.0
Wasson et al. [48]	RCT (2 groups)	Physicians	Primary care (HMO), USA	Internists, residents, and nurses (n = 56)	Adults (≥18 years) Not new (n = 1,564)	Not reported	Feedback plus additional intervention (clinicians)	COOP charts (6 charts, generic)	3.0



Table 1 continued

Reference	Design	Unit of randomization	Setting	Clinicians	Patients (n)	Sample size calculations <sup>a</sup>	Intervention	Intervention instruments	Quality score
Wasson et al. [49]	RCT (2 groups)	Practices (n = 22)	Primary care, USA	General practitioners and internists (n = 45)	Geriatric ( $\geq 70$ years), known (n = 1,651)	Not reported	Feedback plus additional intervention (patients)	30-item geriatric health assessment survey	3.5
White et al. [50]	RCT (2 groups)	Practices (n = 23)	Primary care, UK	General practitioners, nurses, and practice managers	Mixed pediatric and adults, known, asthmatic (n = 818)	Not reported	Feedback plus additional intervention (clinicians)	Questionnaire about patient morbidity, health services and drug use	2.5
Whooley et al. [51]	RCT (2 groups) Screening before randomization	Medical clinics	Primary care, USA	General practitioners and internists	Geriatric ( $\geq 65$ years), new and known (n = 331)	Not reported	Feedback plus additional interventions (patients and clinicians)	Geriatric Depression Scale (15-item, specific)	3.0
Williams Jr. et al. [52]	RCT (2 intervention and 1 control group)	Patients	Primary care practice and 3 general internal medicine clinics, USA	General practitioners, internists, and residents (n = 80)	Adults (n = 969)	Not reported	Feedback	20-item Centre for Epidemiologic Studies Depression Screen; DSM-III-R	5.0
Zung et al. [42]	RCT (2 intervention and 1 control group) Screening before randomization	Patients	Primary care, USA	General practitioners	Depressed (n = 143)	Not reported	Feedback	Zung-SDS (specific, psychiatric)	4.0

RCT randomized controlled trials, SDS Self-rating Depression Scale, GHQ General Health Questionnaire, FSQ Functional Status Questionnaire, HAM-D The Hamilton Rating Scale for Depression, MDAS Modified Dental Anxiety Scale, EORTC QLQ C30 European Organization for Research and Treatment of Cancer Quality of Life Questionnaire, BDI Beck's Depression Inventory, GDI Global Depression Index, DIS Diagnostic Interview Schedule, SIP Sickness Impact Profile, AIMS Arthritis Impact Measurement Scale, MHAQ Modified Health Assessment Questionnaire, SCL-90-R Symptom Check List-90-Revised, ISAR Identification of Seniors at Risk, UCLA FSQ University of California, Los Angeles Functional Status Questionnaire, CAGE, HADS Hospital Anxiety and Depression Scale, SF-36 Medical Outcomes Study Short Form 36, COOP Dartmouth Primary Care Cooperative Information Project

<sup>a</sup> When performed, all simple size calculations based on  $\alpha = 0.0.5$  and  $\beta = 0.20$

**Table 2** Characteristics of the interventions ( $n = 28$ )

	Number	Percent
<b>Intervention</b>		
Single feedback <sup>a</sup> only	8	28.6
Complex intervention <sup>b</sup>	20	71.4
Feedback with additional intervention(s) for the clinicians	19	67.9
Feedback with additional intervention(s) for the patients	6	21.4
Multiple feedback	13	46.4
<b>Content of feedback<sup>b</sup></b>		
Scores only (e.g., each patient's results)	14	50.0
Scores with ranges and/or explanation of individual scores	13	46.4
Diagnosis/treatment recommendations (guidelines) in addition	4	14.3
Other information in addition (summaries, notifications, previous scores, etc.)	10	35.7
<b>Interval (between administration and feedback)</b>		
Feedback on the same visit as PRO measurement	19	67.9
Feedback in more than 1 month	5	17.9
Uncertain	4	14.3
<b>Instruments<sup>b</sup></b>		
(a) Generic	11	39.3
Specific	17	60.7
Mental health	14	50.0
Other <sup>c</sup>	5	17.9
(b) Validated	18	64.3
Both validated and not validated	7	25.0
Validation not ascertained	3	10.7
<b>Administration</b>		
(a) Self-administered	17	60.7
Interviewer-administered, face to face	4	14.3
Combination of self- and interviewer-administered instrument(s)	7	25.0
(b) Completed at home	4	14.3
Completed in the waiting room or office	12	42.9
Mailed on patient's address	6	28.6
(c) Handed and instructed by:		
Research staff (e.g., research assistant)	14	50.0
Clinical staff (e.g., clinician or reception staff)	4	14.3
Both research and clinical staff	1	3.6
Not ascertained	8	28.6

<sup>a</sup> Single feedback was defined as measurement and provision of patient-reported outcome (PRO) results occurring only once. Multiple feedback was accordingly defined as feedback occurring more often than once

<sup>b</sup> Categories not mutually exclusive

<sup>c</sup> European Organization for Research and Treatment of Cancer Quality of Life Questionnaire (neoplasms), CAGE (alcohol screening), Modified Dental Anxiety Scale, Arthritis Impact Measurement Scale

4–6 months later) on coping with depression, led by a psychiatric nurse, to which family members were also invited [51], (3) promotional and educational pamphlet mailed to patient's home prior to the intervention; further explanation by research assistant was available if desired [27], (4) personalized letter with summary of clinically significant results and tailored guidelines and recommendations sent after the

intervention [42], and (5) limited follow-up conducted by a nurse after the intervention to ensure that appointments and services were provided [38, 39].

In general, the instruments used in the trials were well known and validated. Eleven trials used generic measures, such as the Medical Outcomes Study Short-Form 36 (SF-36), the Functional Status Questionnaire (FSQ), Dartmouth

**Table 3** Endpoints of trials assessing the impact of feedback of patient-reported outcomes (PRO)

	Studies ( <i>n</i> )
Process of health care	
I. Patient–physician communication	
1. Offering advice, education, counseling	7
2. Duration of visit or consultation	3
3. Number of physician- or patient-initiated conversations about quality of life	1
II. Diagnosis and recognition	
1. Number (rate) of target diagnoses made (such as diagnosis of depression)	14
2. Number (rate) of target problems identified (conditions, functional status problems, etc.)	7
III. Treatment	
1. Number (rate) of specific target interventions undertaken (e.g., treatment of depression)	8
2. Prescription of any non-target medication	4
3. Treatment changes	2
4. Global score of actions and interventions undertaken	3
IV. Health services and resources use	
1. Outpatient visits (number or rate)	4
2. Consultations or referrals to specialists or social institutions (number or rate)	11
3. Hospitalizations (number or rate)	3
4. Follow-up appointments	2
V. Patient behavior	
1. Compliance with the treatment and/or counseling	2
2. Change in attitudes towards treatment and/or counseling	1
VI. Physician awareness of patients' quality of life	
1. Congruence between clinician and patient assessments	4
Outcomes of health care	
VII. General health outcomes	
1. Mortality	2
2. Morbidity (change in number or rate of diagnoses, days of illness and hospitalization)	2
3. General functional status (change in number or rate of symptoms, complaints)	6
4. Indirect indicators of morbidity—estimated hospital and/or medication cost	1
VIII. Patients' health-related quality of life	
A. Overall score only	5
B. Physical function	
1. Change in score on physical function from generic measure	3
2. Change in score on physical function from specific measure	0
3. Change in score of separate physical function domains (items?)	4
C. Emotional function	
1. Change in score on emotional function from generic measure	2
2. Change in score on emotional function from specific measure	5
3. Change in score of separate emotional function domains	2
Satisfaction with health care	
IX. Patient satisfaction	
1. Overall satisfaction with care (assessment of health care)	4
2. Satisfaction with accessibility and technical aspects of health care	3
3. Satisfaction with communication and interpersonal qualities of physician	5
4. Satisfaction with emotional support	1
5. Satisfaction with time spent for consultation or counseling	1
6. Perceptions about usefulness of intervention, satisfaction with results	6

**Table 3** continued

	Studies ( <i>n</i> )
X. Clinician satisfaction	
1. Attitude and overall satisfaction with intervention	1
2. Satisfaction with communication with patient	1
3. Physician-rated usefulness of information from the health-related quality of life instrument	7
4. Physician-rated impact of intervention on patient management process	4

Primary Care Cooperative Information Project (COOP) charts, or the Modified Health Assessment Questionnaire (MHAQ). The other studies used condition- or disease-specific measures with a high prevalence of popular mental health instruments (General Health Questionnaire, Zung Self-rating Depression Scale, Hospital Anxiety and Depression Scale, etc.) but also used instruments specific for neoplasms [European Organization for Research and Treatment of Cancer Quality of Life Questionnaire (EORTC QLQ-C30)], arthritis [Arthritis Impact Measurement Scale (AIMS)], dental anxiety [Modified Dental Anxiety Scale (MDAS)] and alcohol screening (CAGE Questionnaire). Some studies used more than one questionnaire or additional questions and standardized assessments (for more details, see Table 3).

The modes of instrument administration varied across studies. In more than half (17), the instrument was self-administered, mainly unsupervised, and in one case used touch-screen questionnaires [46]. Seven studies used a combination of self- and interviewer-administered measures; only in four trials was the questionnaire administered face to face to the patient by an interviewer. In 13 studies, patients completed the questionnaires in the waiting room or clinician's office, and in 13, a research assistant handed out the questionnaires and gave instructions on their completion. In four trials, the clinician or some other member of the clinical staff participated actively in the instrument administration, and in one study [42], both research and clinical staff took part in questionnaire

administration. In the rest of the trials (8), this information was not provided.

Investigators measured and reported a wide variety of PRO, which were not completely comparable (Table 4). This heterogeneity hindered assessment of the impact of the interventions [54, 55]. The majority of studies assessed outcomes from more than one aspect of health care (process, outcomes, and satisfaction with care) without taking multiple comparisons into account [56, 57]. The impact of provision of feedback on PRO measurement to clinicians on process of health care was assessed by 23 trials, and 15 (65%) from this group reported a statistically significant difference for at least one of the variables. Seventeen trials studied the effects on outcomes of care, and 12 trials studied the effects on satisfaction with care. Eight (47%) of the former and five (42%) of the latter reported significant improvements. Seven trials studied the effects and reported outcomes only in the area of process of health care: two on effects only on outcomes of health care and one on effects only on satisfaction with care.

The five outcomes selected for evaluation of intervention effects (see "Methods") were: (1) offering advice, education, and counseling during the visit (the most proximal outcome, reflecting the initial patient–physician communication, assessed by seven trials), (2) number of target diagnoses and notations made in the medical chart (the most frequently assessed outcome, assessed in 14 trials), (3) number of consultations or referrals (indicating effect on use of health services, assessed by 11 trials), (4)

**Table 4** Significant results reported by the studies for the main groups and for the selected outcomes (indicators)

Outcomes	Studies ( <i>n</i> )	Significant results (%) <sup>a</sup>
<i>Process of care</i>	23	15 (65%)
Advice, education, counseling	7	3 (43%)
Target diagnoses and notations	14	7 (50%)
Referrals, consultations	11	2 (18%)
<i>Outcomes of care</i>	17	8 (47%)
General functional status	6	3 (50%)
<i>Satisfaction with care</i>	12	5 (42%)
Physician-rated utility of intervention	6	66% (28–97) <sup>b</sup>

<sup>a</sup> Statistically significant difference favoring measurement of patient-reported outcomes for at least 1 variable in the group

<sup>b</sup> Median percentage (range). Measured in the intervention groups only

general functional status of the patient (change of symptoms, complaints) (indicator of the effect on outcomes of care, assessed by six studies), and (5) physician-rated usefulness of information from the PRO instrument (reflecting physician satisfaction and assessed by six trials) (Table 4).

## Discussion

In this systematic review, we present a comprehensive compilation of the evidence on the impact of measuring PRO in clinical practice. Most studies found intervention effects on at least one aspect of the process outcomes assessed; effects on patient health status were less frequently assessed and observed.

We acknowledge some limitations of this review. We performed our search only in two databases (Medline and The Cochrane Collaboration Database). We tried to overcome these limitations by expanding study retrieval based on the references cited in the eligible and reviewed studies. Unfortunately, in most cases when information from the studies was unclear or incomplete, we failed to obtain clarification from the authors. The fact that so many indicators were used in just one or two studies reflects the lack of researchers' consensus on the indicators' relevance and posed yet another challenge to a quantitative summary in this review. Our resulting choice of representative indicators and the criteria we relied on can be subjected to dispute also. However, these were based on explicit criteria and a replicable methodology, increasing the accountability of the procedure. The RCTs analyzed were heterogeneous in the types of settings, participants (both patients and clinicians), intensity of intervention implemented, and diversity of outcomes reported. This fact represents a major challenge to the evaluation of the impact of providing feedback to health professionals and specifically makes a formal quantitative analysis difficult. In addition, many studies were of limited methodological quality.

All these issues prevented us from obtaining a quantitative estimate of the impact of PRO feedback in clinical practice. Although the studies included in our review suggest that it has an as yet unquantified effect on health care (especially on process variables), more research is clearly needed before these types of interventions can be recommended. Specifically, there is a need for well-designed and well-conducted randomized studies that use appropriate statistical methods that consider the unit of randomization and the multiplicity of outcomes and that follow the reporting guidelines currently available.

A number of studies suggest that the ways in which information on PRO is implemented in routine clinical practice and the clinical relevance of this feedback will influence its impact on patient management and outcomes [33, 35, 47, 58–61]. We were, however, unable to find any clear and strong patterns across studies that might have identified intervention characteristics associated with successful outcomes.

A practical limitation that those considering using PRO measures in clinical practice should bear in mind is that, in most of these studies, the intervention (sometimes fairly intensive) was organized and delivered by research staff. In clinical practice, clinical staff would be responsible for implementation.

Our review has a sound methodological basis, and was devised to follow the guidelines published in the Cochrane Reviewers' Handbook 4.2.0 [56]. We compared all previously available systematic reviews [2, 3, 14, 15] on the subject and, in the study selection process, took maximum benefit from the information provided by these previous studies. We implemented a comprehensive search strategy, which showed a slight increase of publications in the past few years. We also achieved a more complete and reliable review process than previous efforts by means of the independent evaluation by two investigators in each stage (screening, eligibility evaluation, validity assessment, and data extraction), as shown by the substantial level of interrater concordance. Furthermore, a recent study has demonstrated the need for a frequent update of systematic reviews relevant to clinical practice [62].

We conclude that whereas there are some grounds for optimism in the possible impact of measurement of PRO in clinical practice (specifically in improving diagnosis and recognition of problems and patient–physician communication), considerable work is still required before clinicians can invest resources in the process and rely on consistent evidence for the benefits for their patients. A number of methodologically stronger trials successfully implementing feasible interventions with clear positive effects are required to provide clear direction for clinicians interested in improving their care through routine use of PRO measures.

**Acknowledgements** This work was supported in part by: Ministerio de Educación, Cultura y Deporte (SAB2001-0088); the Spanish Public Health Network (RCESP) “*Investigación en Epidemiología y Salud Pública: Determinantes, Mecanismos, Métodos y Políticas*” (Exp. C03/09), and the Instituto de Salud Carlos III: Red de Investigación Cooperativa IRYSS (Exp. G03/202), and Ayudas para contratos de Profesionales Post MIR (Exp. CM0300118). A preliminary report of this work was presented at the Annual Conference of the International Society for Quality of Life Research in Hong Kong, China (2004).

## References

- Food and Drug Administration (2006). Draft guidance for industry on patient-reported outcome measures: Use in medicinal product development to support labeling claims. *Federal Register*, 71, 5862–5863.
- Greenhalgh, J., & Meadows, K. (1999). The effectiveness of the use of patient-based measures of health in routine practice in improving the process and outcomes of patient care: A literature review. *Journal of Evaluation in Clinical Practice*, 5, 401–416.
- Espallargues, M., Valderas, J. M., & Alonso, J. (2000). Provision of feedback on perceived health status to health care professionals: A systematic review of its impact. *Medical Care*, 38, 175–186.
- McHorney, C. A. (1999). Health status assessment methods for adults: Past accomplishments and future challenges. *Annual Review of Public Health*, 20, 309–335.
- Lohr, K. N. (1992). Applications of health status assessment measures in clinical practice. Overview of the third conference on advances in health status assessment. *Medical Care*, 30, MS1–MS14.
- Rothwell, P. M., McDowell, Z., Wong, C. K., & Dorman, P. J. (1997). Doctors and patient don't agree: Cross sectional study of patients' and doctors' perceptions and assessments of disability in multiple sclerosis. *British Medical Journal*, 314, 1580–1583.
- Stimson, G. V. (1974). Obeying doctor's orders: A view from the other side. *Social Science & Medicine*, 8(2), 97–104.
- Deyo, R. A., & Carter, W. B. (1992). Strategies for improving and expanding the application of health status measures in clinical settings. A researcher-developer viewpoint. *Medical Care*, 30(5 Suppl), MS176–MS186.
- Deyo, R. A., & Patrick, D. L. (1989). Barriers to the use of health status measures in clinical investigation, patient care, and policy research. *Medical Care*, 27, S254–S267.
- Morris, J., Perez, D., & McNoe, B. (1998). The use of quality of life data in clinical practice. *Quality of Life Research*, 7, 85–91.
- Meadows, K. A., Rogers, D., & Greene, T. (1998a). Attitudes to the use of health outcome questionnaires in the routine care of patients with diabetes: A survey of general practitioners and practice nurses. *The British Journal of General Practice*, 48, 1555–1559.
- Hahn, E. A., Cella, D., Chassany, O., Fairclough, D. L., Wong, G. Y., & Hays, R. D. (2007). Clinical Significance Consensus Meeting Group. Precision of health-related quality-of-life data compared with other clinical measures. *Mayo Clinic Proceedings*, 82(10), 1244–1254.
- Fitzpatrick, R., Fletcher, A., Gore, S., Jones, D., Spiegelhalter, D., & Cox, D. (1992). Quality of Life measures in health care. I: Applications and issues in assessment. *British Medical Journal*, 305, 1074–1077.
- Gilbody, S. M., House, A. O., & Sheldon, T. A. (2001). Routine administered questionnaires for depression and anxiety: Systematic review. *British Medical Journal*, 322, 406–409.
- Gilbody, S. M., House, A. O., & Sheldon, T. A. (2002). Routine administration of Health Related Quality of Life (HRQoL) and needs assessments instruments to improve psychological outcome—a systematic review. *Psychological Medicine*, 32, 1345–1356.
- von Elm, E., Poglia, G., Walder, B., & Tramèr, M. R. (2004). Different patterns of duplicate publication. An analysis of articles used in systematic reviews. *The Journal of the American Medical Association*, 291, 974–980.
- Jadad, A. R., Moore, R. A., Carroll, D., Jenkinson, C., Reynolds, D. J., Gavaghan, D. J., et al. (1996). Assessing the quality of reports of randomized clinical trials: Is blinding necessary? *Controlled Clinical Trials*, 17(1), 1–12.
- Valderas, J. M., Rue, M., Guyatt, G., & Alonso, J. (2005). Systematic Use of Quality of Life Measures in the Clinical Practice Working Group. The impact of the VF-14 index, a perceived visual function measure, in the routine management of cataract patients. *Quality of Life Research*, 14(7), 1743–1753.
- Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20, 37–46.
- Kraemer, H. C. (1992). Measurement of reliability for categorical data in medical research. *Statistical Methods in Medical Research*, 1(2), 183–99. Review.
- Marshall, S., Haywood, K., & Fitzpatrick, R. (2006). Impact of patient reported outcome measures on routine practice: a structured review. *Journal of Evaluation in Clinical Practice*, 12(5), 559–568.
- Brody, D. S., Lerman, C. E., Wolfson, H. G., & Caputo, G. C. (1990). Improvement in physicians' counseling of patients with mental health problems. *Archives of Internal Medicine*, 150, 993–998.
- Calkins, D. R., Rubenstein, L. V., Cleary, P. D., Davies, A. R., Jette, A. M., Fink, A., et al. (1994). Functional disability screening of ambulatory patients: A randomized controlled trial in a hospital-based group practice. *Journal of General Internal Medicine*, 9(10), 590–592.
- Yager, J., & Linn, L. S. (1981). Physician-patient agreement about depression: Notation in medical records. *General Hospital Psychiatry*, 3, 271–276.
- German, P. S., Shapiro, S., Skinner, E. A., Von Korff, M., Klein, L. E., Turner, R. W., et al. (1987). Detection and management of mental health problems of older patients by primary care providers. *Journal of the American Medical Association*, 257, 489–493.
- Dailey, Y. M., Humphris, G. M., & Lennon, M. A. (2002). Reducing patients' state anxiety in general dental practice: A randomized controlled trial. *Journal of Dental Research*, 81(5), 319–322.
- Detmar, S. B., Muller, M. J., Schornagel, J. H., Wever, L. D.V., & Aaronson, N. K. (2002). Health related quality of life assessments and patient physician communication: A randomized controlled trial. *The Journal of the American Medical Association*, 228(23), 3027–3034.
- Dowrick, C. (1995). Does testing for depression influence diagnosis or management by general practitioners? *Family Practice*, 12, 461–465.
- Dowrick, C., & Buchan, I. (1995). Twelve month outcome of depression in general practice: Does detection or disclosure make a difference? *British Medical Journal*, 311, 1274–1276.
- Shapiro, S., German, P. S., Skinner, E. A., et al. (1987). An experiment to change detection and management of mental morbidity in primary care. *Medical Care*, 25, 327–339.
- Goldsmith, G., & Brodwick, M. (1989). Assessing the functional status of older patients with chronic illness. *Family Medicine*, 21, 38–41.
- Hoepfer, E. W., Nycz, G. R., Kessler, L. G., Burke, J. D. Jr., & Pierce, W. E. (1984). The usefulness of screening for mental illness. *Lancet*, 1, 33–35.
- Kazis, L. E., Callahan, L. F., Meenan, R. F., & Pincus, T. (1990). Health status reports in the care of patients with rheumatoid arthritis. *Journal of Clinical Epidemiology*, 43, 1243–1253.
- Linn, L. S., & Yager, J. (1980). The effect of screening, sensitization and feedback on notation of depression. *Journal of Medical Education*, 55, 942.
- Magruder-Habib, K., Zung, W. W.K., & Feussner, J. R. (1990). Improving physicians' recognition and treatment of depression in

- general medical care. Results from a randomized clinical trial. *Medical Care*, 28, 239–250.
36. Mathias, S. D., Fifer, S. K., Mazonson, P. D., Lubeck, D. P., Buesching, D. P., & Patrick, D. L. (1994). Necessary but not sufficient: the effect of screening and feedback on outcomes of primary care patients with untreated anxiety. *Journal of General Internal Medicine*, 9, 606–615.
  37. Mazonson, P. D., Mathias, S. D., Fifer, S. K., Buesching, D. P., Malek, P., & Patrick, D. L. (1996). The mental health patient profile: Does it change primary care physicians' practice patterns? *The Journal of the American Board of Family Practice*, 9, 336–345.
  38. McCusker, J., Verdon, J., Tousignant, P., de Courval, L. P., Dendukuri, N., & Belzile, E. (2001). Rapid emergency department intervention for older people reduces risk of functional decline: Results of a multicenter randomized trial. *Journal of the American Geriatrics Society*, 49(10), 1272–1281.
  39. McCusker, J., Dendukuri, N., Tousignant, P., Verdon, J., Poulin de Courval, J., & Belzile, E. (2003). Rapid two-stage emergency department intervention for seniors: impact on continuity of care. *Academic Emergency Medicine*, 10(3), 233–243.
  40. Moore, J. T., Silimperi, D. R., & Bobula, J. A. (1978). Recognition of depression by family medicine residents: The impact of screening. *The Journal of Family Practice*, 7, 509–513.
  41. Rand, E. H., Badger, L. W., & Coggins, D. R. (1988). Toward a resolution of contradictions. Utility of feedback from the GHQ. *General Hospital Psychiatry*, 10, 189–196.
  42. Zung, W. W. K., Magill, M., Moore, J. T., & George, D. T. (1983). Recognition and treatment of depression in a family medicine practice. *The Journal of Clinical Psychiatry*, 44, 3–6.
  43. Rubenstein, L. V., McCoy, J. M., Cope, D. W., Barrett, P. A., Hirsch, S. H., Messer, K. S., et al. (1995). Improving patient quality of life with feedback to physicians about functional status. *Journal of General Internal Medicine*, 10(11), 607–614.
  44. Saitz, R., Horton, N. J., Sullivan, L. M., Moskowitz, M. A., & Samet, J. H. (2003). Addressing alcohol problems in primary care: A cluster randomized, controlled trial of a systems intervention. The screening and intervention in primary care (SIP) study. *Annals of Internal Medicine*, 138(5), 372–382.
  45. Smith, P. (1998). The role of the general health questionnaire in general practice consultations. *The British Journal of General Practice*, 48(434), 1565–1569.
  46. Velikova, G., Booth, L., Smith, A. B., Brown, P. M., Lynch, P., Brown, J. M., & Selby, P. J. (2004). Measuring quality of life in routine oncology practice improves communication and patient well-being: a randomized controlled trial. *Journal of Clinical Oncology*, 22(4), 714–724.
  47. Wagner, A. K., Ehrenberg, B. L., Tran, T. A., Bungay, K. M., Cynn, D. J., & Rogers, W. H. (1997). Patient-based health status measurement in clinical practice: A study of its impact on epilepsy patients' care. *Quality of Life Research*, 6, 329–341.
  48. Wasson, J., Hays, R., Rubenstein, L., Nelson, E., Leaning, J., Johnson, D., et al. (1992). The short-term effect of patient health status assessment in a health maintenance organization. *Quality of Life Research*, 1(2), 99–106.
  49. Wasson, J. H., Stukel, T. A., Weiss, J. E., Hays, R. D., Jette, A. M., & Nelson, E. C. (1999). A randomized trial of the use of patient self-assessment data to improve community practices. *Effective Clinical Practice*, 2(1), 1–10.
  50. White, P., Atherton, A., Hewett, G., & Howells, K. (1995). Using information from asthma patients: A trial of information feedback in primary care. *British Medical Journal*, 311(7012), 1065–1069.
  51. Whooley, M. A., Stone, B., & Soghikian, K. (2000). Randomized trial of case-finding for depression in elderly primary care patients. *Journal of General Internal Medicine*, 15(5), 293–300.
  52. Williams, J. W. Jr., Mulrow, C. D., Kroenke, K., Dhanda, R., Badgett, R. G., Omori, D., et al. (1999). Case-finding for depression in primary care: A randomized trial. *The American Journal of Medicine*, 106(1), 36–43.
  53. Kramer, M. S., & Feinstein, A. R. (1981). Clinical biostatistics LIV. The biostatistics of concordance. *Clinical Pharmacology and Therapeutics*, 29, 111–123.
  54. Clarke, M., & Oxman, A. D. (Eds.) (2003). *Cochrane Reviewers' Handbook 4.2.0* [updated March 2003]. In: The Cochrane Library, Issue 2. Oxford: Update Software. Updated quarterly.
  55. Puhan, M. A., Soesilo, I., Guyatt, G. H., & Schunemann, H. J. (2006). Combining scores from different patient reported outcome measures in meta-analyses: When is it justified? *Health and Quality of Life Outcomes*, 4, 94.
  56. Goodman, S. N. (1998). Multiple comparisons, explained. *American Journal of Epidemiology*, 147(9), 807–812; discussion 815.
  57. Moye, L. A. (2000). Alpha calculus in clinical trials: Considerations and commentary for the new millennium. *Statistics in Medicine*, 19(6), 767–779.
  58. Rubenstein, L. V., Calkins, D. R., Young, R. T., Cleary, P. D., Fink, A., Koscoff, J., et al. (1989). Improving patient function: A randomized trial of functional disability screening. *Annals of Internal Medicine*, 111(10), 836–842.
  59. Valderas, J. M., Alonso, J., Valderas, J. M., Alonso, J., Prieto, L., Espallargues, M., & Castells, X. (2004). Content-based interpretation aids for health-related quality of life measures in clinical practice. An example for the visual function index (VF-14). *Quality of Life Research*, 13(1), 35–44.
  60. Rebollo, P., Valderas, J. M., & Ortega, F. (2005). Progress in Spain of the described barriers to the use of perceived health status measures in the clinical practice. *Medicina Clínica*, 125(18), 703–705.
  61. Greenhalgh, J., Long, A. F., & Flynn, R. (2005). The use of patient reported outcome measures in routine clinical practice: Lack of impact or lack of theory? *Social Science & Medicine*, 60(4), 833–843.
  62. Shojania, K. G., Sampson, M., Ansari, M. T., Ji, J., Doucette, S., & Moher, D. (2007). How quickly do systematic reviews go out of date? A survival analysis. *Annals of Internal Medicine*, 147(4), 224–233.
  63. Callahan, C. M., Hendrie, H. C., Dittus, R. S., Brater, D. C., Hui, S. L., & Tierney, W. M. (1994). Improving treatment of late life depression in primary care: A randomized clinical trial. *Journal of the American Geriatrics Society*, 42(8), 839–846.
  64. Callahan, C. M., Dittus, R. S., & Tierney, W. M. (1996). Primary care physicians' medical decision making for late-life depression. *Journal of General Internal Medicine*, 11, 218–219.