Patient Related Outcome Measures

Dovepress
open access to scientific and medical research

Open Access Full Text Article

REVIEW

Methodological Quality of PROMs in Psychosocial Consequences of Colorectal Cancer Screening: A Systematic Review

Emma Grundtvig Gram[1,2], Jessica á Rogvi[1], Anders Heiberg Agerbeck[1], Frederik Martiny[1], Anne Katrine Lykke Bie[1], John Brandt Brodersen[1–3]

[1]The Center of General Practice, Department of Public Health, University of Copenhagen, Copenhagen, Denmark; [2]Research Unit for General Practice in Region Zealand, Region Zealand, Denmark; [3]The Research Unit for General Practice, Department of Social Medicine, University of Tromsø, Tromsø, Norway

Correspondence: Emma Grundtvig Gram, Email emma.gram@sund.ku.dk

Objective: This systematic review aimed to assess the adequacy of measurement properties in Patient-Reported Outcome Measures (PROMs) used to quantify psychosocial consequences of colorectal cancer screening among adults at average risk.

Methods: We searched four databases for eligible studies: MEDLINE, CINAHL, PsycINFO, and Embase. Our approach was inclusive and encompassed all empirical studies that quantified aspects of psychosocial consequences of colorectal cancer screening. We assessed the adequacy of PROM development and measurement properties for content validity using The COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) risk of bias checklist.

Results: We included 33 studies that all together used 30 different outcome measures. Two PROMs (6.7%) were developed in a colorectal cancer screening context. COSMIN rating for PROM development was inadequate for 29 out of 30 PROMs (97%). PROMs lacked proper cognitive interviews and pilot studies and therefore had no proven content validity. According to the COSMIN checklist, 27 out of 30 PROMs (90%) had inadequate measurement properties for content validity.

Discussion: The majority of included PROMs had inadequate development and measurement properties. These findings shed light on the trustworthiness of the included studies' findings and call for reevaluation of existing evidence on the psychosocial consequences of colorectal cancer screening. To provide trustworthy evidence about the psychosocial consequences of colorectal cancer screening, editors could require that studies provide evidence of the methodological quality of the PROM. Alternatively, authors should transparently disclose their studies' methodological limitations in measuring psychosocial consequences of screening validly.

Keywords: patient-reported outcome measures, COSMIN, methodology, screening, colorectal cancer, psychometric

Plain Language Summary

Previous research has found that cancer screening is associated with psychosocial consequences, such as anxiety. Measuring psychosocial consequences can be difficult and requires at least a valid questionnaire, so-called Patient-Reported Outcome Measures (PROMs). A PROM should be developed in collaboration with people from the target population and relevant experts. This is important to make sure that: 1) the PROM adequately covers the potential psychosocial consequences, and 2) the PROM is relevant and understandable for the respondent. Also, valid measurement requires that PROMs are statistically tested in accordance with measurement theory. However, many PROMs that are used to measure the psychosocial consequences of cancer screening lack both elements. This results in low-quality evidence and leaves us uncertain about the true magnitude of psychosocial consequences. This review analyzes the quality of the PROMs used in studies that measure the psychosocial consequences of colorectal cancer screening. Twenty-nine out of thirty PROMs included in this review lacked proper patient involvement and had inadequate measurement properties in a screening context. This means that we cannot trust the results of the studies that use these PROMs. Future studies should use PROMs with adequate patient involvement and proper psychometric measurement properties, and existing evidence should be critically evaluated considering potential biases.

Received: 20 October 2022
Accepted: 18 February 2023
Published: 14 March 2023

Introduction

Patient-Reported Outcome Measures (PROMs) are outcomes reported by patients often in the form of standardized questionnaires.[1] PROMS are commonly used in health research to measure latent traits in patients, for example, opinions, behavior, or psychological states, which can then be used to compare or evaluate interventions.[1,2] PROMs have previously been used in screening contexts to measure psychosocial consequences of screening.[3–7] In a screening setting, the word patient in PROM should be read as apparently healthy persons. Any PROM should be rigorously developed and its measurement properties should be assessed to ensure that the PROM validly measures the latent trait and reliably assess changes over time. When this is not done, it is unclear what is measured.

Another aspect of PROMs is that they can be condition-specific or generic. Generic PROMs can be used in general populations to measure broad aspects of latent traits.[2,8,9] Generic PROMs presumably have higher generalizability but at the cost of containing items that are irrelevant for specific conditions (content relevance) and vice versa lack items related to aspects of the construct relevant for the specific condition (content coverage). Generic PROMs might therefore have low content validity in settings for specific conditions. Condition-specific PROMs are developed for a specific group or condition and capture elements of traits that are relevant in the specific context. When condition-specific PROMs are used outside of their intended context or generic PROMs are used without proper pre-testing, the validity and reliability of measurement can potentially become compromised.[9] The PROMs will then measure inaccurately or have low power to detect the specific trait, and thus findings are questionable.[10] Despite these concerns, the use of generic PROMs for specific conditions and condition-specific PROMs outside of their validated context is pervasive.[9,11,12] Arguably, many factors potentially drive this practice: 1) the ease of using available PROMs, thus bypassing the extensive work required for development and psychometric testing, 2) the comparability of findings to research that uses the same PROM, and 3) a spiral effect, where PROMs are used so frequently to assess specific traits in clinical and research settings, that it becomes a dogma that the PROM is valid and reliable even though it has never been tested in a relevant target population. Another widespread practice is the use of shortened versions or subscales of frequently used PROMs which also expose measures to the risk of poor validity and reliability.[11]

This research tendency or methodological unawareness have both scientific and practical implications. Scientifically, studies that use PROMs with inadequate measurement properties can produce biased effect estimates and hence evidence of low quality. Further, reviews and meta-analyses that are based on studies that use inadequate PROMs will not provide a higher rank of evidence. The scientific implications may in turn lead to practical implications. For instance, systematic reviews and meta-analyses are important for policymaking and an essential component of the practice of evidence-based medicine.[13] Systematic reviews and meta-analyses that build on studies using inadequate PROMs might cause practices or policies to be implemented or changed based on biased effect estimates. Consequently, such evidence syntheses, can have negative consequences for patients, care providers, and society, and result in ineffective and harmful policies and interventions. From an ethical viewpoint, this is especially important to keep in mind in regard to screening as it involves the general population and not a group of patients.[14] As a countermeasure to these concerns, different guidelines and checklists have been developed to assess the quality of PROMs, thus promoting valid and reliable measurement in research.[15–19] These guidelines and checklists define the quality of a PROM, which should then be defining for the trustworthiness of the results.

The aim of this study was to systematically assess studies that measure the psychosocial consequences of colorectal cancer screening using PROMs and review their methodological quality.

Methods

Prior to the initial search, we uploaded a protocol for this systematic review to the International Prospective Register of Systematic Reviews (PROSPERO) on November 17, 2016 (Registration number CRD42016051608).[20] The conduct and reporting of this systematic review have followed the Cochrane Handbook,[2] the PRISMA checklist,[21] and relevant methodological literature.[13,16]

Approach and Eligibility

We included all empirical research that studied aspects of psychosocial consequences of colorectal cancer screening.[20] Studies were included regardless of the PROMs that was used to assess the outcome (Appendix 1). Studies that only reported, for example, anxiety as a single item were excluded. We restricted the inclusion to studies that reported on an average screening population, in other words, adults (+18) that did not have any known risk factors of colorectal cancer. Eligible study designs included randomized controlled trials, cohort studies, case-control studies, prognostic studies, and qualitative development or validation studies. We did not perform any restrictions regarding study groups, screening settings, follow-up time, or language. Peer review was not required.

Search Process

Three authors and a librarian scientist developed the literature search. We performed the search in four databases: MEDLINE, CINAHL, PsycINFO, and EMBASE (including PubMed) all on August 22, 2016. We initially developed the search for PubMed and subsequently adapted it to the databases (Appendix 2). We updated the search twice: in August 21, 2019 and November 23, 2021.

Two authors independently screened studies at title, abstract, and full-text levels according to the pre-defined eligibility criteria (Appendix 1). [20] Discrepancies were resolved by discussion. The last author was consulted if consensus could not be reached.

Two authors independently assessed the reference lists of included studies to identify additional studies not found in the systematic search (Snowballing). We kept track of systematic reviews and likewise scrutinized reference lists for relevant literature.

Data Extraction and Synthesis

Data extraction was pre-specified in the protocol.[20] Data extraction included study design, setting and population, content validity, statistical psychometric measurement properties, and information about the PROM. Two authors extracted the data independently, and discrepancies were resolved by discussion. When consensus could not be reached, the last author was consulted. Authors of the included publications were contacted when necessary, for example, if the methodology or use of the PROM were unclear or in case of missing data.

The COSMIN Checklist

We used the COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) Risk of Bias checklist to systematically evaluate the quality of measurement properties in PROMs.[15,22] Risk of bias refers to whether the results of the study are trustworthy in regard to methodological quality.[15,16] In accordance with the COSMIN checklist, two authors independently assessed the PROMs.[15,16] Each domain was assessed by several items rated as “Very good”, “Adequate”, “Doubtful”, or “Inadequate”. Some items could only be graded dichotomously: “Very good” or “Doubtful”.[15] For some standards, “Not applicable” (N) was also an option. The COSMIN checklist grades PROMs according to the principle of “worst score counts” because poor methodological aspects cannot be compensated by good aspects.[23] Any disagreements were resolved through discussion, and in case of non-consensus, the last author was consulted. The COSMIN checklist consists of 10 domains and respective subdomains (Table 1). [15]

The first domain covers the development of the PROM, which includes an overall assessment of the construct and appropriateness of cognitive interviews or pilot testing. The second domain covers the content validity and the degree of patient and professional involvement in the PROM development. This should be evaluated for the context in which the PROM is used. This domain was evaluated based on three aspects: content relevance, content coverage (comprehensiveness), and understandability (comprehensibility). These two domains have been emphasized as the most important properties of a PROM.[15,16,23–26]

The rest of the domains and subdomains were evaluated based on design requirements, use of specific statistical methods, and an assessment of design and methodological flaws. Regarding the dimensionality of the

factors, the COSMIN checklist considers whether the factor structure is validated by exploratory factor analysis (EFA), confirmatory factor analysis (CFA), or item response theory (IRT) models, as well as the adequacy of the sample size. The COSMIN checklist assesses internal consistency through the calculation of Cronbach's alpha or omega. Measurement invariance is confirmed if the researchers have performed differential item function (DIF) analysis or CFA.

We evaluated PROMs based on all the studies that used them and based on their respective method sections and references. If no references were provided, we searched PubMed for original development or validation studies. This approach gave the grading of the PROMs the benefit of the doubt, as we are aware that poor methodological reporting does not always equals poor quality.

The COSMIN checklist is a modular tool and the boxes can be used separately. If a review only focuses on elements of PROM quality or if not all measurement properties are assessed for the respective PROM, it is not necessary to complete the whole checklist.[15,16] We categorized PROMs as condition-specific if they were either developed in the context of cancer or screening. If the PROM was condition-specific, then we used box 1. If not, this box was skipped, and the PROM was immediately rated "Inadequate" in regard to PROM development (Table 1). We continued the grading if the overall score was "Adequate" or "Very good".[15] If not, the grading was

Table 1 The COSMIN Checklist

Box	Domain	Taxonomy/Subdomains
1	PROM development	1. PROM design 2. Cognitive interview study or other pilot tests
2	Content validity	The PROM is an adequate reflection of the construct to be measured 1. Asking patients about relevance 2. Asking patients about comprehensiveness 3. Asking patients about comprehensibility 4. Asking professionals about relevance 5. Asking professionals about comprehensiveness
3	Structural validity	The degree to which the scores of a PROM are an adequate reflection of the dimensionality of the construct to be measured and is usually assessed by factor analysis or exploratory analysis
4	Internal consistency	The degree of interrelatedness among the items and is often assessed by Cronbach's alpha or omega
5	Cross-cultural validity /measurement invariance	The degree to which the performance of the items on a translated or culturally adapted instrument are an adequate reflection of the performance of the items of the original version of the instrument
6	Reliability	The proportion of the total variance in the measurements which is due to "true" differences between patients (Intraclass correlation coefficient, Kappa, etc.).
7	Measurement error	The systematic and random error on an individual patient's score that is not attributed to true changes in the construct to be measured (Stability of patients in interim period, time interval, and test conditions and appropriateness of the statistical methods)
8	Criterion validity	The degree to which the scores of a PROM are an adequate reflection of a "gold standard" (ROC curves, sensitivity, specificity, correlations)
9	Hypothesis testing for construct validity	The degree to which the scores of a PROM are consistent with hypotheses based on the assumption that the PROM validly measures the construct to be measured. For instance, comparison with other outcome measurement instruments (convergent validity) or comparison between subgroups (discriminative or known-groups validity)
10	Responsiveness	The ability of a PROM to detect change over time in the construct to be measured (Criterion approach, Construct approach: comparison with other outcome measurement instruments or comparison between subgroups or before and after intervention)

concluded. This approach applied for each of the ten domains. According to the COSMIN taxonomy, if the study did not report on a domain or measurement property, the respective boxes were skipped (Table 1).[15,23]

Data Synthesis

As pre-specified in the protocol, we anticipated a wide spectrum of PROMs and thus limited scope for meta-analyses. If several PROMs had adequate measurement properties and comparable study designs or subgroups, we would perform meta-analyses. The COSMIN guideline for systematic review of PROMs recommends that evidence is graded according to the GRADE approach if studies' results are analyzed or pooled.[16]

Results

After the removal of duplicates, we overall identified 13687 unique publications whereof 33 were included for review (Figure 1). We excluded 68 studies at full-text level, mostly due to wrong outcome or design (Appendix 3).

Study Characteristics

The majority of the included studies were observational studies (88.6%) and used a control group (74.3%). All studies were conducted in high-income countries: one study was conducted in Taiwan (2.9%), while the rest was from European countries (77.2%), Australia (11.4%), or The United States (8.5%). Most studies reported on adults aged 50–80 (94.3%), despite two studies that included all adults older than 20 or 40 years, respectively (Table 2). Most studies reported on the impact of a positive FOBT (42.9%) or invitation to screening (25.7%) (Table 2).

Most studies defined their primary outcome as psychological consequences (69.7%), the rest aimed to measure psychosocial consequences (12.1%), quality of life (3.0%), or both quality of life and psychological impact (15.2%) (Table 2). The 33 included studies used 30 different PROMs. Eleven of the PROMs (36.7%) were condition-specific (counting all versions of Psychological Consequences Questionnaire (PCQ)), and 16 of the 33 studies (48.5%) used one

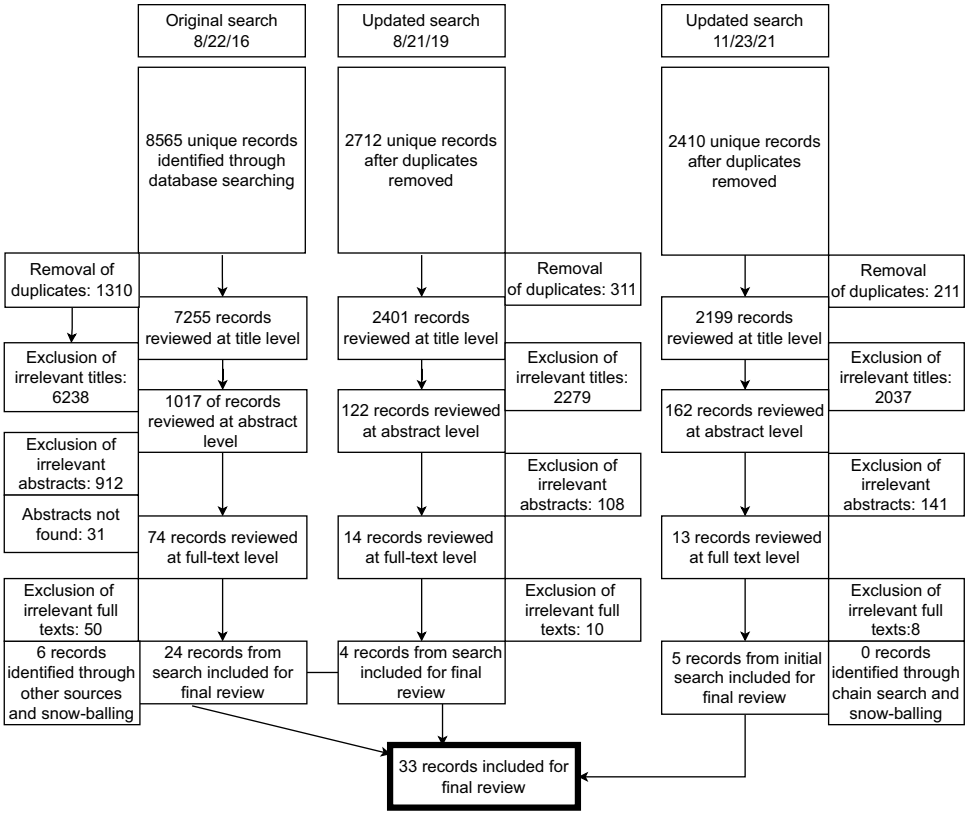

Figure 1 PRISMA flow diagram.

Table 2 Study Characteristics

Authors, Year	Population	Study Size	Study Type	Comparators	Outcome of Interest	Outcome Measure(s)[a]	Frequency/ Burden
Alexander et al 2003[27]	Men and women aged 50–69 screened with FOBT	6601	Cross-sectional	Non-responders, FOBT negative, Phase III negatives, FOBT positive, Cancer positive	Psychological consequences	HADS, STAI	Both
Bobridge et al 2011[28]	Men and women averagely aged 58.4 with a positive FOBT	93	Longitudinal	Normative data	Psychological consequences	STAI	Burden
Bobridge et al 2014[29]	Men and women aged 50–76 with a positive FOBT	301	Longitudinal	Negative result	Psychological consequences	SF-36, STAI, MLoC	Burden
Brasso et al 2010[30]	Men and women aged 50–75 with a positive FOBT	600	Longitudinal	Negative result	Psychological consequences	SCL92 short-form	Both
Chiu et al 2016[31]	Adults aged ≥20 with a positive FOBT	110	RCT	Control group	Psychological consequences	IES, STAI, **FACT-C**	Burden
Christy et al 2018[32]	Men and women aged 50–75 invited to CRC screening	416	Cross-sectional	None	Cancer worry	**Cancer worry variables**	Both
Denters et al 2013[33]	Men and women aged 50–74 invited to CRC screening	10265	Longitudinal	Negative result	Psychological consequences	**PCQ (shortened, 12 items on negative consequences)**	Burden
de Wijkers-looth et al 2011[34,35]	Men and women aged 50–75 invited to or screened with CT colonography or colonoscopy	3310	RCT	CT-colonography invitees and participants	Psychological consequences	EBQ/PBQ	Both
Hagger and Orbell 2006[36]	Men and women averagely 62.1 years with a positive FOBT	1361	Cross-sectional	None	Psychological consequences	STAI-6, STAXI, HADS	Burden
Kapidzic et al 2012[37]	Men and women aged 50–74 with a positive FOBT	2461	Cross-sectional	Negative result	QoL and psychological consequences	STAI-6, SF-12, EQ-5D, **PCQ (shortened, 12 items on negative consequences), CWS**	Burden
Kirkøen et al 2016[38]	Men and women aged 50–74 invited to CRC screening	21944	Longitudinal	Control group	QoL and psychological consequences	HADS, SF-12	Burden
Kirkøen et al 2016[39]	Men and women aged 50–74 invited to CRC screening	21944	Longitudinal	Control group	QoL and psychological consequences	HADS, SF-12	Burden
Laing et al 2013[40]	Men and women aged 50–74 with a positive FOBT	2260	Longitudinal	Negative result	Psychological consequences	STAI	Both
Lindholm et al 1997[41]	Men and women aged 60–64 invited to CRC screening	6366	RCT	Non-attendees	Psychological consequences	**Worry variables**	Frequency
Malmqvist et al 2020[42]	Men and women aged 50–74 screened with FOBT	410	Validation study	Not relevant	Psychosocial consequences	**COS-CRC**	Not relevant
Malmqvist et al 2021[43]	Men and women aged 50–74 with a positive FOBT	3732	Longitudinal	True positive, negative results and not screened	Psychosocial consequences	**COS-CRC**	Burden

(*Continued*)

Table 2 (Continued).

Authors, Year	Population	Study Size	Study Type	Comparators	Outcome of Interest	Outcome Measure(s)[a]	Frequency/ Burden
Malmqvist et al 2021[44]	Men and women aged 50–74 invited to CRC screening	2000	Longitudinal	Control group	Psychosocial consequences	**COS-CRC**	Burden
Miles and Wardle 2006[45]	Men and women aged 55–64 screened with sigmoidscopy	4329	Longitudinal and cross-sectional	Negative result	Psychological consequences	HAQ (shortened, 4 items), STAI-6 (generic anxiety), **PCQ (shortened. 3 items on positive emotions)**	Burden
Miles et al 2009[46]	Men and women aged 55–64 invited to CRC screening	50963	RCT	Negative result	Psychological consequences	GHQ-12, **PCQ (shortened, 3 items on positive emotions)**, STAI-6	Burden
Miles et al 2015[47]	Men and women aged 56–81 with a positive FOBT	675	Cross-sectional	Control group	QoL and psychological consequences	**FACT-C**, CES-D (shortened 10-item version)	Burden
Mountifield et al 2013[48]	Men and women averagely 58.2 years with a positive FOBT awaiting colonoscopy	70	Cross-sectional	Not relevant	QoL and psychological consequences	SF-36, MLoC, STAI	Burden
Mountifield et al 2011[49]	Men and women aged <40 with a positive FOBT going who after underwent colonoscopy	42	Cross-sectional	Not relevant	Psychological consequences	STAI	Burden
Orbell et al 2008[50]	Men and women aged 50–70 with a positive FOBT	1335	Cross-sectional	None	Psychosocial consequences	STAI, **IPQR**, WCQ	Burden
Parker et al 2002[51]	Men and women aged 50–75 1) screened with FOBT or 2) with a positive FOBT	1) 1496 2) 101	Longitudinal	None	Psychological consequences	GHQ, STAI	Both
Robb et al 2012[52]	Men and women aged 58–59 invited to CRC screening	1024	Longitudinal	None	Psychological consequences	STAI-4	Frequency
Sharp et al 2015[53]	Men and women aged 50–74 with a positive FOBT	201	Cross-sectional	None	Psychological consequences	IES	Both
Simon et al 2005[54]	Men and women aged 50–59 screened with sigmoidscopy	5942	Longitudinal	Negative result	Psychological consequences	STAI, GHQ (Shortened, 12 items), **PCQ (shortened, 6 items on positive consequences)**	Burden
Thiis-Evensen et al 1999[55]	Men and women aged 50–59 screened with colonoscopy	876	Longitudinal	None	Psychological consequences	HADS, GHQ (shortened, 28 items)	Burden
Taupin et al 2006[56]	Men and women aged 55–74 screened with colonoscopy	231	Longitudinal	Normative data	Quality of life	SF-36	Burden
Tutein Nolthenius et al 2016[57]	Men and women aged 50–75 with CT-colonography-screen-detected polyps.	78	Longitudinal	Index screening	Psychological consequences	PBQ, IES	Both
van Dam et al 2013[58]	Men and women aged 50–74 screened with CT colonography	609	Longitudinal	Endoscopy	Psychological consequences	QOL questionnaire, STAI-6,	Frequency
Vermeer et al 2020[59]	Men and women aged 55–75 with a positive FOBT	2151	Longitudinal	Index screening and normative data	Psychological consequences	**PCQ (shortened 12 items), CWS**, DRS, SF-36	Both

(*Continued*)

Table 2 (Continued).

Authors, Year	Population	Study Size	Study Type	Comparators	Outcome of Interest	Outcome Measure(s)[a]	Frequency/ Burden
Wardle et al 2003[60]	Men and women 1) aged 55–64 screened with sigmoidscopy or 2) aged 50–65 invited to CRC screening	1) 4153 2) 1951	1) Cross-sectional 2) Longi-tudinal	Negative result	Psychological consequences	STAI-6, **PCQ (Shortened, 6 items)**	Burden

Note: [a]Condition-specific PROMs are marked as bold.
Abbreviations: CWV, Cancer Worry Variables; CES-D, Center for Epidemiologic Studies Depression Scale; COS-CRC, Consequences Of Screening- ColoRectal Cancer; CWS, Cancer Worry Scale; EQ-5D, EuroQol-5 Domain; DRS, EBQ/PBQ, Expected and Perceived Barrier Questionnaire; FACT-C, Functional Assessment of Cancer Therapy – Colorectal; GHQ, General Health Questionnaire; HADS, Hospital Anxiety and Depression Scale; HAQ, Health Assessment Questionnaire; IES, Impact of Event Scale; IPQ-R, Revised Illness Perception Questionnaire; MLoC, Multidimensional Locus of Control; PCQ, Psychological Consequences Questionnaire; QOL, Quality of Life; SCL-92, Symptom Checklist; SF, The Short Form 12/36-Item Health Survey; STAI, State-Trait Anxiety Inventory; STAXI, State-Trait Anger eXpression Inventory; WCQ, Ways of Coping Questionnaire.

of these. Consequences Of Screening – ColoRectal Cancer (COS-CRC) and Cancer Worry Variables (CWV) were the only two measures (6.7%) developed in both a colorectal cancer and screening setting. The studies generally used multiple PROMs with an average of 2.1 (Range 1–4).

Because of the wide spectrum of outcome measures and study designs, we argue that there was no scope for meta-analyses. Therefore, we conducted descriptive data synthesis with focus on the methodological aspects of included PROMs.

The COSMIN-Checklist Grading

The COSMIN grading is presented in the two tables below (Tables 3 and 4). Inter-rater reliability of COSMIN grading was 93%.

The quality of PROM development is presented in Table 3. Most condition-specific PROMs had a clear description of the construct of interest they aimed to assess and of the target population. Only one PROM, COS-CRC, used appropriate construct theory (origin of construct); all other PROMs received the lowest possible score "Doubtful" on this item. PROMs generally received low scores due to a lack of appropriate qualitative data-collection methods to identify relevant items and use of cognitive interviews. For example, both COS-CRC and CWV were developed in the context of screening and colorectal cancer, yet only COS-CRC was qualitatively tested in the context of colorectal cancer screening.

Most studies had sparse methodological reporting and did not reference any development- or validation studies and were thus graded "Inadequate". Except for Worry Variables, all condition-specific PROMs had one or more scores of "Very good". All PROMs, except COS-CRC, received the lowest possible score for the total assessment. COS-CRC was overall graded "Doubtful" because two researchers did not code the qualitative data independently (Table 3).

We assessed the content validity for all PROMs whether they were condition-specific or not (Table 4).

The majority of the PROMs had inadequate involvement of patients and professionals in the development phase. While COS-CRC and CWV were the only two measures developed for both colorectal cancer and screening setting, CWV did not have any patient involvement in the development phase. Three PROMs were overall graded "Doubtful": Health Assessment Questionnaire (HAQ) short-form, EuroQual 5 Domains (EQ-5D), and COS-CRC. COS-CRC was the only PROM to ever receive the best score "Very good" in three out of five subdomains. COS-CRC was overall graded "Doubtful" due to a lack of proper involvement of professionals (Table 4).

No PROMs were assessed beyond this domain as none were overall rated as "Adequate" or "Very good". The low grading was mainly due to the fact that studies had not sufficiently adapted the PROM to a colorectal cancer screening context.

Table 3 Quality of the PROM Development

PROM[a]	PROM Design							Cognitive Interview (CI) Study[c]				Total Prom Development
	General Design Requirements					Concept Elicitation	Total PROM Design	Design Requirements	Comprehen-Sibility	Comprehen-Siveness	Total CI Study	
	Clear Construct	Clear Origin of Construct[b]	Clear Target Population for Which the PROM Was Developed	Clear Context of Use[b]	PROM Developed in Sample Representing the Target Population			CI study Performed in Sample Representing the Target Population				
CWV	V	D	V	V	D	I	I	N	N	N	I	I
CES-D												I
COS-CRC	V	V	V	V	V	D	D	V	V	V	V	D
CWS	V	D	V	V	I	I	I	V	N	I	I	I
EQ-5D												I
DRS												I
EBQ/ PBQ	I	D	V	V	V	I	I	N	N	N	I	I
FACT-C	V	D	V	D	V	I	I	V	I	I	I	I
GHQ												I
GHQ-12												I
GHQ-28												I
HADS												I
HAQ Short-form												I
IES												I
IPQ-R	V	D	I	D	D	I	I	N	I	N	I	I
MLoC												I
PCQ	V	D	V	D	I	I	I	N	I	I	I	I
PCQ-short 1	V	D	V	D	I	I	I	N	I	I	I	I
PCQ-short 2	V	D	V	D	I	I	I	N	N	N	I	I

(*Continued*)

https://doi.org/10.2147/PROM.S394247 DovePress

Table 3 (Continued).

PROM[a]	PROM Design							Cognitive Interview (CI) Study[c]				Total Prom Development
	General Design Requirements					Concept Elicitation	Total PROM Design	Design Requirements	Comprehen-Sibility	Comprehen-Siveness	Total CI Study	
	Clear Construct	Clear Origin of Construct[b]	Clear Target Population for Which the PROM Was Developed	Clear Context of Use[b]	PROM Developed in Sample Representing the Target Population			CI study Performed in Sample Representing the Target Population				
PCQ-short 3	V	D	V	D	I	I	I	N	N	N	I	I
QOL												I
SCL-92												I
SF-12												I
SF-36												I
STAI												I
STAI-4												I
STAI-6												I
STAXI												I
WCQ												I
Worry Variables	I	D	I	D	D	I	I	N	N	N	I	I

Notes: [a]Only condition-specific PROMs were graded and these are marked as bold. [b]Dichotomous score: V or D. [c]N indicates that a cognitive interview study (or part of it) was not performed.
Abbreviations: V, very good; A, Adequate; D, Doubtful; I, Inadequate; N, Not Applicable.

https://doi.org/10.2147/PROM.S394247 DovePress

Table 4 Quality of the Content Validity

PROM[a]	Content Validity					
	Asking Patients			Asking Professionals		Lowest Score on Content Validity
	Relevance	Compre-Hensiveness	Compre-Hensibility	Relevance	Compre-Hensiveness	
CWV	I	I	I	I	I	I
CES-D	I	I	I	I	I	I
COS-CRC	V	V	V	D	D	D
CWS	D	D	I	D	I	I
EQ-5D	D	D	D	D	D	D
DRS	D	D	I	D	I	I
EBQ/PBQ	I	I	I	I	I	I
FACT-C	D	D	I	D	D	I
GHQ	I	I	I	I	I	I
GHQ-12	I	I	I	I	I	I
GHQ-28	I	I	I	I	I	I
HADS	I	I	I	I	I	I
HAQ Short-form	D	D	D	D	D	D
IES	I	I	I	I	I	I
IPQ-R	I	I	I	I	I	I
MLoC	I	I	I	I	I	I
PCQ	I	I	I	I	I	I
PCQ-short 1	I	I	I	I	I	I
PCQ-short 2	I	I	I	I	I	I
PCQ-short 3	I	I	I	I	I	I
QOL	I	I	I	I	I	I
SCL-92	I	I	I	I	I	I
SF-12	I	I	I	I	I	I
SF-36	I	I	I	I	I	I
STAI	I	I	I	I	I	I
STAI-4	I	I	I	I	I	I
STAI-6	I	I	I	I	I	I
STAXI	I	I	I	I	I	I
WCQ	I	I	I	I	I	I
Worry Variables	D	I	I	I	I	I

Note: [a]Condition-specific PROMs are marked as bold.
Abbreviations: V, very good; A, Adequate; D, Doubtful; I, Inadequate.

Discussion

Summary of Findings

This review included 33 studies that all together used 30 different PROMs to measure psychosocial consequences of colorectal cancer screening. Only eleven of these PROMs (36.7%) were developed in the context of either cancer or screening, and only two in both (6.7%). Studies generally used multiple PROMs, while less than half included one that was condition-specific. According to the COSMIN checklist, 29 out of 30 studies (96.7%) had inadequate PROM development. PROMs generally lacked proper cognitive interview and pilot studies. Across all PROMs, three (10%) had doubtful and 27 (90%) had inadequate measurement properties, due to lack of patient and professionals involvement in the development phase.

According to the COSMIN manual, we chose not to conduct meta-analyses, as PROMs generally had inadequate measurement properties and pooling results in one analysis assumes high content and construct validity. By extension, we neither graded the evidence according to GRADE or COSMIN criteria.[61,62]

Comparison to Existing Literature

Using PROMs with inadequate or unknown measurement properties or outside of their intended context has been criticized before.[8,12,63–66] Already in 2004, Brodersen et al argued that the General Health Questionnaire (GHQ), the State-Trait-Anxiety-Inventory (STAI), and the Hospital Anxiety and Depression Scale (HADS) were not adequate in any cancer screening context.[12] Despite these critiques, this questionable use of PROMs remains partly unchanged. A 2002 study illustrates these concerns; post-hoc analyses showed that the generic Short-Form 36 (SF-36) had limited validity as an outcome measure of health status after stroke.[67] The authors highlighted the importance of testing scale assumptions before applying outcome measures to new populations. This study also highlights another concern; that this issue is not unique to cancer screening. For example, a review on PROMs used in sports science showed that about one-third of included studies used PROMs in other contexts than they were intended for.[11] The same was found in a review on PROMs used in Randomized Controlled Trials (RCTs) in sports medicine.[8] Even though a PROM is validated in one context, there is no guarantee that it is valid in another context. Previous studies have reported a misalignment between patients' interpretation, and hence response, and the PROM's intended meaning.[68,69] This misalignment could stem from a lack of proper cognitive interviews or pilot studies.[24] A scoping review found that only 6.7% of included PROMs had proper patient involvement in the development phase and suggests that future research should base their choice of PROM on the level of patient involvement.[25] A finding compatible with our findings. One of the reviews found that when PROMs are used to evaluate conditions that they were not developed for, estimates are biased toward null.[8] This is potentially due to inaccuracy of measurement or low power; lack of content validity and thereby responsiveness. This might also be the case in the included studies of this review, and their respective effect estimates should be evaluated with these biases in mind.

In this review, the grading of PROMs was hindered by poor methodological reporting. Poor methodological reporting on the use of PROMs was also the case in reviews on shared decision-making,[24] pain,[65] and PROMs used in RCTs.[8] This lack of emphasis on methodological reporting also speaks to the lack of attention on PROM development and validity.

Previous reviews on psychosocial consequences in colorectal cancer screening have not assessed the adequacy of measurement properties or quality of PROMs.[70,71] However, based on included studies, van der Velde et al conclude that no psychological impact was sustained three months after a false-positive colorectal cancer screening.[70] This interpretation is questionable as it relies on the quality of the PROMs, which was not assessed and estimates are likely to be biased towards null when based on generic or non-specific PROMs.[8] A review by Selva et al reported that only 7 out of 75 identified PROMs used to measure experience or satisfaction with colorectal cancer screening were (self-reported) validated.[72]

Strengths and Limitations

Initially, we aimed to synthesize the psychosocial consequences of colorectal cancer screening. As we did the formal screening of results, we found that this was not feasible due to number of different PROMs and the varying quality of these. Therefore, we chose to change the main outcomes of our review to the methodological quality of PROMs instead of the psychosocial harms themselves as specified in the original protocol (Appendix 4). Another change was made from the original protocol concerning methods, because the COSMIN checklist was published in the meantime and we wanted to conform to best available methods (Appendix 4).

In this study, a PROM was considered condition-specific if it was developed within a context of either cancer or screening. Our definition was very inclusive to give semi-condition-specific PROMs the benefit of the doubt regarding COSMIN grading. However, what is relevant in the context of cancer patients will not always be relevant for apparently healthy citizens participating in colorectal cancer screening – and will most likely not cover all aspects of psychosocial consequences.[42,73] We did not assess the PROM development for PROMs that were considered non-specific (Box 1). A generic or non-condition-specific PROM can indeed be well-developed, yet that does not mean that it is adequate in a context of colorectal cancer screening. If these non-condition-specific PROMs were evaluated in box 1, it would seem that they were more adequate in the context than they are. Nevertheless, If PROMs had at least adequate content validity, the quality of the PROM development could be relevant for the use of the PROM. Ideally, if researchers wish to use a non-condition-specific PROM, they should conduct a content-validity study in the population of interest. However, if such a study is not conducted or if such information is not available, the PROM development could have some value.

Although the COSMIN checklist is one of the most rigorous and widespread tools for evaluating the adequacy of PROM measurement in health research, the approach has some limitations. First, the COSMIN checklist is a standardized checklist which means that it does not allow for subjective assessment of the specific PROM. The standardized checklist assumes that the domains weigh equally in the overall grading of the PROM. For example, in the context of psychosocial consequences of colorectal cancer screening, we would argue that patient involvement is far more important than the involvement of professionals. The majority of PROMs included in this review failed to involve a proper number of professionals from relevant disciplines, but the importance of this specific item should potentially be downgraded. The COSMIN checklist requires ≥7 professionals for qualitative studies and ≥50 for quantitative studies to reach the grade "Very good" and ≥4 or 30 professionals, respectively, for "Doubtful". This quality-quantification of the qualitative development might falsely grade PROMs better or worse than they actually are. Acknowledging these problems, the COSMIN group has amended a number of standards since the first introduction of the checklist and recommends that the checklist is used as guidance.[15,16]

Limitations of the COSMIN checklist regarding assessment of the dimensionality has previously been discussed in Heiberg Agerbeck et al 2021 and McKenna and Heaney 2021.[26,65] Researchers have also argued for the importance of construct theories in the development of PROMs, which the COSMIN checklist only sparsely emphasizes.[26,74] Other researchers have argued that the COSMIN checklist does not take into account that quality in the development phase can differ across health outcomes.[26]

We graded according to the principle of benefit of the doubt, but we cannot rule out that PROMs might be of better or worse quality than what is graded here. Conclusively, we encourage research that use the COSMIN checklist, to evaluate the risk of bias beyond the checklist, for example, in regard to composite measures, unidimensionality, and biases relevant for the specific health outcome.

Additional Findings

While conducting this review, we noticed additional aspects of the included PROMs beyond the outcomes defined in the protocol.

Only seven articles (21.2%) across three author groups discussed the limitations of the PROMs they used.[42–46,59] However, this reporting was generally deficient and did not focus on the implications for results.

Almost half of the studies (42.2%) used at least one shortened form of a PROM (Table 2). To use only a part of a questionnaire threatens content coverage and construct validity. Using a short-form has an underlying, often implicit,

assumption that the short-form can be a surrogate for the full measure. A statement that should be tested qualitatively or statistically. However, none of the studies that used short-forms had tested whether this was valid in their respective population.

We also noticed a tendency to produce composite outcomes. Summating scales into one composite measure can conceal true changes in the outcome. Individuals might score higher on one domain, lower on another compared to baseline, and thus these changes will be balanced out when scores are summated across domains. McKenna and Heaney argue that composite scales, by principle, should be considered low quality.[26]

Implications for Research and Practice

This review sheds light on the trustworthiness of studies that use inadequate outcome measures and calls for reevaluating the existing evidence on the psychosocial consequences of colorectal cancer screening. The magnitude of measurement bias should be evaluated for each PROM individually.

PROMs should be tested in the population of interest prior to measurement. If the resources are not available, then researchers will have to explore alternative methodologies. Patient involvement is crucial in PROM development to ensure concordance between patients' interpretation and intention, and to gain high content validity. Therefore, this part cannot be left out in proper PROM design. Future grading of PROMs should account for the bias beyond the COSMIN checklist, for example, in regard to composite measures and unidimensionality.

For existing evidence, policymakers, researchers, and clinicians will need to beware of the poor quality and resulting potential biases, so that real-life practices are not affected accordingly. If policies or medical practices are changed as a consequence of biased research it might have unintended harmful implications that can affect patients as well as professionals. Scientific journals should preferably not publish studies that use inadequate outcome measures or should at least demand disclosure of the limitations when doing so.

Unanswered Questions and Future Research

This review focused on the methodological quality of the use of PROMs to measure psychosocial consequences in a colorectal cancer screening context. Future reviews could focus on the overall quality of evidence of the studies' designs and how the quality of measurement affects estimates of psychosocial consequences. Such a review should follow the COSMIN manual for systematic reviews of PROMs and use the GRADE approach and give overall quality ratings of PROMs.

Conclusion

This review included 33 studies that used 30 different PROMs. Studies generally used multiple PROMs, yet less than half included one that was condition-specific. According to the COSMIN checklist, 29 out of 30 PROMs had inadequate PROM development and 27 had inadequate measurement properties. Conclusively, the majority of PROMs used to study psychosocial consequences of colorectal cancer screening have no proven content validity in this context. This grading of methodological quality should be used in the overall grading of the evidence. Evidence of methodological quality should as well be a defining factor for the trustworthiness of studies that report on the psychosocial consequences of colorectal cancer screening.

Abbreviations

PROM, Patient-Reported Outcome Measure; COSMIN, The COnsensus-based Standards for the selection of health Measurement Instruments; RCT, Randomized Controlled Trials.

Acknowledgment

JaR and JBB developed the COS-CRC questionnaire. Neither were involved in the grading of the PROMs.

Disclosure

All authors declare no financial or non-financial competing interests.

References

1. Krogsgaard MR, Brodersen J, Christensen KB, et al. What is a PROM and why do we need it? *Scand J Med Sci Sports*. 2021;31(5):967–971. doi:10.1111/sms.13892
2. Johnston B, Patrick D, Devji D, et al. Patient-reported outcomes. In: *Cochrane Handbook for Systematic Reviews of Interventions*. Wiley Online Library; 2019.
3. Brodersen J, McKenna S, Doward L, Thorsen H. Measuring the psychosocial consequences of screening. Commentary. *Health Qual Life Outcomes*. 2007;5(1). doi:10.1186/1477-7525-5-3
4. Brodersen J, Siersma VD. Long-term psychosocial consequences of false-positive screening mammography. *Ann Fam Med*. 2013;11(2):106–115. doi:10.1370/afm.1466
5. Brodersen J, Thorsen H, Kreiner S. Consequences of screening in lung cancer: development and dimensionality of a questionnaire. *Value Health*. 2010;13(5):601–612. doi:10.1111/j.1524-4733.2010.00697.x
6. Damhus CS, Siersma V, Hansson A, Bang CW, Brodersen J. Psychosocial consequences of screening-detected abdominal aortic aneurisms: a cross-sectional study. *Scand J Prim Health Care*. 2021;39(4):459–465. doi:10.1080/02813432.2021.2004713
7. Cockburn J, Staples M, Hurley SF, De Luise T. Psychological consequences of screening mammography. *J Med Screen*. 1994;1(1):7–12. doi:10.1177/096914139400100104
8. Hansen CF, Jensen J, Brodersen J, Siersma V, Comins JD, Krogsgaard MR. Are adequate PROMs used as outcomes in randomized controlled trials? An analysis of 54 trials. *Scand J Med Sci Sports*. 2021;31(5):972–981. doi:10.1111/sms.13896
9. Churruca K, Pomare C, Ellis LA, et al. Patient-reported outcome measures (PROMs): a review of generic and condition-specific measures and a discussion of trends and issues. *Health Expect*. 2021;24(4):1015–1024. doi:10.1111/hex.13254
10. Wiebe S, Guyatt G, Weaver B, Matijevic S, Sidwell C. Comparative responsiveness of generic and specific quality-of-life instruments. *J Clin Epidemiol*. 2003;56(1):52–60. doi:10.1016/S0895-4356(02)00537-1
11. Krogsgaard MR, Brodersen J, Jensen J, Hansen CF, Comins JD. Potential problems in the use of patient reported outcome measures (PROMs) and reporting of PROM data in sports science. *Scand J Med Sci Sports*. 2021;31(6):1249–1258. doi:10.1111/sms.13888
12. Brodersen J, Thorsen H, Cockburn J. The adequacy of measurement of short and long-term consequences of false-positive screening mammography. *J Med Screen*. 2004;11:39–44. doi:10.1177/096914130301100109
13. Munn Z, Stern C, Aromataris E, Lockwood C, Jordan Z. What kind of systematic review should I conduct? A proposed typology and guidance for systematic reviewers in the medical and health sciences. *BMC Med Res Methodol*. 2018;18(1):5. doi:10.1186/s12874-017-0468-4
14. The National Health Service (NHS). NHS screening. Available from: https://www.nhs.uk/conditions/nhs-screening/. Accessed August 31, 2022.
15. Mokkink LB, de Vet HC, Prinsen CAC, et al. COSMIN risk of bias checklist for systematic reviews of patient-reported outcome measures. *Qual Life Res*. 2018;27(5):1171–1179. doi:10.1007/s11136-017-1765-4
16. Prinsen CAC, Mokkink LB, Bouter LM, et al. COSMIN guideline for systematic reviews of patient-reported outcome measures. *Qual Life Res*. 2018;27(5):1147–1157. doi:10.1007/s11136-018-1798-3
17. Bombardier C, Tugwell P. Methodological considerations in functional assessment. *J Rheumatol*. 1987;14(Suppl 15):6–10.
18. Alrubaiy L, Hutchings HA, Williams JG. Assessing patient reported outcome measures: a practical guide for gastroenterologists. *United Eur Gastroenterol J*. 2014;2(6):463–470. doi:10.1177/2050640614558345
19. Streiner D. A checklist for evaluating the usefulness of rating scale. *Can J Psychiatry*. 1993;38(2):140–148. doi:10.1177/070674379303800214
20. Gram EG, Malmqvist J, Agerbeck A, Martiny F, Bie AK, Brodersen JB. Psychosocial consequences of colorectal cancer screening in the general population: a systematic review on the adequacy of measurement properties (CRD42016051608). PROSPERO; 2022. Available from: https://www.crd.york.ac.uk/prospero/display_record.php?RecordID=51608&VersionID=1300958. Accessed March 2, 2023.
21. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ*. 2021;372:n71. doi:10.1136/bmj.n71
22. Mokkink LB, Terwee CB, Patrick DL, et al. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. *J Clin Epidemiol*. 2010;63(7):737–745. doi:10.1016/j.jclinepi.2010.02.006
23. Terwee CB, Prinsen CAC, Chiarotto A, et al. COSMIN methodology for evaluating the content validity of patient-reported outcome measures: a Delphi study. *Qual Life Res*. 2018;27(5):1159–1170. doi:10.1007/s11136-018-1829-0
24. Barr PJ, Elwyn G. Measurement challenges in shared decision making: putting the 'patient' in patient-reported measures. *Health Expect*. 2016;19(5):993–1001. doi:10.1111/hex.12380
25. Wiering B, de Boer D, Delnoij D. Patient involvement in the development of patient-reported outcome measures: a scoping review. *Health Expect*. 2017;20(1):11–23. doi:10.1111/hex.12442
26. McKenna SP, Heaney A. Setting and maintaining standards for patient-reported outcome measures: can we rely on the COSMIN checklists? *J Med Econ*. 2021;24(1):502–511. doi:10.1080/13696998.2021.1907092
27. Alexander F, Weller D, Orbell S, et al. Evaluation of the UK colorectal cancer screening pilot - final report; 2003.
28. Bobridge A, Young G, Lewis H, Cole S, Bampton P. Does participating in the national bowel cancer screening program have a psychological impact? 2011.
29. Bobridge A, Bampton P, Cole S, Lewis H, Young G. The psychological impact of participating in colorectal cancer screening by faecal immunochemical testing–the Australian experience. *Br J Cancer*. 2014;111(5):970–975. doi:10.1038/bjc.2014.371
30. Brasso K, Ladelund S, Frederiksen BL, Jorgensen T. Psychological distress following fecal occult blood test in colorectal cancer screening–a population-based study. *Scand J Gastroenterol*. 2010;45(10):1211–1216. doi:10.3109/00365521.2010.485355
31. Chiu HC, Hung HY, Lin HC, Chen SC. Effects of a health education and telephone counseling program on patients with a positive fecal occult blood test result for colorectal cancer screening: a randomized controlled trial. *Psychooncology*. 2017;26(10):1498–1504. doi:10.1002/pon.4319
32. Christy SM, Schmidt A, Wang HL, et al. Understanding cancer worry among patients in a community clinic-based colorectal cancer screening intervention study. *Nurs Res*. 2018;67(4):275–285. doi:10.1097/NNR.0000000000000275
33. Denters MJ, Deutekom M, Essink-Bot ML, Bossuyt PM, Fockens P, Dekker E. FIT false-positives in colorectal cancer screening experience psychological distress up to 6 weeks after colonoscopy. *Support Care Cancer*. 2013;21(10):2809–2815. doi:10.1007/s00520-013-1867-7

34. de Wijkerslooth TR, de Haan M, Stoop E, et al. Study protocol- population screening for colorectal cancer by colonoscopy or CT colonography - a randomized controlled trial. *BMC Gastroenterol*. 2010;10(47). doi:10.1186/1471-230X-10-47
35. de Wijkerslooth TR, de Haan MC, Stoop EM, et al. Burden of colonoscopy compared to non-cathartic CT-colonography in a colorectal cancer screening programme: randomised controlled trial. *Gut*. 2012;61(11):1552. doi:10.1136/gutjnl-2011-301308
36. Hagger MS, Orbell S. Illness representations and emotion in people with abnormal screening results. *Psychol Health*. 2006;21(2):183–209. doi:10.1080/14768320500223339
37. Kapidzic A, Korfage IJ, van Dam L, et al. Quality of life in participants of a CRC screening program. *Br J Cancer*. 2012;107(8):1295–1301. doi:10.1038/bjc.2012.386
38. Kirkoen B, Berstad P, Botteri E, et al. Do no harm: no psychological harm from colorectal cancer screening. *Br J Cancer*. 2016;114(5):497–504. doi:10.1038/bjc.2016.14
39. Kirkoen B, Berstad P, Botteri E, et al. Psychological effects of colorectal cancer screening: participants vs individuals not invited. *World J Gastroenterol*. 2016;22(43):9631–9641. doi:10.3748/wjg.v22.i43.9631
40. Laing SS, Bogart A, Chubak J, Fuller S, Green BB. Psychological distress after a positive fecal occult blood test result among members of an integrated healthcare delivery system. *Cancer Epidemiol Biomarkers Prev*. 2014;23(1):154–159. doi:10.1158/1055-9965.EPI-13-0722
41. Lindholm E, Berglund B, Kewenter J, Haglind E. Worry associated with screening for colorectal carcinomas. *Scand J Gastroenterol*. 1997;32(3):238–245. doi:10.3109/00365529709000201
42. Malmqvist J, Siersma V, Bang CW, Brodersen J. Consequences of screening in colorectal cancer (COS-CRC): development and dimensionality of a questionnaire. *BMC Psychol*. 2021;9(1). doi:10.1186/s40359-020-00504-3
43. Malmqvist J, Siersma VD, Hestbech MS, Bang CW, Nicolaisdottir DR, Brodersen J. Short and long-term psychosocial consequences of participating in a colorectal cancer screening programme: a matched longitudinal study. *BMJ Evid Based Med*. 2022;27(2):87–96. doi:10.1136/bmjebm-2020-111576
44. Malmqvist J, Siersma V, Hestbech MS, Nicolaisdottir DR, Bang CW, Brodersen J. Psychosocial consequences of invitation to colorectal cancer screening: a matched cohort study. *J Epidemiol Community Health*. 2021;75(9):867–873. doi:10.1136/jech-2019-213360
45. Miles A, Wardle J. Adverse psychological outcomes in colorectal cancer screening: does health anxiety play a role? *Behav Res Ther*. 2006;44(8):1117–1127. doi:10.1016/j.brat.2005.08.011
46. Miles A, Atkin WS, Kralj-Hans I, Wardle J. The psychological impact of being offered surveillance colonoscopy following attendance at colorectal screening using flexible sigmoidoscopy. *J Med Screen*. 2009;16(3):124–130. doi:10.1258/jms.2009.009041
47. Miles A, McClements PL, Steele RJ, Redeker C, Sevdalis N, Wardle J. The psychological impact of a colorectal cancer diagnosis following a negative fecal occult blood test result. *Cancer Epidemiol Biomarkers Prev*. 2015;24(7):1032–1038. doi:10.1158/1055-9965.EPI-15-0004
48. Mountifield RE, Bampton PA, Prosser R, Bobridge A, Mikocka-Walus AA, Andrews JM. Mo1962 FIT+ and IBD individuals have differing psychological reactions to the need for colonoscopy. *Gastroenterology*. 2013;144(5):S705. doi:10.1016/S0016-5085(13)62614-1
49. Mountifield RE, Moseley A, Prosser R, et al. Colonoscopic bowel cancer screening is associated with more depression and anxiety in previously healthy people than those with inflammatory bowel disease. *Gastroenterology*. 2011;140(5):S433. doi:10.1016/s0016-5085(11)61777-0
50. Orbell S, O'Sullivan I, Parker R, Steele B, Campbell C, Weller D. Illness representations and coping following an abnormal colorectal cancer screening result. *Soc Sci Med*. 2008;67(9):1465–1474. doi:10.1016/j.socscimed.2008.06.039
51. Parker MA, Robinson MH, Scholefield JH, Hardcastle JD. Psychiatric morbidity and screening for colorectal cancer. *J MedScreen*. 2002;9:7–10. doi:10.1136/jms.9.1.7
52. Robb KA, Lo SH, Power E, et al. Patient-reported outcomes following flexible sigmoidoscopy screening for colorectal cancer in a demonstration screening programme in the UK. *J Med Screen*. 2012;19(4):171–176. doi:10.1258/jms.2012.012129
53. Sharp L, Shearer N, Leen R, O'Morain C, McNamara D. Prevalence and predictors of colonoscopy-related distress in individuals undergoing fit-based colorectal cancer screening: a population-based study. Presented at: the Irish Society of Gastroenterology Winter Meeting, November 2013; 2015: The Malton Hotel, Killarney, Co. Kerry. Available from: https://www.ncbi.nlm.nih.gov/pubmed/25686788. Accessed March 2, 2023.
54. Simon AE, Steptoe A, Wardle J. Socioeconomic status differences in coping with a stressful medical procedure. *Psychosom Med*. 2005;67(2):270–276. doi:10.1097/01.psy.0000155665.55439.53
55. Thiis-Evensen E, Wilhelmsen I, Hoff GS, Blomhoff S, Sauar J. The psychologic effect of attending a screening program for colorectal polyps. *Scand J Gastroenterol*. 1999;34(1):103–109. doi:10.1080/00365529950172916
56. Taupin D, Chambers SL, Corbett M, Shadbolt B. Colonoscopic screening for colorectal cancer improves quality of life measures: a population-based screening study. *Health Qual Life Outcomes*. 2006;4(1):82. doi:10.1186/1477-7525-4-82
57. Tutein Nolthenius CJ, Boellaard TN, de Haan MC, et al. Burden of waiting for surveillance CT colonography in patients with screen-detected 6–9 mm polyps. *Eur Radiol*. 2016;26(11):4000–4010. doi:10.1007/s00330-016-4251-4
58. van Dam L, de Wijkerslooth TR, de Haan MC, et al. Time requirements and health effects of participation in colorectal cancer screening with colonoscopy or computed tomography colonography in a randomized controlled trial. *Endoscopy*. 2013;45(3):182–188. doi:10.1055/s-0032-1326080
59. Vermeer NCA, van der Valk MJM, Snijders HS, et al. Psychological distress and quality of life following positive fecal occult blood testing in colorectal cancer screening. *Psychooncology*. 2020;29(6):1084–1091. doi:10.1002/pon.5381
60. Wardle J, Williamson S, Sutton S, et al. Psychological impact of colorectal cancer screening. *Health Psychol*. 2003;22(1):54–59. doi:10.1037/0278-6133.22.1.54
61. Terwee CB, Bot SDM, de Boer MR, et al. Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol*. 2007;60(1):34–42. doi:10.1016/j.jclinepi.2006.03.012
62. Alper BS, Oettgen P, Kunnamo I, et al. Defining certainty of net benefit: a GRADE concept paper. *BMJ Open*. 2019;9(6):e027445. doi:10.1136/bmjopen-2018-027445
63. Bond M, Pavey T, Welch K, et al. Systematic review of the psychological consequences of false-positive screening mammograms. *Health Technol Assess (Rocky)*. 2013;17(13). doi:10.3310/hta17130
64. DeFrank JT, Barclay C, Sheridan S, et al. The psychological harms of screening: the evidence we have versus the evidence we need. *J Gen Intern Med*. 2015;30(2):242–248. doi:10.1007/s11606-014-2996-5

65. Heiberg Agerbeck A, Martiny FHJ, Jauernik CP, et al. Validity of current assessment tools aiming to measure the affective component of pain: a systematic review. *Patient Relat Outcome Meas*. 2021;12:213–226. doi:10.2147/PROM.S304950
66. DeFrank J, Brewer NT. Some more evidence of long-term psychosocial harms from receiving false-positive screening mammography results. *Evid Based Med*. 2014;19(1):38. doi:10.1136/eb-2013-101409
67. Hobart JC, Williams LS, Moran K, Thompson AJ. Quality of life measurement after stroke: uses and abuses of the SF-36. *Stroke*. 2002;33(5):1348–1356. doi:10.1161/01.STR.0000015030.59594.B3
68. Entwistle VA, Skea ZC, O'Donnell MT. Decisions about treatment: interpretations of two measures of control by women having a hysterectomy. *Soc Sci Med*. 2001;53(6):721–732. doi:10.1016/S0277-9536(00)00382-8
69. Davey HM, Lim J, Butow PN, Barratt AL, Redman S. Women's preferences for and views on decision-making for diagnostic tests. *Soc Sci Med*. 2004;58(9):1699–1707. doi:10.1016/S0277-9536(03)00339-3
70. van der Velde JL, Blanker MH, Stegmann ME, de Bock GH, Berger MY, Berendsen AJ. A systematic review of the psychological impact of false-positive colorectal cancer screening: what is the role of the general practitioner? *Eur J Cancer Care (Engl)*. 2017;26(3):e12709. doi:10.1111/ecc.12709
71. Chad-Friedman E, Coleman S, Traeger LN, et al. Psychological distress associated with cancer screening: a systematic review. *Cancer*. 2017;123(20):3882–3894. doi:10.1002/cncr.30904
72. Selva A, Selva C, Alvarez-Perez Y, et al. Satisfaction and experience with colorectal cancer screening: a systematic review of validated patient reported outcome measures. *BMC Med Res Methodol*. 2021;21(1):230. doi:10.1186/s12874-021-01430-7
73. Brodersen J, Thorsen H. Consequences of Screening in Breast Cancer (COS-BC): development of a questionnaire. *Scand J Public Health*. 2008;26:251–256. doi:10.1080/02813430802542508
74. Birney DP, Beckmann JF, Beckmann N, Stemler SE. Sophisticated statistics cannot compensate for method effects if quantifiable structure is compromised. *Front Psychol*. 2022;13:812963. doi:10.3389/fpsyg.2022.812963