Quality of Life Research (2019) 28:567–592
https://doi.org/10.1007/s11136-018-2016-z

REVIEW

A systematic review of randomised controlled trials evaluating the use of patient-reported outcome measures (PROMs)

S. Ishaque[1] · J. Karnon[1] · G. Chen[2] · R. Nair[3] · A. B. Salter[1]

Accepted: 27 September 2018 / Published online: 3 October 2018

Abstract

Background Patient-reported outcome measures (PROMs) could play an important role in identifying patients' needs and goals in clinical encounters, improving communication and decision-making with clinicians, while making care more patient-centred. Comprehensive evidence that PROMS are an effective intervention is lacking in single randomised controlled trials (RCTs).

Methods A systematic search was performed using controlled vocabulary related to the terms: *clinical care setting* and *patient-reported outcome*. English language studies were included if they were a RCT with a PROM as an intervention in a patient population. Included studies were analysed and their methodologic quality was appraised using the Cochrane Risk of Bias tool. The protocol was registered with PROSPERO (CRD42016034182).

Results Of 4302 articles initially identified, 115 underwent full-text review resulting in 22 studies reporting on 25 comparisons. The majority of included studies were conducted in USA (11), among cancer patients (11), with adult participants only (20). Statistically significant and robust improvements were reported in the pre-specified outcomes of the process of care (2) and health care (3). Additionally, five, eight and three statistically significant but possibly non-robust findings were reported in the process of care, health and patient satisfaction outcomes, respectively.

Conclusions Overall, studies that compared PROM to standard care either reported a positive effect or were not powered to find pre-specified differences. There is justification for the use of a PROM as part of standard care, but further adequately powered studies on their use in different contexts are necessary for a more comprehensive evidence base.

Keywords Patient-reported outcome measures · PROMs · Health-related quality of life · HRQL · HRQoL · Quality of life · QOL · Patient outcomes · Patient-reported outcomes · Clinical care

Abbreviations

± PROM studies	Studies that compared patient completion of a PROM with standard care in the control group
PROM ± summary studies	Studies in which all patients completed a PROM and compared the presentation of PROM summary scores to clinicians vs. no presentation of summary scores
FDA	Food and drug administration
HRQL	Health-related quality of life
PRO	Patient-reported outcomes
PROM	Patient-reported outcome measure
QOL	Quality of life

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s11136-018-2016-z) contains supplementary material, which is available to authorized users.

✉ S. Ishaque
sana.ishaque@adelaide.edu.au

[1] School of Public Health, University of Adelaide, Adelaide, Australia

[2] Centre for Health Economics, Monash Business School, Monash University, Clayton, Australia

[3] School of Dentistry, University of Adelaide, Adelaide, Australia

RCT	Randomised controlled trial
SR	Systematic review

Background

According to the *ISOQOL Dictionary of Quality of Life and Health Outcomes Measurement*, a patient-reported outcome (PRO) is 'a measurement of any aspect of a patient's health that comes directly from the patient, without interpretation of the patient's response by a physician or anyone else [1]'. Patients' experiential knowledge of the effects of any intervention is essential for the delivery of high-quality clinical care. All patients in clinical care are unique and therefore may experience different benefits or side effects from the same treatment [2], which cannot be captured by the mere assessment of traditional physiological outcomes. It is therefore important to ask patients about their preferences and values to set self-directed health goals and promote compliance with treatment.

The assessment of PRO requires instruments that are valid and reliable. These instruments are often termed PRO measures (PROMs). It is suggested that regular use of PRO instruments to collect patients' health-related outcomes can affect the health and well-being of patients by improving patient–physician interaction, by focusing on the clinical encounter on patient-directed concerns and by promoting shared clinical decision-making [3]. PROMs are commonly used in comparative effectiveness research, comparative safety analysis and economic evaluations to inform resource allocation [4, 5], with contexts including the ongoing monitoring of PROs for patients with chronic diseases or in palliative care [6].

The existing evidence on the effectiveness of regular assessment of PROs comes from a variety of sources including observational studies and individual randomised controlled trials (utilising qualitative, quantitative and mixed methods) and gold standard literature reviews of randomised controlled trials (RCTs) [7]. The effectiveness of PROMs has been explored in a number of literature reviews [6, 8–25]. Despite having different aims, these synopses [6, 8–25] highlight that the evidence is equivocal. There are several potential reasons for this ambiguity in the field, such as the attempt to aggregate heterogeneous tools under the umbrella term of PROM (which inappropriately considers them equivalent) [26], assessment of different RCT outcomes by various different methods and at different times, as well as a lack of standardised procedures for the provision of PROM results to health care providers and methodological issues with primary studies [27].

The aim of this systematic review was to assess the evidence on the effectiveness of the use of PROMs as an intervention intended to support the representation of patient values and preferences in clinical encounters. This review of RCTs is not limited by disease, age of patient population, nor year of publication. In addition, this is the first systematic review to consider the statistical robustness of results reported, and differentiate between the use of PROMs with and without the formal presentation of completed PROMs to treating clinicians.

Methods

A detailed systematic review protocol outlining the search strategy; methods for relevance and full-text screening; data extraction form; quality assessment method; plan for data analysis, synthesis and statistical issues; sensitivity and subgroup analysis; publication bias and any conflicts of interest was developed and registered with PROSPERO, registration number: CRD42016034182 [28].

Search strategy

With the help of a health research librarian, a systematic search strategy was developed to search three major databases (PubMed, EMBASE and PsycINFO) from inception to February 2017.

The search was conducted using controlled vocabulary and keywords related to the terms: *clinical care setting* and *patient-reported outcomes* [28]. The search strategy for the Medline database is included in the Supplementary 1 and was modified to adapt to variations in indexing among the databases. Reference lists of relevant literature reviews [8–13, 21] were also screened to identify additional articles. Citation searches were performed in *Scopus*.

Eligibility criteria

A publication of a study was eligible for inclusion if it reported on a RCT that applied a PROM to patients with or without providing the patients' PROM score (summary/profile/dimension) to health care providers as an intervention. The review was restricted to studies that were published and reported in English. There were no restrictions on types of PROM, the form in which the PROM was used as an intervention, the health condition being studied, the country or setting in which the study was conducted.

Trials were excluded if they applied PROMs only for screening of psychological disorders such as depression and anxiety, were in the palliative care setting, compared one type of PROM to another type of PROM, compared only paper application to computer application of the same or different PROMs applied PROMs assessing specific constructs such as pain.

Relevance and full-text screening

First, a title, abstract and keyword screening of initially identified articles was performed. In order to pilot the inclusion criteria, (see review protocol [28]), two authors (SI and RN) initially screened a random 10% of the search results. Discrepancies were discussed and inclusion criteria were modified accordingly. Full-text articles identified from this search were retrieved and discussed. Once agreement on the inclusion criteria was achieved, the primary author (SI) completed the relevance screening with the remaining studies. Next, the same authors (SI, RN) independently applied the inclusion criteria to full texts of potentially relevant studies ($n = 115$) to identify studies for final inclusion. Any discrepancies were resolved through discussion. While it was initially planned to contact authors of studies where there was doubt concerning eligibility, this was not necessary as all doubts were resolved by discussion. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines were followed to ensure transparent and comprehensive reporting [29].

Data extraction, analysis, synthesis and statistical issues

The primary author extracted data from all studies including characteristics of the study design, the nature of the PROM, method of intervention and study outcomes.

Consistent with previous systematic reviews in this setting [9, 11, 12], study outcomes were classified into three categories: process of health care, health outcomes and satisfaction with health care. Outcomes relating to how the care was delivered (e.g. consultation time, discussion of quality of life (QOL)/health-related quality of life (HRQL), return visit referral to other health practitioners) were classified as 'process of care'; monitoring of changes in a patient's QOL/HRQL or in any symptoms were classified as 'health outcomes' and finally outcomes relating to patients' satisfaction with health care or feasibility were classified as 'satisfaction with health care'. Given the heterogeneous nature of the data (both for PROMs used and patient populations studied), it was not considered meaningful to perform a meta-analysis.

Positive results (i.e. in favour of the PROM intervention) were considered 'robust' when statistically significant differences in a pre-specified outcome were reported for a study which was adequately powered to determine them. Positive non-significant or significant results for an outcome that was not pre-specified and/or for which the study was not powered to determine were considered 'non-robust'.

Quality assessment

Three authors (SI, RN, AS) performed methodological quality assessments of seven included studies independently using Cochrane's Risk of bias tool [30]; any discrepancies were resolved by discussion. Thereafter, the primary author (SI) performed the quality assessment and discussed it with another team member (AS), both carefully considering the reasons for specified rankings.

Results

After removal of duplicates, 4302 articles were identified from database searches of which, 77 were found eligible for full-text screening. An additional 36 articles (of which 4 were included) were identified from previous literature reviews, and two articles were identified from other sources. After full-text screening of 116 articles, 22 RCTs met the inclusion criteria and were consequently included in this systematic review (Fig. 1) [31–52].

Table 1 summarises the characteristics of included studies and Table 2 presents additional summary details of the RCTs, including the PROMs assessed, the intervention process and whether training was provided for health providers and/or patients. Based on the nature of the intervention evaluated, articles are grouped into two panels in Table 2: 18 studies in which all patients completed a PROM and compared the presentation of the PROM summary scores to clinicians vs. no presentation of summary scores (labelled 'PROM ± summary' studies) [31–38, 40, 41, 45–52]; 7 studies that compared patient completion of a PROM with standard care in the control group (i.e. no use of a PROM) (labelled '± PROM' studies) [39, 42–44, 49]. Studies by Velikova [49] and Rosenbloom [44] compared more than two treatment arms and their data are thus represented in both tables and panels by specific comparisons. For example, the Velikova study (2004) RCT [49] compared PROM with presentation of results to health care providers vs. standard care, and PROM without presentation of results to health care providers vs. standard care.

Publication dates ranged from 1989 to 2016. Most of the studies took place in the USA [31, 32, 34, 35, 38, 40, 44–46, 48, 51, 52], followed by the UK [43, 49, 50], Netherlands [36, 37, 39], Australia [33, 41], Ireland [42] and Norway [47]. More than half of the studies (55%) included cancer patients [31–33, 35, 37, 39, 41, 42, 44, 47, 49, 52] and were performed in tertiary hospitals. Four studies [34, 40, 48, 50] were performed in tertiary care hospitals in subspecialties other than cancer, and the remaining studies took place in GP/internal medicine/family physician offices [36, 38, 43, 45, 46, 51]. Six studies reported enrolling only new patients, five studies enrolled only patients who were

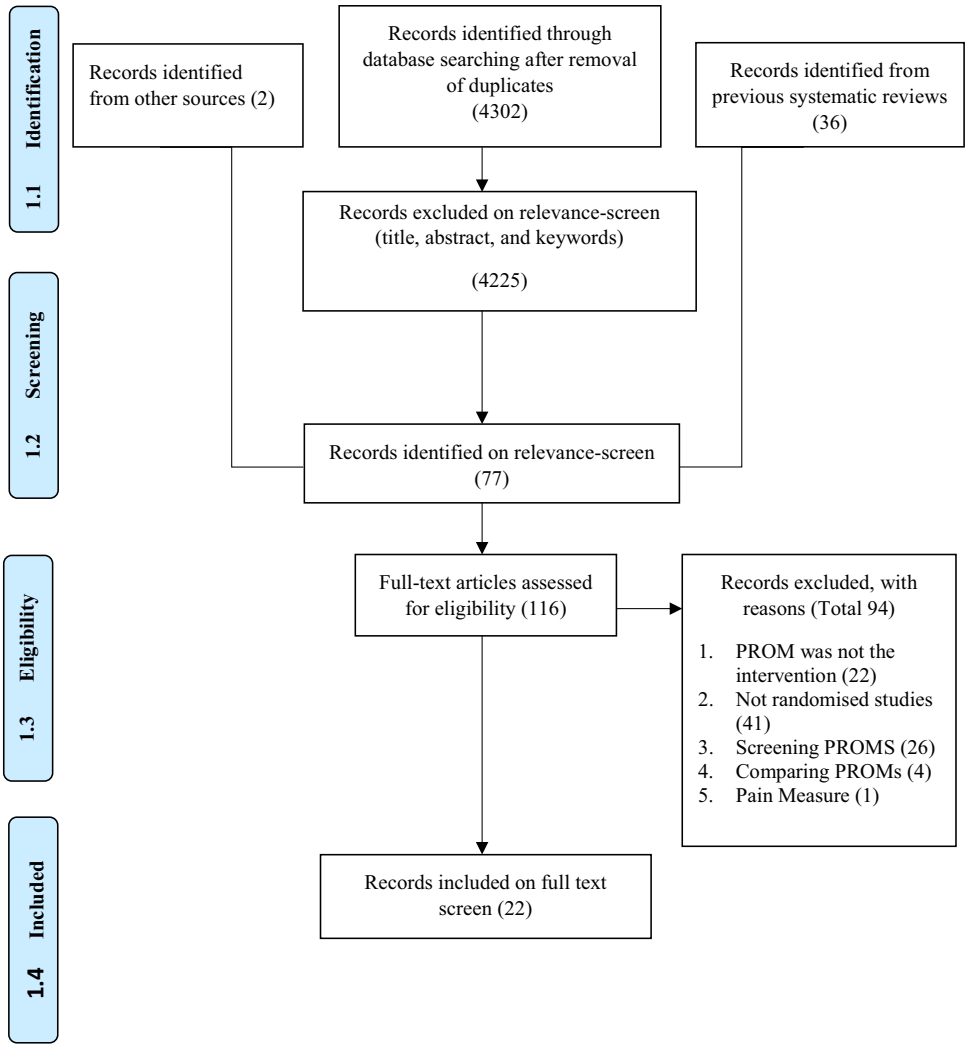

Fig. 1 PRISMA flow diagram

previously known to clinicians and the remaining 11 studies did not specify. Sample size calculations to detect a specified effect size for named primary outcomes were reported in 11 studies [31, 32, 35–37, 40–44, 49]; three studies reported sample size calculations but their named primary outcome had multiple sub-components with no subsequent *P* value adjustments for multiple comparisons [39, 46, 47] and one performed only a post hoc power calculation [40].

A total of 23 PROMs were used in these 22 studies (Table 2). Reference to previous validation work for all PROMs was provided in the RCTs for each PROM in use. However, an evaluation of whether the PROM was a valid choice for the target population of the RCT in which it was in use was not reported. Cancer-specific (9) or generic tools (8) were most commonly applied; four studies used more than one PROM as an intervention [33, 36, 41, 49] [depression-specific and cancer-specific tools (2), generic and cancer-specific tools (1) and generic and diabetes-specific tools (1)] and one applied two arthritis-specific tools [40]. The most commonly utilised tool was the European Organisation for Research and Treatment of Cancer, Quality-of-Life Questionnaire Core 30 (EORTC QLQ-C30) (reported in three studies). The Beth Israel/UCLA Functional Status Assessment Questionnaire, the Hospital Anxiety and Depression Scale (HADS), the SF 36 and PedsQL generic were used in two studies each. The remaining PROM interventions were applied in one study only (Table 2).

In 18 studies [31–33, 36–47, 49–51], PROMs were self-administered by patients on paper (15) or via a computer touch screen (3); in three studies [34, 35, 48], PROMs were completed via telephone with assistance and one study [52] (on paediatric patients) varied the administration method according to the child's age. In this latter trial; children younger than 7 years were allowed a parent-proxy PROM completion, children aged 7–13 years had combined child and parent-proxy PROM completion and children > 13 years self-administered the PROM.

PROMs were mostly completed in waiting rooms/clinics [32, 33, 37, 38, 40, 41, 44, 47, 49–51], followed by home/place of convenience [36, 39, 42, 45, 46], over the telephone

Table 1 Characteristics of studies

Publication ID	Setting (e.g. primary care, emergency room, hospital) Country	Clinicians (e.g. residents, surgeon) Sample size of healthcare providers	Patient Population (sample size, age (years), condition/disease) Sample size calculation reported or not	Patient provider relationship (new, known, not reported)
PROM completed and presented to clinicians vs. PROM completed without presentation to clinicians				
Basch 2016 [31]	Tertiary care cancer hospital USA	Nurse practitioners and treating oncologists Sample size of healthcare providers was not reported	Cancer patients $n=766$ (C 42.42%) C mean age (range) 62 (26–88), I mean age (range) 61 (30–91) Sample size calculation was reported	Not reported
Berry 2011 [32]	Tertiary care cancer hospital USA	Members of clinical team at the hospital $n=262$ (C 53.2%)	Cancer patients $n=660$ (C 49.54%) C mean age (range) 54 (18–86), I mean age (range) 54 (18–89) Sample size calculation was reported	New patients
Boyes 2006 [33]	Medical oncology outpatient clinic at a tertiary care hospital Australia	Oncologists $n=4$	Oncology outpatients $n=80$ (C 47.5%) Sample size calculation was NOT reported	New patients
Calkins 1994 [34]	Tertiary care hospital internal medicine unit USA	Four teams of internal medicine physicians $n=60$ (C 47%)	Patients with functional disability $n=497$ Mean age of all participants at baseline 59 Sample size calculation was NOT reported	Known patients
Cleeland 2011 [35]	Tertiary care cancer clinic at University of Texas Huston USA	Surgical team's advance practice nurse Sample size of healthcare providers was not reported	Patients with primary lung cancer or lung metastasis $n=100$ (C 52%) C mean age (SD) 60.9 (11.8), I mean age 59.2 (13.6) Sample size calculation was reported but there were multiple outcomes and *P* values were not adjusted for the comparisons	Not reported
De Wit et al. 2008 [36]	Four paediatric diabetes outpatient clinics Netherlands	Paediatricians $n=13$ (C 46%)	Patients with diabetes $n=80$ (C 49%) C mean age (SD) 14.9 (1), I mean age (SD) 14.8 (1.1) Sample size calculation was reported	Not reported
Detmar 2002 [37]	Tertiary care oncology inpatients Netherlands	Oncologists $n=10$	Outpatient palliative chemotherapy patients $n=214$ (C 47%) C men age (range) 55(24–81), I mean age (range) 58 (25–84) Sample size calculation was reported	Not reported

Table 1 (continued)

Publication ID	Setting (e.g. primary care, emergency room, hospital) Country	Clinicians (e.g. residents, surgeon) Sample size of healthcare providers	Patient Population (sample size, age (years), condition/disease) Sample size calculation reported or not	Patient provider relationship (new, known, not reported)
Goldsmith 1989 [38]	Family medicine training clinic of Texas Medical Branch USA	General Practitioners $n=27$ (C 48%)	Patient with at least one diagnosed chronic disease $n=62$ C mean age (SD) 66.9 (9.2), I mean age (SD) 70.3 (8.7) Sample size calculation was NOT reported	Not reported
Kazis 1990 [40]	Two independent arthritis centres at tertiary care hospitals USA	Rheumatologists	Arthritis patients Centre 1:356 I, 176 attention-placebo and 175C Centre 2: 614 I, 306 attention-placebo and 290 C Average age (centre 1 = 56, centre 2 = 57) Post hoc power calculation	Not reported
Mclachlan 2001 [41]	Ambulatory clinics at Peter MacCallum Cancer Institute tertiary care hospital Australia	Oncologists and coordination nurse	Cancer patients $n=450$ Median age of all participants (range) 61 (18–92) Sample size calculation was reported	Known patients
Rubenstein 1989 [45]	Community internal medicine practices, Greater Los Angeles Area USA	Internists in community office practices $n=76$ (C 49%)	$n=649$ (C 50%) Sample size calculation was NOT reported	Known patients
Rubenstein 1995 [46]	University Primary Care Clinic Community internal medicine practices, Greater Los Angeles Area USA	Internal medicine house officers $n=73$ (C 45%)	Primary care patients $n=557$ (C 45%) Sample size calculation was NOT reported	New patients
Ruland 2010 [47]	Outpatient and inpatient clinics of a Tertiary care hospital Norway	Physicians and nurses treating leukaemia patients Sample size of healthcare providers was not reported	Patients starting leukaemia or lymphoma treatment $n=145$ (C 48%) Sample size calculation was reported but there were multiple outcomes and *P* values were not adjusted for the comparisons	New patients
Street 1994 [48]	Department of Obstetrics and Gynecology at Scott and White clinic Texas USA	Resident physicians $n=7$	Prenatal patients $n=58$ visits (C 47%) Mean age of all participants (range) 21.9 (17–37) Sample size calculation was NOT reported	New patients
Velikova 2004 [49] (Comparison 1)	Tertiary care oncology hospital UK	Oncologists $n=28$	Cancer patients $n=214$ (C 33%) Mean age of all participants 54.9 SD (12.52) Sample size calculation was reported	New patients

Table 1 (continued)

Publication ID	Setting (e.g. primary care, emergency room, hospital) Country	Clinicians (e.g. residents, surgeon) Sample size of healthcare providers	Patient Population (sample size, age (years), condition/disease) Sample size calculation reported or not	Patient provider relationship (new, known, not reported)
Wagner 1997 [50]	Outpatient epilepsy clinic of a tertiary care hospital UK	Neurologists $n=2$	Epilepsy patients $n=163$ patients (C 23%) 210 clinical encounters (C 21%) C mean age (SD) 45(16), I mean age (SD) 43(13) Sample size calculation was NOT reported	Known patients
Wasson 1992 [51]	Three large urban health maintenance organisations internal medicine clinical sites USA	Internists ($n=41$) and nurse practitioners($n=15$) $n=56$ (C 48%)	Internal medicine patients $n=1522$ Sample size calculation was NOT reported	Known patients
Wolfe 2014 [52]	Three large tertiary care paediatric cancer centres USA	Oncologists and nurses $n=69$	Progressive, recurrent or non-responsive cancer patients were included $n=104$ (control 49%) Sample size calculation was not feasible	Not reported
PROM completed vs. PROM not completed				
Hoekstra 2006 [39]	Two hospitals in the Amsterdam region and general practitioners in the catchment area caring for cancer patients in their palliative phase Netherlands	General practitioners $n=89$	Non-curable cancer with life expectancy of 1–12 months as judged by their physician $n=159$ (C 53%) C mean age 64.6, I mean age 64.1 Sample size calculation was reported but there were multiple outcomes and *P* values were not adjusted for the comparisons	Not reported
Mills 2009 [42]	Three tertiary care cancer hospitals Northern Ireland	Not reported	Inoperable lung cancer patients $n=115$ (C 50%) I age strata (n): <60 (21), 61–70 (18), 70+ (18) C age strata (n): <60 (20), 61–70 (18), 70+ (20) Sample size calculation was reported	Not reported
Qureshi 2001 [43]	Single primary care clinic with mix catchment area in UK	General practitioners Sample size of healthcare providers was not reported	Participants in the general practitioners' list for 2 years $n=100$ (C 50%) C mean age (range, median) 43.14(30.75–53.5, 45.5) I mean age (range, median) 44.96 (36–53, 48) Sample size calculation was reported	Known patients
Rosenbloom 2007 [44] (Comparison 1)	Cancer tertiary care Chicago USA	Nurse practitioners Sample size of healthcare providers was not reported	Advance lung, breast or colorectal cancer $n=144$ (C 49.3%) C mean age 60.4, I mean age 60.2 Sample size calculation was reported	Not reported

Table 1 (continued)

Publication ID	Setting (e.g. primary care, emergency room, hospital) Country	Clinicians (e.g. residents, surgeon) Sample size of healthcare providers	Patient Population (sample size, age (years), condition/disease) Sample size calculation reported or not	Patient provider relationship (new, known, not reported)
Rosenbloom 2007 [44] (Comparison 2)	Cancer tertiary care Chicago USA	Nurse practitioners Sample size of healthcare providers was not reported	Advance lung, breast or colorectal cancer $n = 140$ (C 51%) Mean age of all participants 59 Sample size calculation was reported	Not reported
Velikova 2004 [49] (Comparison 2)	Tertiary care oncology hospital UK	Oncologists $n = 28$	Cancer patients $n = 142$ (C 50%) Mean age of all participants 54.8 SD (12.52) Sample size was not powered to detect this change	New patients
Velikova 2004 [49] (Comparison 3)	Tertiary care oncology hospital UK	Oncologists $n = 28$	Cancer patients $n = 216$ (C 33%) Mean age 54.9 SD (12.52) Sample size calculation was reported	New patients

C control group, *I* intervention group, *SD* standard deviation

[34, 48], first in clinic then subsequently over the telephone [35], a mixture of clinic and telephone interview [52], computer/paper-based completion according to patients' preferences [31] or unspecified [43].

Of the 18 studies that compared presentation of PROM ± summary studies (completion of a PROM with presentation of results to clinicians vs. completion of a PROM alone), [34, 36–38, 45–52] provided training to clinicians/patients about interpretation of PROM scores. One [42] of the five studies that compared completion of a PROM with no PROM provided training to physicians about the layout of the PROM. The content of the training/education sessions provided to physicians and patients concerning the interpretation of the PROM varied substantially between studies (Table 2).

Table 3 summarises outcomes reported in the included studies classified into the three categories: process of care, health and satisfaction with health care. Pre-specified primary outcomes of studies are italicised.

In PROM ± summary studies, a *process of care* outcome was reported in nine studies [31, 32, 37–40, 45–47, 51], of which two were primary outcomes [32, 37]. In both studies that specified a *process of care* primary outcome intervention, patients reported a significant increase in the discussion of HRQL issues with their clinician [32, 37]. Four studies [31, 38, 47, 51] either did not report if their *process of care* outcomes were of primary or secondary interest, or claimed to power their study around a primary outcome that consisted of multiple sub-components. While these studies did report statistically significant results, their findings are considered non-robust due to the lack of adjustment for multiple comparisons.

Health outcomes were reported in 13 PROM ± summary studies [31, 33–37, 40, 41, 45–47, 49, 52], of which four were primary outcomes [31, 36, 41, 49]. A significant improvement in HRQL and psychosocial health was reported in the Basch [31] and De Wit [36] studies, respectively, whereas Velikova et al. [49] reported no significant difference in self-reported HRQL and McLachlan et al. [41] reported that the intervention did not significantly reduce the need for patient information regarding psychological and other health conditions. McLachlan et al. [41] reported a significant decrease in the spiritual needs of intervention patients, but without power for this comparison it was thus considered non-robust (Table 4). Cleeland [35] and Ruland [47] reported on studies whose outcomes had multiple sub-components, of which some showed significant improvement in intervention patients (Table 3). These effects were similarly considered non-robust given no *P* value adjustments for multiple comparisons (of sub-components) were made. Another seven studies [33, 34, 37, 40, 45, 46, 52] reported *health outcomes* without adequate power calculations for these comparisons, of which four reported significant

Table 2 PROM used and the process of intervention

Publication ID	Intervention PROM	Type of PROM (generic, disease specific, individualised)	Process of intervention	Method of administration of PROM	Description of physician training
PROM completed and presented to clinicians vs. PROM completed without presentation to clinicians					
Basch [31]	STAR	Cancer specific	Patients were randomised to control and intervention arms with stratification by 'computer experienced', (CE) vs 'computer inexperienced' (CIE) groups The CIE intervention patients were required to complete the questionnaire at the time of visit, whereas the CE intervention patients were sent weekly reminders encouraging (but not requiring) them to complete them between visit reports. Printed reports of the PROM were provided to clinic nurse and oncologist	Self-administered	No training was provided
Berry [32]	ESRA-C	Cancer specific	Patients were randomised to control and intervention groups after second application of PROM. Clinicians of intervention group patients received a summary of participants' self-reported SQLIs	Self-administered	No training was provided
Boyes [33]	Touch screen surveys: HADS, SCNS	Disease and cancer specific	Each time an intervention group patient completed the survey, a summary graphical representation of anxiety and depression scores, list of debilitating physical symptoms in the past week and supportive care needs plus management strategies was generated. Summary scores were added to patient charts	Self-administered	No training was provided
Calkins [34]	FSQ	Generic	FSQ was applied to all patients. Summary report was only provided to intervention group physicians	Phone	A 2-h seminar with I group physicians regarding PROM report's interpretation and management of functional disability

Table 2 (continued)

Publication ID	Intervention PROM	Type of PROM (generic, disease specific, individualised)	Process of intervention	Method of administration of PROM	Description of physician training
Cleeland [35]	MDASI	Cancer specific	On the occurrence of one or more symptom threshold in intervention group email with PROM response was sent to the surgical team's advanced practice nurse	Phone	No training was provided
De Wit et al. [36]	PedsQL generic, PedsQL diabetes specific	Disease specific	HRQL was monitored with PedsQL generic and PedsQL diabetes specific completed before 3-month appointment AND HRQL scores were discussed with patient during the appointment; Control group completed a lifestyle questionnaire instead	Self-administered	Training to interpret and discuss HRQL scores
Detmar [37]	EORTC QLQ-C30 (Version 3.0)	Cancer specific	Intervention patients' responses to QLQ-C30 were scored and printed as a graphical summary and were given to physicians and patients immediately before consultation; and were added to patient charts	Self-administered	Intervention group physicians had 30 min educational session on interpretation of QLQ-30 summary score Patients in the intervention group received a similar explanation via mailed pamphlet. If desired, a research assistant provided further explanation of the summary No specific guidelines were provided for the use of summary during the medical consultations
Goldsmith [38]	SIP	Generic	SIP scores were provided to the intervention group family physicians immediately before patient consultation	Self-administered	The intervention group ($n = 14$) physicians received a 1-h introduction and written instructions on using SIP in clinical setting
Kazis [40]	AIMS, MHAQ	Disease specific	Intervention consisted of application of PROMS followed by sending health status summary to clinicians on a quarterly basis over a 1-year period	Self-administered	No training was provided

Table 2 (continued)

Publication ID	Intervention PROM	Type of PROM (generic, disease specific, individualised)	Process of intervention	Method of administration of PROM	Description of physician training
Mclachlan [41]	CNQ, EORTC QLQ-C30, BDI short form	CNQ and EORTC QLQ-C30 are cancer specific; BDI short form depression specific	A self-reported questionnaire was completed and summary reports were available immediately for consideration during clinical consult. Study coordination nurse discussed the summary with patient and physicians during consultation	Self-administered	No training was provided
Rubenstein [45]	The Beth Israel/UCLA Functional Status Assessment Questionnaire	Generic	Questionnaire was applied to all patients; a 1-page summary report resembling laboratory test result was provided only to physicians of intervention group patients	Self-administered	I group physicians attended a single 2-h educational and structured discussion program
Rubenstein [46]	Beth Israel-UCLA Functional Status Questionnaire	Generic	Both the experimental and control groups completed the questionnaire The experimental group physicians were given functional status summary reports and management guidelines regarding patient deficit attached to the front of each new patient's medical record	Self-administered	The I group physicians attended an initial and booster (at three months) half-hour educational session
Ruland [47]	ITPAs	Cancer specific	All patients completed baseline questionnaire and a choice interactive tailored patient assessment (ITPA). Summary of intervention group patients added to patients' charts for physicians and nurses	Self-administered	Training provided, but details were not reported
Street [48]	SF-36	Generic	All patients completed SF36 over phone 1–2 days before the consultation Experimental group's information was included to in their medical record	Phone	I physicians were educated about the PROM report structure and told that it would be included in patient charts

Table 2 (continued)

Publication ID	Intervention PROM	Type of PROM (generic, disease specific, individualised)	Process of intervention	Method of administration of PROM	Description of physician training
Velikova [49] (Comparison 1)	EORTC QLQ-C30 and HADS	Cancer specific	Both groups completed PROM questionnaires on a touch screen computer program and the graphic printouts of results were provided to intervention physicians	Self-administered	Physicians were individually trained about interpretations of PROM scores, and to review examples of HRQL and clinical details of real patients. They were asked to review and use the HRQL results, unless totally inappropriate Posters with interpretative information were displayed in clinics No recommendations for specific responses were made
Wagner [50]	SF-36	Generic	Baseline SF36 was completed by all patients (no feedback was provided at that time) At 6 months f/u the health status profiles of intervention group were printed and handed to the neurologist and added to the patient charts	Self-administered	Two 1-h training sessions were conducted by principal investigator and a nephrologist with previous experience of using health status instruments in clinical care Another session to discuss the experiences of physicians was conducted three months after the start of I arm
Wasson [51]	COOP charts	Generic	COOP charts were applied to intervention group patients who were instructed to hand the results to their clinicians	Self-administered	I group clinicians were educated about the use of PROM in a 10-min face-to-face session
Wolfe [52]	PediQUEST system included three tools: PQ-MSAS, PedsQL4.0 and an overall sickness question	PedsQL4.0 is generic, PQ-MSAS is cancer specific	Summary report was provided to families immediately after completion, and to clinicians before the clinic visit; email alerts were sent to oncologist, nurse and psychosocial clinician; the palliative care service and when pain was reported to pain service	Varied as per child's age	Training on interpretation of PROM report was provided to families at enrolment and annually to health care providers No training was provided to health care providers about how to respond to email alerts

Table 2 (continued)

Publication ID	Intervention PROM	Type of PROM (generic, disease specific, individualised)	Process of intervention	Method of administration of PROM	Description of physician training
PROM completed vs. PROM not completed					
Hoekstra [39]	The symptom monitor	Cancer specific	The symptom monitor was completed by intervention group only and patients were encouraged to take the questionnaire to the clinical consultation with their general practitioners	Self-administered	Not applicable
Mills [42]	Diary completed at home/week that included EORTC-C30 plus related lung cancer module LC13	Cancer specific	EORTC-C30 plus related lung cancer module LC13; patients could share their diary information to physicians if they wished	Self-administered	Basic training on layout of diary was provided
Qureshi [43]	Family history questionnaire	Generic	Family history questionnaire was applied to the intervention group patients., They were told that the questionnaire would be reviewed by general practitioner and clinical geneticist	Self-administered	Not applicable
Rosenbloom [44] (Comparison 1)	FACT-G, including cancer-specific 9-item subscales	Cancer specific	Intervention patients completed PROM at baseline and 1, 2, 3 and 6 months and their PROM scores were shared with treating nurses before consultation with patients	Self-administered	No training was provided
Rosenbloom [44] (Comparison 2)	FACT-G, including cancer-specific 9-item subscales	Cancer specific	Intervention patients completed PROM at baseline and 1, 2, 3 and 6 months; PROM scores were shared with treating nurses and a structured interview (by a trained interviewer) about the patients' responses was conducted at baseline, 1 and 2 months	Self-administered	No training was provided
Velikova [49] (Comparison 2)	EORTC QLQ-C30 and HADS	Cancer specific	Intervention group completed PROM questionnaires on a touch screen computer program	Self-administered	Not applicable

Table 2 (continued)

Publication ID	Intervention PROM	Type of PROM (generic, disease specific, individualised)	Process of intervention	Method of administration of PROM	Description of physician training
Velikova [49] (Comparison 3)			Intervention group completed the PROM questionnaire and the graphic printouts of results were provided to their physicians		Physicians were individually trained about interpretations of PROM scores, and to review examples of HRQL and clinical details of real patients. They were asked to review and use the HRQL results, unless totally inappropriate Posters with interpretative information were displayed in clinics

I intervention, *STAR* symptom tracking and reporting—a web-based interface with questions about 12 common symptoms experienced during chemotherapy, *ESRA-C* electronic self-report assessment-cancer, *HADS* 14-item hospital anxiety and depression scale, *SCNS* supportive care needs survey, *FSQ* Functional Status Questionnaire, *MDASI* M. D. Anderson symptom inventory, *SQLIs* symptoms and quality-of-life issues, *HRQL* health-related quality of life, *EORTC QLQ-C30* European organisation for research and treatment of Cancer Quality-of-Life Questionnaire C30, *QLQ-C30* Cancer Quality-of-Life Questionnaire C30, *SIP* sickness impact profile, *AIMS* arthritis impact measurement scales, *MHAQ* Modified Health Assessment Questionnaire, *CNQ* Cancer Needs Questionnaire short form, *BDI short form* beck depression inventory short form, *ITPA* interactive tailored patient assessment, *SF-36* short form-36 Health Status Questionnaire, *COOP* Dartmouth cooperative functional assessment charts, *PediQUEST* paediatric quality of life and evaluation of symptoms technology, *PQ-MSAS* PediQUEST-memorial symptom assessment scale, *FACT-G* functional assessment of cancer therapy-general, *PedsQL4.0* pediatric quality-of-life inventory, *PROM* patient-reported outcome measure, *UCLA* The University of California Los Angeles

Table 3 Outcomes reported

Study identifier	Outcomes assessed		
	Process of health care	Health outcomes	Satisfaction with health care
PROM completed and presented to clinicians vs. PROM completed without presentation to clinicians			
Basch [31]	Intervention patients had significantly fewer emergency visits at 1 year ($P=0.02$) Intervention patients received significantly longer duration of palliative chemotherapy ($P=0.002$) There was moderate evidence of a smaller proportion of intervention patients hospitalised at 1 year (45% vs. 49%; $P=0.08$) No significant difference in the number of nursing calls to patients during participation	*Clinically meaningful HRQL (evaluated by EQ-5D) improved in the I vs. C 21% vs. 11%* ($P=0.0059$) At 1 year, quality-adjusted survival was significantly different between I and C (mean of 8.7 vs. 8.0 months) $P=0.004$	73% of I participants completed the PROM at any given clinic visit
Berry [32]	*Intervention patients had increased likelihood of SQLI, that were reported problematic at first, being discussed* ($P=0.032$) No significant difference in the average length of clinic visit	Not reported	Of enrolled clinicians, 43.1% completed questionnaire on PROM usability Clinicians found PROM useful in identifying appropriate areas of SQLI (67.8%), guiding the interview (64.3%), promoting communication (50%), identifying appropriate areas for referral (53.6%)
Boyes[a] [33]	Not reported	Intervention patients reporting debilitating symptoms at visit 2 were significantly less likely to report debilitating symptom at visit 3 ($P=0.04$), and also reported significantly lower mean depression scores ($P=0.02$) Intervention patients reported non-significant reduction in mean anxiety scores, in number of patients classified as clinically anxious or clinically depressed or with moderate or high psychological needs	Patients found PROM easy to complete, a good way for doctors to get patients' well-being information and were willing to complete the PROM at each visit Oncologists felt that the feedback was useful, helped to provide good patient care and promoted communication Half ($n=2$) of the oncologists felt that the discussion of feedback report with their patients increased the consultation time by 3–5 min
Calkins[a] [34]	Not reported	FSQ subscales Intervention patients had significantly fewer bed days ($P<0.05$) Intervention patients had non-significant improvement in restricted activity days, basic activities of daily living, intermediate activities of daily living, mental health and quality of interaction; and work performance Intervention patients had non-significant decline in social activities	Intervention patients were non-significantly more satisfied with their health status
Cleeland [35]	Not reported	Intervention patients had significant reduction in the number of symptom threshold events (pain, distress, sleep, shortness of breath, constipation) $P=0.003$	Intervention patients were significantly more comfortable with the PROM reporting system ($P=0.03$), and rated the system easy to use ($P=0.01$)

Table 3 (continued)

Study identifier	Outcomes assessed		
	Process of health care	Health outcomes	Satisfaction with health care
De Wit et al. [36]	Not reported	*Significant improvement in HRQL at follow- up compared to baseline in I patients CHQ-CH87* ($P=0.006$), *vs C I vs. C no significant difference on physical health* (*CHQ-CH87*), *family conflict* (*DFCS*) *or depression* (*CES-D*) *I vs. C no significant difference in Hb A1C level* Note: baseline A1C level had confounding effect on psychosocial summary scale. For lower A1C level patients, psychological outcome significantly improved in the intervention group and remained stable for controls. Whereas at highest A1C level > 9.5, there was no difference in the baseline and f/u scores of I &C	At 1 year f/u: I vs. C significantly more satisfied with their care ($P=0.009$)
Detmar [37]	*HRQL issues were discussed more frequently in intervention group. Significant differences in discussion of fatigue* ($P=0.02$), *dyspnea* ($P=0.02$) *and social functioning* ($P=0.05$) Intervention patients received more counselling on managing their health problems ($P<0.05$) No significant differences in medication prescription, ordering of tests, referral to other providers and mean duration of visits	No significant difference in any of the SF-36 scales Larger percentage of I vs. C showed improvement of 0.5 SD* or more in mean health (43% vs. 30% $P=0.04$) and role function (22% vs. 11% $P=0.05$) *A change of 0.5 SD is considered a clinically important difference	Intervention patients had significantly greater satisfaction with the degree of emotional support received ($P<0.05$) No significant difference in physician satisfaction
Goldsmith[a] [38]	No significant difference in return visits to the family physician; referrals to other physicians or allied health professionals or to social or community services	Not reported	Not reported
Kazis[b] [40]	AIMS study group (one of the centres): Intervention patients had non-significant decrease in the number of visits to doctors in the previous 3 months Intervention patients had non-significant increase in the drug category change over 1 year, and more referrals to other arthritis health professionals over 1 year MHAQ group: Intervention patients had non-significant reduction in the number of visit to doctors in previous 3 months, and referrals to other arthritis health professionals over 1 year Intervention patients had a non-significant increase in drug category change over 1 year	Intervention patients in the MHAQ group had non-significant increase in compliance Control group had non-significantly better compliance in AIMS group	No significant differences in patient satisfaction with care Physicians found PROM report helped in patient management (79%), had moderate to substantial value (55%), contributed to doctor–patient relationship (no percentage reported) 2/3rd physicians always examined the report and filed them in on patient chart, 50% discussed it with patients most of the time, 38% never discussed it

Table 3 (continued)

Study identifier	Outcomes assessed		
	Process of health care	Health outcomes	Satisfaction with health care
Mclachlan [41]	Not reported	*Intervention patients had non-significant reduction in their psychologic and health information needs* Intervention patients had significant reduction in spiritual needs at 6 months ($P<0.02$)	No significant difference found in patient satisfaction
Rubenstein[c] [45]	No significant difference in the number of office visits to physicians; hospitalisations; contacts with nurses, physical therapists or other health professionals; new medications started; medical equipment purchased or new diet or exercise regimens	Intervention patients non-significantly improved on three subscales of PROM questionnaire (intermediate activities of daily living, frequency of social contact and sexual satisfaction) Control group non-significantly improved on eight subscales	Physicians found the PROM accurate and useful (97%), felt that it would improve patient health and physician–patient communication (43%), used the PROM results to change patient therapy (43%)
Rubenstein [46]	Intervention patients were more likely to have a specific treatment plan for their symptoms ($P=0.05$), had more medical and functional status problems listed in the visit notes ($P<0.01$), were more likely to be identified as having physical, psychological, social or functional status problems ($P<0.05$) and had more diagnosis of depression ($P<0.05$) and anxiety ($P<0.001$)	Intervention patients had significantly better Mental Health scores ($P<0.03$), and social activities scores for people > 70 years ($P=0.03$) Intervention patients' scores for basic activities of daily living, social activities and work performance improved non-significantly Control group had non-significant improvement in the intermediate activities of daily living	Not reported
Ruland [47]	Significantly increased number of symptoms addressed by physicians and nurses in the inpatient and outpatient records of intervention patients ($P<0.05$)	Intervention patients had significantly reduced symptom distress in pain, eating/drinking, bowel/bladder, energy, sleep/rest, concentration/memory, activities of daily living/self-care and worries/concern ($P<0.01$); bleeding/infection and sexuality ($P<0.05$) Control patients had significant reduction in symptom distress in worries/concern ($P=0.03$)	Not reported
Street[a] [48]	Not reported	Not reported	Non-significant increase in intervention patients' perception of physician inquiry about their health status
Velikova [49] (Comparison 1)	Not reported	No significant difference was found in the emotional well-being, physical well-being, functional well-being and social or family well-being Control patients had non-significant improvement in HRQL 40% Intervention patients had clinically meaningful improvement, whereas 32% of controls	Not reported

Table 3 (continued)

Study identifier	Outcomes assessed		
	Process of health care	Health outcomes	Satisfaction with health care
Wagner [50]	Not reported	Not reported	Physicians reported that availability and discussion of PROM: resulted in change in therapy (12.3%), was at least moderately useful for communication with patients (14.2%) and helped with patient management (8.4%) Physicians indicated the PROM result lengthened 67% of the encounters Intervention patients' perception that their doctor considered how they felt emotionally and considered their usual daily activities when advising them increased non-significantly Control patients showed non-significant increase in patients' satisfaction with clinical encounter, and the concerns shown by their doctor for their feelings
Wasson [51]	Female intervention patients had significant increase in the ordering of tests and procedures ($P<0.001$) Male intervention patients had non-significantly increased ordering of tests and procedures	Not reported	Male intervention patients significantly reported receiving greater help with functional problems related to pain ($P=0.016$) Male Intervention patients were significantly more satisfied with management of pain ($P=0.02$)
Wolfe[a] [52]	Not reported	No significant change in average scores of PQ-MSAS, PedsQL4.0 or sickness scores during 20 weeks f/u Post hoc analysis: survivors beyond 20 weeks I vs. C PedsQL4.0 emotional score (by an average of +6 points) and overall sickness scores (average −5.3 points) significantly improved Post hoc analysis: children ≥8 years I vs. C PedsQL4.0 emotional score (by an average of +8.1 points) and overall sickness scores (average −8.2 points) significantly improved	52% children and 71% parents found reports easy to understand 28% children and 54% parents thought reports helped them quite a bit/very much to talk to their doctors 75% parents agreed that PROM reports helped them understand their child's feelings 21/34 providers completed satisfaction survey. 50% found reports useful when speaking to their patients and did not cause increase in consultation time Reports provided new information (61% on psychosocial issues; and 22% physical symptoms) Reports contributed at least sometimes to their decision to initiate a psychosocial (56%), pain (34%), social work (33%) or palliative care (29%) consult and to discuss goals with families (36%)

Table 3 (continued)

Study identifier	Outcomes assessed		
	Process of health care	Health outcomes	Satisfaction with health care
PROM completed vs. PROM not completed			
Hoekstra [39]	Not reported	Intervention patients had significant reduction in the prevalence of constipation ($P<0.008$) and vomiting ($P<0.01$); and a non-significant reduction in fatigue, pain, lack of appetite, shortness of breath, sleeplessness, nausea and diarrhoea Intervention patients had non-significant increase in the prevalence of cough Intervention patients displayed reduced severity of all reported symptoms, except for constipation and vomiting ($P<0.05$)	Not reported
Mills[d] [42]	No significant associations at 2 and 4 months regarding the discussion of patient problems with health care professionals	*Intervention patients QOL declined non-significantly* (*and non-clinically meaningful*) ($P=0.14$) 47% Intervention vs. 32% controls had clinically meaningful declined QOL; this difference was non-significant 52% Intervention vs. 26% controls had clinically meaningful* decline in lung cancer-specific QOL ($P=0.03$) *A difference of 6 or more is considered clinically meaningful	Control patients were non-significantly more satisfied
Qureshi [43]	Not reported	No difference in anxiety scores at 3 months	The only significant difference in the perception of self-health was found in response to the question 'what do you think is your risk of developing something wrong in the future?' 26% of the intervention and 7% of the control group patients gave a negative response to this (Fisher's exact test, two tailed $P=0.025$) There was no significant difference in having concerns about family history
Rosenbloom [44] (Comparison 1)	No significant difference in clinical treatment changes between the groups	There was no significant difference in HRQL across the groups	No significant difference in general satisfaction, and satisfaction with communication
Rosenbloom [44] (Comparison 2)	No significant difference in clinical treatment changes between the groups	There was no significant difference in HRQL across the groups	No significant difference in general satisfaction, and satisfaction with communication

Table 3 (continued)

Study identifier	Outcomes assessed		
	Process of health care	Health outcomes	Satisfaction with health care
Velikova [49][e] (Comparison 2)	Not reported	Intervention patients had significant improvement in HRQL ($P=0.01$), physical well-being ($P=0.003$) No significant difference was found on emotional well-being, functional well-being and social or family well-being 32% of Intervention patients and 24% of control patients showed clinically meaningful improvement in HRQL	Not reported
Velikova [49] (Comparison 3)	Intervention patients had significantly more HRQL symptoms mentioned during the clinical encounter ($P=0.03$) No significant difference was found in the number of other symptoms discussed, and length of clinical encounter, medical decisions and non-medical decisions	Intervention patients had significant improvement in HRQL ($P=0.006$), emotional well-being ($P=0.008$), physical well-being ($P=0.006$), functional well-being ($P=0.03$) No significant difference was found in social or family well-being 40% of intervention patients and 24% of control patients showed clinically meaningful improvement in HRQL	Physicians explicitly mentioned/referred to HRQL data in 64% of intervention encounters, 43% found HRQL information clinically very useful, 28% found it somewhat useful, 21% found it a little useful and 9% found it not useful Physicians found that HRQL data provided an overall assessment of patients (69%), additional information (33%), identified problems for discussion (27%) and contributed to patient management (11% encounters) Oncologists did not use the HRQL information if the 'data were irrelevant for the purpose of the encounter or irrelevant to patients' major problems'

AIMS arthritis impact measurement scales, *CHQ-CF87* 87-item child report version of the Child Health Questionnaire, *CES-D* 20-item centre for epidemiological studies scale for depression, *C* control group, *DFCS* diabetes-specific family conflict scale, *HRQL* health-related quality of life, *I* intervention group, *MHAQ* Modified Health Assessment Questionnaire, *PEQ-D* patient evaluation of the quality of diabetes care, *PedsQL4.0* pediatric quality-of-life inventory, *STAI* Spielberger state trait anxiety inventory, *PQ-MSAS* PediQUEST-memorial symptom assessment scale, *PCQ* Psychological Consequences Questionnaire, *PROM* patient-reported outcome measure, *QOL* quality of life, *SD* standard deviation, *SQLI* symptoms and quality-of-life issues, *FSQ* Functional Status Questionnaire, *Hb A1C* haemoglobin A1C, *EQ-5D*: EuroQoL 5D, *HRQL* health-related quality of life

Sections are italicised when pre-specified primary outcomes of studies were detailed

[a]No sample size calculation was reported

[b]A post hoc power calculation was performed, no pre-specified primary outcome

[c]No pre-specified primary outcome

[d]Sample size was calculated but was not achieved for the comparison

[e]Study was not powered to detect differences in these groups

Table 4 Summarises the reported results for comparisons in both of the two panels

	PROM results to clinician vs. no PROM results to clinician ($n=18$)			PROM vs. standard care ($n=7$)		
	No. studies reporting no evidence of effect	No. studies reporting non-robust effect(s)	No. studies reporting robust effect(s)	No. studies reporting no evidence of effect	No. studies reporting non-robust effect(s)	No. studies reporting robust effect(s)
Processes of care	3	4	2	3	1	0
Health outcomes	4	7	2	4	1	1
Satisfaction	4	3	0	3	0	0

Reported effects were considered 'robust' if they were statistically significant and pertained to a single reported comparison or there was evidence that the study was adequately powered for more than one comparison. Other positive effects were classified as 'non-robust'

PROM patient-reported outcome measure

non-robust results. In total, seven significant results for health outcomes were considered non-robust.

Satisfaction with health care was reported in seven [34, 36, 37, 40, 41, 50, 51] PROM ± summary studies as one of the outcomes of interest, with none of these studies explicitly specifying it as their primary or secondary outcome. Significantly more intervention patients were reported to be satisfied with their emotional support [37], overall care [36] and management of pain [51] (male patients only); with results categorised as non-robust. Three studies [34, 40, 41] reported no significant difference between the groups, and one [50] reported a greater (but non-significant) satisfaction with care in the control group.

In ± PROM studies (PROM completion with or without presentation of scores to clinicians vs. Standard Care), four studies [42, 44, 49] reported on *process of care* variables as one of their non-primary outcomes of interest. Of those, only Velikova [49] reported significantly (but non-robust) greater discussion of intervention patients' HRQL issues.

Health outcomes were reported in five ± PROM studies [42–44, 49] as primary outcomes. Significant improvement in HRQL of intervention patients was reported by Velikova [49]. Four studies did not report any significant differences in health outcomes [42–44]. *Health outcome* variables were reported as secondary outcomes in Mills [42] and Velikova [49]. Mills [42] reported non-significant (and non-clinically meaningful) poorer lung cancer-specific QOL in the intervention group, whereas Velikova [49] reported non-robust significantly higher physical, emotional, functional well-being and HRQL of participants in the intervention group. The study by Hoekstra [39] reported on change in the prevalence and severity of several symptoms, some of which were reduced significantly, but the lack of adjustment for multiple comparisons rendered them non-robust.

Comparisons on *satisfaction with care* for ± PROM studies were reported in three publications of RCTs [42, 44], none of which were positive or statistically significant, thus there was no significant evidence that intervention patients were more satisfied than their comparator group regarding the health care that they received.

Feasibility data (including physician satisfaction) on acceptance and the perceived usefulness of the PROM intervention tools were collected in nine studies [32–34, 37, 40, 45, 49, 50, 52] with largely positive results (Table 3).

Methodological quality was evaluated using the Cochrane Collaboration's Risk of Bias tool [30], with detailed assessment reported in Supplementary 2. The potential for bias was assessed in the domains of random sequence generation, allocation concealment, detection (blinding of outcome assessment), attrition (incomplete outcome data) and reporting (selective outcome reporting). The risk of introducing systematic error was found to be high in two studies on random sequence generation [46, 48], two studies on allocation concealment [46, 48], none on detection bias, two on attrition bias [40, 46] and six on reporting bias [34, 40, 46–48, 50]. Some studies had missing information, noted by the categorisation of domains as uncertain. Information regarding the likelihood of detection bias (blinding of outcome assessors/data analyst) and allocation concealment was missing in a large number of studies, 13 [32–36, 39–41, 43–45, 48, 50] and 10 [34, 37–40, 43–45, 50, 51], respectively.

We considered the potential for performance bias (blinding of participants and personnel) difficult to avoid due to the nature of the PROM interventions. However, in one study, authors acknowledged that they were able to blind patients and staff to the study hypothesis [44], but not to the interventions, and as such were considered to have low risk of performance bias.

Apart from performance bias, reporting bias was the most common domain, being present in six studies [34, 40, 46–48, 50]. Of those, two were conducted in the internal medicine units: one in each of an arthritis, obstetrics and gynaecology and neurology clinic, and one in a tertiary care cancer hospital.

Discussion

This systematic review of results from RCTs evaluating the use of PROMs in clinical practice categorised the reported comparisons into two groups. While in the first group of 18 studies [31–38, 40, 41, 45–52], the intervention participants completed the PROM and had their PROM results presented to the clinicians providing their clinical care (PROM ± summary), in the remaining seven studies [39, 42–44, 49] participants in the intervention group were simply asked to complete the PROM (± PROM). Reported results were grouped in one of three outcome categories: process of care, health and satisfaction with care (Table 4). Analysis of tabulated results led to the following findings: more positive results were reported for health outcomes, compared to those for the process of care or satisfaction with care; PROM interventions worked better when PROM results were provided to clinicians and the inclusion of PROM training to clinicians prior to a trial commencement appeared to result in no obvious differences in positive results.

Reviewed studies focused predominantly on statistically significant results, without typically mentioning whether they were clinically meaningful. If the results were positive but non-significant, there was no consideration in the publication of whether this may have been the result of a smaller than necessary sample size. Equally, when results were positive and significant, there was no discussion of whether this was possibly due to an inflated Type I error resulting from multiple comparisons. This concurs with results from the methodological quality assessment, indicating that the most common form of bias was that of reporting bias, regardless of the study context. Indeed, when considering characteristics of studies that may have been more prone to bias in any one (or more) domains, no one type of study appeared more prone.

All positive results were reported in this systematic review, regardless of significance, with Table 4 presenting a summary of the results differentiated by robustness, which includes studies reporting no evidence of a difference in treatment groups.

For the 18 studies classified as PROM ± summary, six [31, 32, 36, 37, 41, 49] were powered to detect an effect for their pre-specified primary outcome (two *process of care* [32, 37] and four *health outcome* variables [31, 36, 41, 49]), with the remaining studies either not pre-specifying a primary outcome, not reporting on power calculations for pre-specified outcomes or reporting that power calculations were performed for outcomes with multiple subcomponents without evidence of this. Of the four studies with *health outcomes* as their primary focus, two reported significant results in favour of the intervention group [31, 36]. Non-significant improvement was reported in one study for the intervention group [41] and for the control group in another study [49]. Comparing the characteristics of studies reporting significant vs. non-significant effects, there were no identifiable differences in sample size, disease area, clinician training and mean age of participants or risk of bias.

Among seven studies [39, 49, 4244, 49] in the group classified as ± PROM, none reported a *process of care* primary outcome. Five studies [42–44, 49] reported a *health outcome* of primary interest with only one [49] reporting results in favour of the intervention, three reporting no significant evidence of a difference [43, 44] and one reporting a non-significant poorer effect in the intervention patients [42]. Studies that failed to show any significant difference or reported poorer effects in intervention patients either stated that they did not achieve their desired sample size [42, 44] or did not state this but appeared to have a relatively small sample size [43].

A total of 15 comparisons (in all identified studies) were on HRQL/health status/PROM score outcomes [31, 33, 34, 36, 37, 42–46, 49, 52]. One key observation noted in over half of these studies was the lack of discussion on what constitutes a clinically important difference. A predefined 'clinically important' and statistically significant difference in HRQL was reported in only three comparisons [31, 49]. The Mills study [42] reported non-significant non-clinically important poorer HRQL in intervention patients, but while the Detmar study [37] found no significant difference in health status measured by SF-36; a significantly larger percentage of people in the intervention group had clinically meaningful improved SF-36 scores [53]. The Velikova et al. PROM ± summary comparison did not find any significant and clinically important HRQL differences [49]. The remaining studies (8/15) [33, 34, 36, 43–46, 52], simply referred to *P* values to oppose or support the PROM intervention without reference to whether differences were considered to be clinically meaningful. While *P* values can provide important evidence of a difference in average outcome scores, they indicate only the probability that study findings such as those reported (or more extreme) could have occurred due to chance alone if there really was no difference in the two groups in the underlying population [54, 55]. As such, they lack the ability to inform clinicians of whether (in general) the difference really matters to their patients, i.e. if it was clinically meaningful [54, 55].

There are a large number of validated PROMs available for use (generic and disease specific) and their selection for a particular clinical population can be challenging [56]. The fact that different PROMs are often designed for and used in different populations means that the recommendation of one particular PROM over another in any given scenario is generally not possible. Given that studies in

this review reported some positive and robust effects of PROM interventions, there is likely to be value in the use of PROMs in clinical care.

Thirteen studies provided training sessions to clinicians (and patients/families in some cases) on the interpretation and understanding of the PROM [34, 36–38, 42, 45–52]. Contrary to our expectation that clinicians would be more engaged in the use of the PROMs if training was provided [57–59], there appeared to be no obvious difference in positive outcomes when this took place. Some studies assessed the feasibility of PROM use, but none evaluated changes in clinicians' perceptions before and after the intervention and thus the role of the clinician in the use of PROMs may require further study to be understood.

Compared to the previous literature [6, 8–25], this systematic review provides more contemporary evidence on the use of PROMs in RCTs and additional information about the use of PROMs in children, by including two RCTs with patients under 18 years old. The fact that only two paediatric studies [36, 52] met the inclusion criteria and were included in this review highlights the fact that little has been done to understand the value of PROMs in this context. It is well documented that children often struggle to communicate their health issues with parents and clinicians, and so a rationale for the use of PROMs to provide a voice for children is strong. Wolfe et al. [52] reported improvements in children's and parents' perceptions of talking to doctors, and in parents' understanding of their child's feelings. Clinicians also found PROM reports by children provided useful and new information in many cases [52]. While the study by Wolfe [52] did not report significant effects on primary health outcomes from the PROM intervention, a post hoc analysis of survivors beyond 20 weeks showed significant improvement in the emotional subscale of the PROM and in overall sickness scores. These findings were even stronger in children aged 8 years and older who were more likely to have completed the PROM without a proxy. Hemmingsson [60] recommended the use of self-reporting paediatric PROMs whenever possible, or observable components of HRQL when parent-proxy assessment is unavoidable, e.g. for young children. The De Wit [36] study on diabetic adolescents also reported a positive effect of PROMs on patients' HRQL and patient satisfaction with health care. Also, provided that haemoglobin A1C levels were kept under control, a positive effect of PROMs on psychological outcomes was reported [36].

Methodological quality was evaluated using the Risk of Bias tool [30]. Detection bias occurs when outcome assessors are not blinded to the group allocation and study hypothesis. Although the studies in this review are considered pragmatic trials, blinding of the data analyst could have been achieved relatively easily.

This review has several methodological strengths. To minimise bias and to ensure that the systematic review was conducted in a transparent manner, a review protocol was submitted to PROSPERO. There are several benefits of developing a review protocol a priori as outlined in the PRISMA statement on reporting items for systematic review and meta-analysis protocols [61], and they include assurance that methods are replicable and in line with current recommendations. A comprehensive systematic search of three major databases was performed to reduce the possibility of missing relevant studies. An especially wide time frame, age range of target population and disease type were considered.

We acknowledge that there are some limitations of this review. Firstly, the inclusion criteria were restricted to studies published in the English language only. While this means we may have missed some important studies, it is unlikely that our conclusions would change given that previously documented systematic reviews included articles in multiple languages (in French, German, Italian, Russian or Spanish) and yielded similar results [9, 10]. Secondly, the relevance screening was performed by a single author, and data were extracted by the same author, but the title, abstract and keywords of a random 10% of the search were independently screened and inclusion criteria were modified after discussion with a second author (RN) and a more cautious (inclusive) approach was taken when screening. We acknowledge that single data extraction could result in missed items, and thus could affect the conclusions of a review. However, in an effort to minimise issues with interpretation, double extraction for the quality assessment and risk of bias analysis was performed. Thirdly, unlike previous systematic reviews, we excluded trials that claimed to use PROMs for screening of psychological conditions. Studies on psychological and mental health conditions were excluded from our systematic review. Patient-reported outcomes in these contexts are typically used for the purpose of diagnosing psychological or mental health conditions. Given that our focus was on the use of PROMs to effectively incorporate patient values and preferences in clinical encounters, these studies did not meet with the aim of our systematic review. Therefore, despite a much larger time frame for inclusion, there were fewer studies in this review compared to previous systematic reviews [6, 9–22]. Our justification for this decision was our interest in the use of PROMs as a behavioural intervention and not as a screening tool. Given that the results of screening tools are typically reviewed by clinicians as part of a process to identify a disease at an earlier stage for secondary prevention, and thus identify patients who require follow-up, their use in this context is not optional. Hence, the use of PROMs as part of diagnostic tools for mental health disorders is different from proposing their use to inform patient-centred decision-making related to the choice of approach to care or treatment, and thus does not add to our evidence base.

Conclusions

Overall, positive findings in favour of the PROM intervention were reported on 21 occasions but the reported effects were robust in only five cases, i.e. statistically significant and adequately powered. Despite explicit CONSORT guidelines [27], many trials on PROM interventions failed to pre-specify their primary and secondary outcomes and/or adequately power their comparisons for clinically meaningful differences. Despite this, the combined evidence appears to support the use of PROMs to improve communication and decision-making in clinical practice. It is vital that future trials on PROM interventions follow CONSORT guidelines and continue to contribute robust evidence on the use of PROMs in clinical practice.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Ethical approval This review does not contain any studies with human participants performed by any of the authors.

Informed consent Informed Consent was not applicable to this review as no primary data were collected.

References

1. Mayo, E. N., et al. (2015). ISOQOL dictionary of quality of life and health outcomes measurement. First Edition.
2. Varadhan, R., Segal, J. B., Boyd, C. M., Wu, A. W., & Weiss, C. O. (2013). A framework for the analysis of heterogeneity of treatment effect in patient-centered outcomes research. *Journal of Clinical Epidemiology, 66*, 818–825.
3. Greenhalgh, J., & (Feb (2009). The applications of PROs in clinical practice: What are they, do they work, and why? *Quality of Life Research: An International Journal of Quality of Life Aspects of Treatment, Care & Rehabilitation, 18*(1), 115–123.
4. Ahmed, S., Berzon, R. A., Revicki, D. A., Lenderking, W. R., Moinpour, C. M., Basch, E., et al. (2012). The use of patient-reported outcomes (PRO) within comparative effectiveness research: Implications for clinical practice and health care policy. *Medical Care, 50*(12), 1060–1070.
5. Wu, A. W., Snyder, C., Clancy, C. M., & Steinwachs, D. M. (2010). Adding the patient perspective to comparative effectiveness research. *Health Affairs, 29*(10), 1863–1871.
6. Antunes, B., Harding, R., Irene, J. H., & on behalf EUROIMPACT (2014). Implementing patient-reported outcome measures in palliative care clinical practice: A systematic review of facilitators and barriers. *Palliative Medicine, 28*(2), 158–175.
7. Burns, P. B., Rohrich, R. J., & Chung, K. C. (2011). The levels of evidence and their role in evidence-based medicine. *Plastic and Reconstructive Surgery, 128*(1), 305–310. https://doi.org/10.1097/PRS.0b013e318219c171.
8. 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*(4), 401–416.
9. 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*(2), 175–186.
10. Gilbody, S. M., House, A. O., & Sheldon, T. (2002). Routine administration of health related quality of life (HRQoL) and needs assessment instruments to improve psychological outcome—A systemic review. *Psychological Medicine, 32*(8), 1345–1356.
11. Kotronoulas, G., Kearney, N., Maguire, R., Harrow, A., Di Domenico, D., Croy, S., et al. (2014). What is the value of the routine use of patient-reported outcome measures toward improvement of patient outcomes, processes of care, and health service outcomes in cancer care? A systematic review of controlled trials. *Journal of Clinical Oncology, 32*(14), 1480–1501.
12. Valderas, J. M., Kotzeva, A., Espallargues, M., Guyatt, G., Ferrans, C. E., Halyard, M. Y., et al. (2008). The impact of measuring patient-reported outcomes in clinical practice: A systematic review of the literature. *Quality of Life Research, 17*(2), 179–193.
13. 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.
14. Alsaleh, K. (2013). Routine administration of standardized questionnaires that assess aspects of patients' quality of life in medical oncology clinics: A systematic review. *Journal of the Egyptian National Cancer Institute, 25*(2), 63–70. https://doi.org/10.1016/j.jnci.2013.03.001.
15. Boyce, M. B., & Browne, J. P. (2013). Does providing feedback on patient-reported outcomes to healthcare professionals result in better outcomes for patients? A systematic review. *Quality of Life Research, 22*(9), 2265–2278. https://doi.org/10.1007/s11136-013-0390-0.
16. Chen, J., Ou, L., & Hollis, S. J. (2013). A systematic review of the impact of routine collection of patient reported outcome measures on patients, providers and health organisations in an oncologic setting. *BMC Health Services Research*. https://doi.org/10.1186/1472-6963-13-211.
17. Greenhalgh, J., Dalkin, S., Gooding, K., Gibbons, E., Wright, J., Meads, D., et al. (2017). Functionality and feedback: A realist synthesis of the collation, interpretation and utilisation of patient-reported outcome measures data to improve patient care. *Health Services and Delivery Research, 5*(2), 1–280.
18. Etkind, S. N., Daveson, B. A., Kwok, W., Witt, J., Bausewein, C., Higginson, I. J., et al. (2015). Capture, transfer, and feedback of patient-centered outcomes data in palliative care populations: Does it make a difference? A systematic review. *Journal of Pain and Symptom Management, 49*(3), 611–624. https://doi.org/10.1016/j.jpainsymman.2014.07.010.
19. Gutteling, J. J., Darlington, A. S., Janssen, H. L., Duivenvoorden, H. J., Busschbach, J. J. V., & De Man, R. A. (2008). Effectiveness of health-related quality-of-life measurement in clinical practice: A prospective, randomized controlled trial in patients with chronic liver disease and their physicians. *Quality of Life Research, 17*(2), 195–205. https://doi.org/10.1007/s11136-008-9308-7.
20. Lambert, M. J., Whipple, J. L., Hawkins, E. J., Vermeersch, D. A., Nielsen, S. L., & Smart, D. W. (2003). Is it time for clinicians to routinely track patient outcome? A meta-analysis. *Clinical Psychology: Science and Practice, 10*(3), 288–301. https://doi.org/10.1093/clipsy/bpg025.
21. Luckett, T., Butow, P. N., & King, M. T. (2009). Improving patient outcomes through the routine use of patient-reported data in cancer clinics: Future directions. *Psycho-oncology, 18*(11), 1129–1138.
22. Guyatt, G. H., Van Zanten, S. V., Feeny, D. H., & Patrick, D. L. (1989). Measuring quality of life in clinical trials: A taxonomy and review. *CMAJ: Canadian Medical Association Journal, 140*(12), 1441–1448.

23. Williams, K., Sansoni, J., Morris, D., Grootemaat, P., & Thompson, C. (2016). *Patient-reported outcome measures: Literature review*. Sydney: ACSQHC.
24. Boyce, M. B., Browne, J. P., & Greenhalgh, J. (2014). The experiences of professionals with using information from patient-reported outcome measures to improve the quality of healthcare: A systematic review of qualitative research. *BMJ Quality and Safety, 23*(6), 508–518. https://doi.org/10.1136/bmjqs-2013-002524.
25. Duncan, E. A. S., & Murray, J. (2012). The barriers and facilitators to routine outcome measurement by allied health professionals in practice: A systematic review. *BMC Health Services Research*. https://doi.org/10.1186/1472-6963-12-96.
26. Valderas, J. M., & Alonso, J. (2008). Patient reported outcome measures: A model-based classification system for research and clinical practice. *Quality of Life Research, 17*(9), 1125–1135. https://doi.org/10.1007/s11136-008-9396-4.
27. Schulz, K. F., Altman, D. G., & Moher, D. (2010). CONSORT 2010 Statement: Updated guidelines for reporting parallel group randomised trials. *Journal of Clinical Epidemiology, 63*(8), 834–840. https://doi.org/10.1016/j.jclinepi.2010.02.005.
28. Ishaque, S., Salter, A., Karnon, J., Chan, G., Nair, R. (2016). *A systematic review of randomized clinical trials evaluating the use of patient reported outcome measures (PROMs) to improve patient outcomes*. PROSPERO CRD42016034182. Available from: http://www.crd.york.ac.uk/PROSPERO/display_record.php?ID=CRD42016034182.
29. Moher, D., Shamseer, L., Clarke, M., Ghersi, D., Liberati, A., Petticrew, M., et al. (2015). Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. *Systematic Reviews*. https://doi.org/10.1186/2046-4053-4-1
30. Higgins, J. P. T., Altman, D. G., Gøtzsche, P. C., Jüni, P., Moher, D., Oxman, A. D., et al. (2011). The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. *BMJ, 343*, d5928.
31. Basch, E., Deal, A. M., Kris, M. G., Scher, H. I., Hudis, C. A., Sabbatini, P., et al. (2016). Symptom monitoring with patient-reported outcomes during routine cancer treatment: A randomized controlled trial. *Journal of Clinical Oncology, 34*(6), 557–565.
32. Berry, D. L., Blumenstein, B. A., Halpenny, B., Wolpin, S., Fann, J. R., Austin-Seymour, M., et al. (2011). Enhancing patient-provider communication with the electronic self-report assessment for cancer: A randomized trial. *Journal of Clinical Oncology, 29*(8), 1029–1035.
33. Boyes, A., Newell, S., Girgis, A., McElduff, P., & Sanson-Fisher, R. (2006). Does routine assessment and real-time feedback improve cancer patients' psychosocial well-being? *European Journal of Cancer Care, 15*(2), 163–171.
34. 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.
35. Cleeland, C. S., Wang, X. S., Shi, Q., Mendoza, T. R., Wright, S. L., Berry, M. D., et al. (2011). Automated symptom alerts reduce postoperative symptom severity after cancer surgery: A randomized controlled clinical trial. *Journal of Clinical Oncology, 29*(8), 994–1000.
36. De Wit, M., Delemarre-van De Waal, H. A., Bokma, J. A., Haasnoot, K., Houdijk, M. C., Gemke, R. J., et al. (2008). Monitoring and discussing health-related quality of life in adolescents with type 1 diabetes improve psychosocial well-being: A randomized controlled trial. *Diabetes Care, 31*(8), 1521–1526.
37. 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. *Journal of the American Medical Association, 288*(23), 3027–3034.
38. Goldsmith, G., & Brodwick, M. (1989). Assessing the functional status of older patients with chronic illness. *Family Medicine, 21*(1), 38–41.
39. Hoekstra, J., de Vos, R., van Duijn, N. P., Schade, E., & Bindels, P. J. (2006). Using the symptom monitor in a randomized controlled trial: The effect on symptom prevalence and severity. *Journal of Pain & Symptom Management, 31*(1), 22–30.
40. 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*(11), 1243–1253.
41. McLachlan, S. A., Allenby, A., Matthews, J., Wirth, A., Kissane, D., Bishop, M., et al. (2001). Randomized trial of coordinated psychosocial interventions based on patient self-assessments versus standard care to improve the psychosocial functioning of patients with cancer. *Journal of Clinical Oncology, 19*(21), 4117–4125.
42. Mills, M. E., Murray, L. J., Johnston, B. T., Cardwell, C., & Donnelly, M. (2009). Does a patient-held quality-of-life diary benefit patients with inoperable lung cancer? *Journal of Clinical Oncology, 27*(1), 70–77.
43. Qureshi, N., Standen, P. J., Hapgood, R., & Hayes, J. (2001). A randomized controlled trial to assess the psychological impact of a family history screening questionnaire in general practice. *Family Practice, 18*(1), 78–83.
44. Rosenbloom, S. K., Victorson, D. E., Hahn, E. A., Peterman, A. H., & Cella, D. (2007). Assessment is not enough: A randomized controlled trial of the effects of HRQL assessment on quality of life and satisfaction in oncology clinical practice. *Psycho-oncology, 16*(12), 1069–1079.
45. Rubenstein, L. V., Calkins, D. R., Young, R. T., Cleary, P. D., Fink, A., Kosecoff, J., et al. (1989). Improving patient function: A randomized trial of functional disability screening. *Annals of Internal Medicine, 111*(10), 836–842.
46. 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.
47. Ruland, C. M., Holte, H. H., Røislien, J., Heaven, C., Hamilton, G. A., Kristiansen, J., et al. (2010). Effects of a computer-supported interactive tailored patient assessment tool on patient care, symptom distress, and patients' need for symptom management support: A randomized clinical trial. *Journal of the American Medical Informatics Association, 17*(4), 403–410.
48. Street, R. L., Gold, W. R., & McDowell, T. (1994). Using health status surveys in medical consultations. *Medical Care, 32*(7), 732–744.
49. Velikova, G., Booth, L., Smith, A. B., Brown, P. M., Lynch, P., Brown, J. M., et al. (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.
50. 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*(4), 329–341.
51. 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.
52. Wolfe, J., Orellana, L., Cook, E. F., Ullrich, C., Kang, T., Geyer, J. R., et al. (2014). Improving the care of children with advanced cancer by using an electronic patient-reported feedback intervention: Results from the PediQUEST randomized controlled trial. *Journal of Clinical Oncology, 32*(11), 1119–1126.

53. Measuring Impact—SF-36. (2018). Accessed September 2, 2018, from http://www.measuringimpact.org/s4-sf-36.
54. Skelly, A. C. (2011). Probability, proof, and clinical significance. *Evidence-Based Spine-Care Journal, 2*(4), 9–11.
55. Hays, R. D., & Woolley, J. M. (2000). The concept of clinically meaningful difference in health-related quality-of-life research: How meaningful is it? *PharmacoEconomics, 18*(5), 419–423.
56. El Gaafary, M. (2016). A guide to PROMs methodology and selection criteria. In Y. El Miedany (Ed.), *Patient reported outcome measures in rheumatic diseases* (pp. 21–58). Cham: Springer.
57. Cappelleri, J. C., & Bushmakin, A. G. (2014). Interpretation of patient-reported outcomes. *Statistical Methods in Medical Research, 23*(5), 460–483. https://doi.org/10.1177/0962280213476377.
58. Coon, C. D., & Cappelleri, J. C. (2016). Interpreting change in scores on patient-reported outcome instruments. *Therapeutic Innovation and Regulatory Science, 50*(1), 22–29. https://doi.org/10.1177/2168479015622667.
59. Dreyer, R. P., Jones, P. G., Kutty, S., et al. (2016). Quantifying clinical change: Discrepencies between patient's and providers' perspectives. *Quality of Life Research, 25*(9), 2213–2220.
60. Hemmingsson, H., Ólafsdóttir, L. B., & Egilson, S. T. (2017). Agreements and disagreements between children and their parents in health-related assessments. *Disability and Rehabilitation, 39*(11), 1059–1072.
61. Shamseer, L., Moher, D., Clarke, M., Ghersi, D., Liberati, A., Petticrew, M., et al. (2015). Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015: Elaboration and explanation. *BMJ: British Medical Journal*. https://doi.org/10.1136/bmj.g7647.