Quality of Life Research (2019) 28:593–607
https://doi.org/10.1007/s11136-018-2045-7

Is quality of life a suitable measure of patient decision aid effectiveness? Sub-analysis of a Cochrane systematic review

Claudia Rutherford[1] · Madeleine T. King[1,2] · Phyllis Butow[1] · France Legare[3] · Anne Lyddiatt[4] · Intissar Souli[5,6] · Orlando Rincones[7] · Dawn Stacey[5,6]

Accepted: 2 November 2018 / Published online: 13 November 2018

Abstract

Purpose Patient decision-aids (PtDAs) help patients make informed treatment decisions incorporating their values. Health-related quality of life (HRQOL) is sometimes an outcome of PtDA effectiveness trials, but its suitability for this purpose is unclear. We sought to provide insights into this question by critically appraising how randomized controlled trials (RCTs) evaluating PtDA effectiveness measure and report HRQOL.

Methods We conducted a sub-analysis of RCTs included in the 2017 Cochrane review of PtDAs. Trials assessing HRQOL at baseline and post-PtDA, and comparing PtDA with comparison groups were included. Two reviewers independently extracted data and assessed study quality. Analysis was descriptive.

Results Of 105 RCTs, 11 were eligible for inclusion. Patients randomized to PtDAs did not report better HRQOL than those randomized to usual care. While all 11 RCTs adequately described baseline sample characteristics and reported HRQOL results for study groups, few stated *a priori* HRQOL expectations or hypotheses (36%); made a link between HRQOL and the decision (18%); provided a rationale or justification for HRQOL assessment (18%); provided reason for choice of HRQOL assessment time-points (9%); or adjusted *p*-values for multiple HRQOL domains and time-points (0%).

Discussion PtDAs did not conclusively impact HRQOL. If this holds generally, then HRQOL is an uninformative endpoint for PtDA effectiveness trials. When planning trials of PtDAs, investigators considering HRQOL endpoints should consider whether and why their PtDA is likely to affect HRQOL in their context, and if so, which specific aspect(s) of HRQOL and at which time-point(s), and ensure HRQOL is assessed accordingly.

Keywords Quality of life · Patient decision aids · Shared decision making · Randomized controlled trials · Cochrane, systematic review

Introduction

Patient decision aids (PtDAs) are evidence-based tools that help patients participate in making specific and deliberated choices among healthcare options [1, 2], particularly when

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s11136-018-2045-7) contains supplementary material, which is available to authorized users.

✉ Claudia Rutherford
claudia.rutherford@sydney.edu.au

[1] Quality of Life Office, Psycho-Oncology Co-Operative Research Group, School of Psychology, University of Sydney, Level 6 North, Chris O'Brien Lifehouse C39Z, Sydney, NSW 2006, Australia

[2] Central Clinical School, Sydney Medical School, University of Sydney, Sydney, NSW 2006, Australia

[3] Université Laval Primary Care Research Centre (CERSSPL-UL), Quebec City, Canada

[4] Health Consumer, Ingersoll, ON, Canada

[5] School of Nursing, University of Ottawa, Ottawa, Canada

[6] Ottawa Hospital Research Institute, Ottawa, Canada

[7] Centre for Oncology Education and Research Translation (CONCERT), Ingham Institute for Applied Medical Research & University of New South Wales, Sydney, Australia

there is insufficient or conflicting evidence about outcomes or there is a need to trade off known benefits and harms [3]. To be defined as a PtDA, they need to specify the decision, provide information on options including benefits and harms, and help patients clarify their values for outcomes of options [2]. They do not recommend a particular option, rather they inform and clarify an individual's preferences [4]. Compared to usual care, a Cochrane review of PtDAs for treatment or screening decisions found patients exposed to PtDAs had improved knowledge/understanding of their options, better perceptions of benefits and risks, lower decisional conflict, decreased practitioner controlled decision making; and were more likely to choose an option consistent with their values [3].

It is well known that disease and treatment can have a range of impacts on quality of life. These have been widely studied, providing evidence for health care decision making at all levels, from policy makers to health care providers and patients [5]. While there are many approaches to defining quality of life, a definition of health-related quality of life (HRQOL) widely used in clinical trials and health services research is: "a multidimensional construct encompassing perceptions of both positive and negative aspects of dimensions, such as physical, emotional, social, and cognitive functions, as well as the negative aspects of somatic discomfort and other symptoms produced by a disease or its treatment" [6]. Assessment of HRQOL and other patient-reported outcomes (PROs) in clinical trials has been particularly useful for elucidating the effects of various diseases and their treatments on patients' lives, augmenting clinical indicators of benefits and harms of treatment [7], and evaluating the patients' experience of treatment toxicity and quality of survival [8]. HRQOL has also been included in trials of PtDA effectiveness. However, a Cochrane systematic review of RCTs of PtDAs found that PtDAs do no better than alternative interventions when assessed using either generic HRQOL questionnaires (i.e. applicable to any health condition) or condition-specific HRQOL questionnaires [3].

A recent conceptual model of shared decision making proposed that improving patient-clinician communication about HRQOL may lead to quicker and better identification and treatment of symptoms and side-effects [9]. Thus HRQOL has some face validity as an outcome measure in PtDA trials. However, given that PtDAs are not intended to directly improve HRQOL but rather facilitate informed values-based decision making, it is not clear that HRQOL is an informative endpoint for PtDA effectiveness trials. Notably, when PtDAs are used during the clinical consultation, HRQOL assessment has been found to improve communication but not necessarily patient outcomes [10].

The concept of a proximal–distal continuum in health outcomes may provide a rationale for including HRQOL outcomes in PtDA trials. This illustrates that proximal effects occur directly as a consequence of the health condition and/or treatment for the condition, such as symptoms of the condition itself (e.g. pain, fatigue) and side-effects and toxicities from treatment (e.g. nausea, vomiting) [11]. These may consequently affect a person's ability to function and their overall sense of wellbeing, i.e. cause distal effects. Many decisions require trade-offs between benefits and harms across options; different treatment options may result in different potential benefits and harms (e.g. symptoms and toxicity) that may impact upon HRQOL. But these are future impacts, occurring only after the treatment decision is made and implemented. During decision making, the process of considering one's values and preferences may not immediately affect HRQOL. It is possible but not inevitable that choosing a treatment option that matches one's values and preferences may contribute to improved HRQOL outcomes during and after the chosen treatment.

Given the lack of certainty about whether HRQOL is affected by PtDAs per se, as opposed to the effects of subsequent treatments, we questioned whether HRQOL is a suitable endpoint for PtDA effectiveness. Our aims were therefore to explore in depth the use of HRQOL as an outcome measured in PtDA effectiveness trials to gain insights into whether it is the right effectiveness outcome. Specifically, we aimed to critically appraise how randomized controlled trials (RCTs) evaluating PtDA effectiveness measure HRQOL, what they found, and how they report HRQOL; these aspects were not considered in the 2017 Cochrane review of decision aids [3].

Methods

We conducted a sub-analysis of 105 RCTs included in the 2017 Cochrane review of decision aids [3] that assessed HRQOL outcomes. The 2017 review assessed the effects of PtDAs in people facing treatment or screening decisions. It included RCTs comparing PtDAs to usual care and/or alternative interventions retrieved from MEDLINE, CENTRAL, EMBASE, CINAHL, and PsycINFO from their inception to April 2015. A detailed description of the methods used can be found in the full review available from the Cochrane library [3]. Importantly, while the review included HRQOL as one of many secondary outcomes of interest, it did not scrutinize HRQOL specifically. In this sub-analysis, we scrutinized the PRO components of the design, conduct and reporting aspects of trials that included PRO endpoints and rationales for why HRQOL was assessed. We included RCTs that assessed HRQOL with a validated PRO measure at baseline and at least one other assessment after PtDA use, and trials that compared HRQOL between PtDA and comparison groups (e.g. usual care). We excluded studies that assessed HRQOL at only one time-point after PtDA

use as they did not provide any information about the impact of PtDAs on HRQOL outcomes. Although the EQ-5D was developed as a preference-based measure to support economic analysis, it can also be used as an outcome measure of HRQOL. Studies that used it in the latter way were included in our sub-analyses.

Quality appraisal of HRQOL assessment methods

Existing quality appraisal tools for RCTs do not adequately address the key issues for assessing the quality of PRO assessment methods in clinical research. Therefore, lead authors (CR, MTK and DS), used 18 items from the CONSORT-PRO reporting statement that assess design, conduct and reporting aspects of trials with PRO endpoints (Supplement 1) to provide insights into the adequacy and quality of the PRO findings they produce.

Study quality was assessed by CR and IS independently. Each quality criterion was scored as follows: fully met (2)/partially met (1)/not met (0). Inter-assessor consistency was cross-checked, and when scores were discrepant, a third assessor independently adjudicated (MTK). Total quality scores were calculated as a percentage of the total possible score per study (as not all items applied to all studies).

Data extraction and analysis

Pre-defined standard data items were extracted by one reviewer (OR) and cross-checked for errors and omissions by a second reviewer (CR). Discrepancies were resolved by discussion between reviewers. Data pertaining to: study aims; sample characteristics; design and methods; outcomes and outcome measures; PRO results; and study methodological quality relating to how HRQOL was assessed and reported were extracted for included studies. Meta-analysis was not feasible given our review aims and the heterogeneity of study populations and designs, PRO measures and analytic approaches. Instead, data were synthesized descriptively [14].

Results

Of 105 RCTs of PtDAs, 11 trials (10%) involving 2684 patients making a health-related decision, measured HRQOL as an outcome and were included in this sub-analysis (Fig. 1). RCTs were conducted in England (4 RCTs), USA (3), Finland (1), The Netherlands (1), Australia (1) and Canada (1). The focus of the PtDA interventions ranged across a variety of treatment decisions for chronic (5), acute (1), or temporary (4) health conditions, and cervical screening (1). Nine trials assessed only generic HRQOL/health status, one trial assessed generic HRQOL/health status and condition-specific HRQOL, and one trial assessed only condition-specific HRQOL (i.e. menorrhagia-related HRQOL) (Table 1).

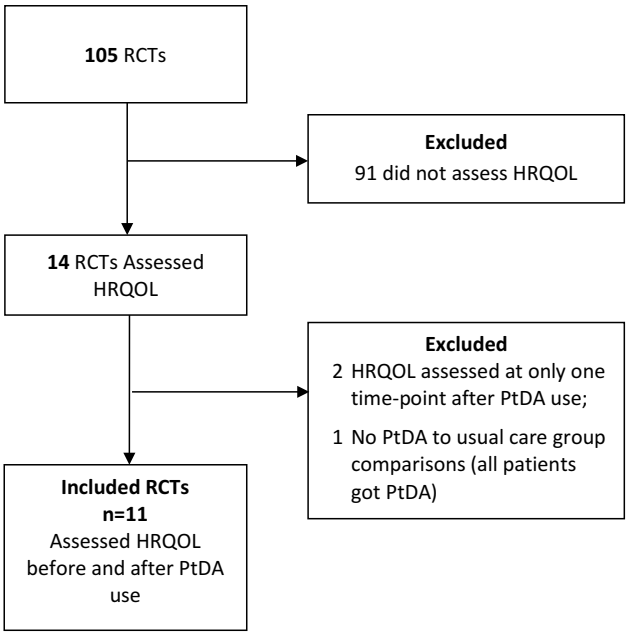

Fig. 1 Flow of studies through the search and review process

HRQOL as a primary endpoint

HRQOL was a primary endpoint in four RCTs. All four used a generic HRQOL measure and all included an *a priori* sample size calculation based on effect size or mean difference [15–18]. All four had sufficient power to detect clinically meaningful HRQOL differences, but only two found significant differences. Significant differences were observed on physical, emotional and social function, energy, and pain domains favouring the PtDA group at 1-year [18] and physical function at 2-years post-PtDA [15]. Differences between groups at earlier time-points, although assessed, were not reported. None of the four trials adjusted analyses for multiple comparisons, so may be prone to false positive findings through accumulation of type 1 errors.

Generic HRQOL

Generic HRQOL was assessed in 10 of the 11 trials with the SF-36 (7 RCTs), SF-12 (2), or EQ-5D (3) measures. The HRQOL outcomes from these three measures were generally reported as higher level outcomes (e.g. physical and mental composite score of SF36, overall HRQOL from EQ-5D) or only selected domains were assessed or reported (3 trials). Two trials reported all questionnaire domains assessed, while, in another two trials, the domains assessed were not

Table 1 Study characteristics of included studies ($n=11$)

First author, year (Country)	Participants	Nature of the decision	Intervention and controls	HRQOL assessments and analysis	HRQOL outcomes (measures)	HRQOL results
Barry 1997 (USA) [25] Chronic condition	162 men with benign prostate hyperplasia in 3 urologic practices	Watchful waiting vs active treatment (prostatectomy or medication)	Intervention ($n=78$): Patient decision aid formatted as video with graphics Controls ($n=84$): information brochure	Baseline, 3, 6 and 12-months Mean change scores from baseline	Secondary endpoint: QOL (SF36)	No significant difference in social function between groups (trend favoring intervention group) at 12-months. Significant difference on general health and physical function favoring intervention group ($p=0.02$ for both) at 12-months Presented HRQOL results for 4 time-points assessed
Bernstein 1998 (USA) [21] Chronic condition	217 coronary angiography (CA) from 2 hospital outpatients (analysis $n=118$)	No treatment decision; pts scheduled for CA	Intervention ($n=65$): video PtDA. Controls ($n=53$): usual consult with surgeon	QOL before CA and 3-months post Mean change scores from baseline	Secondary endpoint: Physical and Mental function (SF-12) Cardiac-specific health (SAQ)	SF12: No significant difference between PtDA and controls at 3-months Low physical function at baseline (mean 35.7) compared to population norms (mean 46.7); mental function similar to population SAQ: Controls lower physical capacity (57 vs 65; $p=0.11$); no significant difference between groups on angina stability or frequency, treatment satisfaction and disease perception over time Significant improvements in anginal stability and disease perception for both groups ($p<0.001$); and anginal frequency ($p=0.002$ control, $p=0.014$ intervention) No significant effect of intervention on HRQOL. Presented HRQOL results for 2 time-points assessed

Table 1 (continued)

First author, year (Country)	Participants	Nature of the decision	Intervention and controls	HRQOL assessments and analysis	HRQOL outcomes (measures)	HRQOL results
Kennedy 2002 (England) [16] Temporary condition	894 with un-complicated menorrhagia (analysis $n=625$) in 6 centres	Medical (drug therapy) vs surgical (endometrial destruction or hysterectomy)	Interview Group ($n=215$): PtDA (booklet and video) plus interview with a trained nurse Information Group ($n=206$): PtDA (booklet and video) only Controls ($n=204$): usual consult with surgeon	QOL at baseline, 6, 12 and 24-months; Mean change from baseline and 2-years; difference between groups at 2-years	Primary endpoint: HRQOL (SF36) Secondary endpoint: Menorrhagia severity (Menorrhagia Health Status)	SF36: Significant difference between interview and control groups on physical function ($p=0.04$) at 2-years. No other differences reached statistical significance; lack of consistent effect across the 3 groups Considerable improvement in HRQOL from baseline to 2-years Menorrhagia HRQOL: no significant difference between groups at 2-years Difference between groups presented for only the 2-year assessment; HRQOL results not presented for all assessment time-points
Knops 2014 (The Netherlands) [20] Acute condition	178 asymptomatic abdominal aortic aneurysm patients in 6 centres	Surgery vs watchful waiting	Intervention ($n=91$): PtDA computer program Controls ($n=87$): usual consult with surgeon	At 1-month FU all pts had decided treatment QOL at baseline before PtDA and then at 1, 4 and 10-months post randomisation Mean change scores from baseline	Secondary endpoint: HRQOL (SF-12)	No significant difference in HRQOL between groups at 1, 4 or 10-months Presented HRQOL results for 4 time-points assessed

Table 1 (continued)

First author, year (Country)	Participants	Nature of the decision	Intervention and controls	HRQOL assessments and analysis	HRQOL outcomes (measures)	HRQOL results
Le Blanc 2015 (USA) [17] Chronic condition	79 women over 50 with osteopenia or osteoporosis in 4 centres (analysis $n=70$)	Yes to start treatment with bisphosphonates vs not to start	Intervention PtDA ($n=32$): PtDA during the clinic visit Intervention FRAX ($n=33$) individualized 10-year risk of bone fracture estimated using the FRAX calculator before visit for use during the clinical encounter Controls ($n=14$): usual care with clinician	Baseline QOL immediately after consultation. FU at 6-months Difference between groups at 6-months; pooled FRAX and usual care groups for analysis	Primary endpoint: QOL (EQ-5D)	No significant difference in HRQOL between PtDA vs FRAX/control groups at 6-months HRQOL results presented for only the 6-mth assessment. Difference between groups presented but not considered change from baseline
McCaffrey 2010 (Australia) [18] Screening/prevention	314 women in routine cervical screening received smear results categorized as non-specific minor changes with or without HPV effect from 18 family planning clinics	No treatment choice assessed	Intervention (DA) ($n=104$): PtDA Intervention HPV triage ($n=104$): asked to arrange HPV testing as soon as possible Controls ($n=106$): repeat smear test in 6-months	Non-blinded RCT; HRQOL soon after receipt of first abnormal test (T0) then at 2-weeks (T1), 3 (T2), 6 (T3) and 12-months (T4) after triage test Mean difference between groups, and mean change from baseline	Primary endpoint: mental health component (SF36) Secondary endpoint: vitality, social function, role emotional and mental health (SF36)	Mental health summary score, vitality, social function, role emotional and mental health worse in HPV group at 2-weeks; weak effect on vitality scale for HPV group compared to PtDA and control groups Over a year, consistently better HRQOL outcomes in PtDA group than both other groups; however not statistically significant Presented HRQOL results for T1 and T4 time-points only

Table 1 (continued)

First author, year (Country)	Participants	Nature of the decision	Intervention and controls	HRQOL assessments and analysis	HRQOL outcomes (measures)	HRQOL results
Morgan 2000 (Canada) [22] Chronic condition	240 patients with ischemic heart disease undergoing elective coronary angiography from 1 centre (analysis $n=187$)	Revascularization (performed at 6-months) vs medical therapy alone (not stated explicitly)	Intervention ($n=120$): interactive video PtDA Controls ($n=120$): usual consultation with physician	PtDA viewed within 4-weeks post-angiography QOL at randomisation (T1), then at time of treatment decision (T2) and at 6-months (T3) Difference between groups at 6-months	Secondary endpoint: QOL (SF36)	No statistical difference in HRQOL between groups at 6-months Patients who withdrew from the study had poorer general health, physical function and pain scores ($p=<0.05$) compared to the study patients HRQOL results presented for only the 6-month assessment; not considered change from baseline over-time
Murray 2001a (England) [23] Chronic condition	112 men with benign prostatic Hypertrophy in 33 general practices	Surgical (prostatectomy or transurethral incision), balloon dilation, medication (alpha blockers or reductase inhibitors) or watchful waiting	Intervention ($n=57$): interactive multimedia PtDA with booklet and printed summary Controls ($n=55$): usual care	QOL before randomization (T1), then 3-(T2) and 9-months (T3) after randomization Mean change scores from baseline over time	Secondary endpoint: Health status and physical function (SF-36 Heath status (EQ5D)	SF36/EQ5D: No significant difference in health status and physical function between groups over time Not presented HRQOL results for all 3 time-points assessed but reported results for trends over time
Murray 2001b (England) [24] Temporary condition	205 peri-menopausal or menopausal women in 26 general practices	Whether to start, stop or continue hormone replacement therapy	Intervention ($n=103$): interactive multimedia PtDA with booklet and printed summary discussing treatment options. Controls ($n=102$): standard care. No details provided	QOL before randomization (T1), then 3-(T2) and 9-months (T3) after randomization Mean change scores from baseline at T3	Secondary endpoint: Health status and physical function (SF-36) Health status (EQ5D)	SF36/EQ5D: No significant difference in QOL between groups from baseline and 9-months. Not presented HRQOL results for all 3 time-points assessed and only reported results for comparisons between T1 and T3

Table 1 (continued)

First author, year (Country)	Participants	Nature of the decision	Intervention and controls	HRQOL assessments and analysis	HRQOL outcomes (measures)	HRQOL results
Protheroe 2007 (England) [26] Temporary condition	146 women aged 30–55 years presenting with menorrhagia in 19 general practices (analysis $n=144$)	Watchful waiting, non-hormonal therapy, hormonal therapy, intrauterine system or surgery	Intervention ($n=74$) information booklet immediately after randomization plus computerized, self-directed PtDA, which provides information about treatment and patient preference clarification Control ($n=72$): information booklet	QOL assessed at baseline (T1), 2-weeks post-intervention (T2) and 6-months (T3); Mean change scores from T2 to T3	Secondary endpoint Menorrhagia-related QOL (Menorrhagia Specific Utility Scale)	At 6-months, the intervention group had significantly higher HRQOL outcomes compared with controls at 2-weeks Not presented HRQOL results for all 3 time-points assessed and only reported results for comparisons between T2 and T3; both time-points after PtDA

Table 1 (continued)

First author, year (Country)	Participants	Nature of the decision	Intervention and controls	HRQOL assessments and analysis	HRQOL outcomes (measures)	HRQOL results
Vuorma 2003 (Finland) [19] Temporary condition and chronic condition	569 women aged 35–54 years with menorrhagia or fibroids in 14 hospitals	Watchful waiting, non-hormonal therapy, hormonal therapy, intrauterine system, removal of copper intrauterine device and progestin or surgery	Intervention ($n=184$): PtDA booklet about heavy menstruation and treatment options 7 days before clinical appointment Controls ($n=179$): standard care. No details provided Pre-trial cohort ($n=206$): comparison group to assess changes in treatment and any carry-over effect of information dissemination; no intervention	RCT plus pre-trial prospective cohort; QOL at baseline (pre-randomisation; T1) then at 3-months (T2) and 12-months after PtDA (T3) Mean change scores from baseline and T3	Primary endpoint: QOL (RAND-36) Secondary endpoint: Menstrual symptoms (Menstrual Health Questionnaire; MHQ)	Within group differences: At T3, physical, emotional and social function, energy, pain, role physical, and role emotional improved from T1 in the intervention group; general health and perceived health did not change significantly At T3, general health, emotional and social function, energy, pain, and perceived health improved from T1 in the control group; physical, role physical and role emotional did not change significantly At T3, all QOL outcomes except physical function improved from T1 in the pre-trial group Between group differences; RAND-36: the intervention group reported an increase in emotional role function compared to control group (12.6 vs 1.9, $p=0.01$) at T3. No other differences on QOL outcomes between groups Not presented HRQOL results for all 3 time-points assessed and only reported results for comparisons between T1 and T3 MHQ: No difference in menstrual bleeding and pain between groups at 12-months

stated, rather only a general statement that no change was found on the questionnaire. Of the 11 trials, six reported no significant difference in HRQOL outcomes over time or when comparing PtDA and usual care groups [16, 19–23].

HRQOL findings: PtDA versus usual care or other intervention

When HRQOL outcomes were compared between PtDA and usual care or other intervention groups, the study of women deciding about cervical screening found a trend for higher scores in the PtDA group on all SF36 domains across repeated assessments up to 1 year compared to controls, but findings were not statistically significant [17] (Table 1). In a study of women with menorrhagia, the PtDA group reported an increase in emotional role function at 1 year compared to controls (usual care; 12.6 vs 1.9, $p = 0.01$). No other differences in HRQOL outcomes were observed between groups [18]. One study of men with prostate hyperplasia found general health and physical function deteriorated over time for controls (information booklet) but remained stable for the PtDA group 1-year post-decision [24].

HRQOL changes over time

Within PtDA group comparisons found significant improvement in physical, emotional and social function, energy, pain, role physical, and role emotional outcomes at 1 year compared to baseline in women with menorrhagia [18] and also in HRQOL outcomes at 2 years compared to baseline in another sample of patients with menorrhagia [15] (Table 1). Another study found a trend for greater improvement in all SF36 domains compared to usual care at multiple time-points across 1 year, however, HRQOL outcomes improved in all study groups over 1 year [17] (Table 1). Where significant changes were observed in the trials, they were on specific domains: general health [24], physical function [15, 24], and emotional role function [18].

Condition-specific HRQOL

Of 11 studies, four assessed condition-specific HRQOL (Table 1). Two studies found differences in some condition-specific HRQOL domains in favour of the PtDA group [20, 25]. One study of coronary angiography patients reported better physical function in the PtDA group relative to controls (usual care) but no differences in other domains of the Seattle Angina Questionnaire; both control and PtDA groups had improved in cardiac-specific HRQOL at 3-month post-decision [20]. A study of women with menorrhagia found that the PtDA group had significantly better menorrhagia-related HRQOL (Menorrhagia Specific Utility Scale) compared with controls (information booklet) at 3-month and 6-month post-decision (the only time-points reported) [25], but by 2 years, there were no differences (6- and 12-month results not reported [15]), nor were there any group differences on menstrual symptoms at 12 months [18].

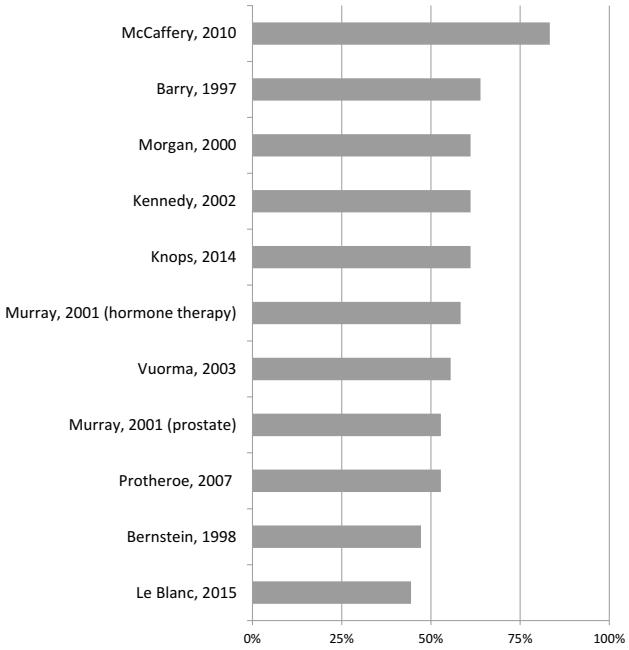

Fig. 2 Quality assessment score (%) for 11 papers, ranked by total score

Critical appraisal of study HRQOL methods

Our critical appraisal of RCTs' PRO assessment methods, conduct and reporting yielded quality scores ranging from 44 to 83%; higher scores indicate better quality (Fig. 2). Figure 3 presents the extent to which each quality criterion was met across all 11 RCTs.

HRQOL methods adequately designed and reported

All trials included a control or comparison group, adequately described baseline characteristics for the HRQOL sample, and reported HRQOL results for all study groups. Most trials adequately described how HRQOL was analysed, and reported results for all HRQOL outcomes (i.e. scales) assessed in the trial (Fig. 3).

HRQOL method and reporting limitations

Only two of the 11 trials stated they expected the PtDA to shift the actual health decision made and therefore expected that HRQOL would be impacted (Fig. 3); one related to deciding whether to start hormone treatment and the other deciding about cervical cancer screening. However, neither

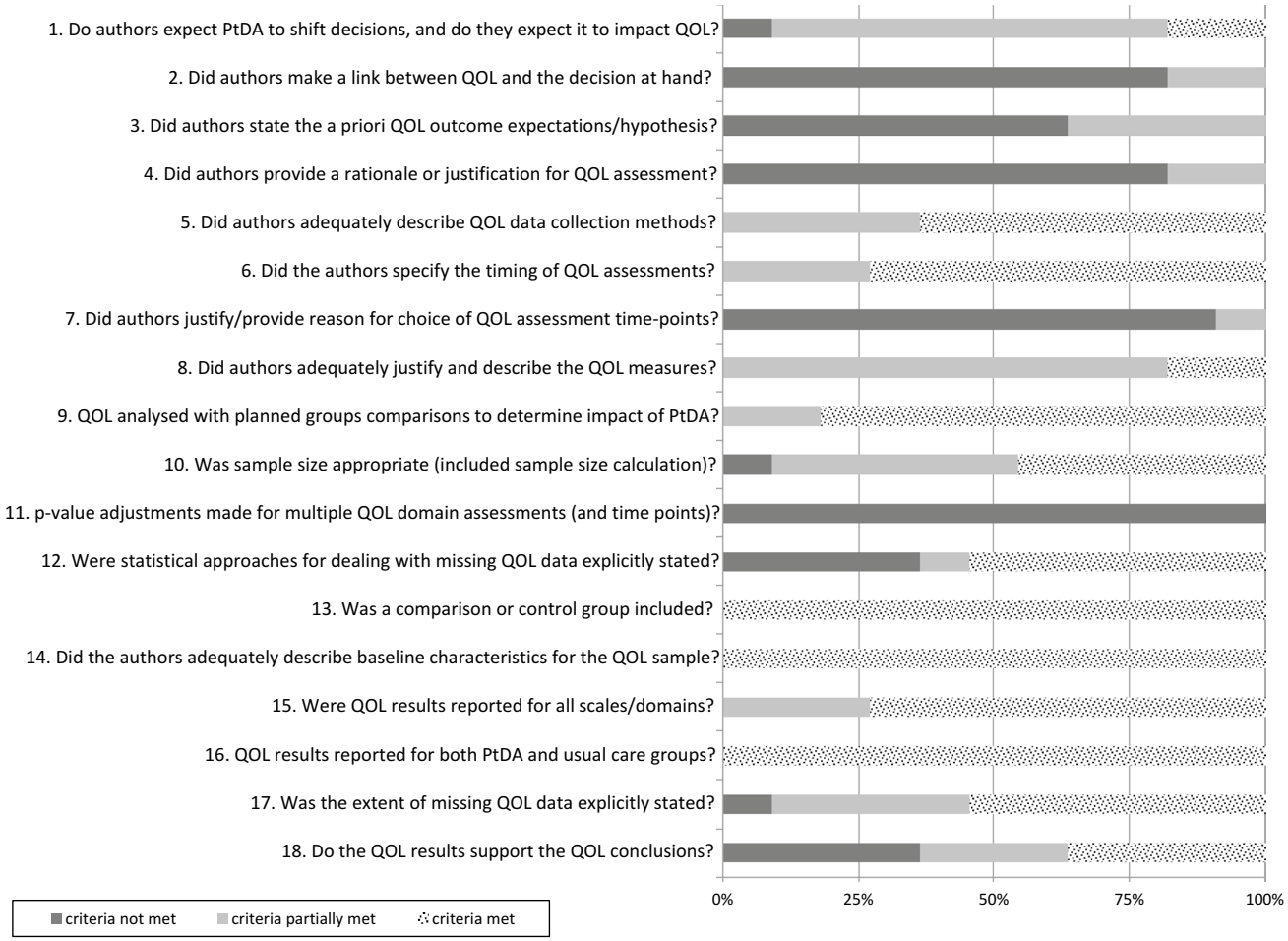

Fig. 3 Percentage of papers meeting each criteria

study found a significant difference in HRQOL between groups over time [17, 23] (Table 1). No studies adequately made a link between HRQOL and the decision at hand, stated a priori HRQOL expectations or hypotheses, provided a rationale or justification for assessing HRQOL, provided a reason for choice of HRQOL assessment time-points, or adjusted *p*-values for assessment of multiple HRQOL domains and time-points (Fig. 3).

Considering HRQOL assessment time-points, all 11 RCTs reported a baseline HRQOL assessment. However, only five assessed baseline HRQOL prior to the PtDA, one assessed it after the PtDA [16], and in five trials it was assumed that baseline was before the PtDA or very close after PtDA use but timing of baseline was not explicitly stated [17–21]. Despite all 11 RCTs having assessed HRQOL at baseline and up to three other time-points after PtDA use, only three RCTs reported HRQOL results for all time-points assessed. Of the other trials, HRQOL results were not reported for all time-points that HRQOL was assessed. Three only reported long-term outcomes: one only mean change scores from baseline and the 9-month assessment [23]; two only mean change scores from baseline and the 12-month assessment [17, 18] (although assessed, these 3 trials did not report HRQOL results for the acute and recovery from treatment periods); and one trial only reported change scores from 2 weeks and 6 months post-PtDA, both time-points after the intervention [25]. Similar selected reporting was observed for between group comparisons. Three trials only reported HRQOL differences between PtDA and comparison groups at one time-point, despite HRQOL having been assessed at multiple time-points (Table 1).

Discussion

HRQOL has been used as an outcome in eleven PtDA effectiveness trials. Four RCTs assessed HRQOL as a primary outcome; two reported improved HRQOL outcomes while two found no difference. In the other seven trials, there were mixed findings but none reported worse HRQOL outcomes in those exposed to the PtDA. Our critical appraisal noted several HRQOL study method and reporting limitations: the

trials failed to provide explicit rationales about why and how HRQOL had been assessed. No studies adequately made a link between HRQOL and the decision at hand, stated a priori HRQOL expectations or hypotheses, or provided a rationale or justification for assessing HRQOL. Further, only 45% of trials assessed baseline HRQOL before PtDA use, and not all studies reported HRQOL change scores for all HRQOL time-points. Reasons for selective reporting are not clear. Comparing change over time to a baseline score after PtDA use obscures any PtDA impact on HRQOL and patterns of acute impact vs long-term recovery in HRQOL outcomes. Further, inadequate reporting of study methods and results for the HRQOL components raises questions about the quality of the HRQOL data collected and subsequently confidence in the results and conclusions they produce. Several resources have been developed to improve the design [26] and reporting [13] of HRQOL and other PRO components of clinical research. While it is possible that limitations in HRQOL study design and reporting obscured evidence of the impact of PtDAs on HRQOL, perhaps it is more likely that there is no difference to be found.

HRQOL has been defined in many ways, but few, if any definitions and conceptualisations provide outcomes that could be directly improved by a PtDA. Many definitions of HRQOL posit it as multidimensional, with core domains of physical, role, social and emotional function and well-being, and common symptoms such as pain and fatigue [6, 27, 28]. Many questionnaires developed to assess HRQOL are based on this definition, and these provide information about physical consequences of disease and treatments (symptom burden and decreased function), effect on a person's emotional state and feelings (psychological functioning), and a person's ability to interact with others and participate socially (work, social interaction and relationships, role functioning). These outcomes are important in judging the effectiveness of a treatment [7, 28–30] but not necessarily the effectiveness of a PtDA. However, if the PtDA increases participation in decision making and facilitates choosing a treatment option that matches one's values and preferences, the patient may have a more positive perception of the impact of the treatment; thus, the PtDA may indirectly contribute to improved HRQOL outcomes. Support for this notion is found in the broader shared decision-making literature that provides substantial evidence of an association between shared decision making and improved decision quality, patient knowledge and patient risk perception [31, 32], all of which may indirectly contribute to improved HRQOL outcomes. However, reviews exploring the association between shared decision making and HRQOL outcomes in cancer and diabetes found weak [31] or no [32] evidence for such an association.

Several factors may explain why many studies failed to find a difference in HRQOL outcomes between PtDA and usual care groups, while some reported significant differences in at least one HRQOL domain in favour of PtDAs. Perhaps the most informative are the four RCTs that included HRQOL as a primary endpoint: all four had sufficient power to detect HRQOL changes, but only two found significant differences, making the evidence equivocal. Where significant differences were observed they were on general health, physical function, and emotional role function, favouring the PtDA group. However, some of these may have been false positive findings due to large number of domains analysed without adjusting p value for multiple hypothesis tests. Within PtDA group comparisons found significant improvement in certain HRQOL domains (e.g. physical, emotional and social function) at 1 year compared to baseline and also in global HRQOL at 2 years compared to baseline. These findings are based on HRQOL defined as multi-dimensional, and consistent with HRQOL research that concludes that many distal effects (such as domains of HRQOL) generally return to pre-treatment levels by 1 or 2 years post treatment, assuming no persistent or enduring impairment due to the disease or treatment.

Inconsistent results could in part be due to major limitations in how HRQOL was assessed and reported, and could also be due to the heterogeneity in the nature of health decisions, populations, and consequent impairments due to the treatments patients had; these factors were not considered by any study when interpreting their HRQOL results. However, an important consideration is whether PtDAs are expected to directly improve HRQOL outcomes in patients. One consideration might be whether patients get the outcomes they value. Many treatment options have associated benefits and harms that patients' trade off when they choose a treatment. Exposure to a PtDA may not directly affect the core domains of HRQOL, but rather, having the opportunity to consider one's values and preferences when trading off treatment benefits and harms may contribute to improved HRQOL outcomes. Choosing a treatment that matches ones values and preferences may result in better psychological and functioning outcomes compared to individuals who have not considered harms and benefits of treatment in light of their values and preferences but this needs empirical testing. Similarly, being involved in the decision making process may improve overall HRQOL [31], and there may be a correlation between getting the outcomes one anticipates or wants with improved HRQOL, but these hypotheses have not been tested. We cannot say conclusively whether use of a PtDA improves HRQOL outcomes, but the evidence suggests that they do not lead to worse HRQOL outcomes. The relationship between PtDAs and HRQOL is confounded by the decision at hand and whether any impairment due to the disease or treatment that one chooses is experienced rather than the PtDA itself. None of the trials included methods to assess the impact of any treatments received, rather, the focus was on PtDA effectiveness. One circumstance in which

a PtDA could change HRQOL is where the PtDA influenced more people to have less invasive or no treatment compared to no-PtDA groups. Ductal carcinoma in situ and prostate cancer are two examples where watchful waiting may be a reasonable management option, and would avoid negative impacts of unnecessary treatments on HRQOL [33].

A strength of our sub-analysis was the inclusion of RCTs included in the Cochrane systematic review of PtDAs [3], which followed a rigorous methodology and is therefore exhaustive of the PtDA research. However, given the aim of the included RCTs was to determine the effectiveness of PtDAs and not specifically to explore the relationship between PtDAs and HRQOL, future systematic reviews could broaden the inclusion criteria to include other research designs that explored the relationship between PtDAs and HRQOL through analyses such as multivariate analysis. Broadening the inclusion criteria to include other study designs in future reviews may provide some additional useful information about the relationship between HRQOL and PtDAs.

Implications for research

In future PtDA effectiveness trials, measures that reflect the decision making process and outcomes likely to be improved by PtDA, for example, decisional comfort, decisional regret, involvement in decision making, and satisfaction with the decision making process, should be used. A global measure of HRQOL may not be the right measure when patients have to trade off benefits and harms between options as it may not be sensitive to this. The more distal the measure, the smaller the effects of disease and treatment will be [11]. Therefore, when the objective is to assess the effects of disease and treatment, proximal outcomes such as symptom severity may be more appropriate as they will be more sensitive to differences and responsive to change than those measures that are more distal [34].

Given that HRQOL seems an inappropriate direct measure of PtDA effectiveness, future studies that want to include this outcome should interpret HRQOL results in relation to proximal outcomes (i.e. effects of disease and treatment(s) received), decision making quality and process outcomes, and the context and nature of the decision (e.g. prevention vs acute vs chronic treatment decisions) [9]. Variables that may moderate the effect of decision making on HRQOL, such as age, income, disease severity, and level of involvement in decision making should also be controlled [31]. Finally, the risks of using HRQOL as a PtDA effectiveness trial endpoint, for example additional hypothesis tests that are likely to find no effect of the PtDA and distract from the more likely effects of the PtDA, should also be considered.

In future PtDAs effectiveness trials, researchers should clearly state whether they expect the PtDA to shift the actual decision made, and if so, whether that would be expected to impact HRQOL, and if so, why and how. And if HRQOL is an outcome used to assess PtDA effectiveness, then a link between HRQOL and the decision at hand should be made. Future research needs to investigate whether choosing a treatment option that is in line with one's values and preferences leads to better HRQOL outcomes in patients who use PtDAs.

Conclusion

Our critical appraisal of PtDA effectiveness trials suggests that PtDAs do not directly affect HRQOL so may not be a suitable endpoint for PtDA effectiveness trials. However, the HRQOL findings may be obscured by limitations in study planning and reporting. Therefore, when planning trials of PtDAs, investigators considering HRQOL as an endpoint should carefully consider where HRQOL sits in conceptual models of shared decision making, and whether and why their PtDA is likely to affect HRQOL in their clinical context. If investigators feel that HRQOL, or some other specific PROs, are likely to be impacted by their PtDA, then it is essential to choose a measure that assesses those specific aspect(s) of HRQOL expected to be affected by the PtDA, and to assess HRQOL at time-point(s) when maximal effects are expected. Importantly, a rational for all of these should be provided.

Acknowledgements We thank the authors of the Cochrane review of decision aids for people facing health screening and treatment decisions.

Funding None.

Compliance with ethical standards

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

References

1. Elwyn, G., O'Connor, A., Stacey, D., Volk, R., Edwards, A., Coulter, A., et al. (2006). Developing a quality criteria framework for patient decision aids: Online international Delphi consensus process. *BMJ, 333*(7565), 417. https://doi.org/10.1136/bmj.38926.629329.AE.
2. Volk, R., & Llewellyn-Thomas, H. (2012). The 2012 IPDAS Background Document: An Introduction. In R. Volk & H. Llewellyn-Thomas (Eds.), *2012 Update of the International patient Decision Aids Standards (IPDAS) Collaboration's Background Document*.
3. Stacey, D., Légaré, F., Lewis, K., Barry, M. J., Bennett, C. L., Eden, K. B., et al. (2017). Decision aids for people facing health treatment or screening decisions. *Cochrane Database of Systematic Reviews*, https://doi.org/10.1002/14651858.CD001431.pub5.

4. Abhyankar, P., Volk, R., Blumenthal-Barby, J., Bravo, P., Buchholz, A., Ozanne, E., et al. (2013). Balancing the presentation of information and options in patient decision aids: An updated review. *BMC Medical Informatics and Decision Making, 13*(Suppl 2), S6.
5. 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 Service Delivery Research, 5*(2), 1–280. https://doi.org/10.3310/hsdr05020.
6. Osoba, D. (1994). Lessons learned from measuring health-related quality of life in oncology. *Journal of Clinical Oncology, 12*(3), 608–616. https://doi.org/10.1200/jco.1994.12.3.608.
7. Au, H.-J., Ringash, J., Brundage, M., Palmer, M., Richardson, H., & Meyer, R. M. (2010). Added value of health-related quality of life measurement in cancer clinical trials: the experience of the NCIC CTG. *Expert Review of Pharmacoeconomics & Outcomes Research, 10*(2), 119–128. https://doi.org/10.1586/erp.10.15.
8. Food and Drug Administration. (2009). *Patient reported outcome measures: Use in medical product development to support labelling claims*. MD: US Department of Health & Human Support Food & Drug Administration.
9. Shay, L. A., & Lafata, J. E. (2015). Where is the evidence? A systematic review of shared decision making and patient outcomes. *Medical Decision Making, 35*(1), 114–131. https://doi.org/10.1177/0272989x14551638.
10. Velikova, G., Booth, L., Smith, A. B., Brown, P., Lynch, P., & Brown, J. (2004). Measuring quality of life in routine oncology practice improves communication and patient well-being: A randomised controlled trial. *Journal of Clinical Oncology, 22*(4), 714–724.
11. Brenner, M. H., Curbow, B., & Legro, M. W. (1995). The proximal-distal continuum of multiple health outcome measures: The case of cataract surgery. *Medical Care, 33*(4 Suppl), As236–A244.
12. King, M. T., Winters, Z. E., Olivotto, I. A., Spillane, A. J., Chua, B. H., Saunders, C., et al. (2017). Patient-reported outcomes in ductal carcinoma in situ: A systematic review. *European Journal of Cancer, 71*, 95–108. https://doi.org/10.1016/j.ejca.2016.09.035.
13. Calvert, M., Brundage, M., Jacobsen, P. B., Schunemann, H. J., & Efficace, F. (2013). The CONSORT Patient-Reported Outcome (PRO) extension: implications for clinical trials and practice. *Health Qual Life Outcomes, 11*, 184. https://doi.org/10.1186/1477-7525-11-184.
14. NHS Centre for Reviews and Dissemination. (2001). *Undertaking systematic reviews of research on effectiveness: CRD's guidance for those carrying out or commissioning reviews* (2nd edn.). Vol. CRD Report 4). York: University of York.
15. Kennedy, A. D., Sculpher, M. J., Coulter, A., Dwyer, N., Rees, M., Abrams, K. R., et al. (2002). Effects of decision aids for menorrhagia on treatment choices, health outcomes, and costs: a randomized controlled trial. *JAMA, 288*(21), 2701–2708.
16. LeBlanc, A., Wang, A. T., Wyatt, K., Branda, M. E., Shah, N. D., Van Houten, H., et al. (2015). Encounter decision aid vs. clinical decision support or usual care to support patient-centered treatment decisions in osteoporosis: The osteoporosis choice randomized trial II. *PLOS ONE, 10*(5), e0128063. https://doi.org/10.1371/journal.pone.0128063.
17. McCaffery, K. J., Irwig, L., Turner, R., Chan, S. F., Macaskill, P., Lewicka, M., et al. (2010). Psychosocial outcomes of three triage methods for the management of borderline abnormal cervical smears: An open randomised trial. *BMJ, 340*, b4491. https://doi.org/10.1136/bmj.b4491.
18. Vuorma, S., Rissanen, P., Aalto, A. M., Hurskainen, R., Kujansuu, E., & Teperi, J. (2003). Impact of patient information booklet on treatment decision—a randomized trial among women with heavy menstruation. *Health Expectations, 6*(4), 290–297. https://doi.org/10.1046/j.1369-7625.2003.00225.x.
19. Knops, A. M., Goossens, A., Ubbink, D. T., Balm, R., Koelemay, M. J., Vahl, A. C., et al. (2014). A decision aid regarding treatment options for patients with an asymptomatic abdominal aortic aneurysm: A randomised clinical trial. *European Journal of Vascular and Endovascular Surgery, 48*(3), 276–283. https://doi.org/10.1016/j.ejvs.2014.04.016.
20. Bernstein, S. J., Skarupski, K. A., Grayson, C. E., Starling, M. R., Bates, E. R., & Eagle, K. A. (1998). A randomized controlled trial of information-giving to patients referred for coronary angiography: Effects on outcomes of care. *Health Expectations, 1*(1), 50–61.
21. Morgan, M. W., Deber, R. B., Llewellyn-Thomas, H. A., Gladstone, P., Cusimano, R. J., O'Rourke, K., et al. (2000). Randomized, controlled trial of an interactive videodisc decision aid for patients with ischemic heart disease. *Journal of General Internal Medicine, 15*(10), 685–693.
22. Murray, E., Davis, H., Tai, S. S., Coulter, A., Gray, A., & Haines, A. (2001). Randomised controlled trial of an interactive multimedia decision aid on benign prostatic hypertrophy in primary care. *BMJ, 323*(7311), 493–496.
23. Murray, E., Davis, H., Tai, S. S., Coulter, A., Gray, A., & Haines, A. (2001). Randomised controlled trial of an interactive multimedia decision aid on hormone replacement therapy in primary care. *BMJ, 323*(7311), 490–493.
24. Barry, M., Cherkin, D., Chang, Y., Fowler, F., & Skates, S. (1997). A randomized trial of a multimedia shared decision-making program for men facing a treatment decision for benign prostatic hyperplasia. *Disease Management and Clinical Outcomes, 1*(1), 5–14.
25. Protheroe, J., Bower, P., Chew-Graham, C., Peters, T. J., & Fahey, T. (2007). Effectiveness of a computerized decision aid in primary care on decision making and quality of life in menorrhagia: Results of the MENTIP randomized controlled trial. *Medical Decision Making, 27*(5), 575–584. https://doi.org/10.1177/0272989x07306785.
26. Calvert, M., Kyte, D., Mercieca-Bebber, R., Slade, A., Chan, A. W., King, M. T., et al. (2018). Guidelines for inclusion of patient-reported outcomes in clinical trial protocols: The SPIRIT-PRO extension. *JAMA, 319*(5), 483–494. https://doi.org/10.1001/jama.2017.21903.
27. Cella, D. F. (1995). Measuring quality of life in palliative care. *Seminar in Oncology, 22*(2 Suppl 3), 73–81.
28. Revicki, D. A., Osoba, D., Fairclough, D., Barofsky, I., Berzon, R., Leidy, N. K., et al. (2000). Recommendations on health-related quality of life research to support labeling and promotional claims in the United States. *Quality of Life Research, 9*(8), 887–900.
29. Gilbert, A., Sebag-Montefiore, D., Davidson, S., & Velikova, G. (2015). Use of patient-reported outcomes to measure symptoms and health related quality of life in the clinic. *Gynecologic Oncology, 136*(3), 429–439.
30. 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. https://doi.org/10.1007/s11136-007-9295-0.
31. Kashaf, M. S., & McGill, E. (2015). Does shared decision making in cancer treatment improve quality of life? A systematic literature review. *Medical Decision Making, 35*(8), 1037–1048. https://doi.org/10.1177/0272989x15598529.

32. Saheb Kashaf, M., McGill, E. T., & Berger, Z. D. (2017). Shared decision-making and outcomes in type 2 diabetes: A systematic review and meta-analysis. *Patient Education Counseling, 100*(12), 2159–2171. https://doi.org/10.1016/j.pec.2017.06.030.
33. McCaffery, K. J., Jansen, J., Scherer, L. D., Thornton, H., Hersch, J., Carter, S. M., et al. (2016). Walking the tightrope: Communicating overdiagnosis in modern healthcare. BMJ, https://doi.org/10.1136/bmj.i348.
34. Luckett, T., & King, M. T. (2010). Choosing patient-reported outcome measures for cancer clinical research–practical principles and an algorithm to assist non-specialist researchers. *European Journal of Cancer, 46*(18), 3149–3157. https://doi.org/10.1016/j.ejca.2010.08.002.