Ghanouni, A; Smith, SG; Halligan, S; Plumb, A; Boone, D; Yao, GL; Zhu, S; Von Wagner,C.(2013) Public preferences for colorectal cancer screening tests: a review of conjoint analysis studies. **Expert Rev Med Devices** , 10 (4) 489 - 499. 10.1586/17434440.2013.811867.

Article

Public preferences for colorectal cancer screening tests: A review of conjoint analysis studies

Alex Ghanouni[1] BSc MSc; Samuel G. Smith[1] BSc MSc; Steve Halligan[2] MD FRCR FRCP; Andrew Plumb[2] MRCP FRCR; Darren Boone[2] BSc MRCS FRCR; Guiqing Lily Yao[3] BSc MSc PhD; Shihua Zhu[3] BSc MSc PhD; Richard Lilford PhD FRCOG FRCP FFPH; Jane Wardle[1] BA MPhil PhD; Christian von Wagner[1] BSc PhD

[1]Department of Epidemiology and Public Health, University College London, London, UK
[2]Centre for Medical Imaging, University College London, London, UK
[3]Public Health, Epidemiology and Biostatistics, University of Birmingham, Birmingham, UK

Corresponding author at:
Christian von Wagner
Health Behaviour Research Centre
Department of Epidemiology & Public Health
UCL, Gower Street, London
WC1E 6BT, UK
Email: c.wagner@ucl.ac.uk
Tel: +44 (0)20 7679 1940
Fax: +44 (0)20 7679 8354

Word Count: 132 in abstract; 4835 in main body

Abstract

A wide range of screening technologies is available for colorectal cancer (CRC) screening. There is demand to discover public preferences for these tests on the rationale that tailoring screening to preferences may improve uptake. This review describes a type of study (conjoint analysis) used to assess people's preferences for CRC screening tests and critically evaluates research quality using a recently published set of guidelines. Most primary studies assessed preferences for colonoscopy and faecal occult blood testing but newer technologies (e.g. capsule endoscopy) have not yet been evaluated. Although studies often adhered to guidelines, there was limited correspondence between stated preferences and actual screening behaviour. Future research should investigate how studies can go beyond the guidelines in order to improve this and also explore how test preferences may differ by important population subgroups.

Keywords: Colorectal cancer; colon cancer; rectal cancer; screening; conjoint analysis; discrete choice experiment; preferences; colonoscopy; stool test; faecal occult blood test; faecal immunochemical test; CT colonography; barium enema; flexible sigmoidoscopy

Background

Consensus guidelines have recommended colorectal cancer (CRC) screening for well over a decade [1] and Level 1 evidence has shown a consistent reduction in incidence [2-4] and disease-specific mortality [5]. However, screening uptake continues to languish behind other more established forms of screening such as for breast and cervical cancer. In England, uptake of biennial faecal occult blood testing (FOBt) was approximately 54% after the first round of screening invitations in the Bowel Cancer Screening Programme [6] compared to 70-80% uptake of breast and cervical screening [7-8]. Similar trends are apparent in other countries [9]. Given that uptake is the key determinant of the health benefits of screening, there is clearly scope to reduce incidence and mortality further by enhancing uptake. Furthermore, uptake is affected by socioeconomic status, ethnicity and gender, suggesting that the benefits of CRC screening are disproportionally distributed among the population, leading to inequities in which those who are less affluent, from a non-white ethnic background or male do not benefit as much as others [6].

A unique feature of CRC screening is the wide range of screening tests available; US guidelines currently endorse no fewer than seven modalities, each of which has received some level of empirical support for their effectiveness [10]. Test characteristics are diverse and it is unclear which is superior. For example, FOBt has a strong evidence base for reducing CRC mortality, is convenient and cheap (the test is administered at home), but its relative lack of sensitivity for pre-cancerous adenomas means there is no confirmed potential for prevention [11]. At the other extreme, colonoscopy is believed to have the most potential for both CRC prevention and early detection due to its ability to visualise the entire colorectum directly and perform therapeutic polypectomy. However, it is time-consuming for patients, technically demanding, expensive, and commonly involves sedation. Furthermore, colonoscopy has a small but well-documented morbidity [12].

Why is measuring preferences important?

There are three principal reasons why measuring preferences has been advocated in the context of CRC screening. First, knowledge of public preferences is particularly important for health care systems that offer a specific test via an organised screening programme. In this context, offering the specific test that would be preferred by the greatest number of the eligible population is expected to optimise uptake. There may be practical barriers to tailoring screening to preferences to this degree; for example, the most preferred test may be significantly less effective or have limited availability. However, knowledge of preferences has important applications since it allows estimates of screening uptake, which is necessary for calculating the cost-effectiveness of particular screening strategies.

Secondly, preferences may also be important for systems where individuals can choose from a range of available tests, for example via a process of shared-decision making with a clinician. In this case, knowledge of public preferences could be valuable to guide consultations (e.g. by allowing predictions of what a given patient is likely to prefer).

A third reason used to argue for a greater understanding of preferences for different screening modalities is that this allows public health policy makers to fulfil ethical duties of ensuring that screening is as acceptable as possible to screening participants, improving people’s satisfaction with the process.

How are preferences measured?

Revealed preference studies
One of the most intuitive and commonly used methods to establish preferences is to measure uptake after randomising individuals to receive an offer of a single test. A meta-analysis of 100 such studies [13] found that tests which may have been perceived as less invasive (e.g. stool tests or flexible sigmoidoscopy) achieved higher uptake than alternatives (e.g. colonoscopy or CT colonography [CTC]). Pooled uptake of stool tests was 47% and 42% (for FOBt and faecal immunochemical testing [FIT], respectively), 35% for flexible sigmoidoscopy, 28% for colonoscopy and 22% for CTC. A less commonly used method to establish preferences is to measure uptake after offering individuals a choice of different modalities. For example, a trial by the Multicentre Australian Colorectal-neoplasia Screening (MACS) Group offered participants a choice of FOBt, FOBt combined with flexible sigmoidoscopy, colonoscopy, or CTC [14]. The most frequently preferred test was FOBt, a finding consistent with between-subjects designs. The rationale for choosing one of these so-called revealed preference designs over the other is guided by what would be most applicable to a particular healthcare system. Revealed preference studies have the advantage of generating reliable and easily interpreted data on preferences and the relative acceptability of different tests. However, the practical and financial barriers to administering one or more screening tests are significant and so it is often expedient to use alternative methodologies.

One such approach that avoids these issues is the use of interviews and focus groups that lend themselves to qualitative methods of analysis. The open-ended nature of such designs allows considerable flexibility to explore a wide range of issues relating to screening uptake. For example, qualitative studies can be used to ask individuals with experience of screening (as it might already be available in a particular healthcare system) about aspects of tests that they found difficult or unpleasant, allowing researchers to consider methods of screening that might be preferable (e.g. [15]). Such studies can also be carried out with screening-naïve individuals and may provide participants with information on different aspects of a particular test (e.g. its ability to detect CRC or the practicalities of what it involves) and investigate aspects that are likely to dissuade or encourage them to undergo it, as well as participants' underlying reasons (e.g. [16]).

Qualitative approaches are an effective way of allowing (potential) healthcare users to raise issues that researchers had not previously considered and can yield extremely detailed data to guide future research. However, it is normally the case that subsequent quantitative research is needed to confirm any findings.

Quantitative stated preferences – Conjoint analysis
Conjoint analysis aims to quantify potential screening participants' willingness to undergo a test in a hypothetical context, without running an expensive and potentially futile trial. It also aims to quantify the importance of key features that determine users' preferences.

Conjoint analysis applies principles from health economics. It begins from the premise that the value ("utility") of a good or service can be defined in terms of several key characteristics ("attributes"). For screening, one possible attribute is the preparation before the test. This can differ between (or within) tests and the conjoint design will specify several different values with the intention of capturing the full range that might be relevant (e.g. none vs. enema vs. oral laxatives and low fibre diet).

Participants are presented with a number of scenarios involving a configuration of attributes, each with different levels, as part of a survey (see Figure 1. for an example). Participants are typically asked to make a choice between these options, which is why a common method of

conjoint analysis is known as a discrete choice experiment. Statistical modelling is then used to quantify the relative importance of different attributes. In order to estimate uptake for options defined by a given combination of attribute levels, an “opt-out” is added to the list of options, allowing participants to state a preference for no testing (option C in Figure 1).

----------Place Figure 1 here----------

The use of conjoint analysis studies has been gathering momentum for over a decade (Figure 2) and has led to the recently published set of guidelines, written by the International Society for Pharmaceconomics and Outcomes Research (ISPOR) on how to plan, conduct and analyse conjoint analysis studies [17].

Briefly, the guidelines consist of ten primary questions (each with three further secondary questions) that researchers are encouraged to take into account when designing conjoint analysis studies. They cover important issues such as study design and data collection strategies (Table 1).

----------Place Table 1 here----------

----------Place Figure 2 here----------

Review aims

We used the ISPOR guidelines to describe and evaluate key features of conjoint analysis studies of CRC screening tests. Primary studies were identified through a systematic search of several potentially relevant databases (Pubmed, CINAHL, Web of Knowledge, Embase, PsycINFO) for English language articles (see Table 2. for search terms).

Plurals were searched for, as were both United States and United Kingdom spellings. Titles and abstracts were reviewed and studies were excluded if they did not refer to conjoint analysis, colorectal cancer or screening. Studies that did not meet these exclusion criteria were read in full and reference lists were scanned to identify further relevant articles. Remaining articles were excluded if they were reviews or evaluations of conjoint analysis methodology itself.

----------Place Table 2 here----------

Overview of studies

----------Place Table 3 here----------

Expert commentary

Heterogeneity between the seven primary studies made direct comparisons difficult, despite the common theme of CRC screening. Most obviously, there was considerable variation in the specific tests that were compared and the attributes used to describe them, reflecting the large number of different testing options available. Furthermore, rapid development within the field also meant that some potential screening modalities (e.g. capsule endoscopy) have yet to be evaluated using this technique.

Four studies estimated and compared uptake between real screening tests. The two that included CTC as an option found it to be the most commonly preferred modality [21-22]. Conversely, the two studies that did not evaluate CTC [18,20] estimated flexible sigmoidoscopy to have the greatest potential to increase uptake. However, flexible sigmoidoscopy was the least preferred test in one study [21]. In addition, stool-tests (FOBt or faecal DNA testing) were least preferred in three studies [18,20,22]. This illustrates the difficulties of trying to generalise findings across studies.

Given this lack of consistency within stated preference studies, it is not surprising that similar discordance exists between results of stated preference findings and revealed preference studies. The meta-analysis described above showed that uptake of CTC compares poorly with other tests, despite the positive findings from two of these stated preference studies. Similarly, stool tests have comparatively high uptake in, despite being least preferred in three stated preference studies.

The recently published ISPOR guidelines are a useful framework for exploring specific methodological strengths and weaknesses in individual conjoint analysis studies and a starting point for identifying some of the reasons for such discrepancies.

Study evaluation using the ISPOR guidelines

Study aims and perspectives

Most primary studies were oriented around optimising uptake, either by estimating which of several specific screening tests (as defined by a particular set of attributes and levels) was preferred most or by understanding preference structures (i.e. determining the order and magnitude of importance of attributes). One early study aimed to determine whether it was possible to model utility and disutility based on screening interval (i.e. the period between screening invitations), CRC mortality risk reduction, test specificity (the proportion of people without abnormalities who are correctly categorised as such) and financial costs [24]. Other

perspectives included improving service users' satisfaction [19] or fulfilling an ethical duty to involve service users in healthcare decisions [23].

Articles were generally good in terms of defining the research question and although most did not state explicit testable hypotheses, aims were expressed with enough clarity that hypotheses could be inferred. Most studies gave a clear rationale for using conjoint analysis to address the research question.

Attributes and levels

Frequently used attributes included descriptions of what the test involves (in terms of what a screening participant would have to go through) [18,21-22], preparation [20-22] and pain or discomfort [20-22]. All studies included performance-related attributes that related to either accuracy (e.g. sensitivity, "unnecessary colonoscopies" or "accuracy" [21-23]) or CRC-mortality risk reduction [18,20]). Two studies provided information on both accuracy and CRC-mortality risk reduction [19,23].

Three studies included participants' out-of-pocket costs as an attribute in the final survey [19,22,24]. Marshall et al. included this attribute in order to estimate the effect that varying costs would have on uptake for each test under investigation. Gyrd-Hansen and Søgaard and Nayaradou et al. included cost with the rationale that it is a pertinent factor in healthcare programmes. Salkeld et al. included this attribute during piloting but removed it as an insufficient number of participants expected it to encourage their participation. Remaining studies gave no explicit rationale for not including cost.

Several attributes were defined using just two levels (e.g. location: home vs. hospital [20]; discomfort: none vs. some [21]). A single study defined an attribute with six levels (cost: 0; 100; 500; 1000; 2000; 5000 Danish krone per test [24]). All attributes were defined using this range of 2 to 6 levels.

Four studies used attributes and levels in order to define real tests [18,20-22]. Of these four, all defined real stool tests (guaiac FOBt, faecal immunochemical testing [FIT], or faecal DNA sampling), flexible sigmoidoscopy and colonoscopy. Two defined barium enema and two defined CTC. The remaining three studies used hypothetical tests that did not clearly relate to an existing screening modality, although one study described a stool test that did not specifically relate to guaiac FOBt, faecal immunochemical testing or faecal DNA testing.

Most studies stated that literature reviews, expert advice, and/or pilot work (usually qualitative) were used in order to identify and select the attributes used and their levels. However, a common weaknesses was a lack of detail on the specifics of how attributes were selected based on this evidence. Similarly, studies were mixed in terms of whether they justified particular levels. Several provided explicit references for sources used to derive levels; others did not refer to specific literature. This lack of clarity also had the effect of reducing reproducibility.

Study design

Study design refers to considerations regarding how tasks are constructed (e.g. the number of alternatives per choice task), the experimental design (such as statistical considerations used to determine which combinations of attribute levels should make up the survey questions) and how preferences are elicited (e.g. asking participants to state a preference for a single option or to rank tests from most preferred to least preferred).

Most studies added considerable value to their findings by adding an opt-out question which allowed an estimation of how preferences could be projected in terms of actual screening participation. Most of these studies presented participants with at least two profiles

(examples of possible tests) per task and an opt-out profile representing no testing; others included an opt-out option without presenting it as a profile. It is possible that presenting the opt-out question separately, after stating a preference from a choice of tests (e.g. “would you have your preferred test?”) leads to more valid results compared to presenting it alongside the test options themselves (as in Figure 1). Breaking down the decision-making process into two parts may reduce participant burden and “naysaying” where participants state a preference for no screening as a means to avoid engaging with the question.

All studies were mindful of the number of tasks they presented to participants which ranged from 1 to 16. The best studies conducted pilot studies to ensure that the data collection instrument was comprehensible and acceptable to the target population.

Most studies used discrete choice tasks in which participants were asked to express a preference for a specific option. There were two notable exceptions that used ranking or rating scales [21,24]. In both cases, these atypical choices were justified. In the case of Hawley et al., using a rating scale was anticipated to be easier for those of lower education. Gyrd Hansen and Søgaard were able to carry out a more detailed analysis by combining a ranking task with a discrete choice task. Participants were asked whether they would have one of three alternatives, which one they would prefer and which their second and third preferences were.

It was generally difficult to assess studies regarding the quality of their instructions for completing surveys since data collection instruments and interview schedules were not supplied in full. Furthermore, despite the increasing application of conjoint analysis, it is often necessary to provide explanations of concepts for readers of clinically based publications and so it may be impractical for articles to include complete experimental details. However, authors are now encouraged to make use of online resources to include appendices that enable other researchers to evaluate and replicate study methodology.

Instrument design

Instrument design refers to the actual full survey that participants are asked to complete. This includes the conjoint tasks and also any preliminary information that describes the study context and attributes for participants, qualifying questions (e.g. assessments of how difficult participants found the survey) and demographic questions (e.g. gender and age). Most studies introduced the study with an explanation of the context of screening as well as the attributes and levels under investigation.

Most studies also collected appropriate respondent information. Key demographic characteristics were recorded on participants’ ages and gender and six studies reported further results that usefully characterised their sample including race or ethnicity, and socioeconomic status, or SES proxies such as income or education.

Several studies also reported data regarding participants’ previous CRC screening history or other measures of exposure to CRC such as whether participants knew someone who had been affected. Collecting detailed information about the background of participants is useful when trying to determine the extent to which stated preferences are influenced or moderated by factors other than test characteristics.

Data collection

Data collection primarily refers to the mode by which the survey is administered and the sampling strategy. Five of the primary studies collected data from participants using pen and paper surveys (e.g. Hol et al.; Marshall et al. [18,22]) although two used a face-to-face interview [23-24]. Participants were usually identified through primary care networks or

population registries and invited to participate via mail. Sample sizes varied from 205 to 656. One study [20] assessed two subgroups of screening-naïve (N=156) and previously screened participants (N=124).

Very few studies described a power calculation and justification of sample size was uncommon. Although power calculations are challenging for conjoint analysis studies, some guides are available [25]. In addition, studies did not justify the age range of recruited participants. Although it could be assumed that studies sampling of older adults was related to the age at which risk of CRC increases (and consequently the age at which CRC screening is first offered), it was not always apparent whether individuals were eligible/recommended for screening or approaching screening age. There is evidence that perceptions of screening are influenced by whether an invitation is expected soon or in the distant future [26], which may affect the comparability of study samples.

Studies were also poor in terms of justifying the chosen method of administering the data collection instrument. Given that there are advantages and disadvantages to various approaches (e.g. online or postal questionnaires vs. interview surveys), and that the best approach will vary depending on a particular study's aims and priorities, it is important to explain why a particular approach was chosen for an individual study.

Statistical analyses

Studies were generally much better at ensuring the internal validity of their results. Almost all studies reported data on respondent characteristics and most assessed the validity of preference data by testing either 'rationality' (i.e. whether participants chose an option that was superior to another on all attributes over a comparator), willingness to trade (i.e. whether people always stated a preference for a test that was superior on a specific attribute and were unwilling to accept an inferior level of that attribute in exchange for a superior level of another attribute), and general comprehension. Most studies were good at assessing the quality of participants' responses. One study assessed willingness to trade and five used rationality tests (or a similar comprehension test); some went further by running analyses with and without irrational responders.

However there was clear scope for studies to explore these latter two issues in greater depth. For example, unwillingness to trade may be the result of participants not engaging with the question or a genuinely high valuation of a particular attribute. In addition, qualitative research has found that responses that are "irrational" under economic theory can have a rational basis [27]. Studies should aim to explore the nature of such responses, wherever practical.

Most studies could have been strengthened by analysing respondent characteristics in detail. Crucially, most studies did not assess the extent to which the sample was representative of the population of interest, meaning that external validity was difficult to gauge. One method of testing this (used by Marshall et al. [22]) is to compare demographics from the sample (e.g. the ratio of males to females) with what would be expected from a population-representative sample based on census data. Similarly, characteristics of study participants and non-participants were rarely compared.

Response rates could be extrapolated from all but one study [21] and showed wide variation from 31% [20]) to 83% [24]. It was notable that only two studies reported higher response rates among participants with experience of CRC screening compared to screening-naïve participants [18,20]. However, this is likely to be a common trend that is worth assessing since it is likely that the preferences of screening participants differ to those of non-participants (e.g. they may have fewer concerns over practical barriers since they have previously overcome them).

Results and conclusions
Most studies reported models that reflected their respective hypotheses and were able to frame their findings in the context of wider policy or decision-making. Similarly, studies were generally good at reporting relevant limitations.

In addition to the findings noted above, accuracy- and process-related attributes (e.g. what participants would have to go through) were the characteristics most commonly found to be the primary predictors of preferences (Table 4.).

The preference structures of particular sub-groups were compared in most studies. For example, studies found differences between racial and ethnic groups [21]; white participants were more likely to rate accuracy as the most important attribute compared with African Americans and Hispanic participants, and African Americans were more likely to rate test frequency as most important compared with white and Hispanic participants.

Differences were also found between previously screened and unscreened participants [18,20]. For example, screening-naïve participants valued the risk reduction of CRC more highly than those who had been screened. In addition, participants with endoscopy experience rated pain and hospital testing as more negative and positive respectively, compared to participants without endoscopy experience (full results are available in Table 5). There were also differences based on socioeconomic status [19] whereby participants with higher self-rated social class or more education valued sensitivity more than those of lower social class or less education.

----------Place Table 4 here----------

----------Place Table 5 here----------

Going beyond the ISPOR guidelines
The ISPOR guidelines offer a detailed list of important considerations which, if they are adhered to, will invariably improve the conduct of conjoint analysis studies. The studies in this review performed well on a majority of aspects although the issues regarding external validity were among the most significant shortcomings. However, these collective weaknesses were not a plausible explanation for the clear discrepancy between the findings of stated and revealed preference studies noted above. Consequently, there are likely to be several important aspects that are not addressed by the guidelines but which might need to be explored further.

Oversimplified and inaccurate information.
Conjoint analysis studies simplify the information provided to participants regarding study context, attributes and levels in order to reduce participant burden; it is a matter for researchers' judgment to determine the level of detail provided. For example, FOBt and FIT

differ in terms of the number of stool samples required per test, dietary restrictions, medicinal use, sampling and storage methods, and these factors may contribute to differences in uptake between the tests [28]. However, both may be described in conjoint analysis as simply “stool sampling”, disguising key information that could influence preferences and uptake. The diverse and multifaceted nature of screening tests inevitably means that some simplification is necessary. However, future research should explore whether the external validity of conjoint analysis studies could be improved by providing participants with more detailed information regarding the concept of screening and test attributes. It is also difficult to evaluate studies when the information presented to participants is not fully documented; data collection instruments should be made available wherever possible.

In addition to oversimplification, studies may also present information which is inaccurate, leading to unreliable estimates. For example, although Marshall et al. [22] used existing literature to select levels of sensitivity for CRC that approximated various screening tests, specificity was assumed to be 100% for all modalities. However, false positives are an appreciable issue with all radiographic and stool tests [29] and since they result in unnecessary and inconvenient endoscopy, failing to account for this will lead to overestimates of value and uptake for these tests.

Providing inaccurate information may also be an issue with revealed preference studies and population screening. For example, one trial provides evidence that participants expected a reduced-laxative preparation to be less burdensome than a powerful-laxative preparation but that experiences were more burdensome, despite the effects most likely being objectively less severe [30]. This suggests that the participants may have received information on reduced-laxative preparation that was overly positive, leading to unrealistic expectations that detracted from their experience. Although this study reported that a large majority of participants stated a willingness to repeat the experience in future, inadequate information may negatively affect people’s willingness to participate in future rounds of screening.

Heuristics and biases.

An issue overlapping with oversimplification is that study designs may inadvertently potentiate various heuristics and biases in participants, leading to less valid stated preferences. For example, studies which describe sensitivity and specificity in terms of percentages (of lesions found out of all possible lesions and false positives out of all patients without lesions) are vulnerable to “base rate neglect” [31] in which participants are not cognisant of the prevalence of lesions in the population to begin with. This means participants are unable to gauge the probability of the various outcomes of testing, leading to preferences based on ideas of prevalence that are either unrealistically high or not mentally quantified at all [32].

There are also well-recognised biases that affect how people appraise identical information, depending on whether that information is framed as a loss or a gain [33]. Consequently, an attribute like specificity may be valued differently, depending on whether it is presented as people correctly reassured with a normal result (a gain frame) or subjected to an unnecessary follow-up investigation (a loss-frame). Conjoint analysis studies cannot take account of all possible biases. However, awareness of these issues can allow researchers to take steps to minimise them, where possible, such as by using an icon array showing a given number of people undergoing a test in which the number of people experiencing each of the possible outcomes is highlighted using colour-coding [34]. As with the issues around inaccurate information, this is relevant beyond conjoint analysis and can be applied to revealed preferences studies or population screening.

Hypothetical scenarios.
One of the shortcomings of stated preference studies is that they necessarily treat intention to be screened as a proxy of actual behaviour. However, the gap between intentions and behaviour has been well documented [35]. Assessing preferences in a hypothetical context could also result in inaccurate conclusions regarding preference structures because people may value test attributes differently in a real world setting. For example, participants may state a preference based primarily on outcome characteristics of a test (e.g. sensitivity) in hypothetical situations and underestimate the extent to which process attributes would be barriers since they are not currently having to face them (e.g. out-of-pocket expenses, time expenses or discomfort). Such a bias would go some way to explain why an accurate but inconvenient modality such as CTC is often preferred over less onerous tests such as FOBt in stated preference tasks, in contrast with revealed preference studies. An important challenge for conjoint analysis will be to refine questions to make sure people consider practical barriers when stating test preferences. One possible method of exploring these issues in further detail is the use of qualitative methods within conjoint analysis studies. Gyrd-Hansen & Søgaard provide an example of how a qualitative component can be used to improve understanding of quantitative responses; this could be extended to explore how participants incorporate complex information into their decision-making. Similarly, it would be useful to apply these methods to screening participants for comparison. It may also be helpful to learn more about why research in other healthcare contexts (e.g. chlamydia screening) has found much better concordance between stated and revealed preferences [36].

Understanding sub-group differences.
Future conjoint analysis research may also investigate how test preferences vary by relevant population demographics, in particular race, socioeconomic status and gender. For example, Hawley et al. [21] found that white participants were more likely to rate accuracy as the most important attribute compared to other racial and ethnic groups, whereas other racial/ethnic groups were more likely to rate frequency as most important. Consistent with this, a study of revealed preferences by Inadomi et al. [37] found that white participants randomised to either an offer of FOBt, colonoscopy, or choice between the two, adhered more often to the more accurate colonoscopy whereas non-white invitees adhered more often to FOBt. Future research regarding subgroup preferences will be directly applicable to national screening policy.

Although the focus of this review has been CRC screening tests, all the issues described above could apply to conjoint analysis in general and so concerns may be equally applicable to studies investigating preferences for other medical technologies or healthcare delivery [38].

Five-year view

CRC screening strategies are in a state of flux globally as new tests and evidence continue to emerge and the policy landscape continues to shift. In the United Kingdom, several changes to the existing screening programme are taking place. Most notably, pilots are being established that offer one-off flexible sigmoidoscopy to the population aged 55-59, following a randomised trial that found reductions in both disease-specific incidence and mortality [4]. These results have since been supported by a trial in the United States [2] and Italy [3]. Furthermore, the existing UK programme of biennial stool testing is converting from guaiac FOBt to FIT and this change is planned to occur parallel with the incoming flexible-sigmoidoscopy programme. The upper age range at which people will be invited to have stool testing is also being extended to 74, beyond the current upper age range of 69.

In the United States, the Patient Protection and Affordable Care Act is likely to have wide-ranging impact on CRC screening. It is estimated that an additional 32-50 million individuals will receive health insurance because of the Act, substantially reducing barriers to access of screening tests [39]. In addition new colorectal tests such as capsule endoscopy are continuing to emerge and may warrant consideration for screening [40].

Measuring test preferences will continue to be an important part of meeting policy goals while these changes are implemented and conjoint analysis is likely to play a role. In order to do so effectively, conjoint analysis should be as robust as possible.

Key issues

- Screening for colorectal cancer has been shown to be effective at reducing disease specific mortality and a number of viable tests have been mooted. However, uptake of CRC screening is poor compared with other more established cancer screening strategies.

- Offering screening tests that are consistent with service users' preferences and values may improve uptake, optimising reductions in disease-specific incidence and mortality.

- Stated preference studies such as conjoint analysis are a low cost means of assessing users' preferences.

- When reviewed in accordance with recent ISPOR guidelines, previous conjoint analysis studies of colorectal cancer screening demonstrate a number of strengths and weaknesses.

- Inconsistencies between revealed and stated preferences are an important issue that needs to be resolved, requiring considerations beyond the guidelines.

- Studies that evaluate preferences for CRC screening tests using conjoint analysis should explore whether stated preferences would be more consistent with revealed preferences if a greater level of detail regarding tests and screening context were provided, strategies to reduce heuristics and bias were introduced and participants were encouraged to consider important motivators and barriers fully.

References

*Of interest

**Of considerable interest

1. Levin B, Bond J. Colorectal cancer screening: recommendations of the U.S. Preventive Services Task Force. American Gastroenterological Association. *Gastroenterology*. 111(5), 1381–1384 (1996).

2. Schoen RE, Pinsky PF, Weissfeld JL et al. Colorectal-cancer incidence and mortality with screening flexible sigmoidoscopy. *N Engl J Med*. 366(25), 2345–2357 (2012).

3. Segnan N, Armaroli P, Bonelli L et al. et al. Once-only sigmoidoscopy in colorectal cancer screening: Follow-up findings of the Italian randomized controlled trial--SCORE. *J Natl Cancer I*. 103(17), 1310–1322 (2011).

4. Atkin WS, Edwards R, Kralj-Hans I et al. Once-only flexible sigmoidoscopy screening in prevention of colorectal cancer: a multicentre randomised controlled trial. *Lancet*. 375(9726), 1624–1633 (2010).

5. Hewitson P, Glasziou PP, Irwig L, Towler B, Watson E. Screening for colorectal cancer using the faecal occult blood test, Hemoccult (Review). The Cochrane Database of Systematic Reviews (2011).

6. von Wagner C, Baio G, Raine R et al. Inequalities in participation in an organized national colorectal cancer screening programme: results from the first 2.6 million invitations in England. *Int J Epidemiol*. 40(3), 712–718 (2011).

7. The NHS Information Centre, Screening and Immunisations team. *Breast Screening Programme, England 2010-2011* (2012). http://www.ic.nhs.uk/webfiles/publications/008_Screening/brstscreen1011/brst_scr_prog_eng_2010_11_rep.pdf, accessed 26 October 2012.

8. The NHS Information Centre, Screening and Immunisations team. *Cervical screening Programme, England 2010-11* (2011). http://www.ic.nhs.uk/webfiles/publications/008_Screening/cervscreen1011/Cervical_Bulletin_2010_11_v1_1.pdf, accessed 26 October 2012.

9. Power E, Miles A, von Wagner C, Robb K, Wardle J. Uptake of colorectal cancer screening: system, provider and individual factors and strategies to improve participation. *Future Oncol*. 5(9), 1–18 (2009).

10. Smith RA, Cokkinides V, Brooks D, Saslow D, Shah M, Brawley OW. Cancer Screening in the United States. *CA - Cancer J Clin*. 61(1), 8–30 (2011).

11. Scholefield JH, Moss SM, Mangham CM, Whynes DK, Hardcastle JD. Nottingham trial of faecal occult blood testing for colorectal cancer: a 20-year follow-up. *Gut*. 61(7), 1036–1040 (2012).

12. Weller D, Alexander F, Orbell S et al. *Evaluation of the UK Colorectal cancer screening pilot: Final Report* (2003). http://www.cancerscreening.nhs.uk/bowel/pilot-evaluation.html, accessed 26 October 2012.

13. **Khalid-de Bakker C, Jonkers D, Smits K, Mesters I, Masclee A, Stockbrügger R. Participation in colorectal cancer screening trials after first-time invitation: a systematic review. *Endoscopy*. 43(12), 1059–1086 (2011). *This meta-analysis compares uptake (revealed preferences) of CRC screening tests.*

14. The Multicentre Australian Colorectal-neoplasia Screening Group. A comparison of colorectal neoplasia screening tests: a multicentre community-based study of the impact of consumer choice. *Med J Australia*. 184(11), 546–550 (2006).

15. Chapple A, Ziebland S, Hewitson P, McPherson A. What affects the uptake of screening for bowel cancer using a faecal occult blood test (FOBt): a qualitative study. *Soc Sci Med*. 66(12), 2425–2435 (2008).

16. Ghanouni A, Smith SG, Halligan S et al. Public perceptions and preferences for CT colonography or colonoscopy in colorectal cancer screening. *Patient Educ Couns*. 89(1) 116-121 (2012).

17. **Bridges JFP, Hauber AB, Marshall D et al. Conjoint analysis applications in health - a checklist: A report of the ISPOR Good Research Practices for Conjoint Analysis Task Force. *Value Health*. 14(4), 403–413 (2011). *Guidelines for conducting conjoint analysis studies.*

18. *Hol L, de Bekker-grob EW, van Dam L et al. Preferences for colorectal cancer screening strategies: a discrete choice experiment. *Brit J Cancer*. 102(6), 972–980 (2010). *Conjoint analysis study of preferences for real CRC screening tests.*

19. *Nayaradou M, Berchi C, Dejardin O, Launoy G. Eliciting population preferences for mass colorectal cancer screening organization. *Med Decis Making*. 30(2), 224–233 (2010). *Conjoint analysis study of preferences for hypothetical CRC screening tests.*

20. *van Dam L, Hol L, de Bekker-grob EW et al. What determines individuals' preferences for colorectal cancer screening programmes? A discrete choice experiment. *Eur J Cancer*. 46(1), 150–159 (2010). *Conjoint analysis study of preferences for real CRC screening tests.*

21. *Hawley ST, Volk RJ, Krishnamurthy P, Jibaja-Weiss M, Vernon SW, Kneuper S. Preferences for colorectal cancer screening among racially/ethnically diverse primary care patients. *Med Care*. 46(9 Suppl 1), S10–S16 (2008). *Conjoint analysis study of preferences for real CRC screening tests.*

22. *Marshall DA, Johnson FR, Phillips KA, Marshall JK, Thabane L, Kulin NA. Measuring patient preferences for colorectal cancer screening using a choice-format survey. *Value Health*. 10(5), 415–430 (2007). *Conjoint analysis study of preferences for real CRC screening tests.*

23. *Salkeld G, Solomon M, Short L, Ryan M, Ward JE. Evidence-based consumer choice: a case study in colorectal cancer screening. *Aust NZ J Publ Health*. 27(4), 449–455 (2003). *Conjoint analysis study of preferences for hypothetical CRC screening tests.*

24. *Gyrd-Hansen D, Søgaard J. Analysing public preferences for cancer screening programmes. *Health Econ*.10(7), 617 – 634 (2001). *Conjoint analysis study of preferences for hypothetical CRC screening tests.*

25. Johnson R, Orme B. *Getting the most from CBC* (2003). Sawtooth Software Research Paper Series. http://www.sawtoothsoftware.com/download/techpap/cbcmost.pdf, accessed 26 October 2012.

26. Von Wagner C, Semmler C, Power E, Good A. What matters when deciding whether to participate in colorectal cancer screening? The moderating role of time perspective. *J Appl Biobehav Res*. 15(1), 20–30 (2010).

27. Ryan M, Watson V, Entwistle V. Rationalising the "irrational": A think aloud study of discrete choice experiment responses. *Health Econ*. 18(3), 321–336 (2009).

28. Vart G, Banzi R, Minozzi S. Comparing participation rates between immunochemical and guaiac faecal occult blood tests: A systematic review and meta-analysis. *Prev Med*. 55(2), 87–92 (2012).

29. Whitlock EP, Lin J, Liles E et al. Screening for colorectal cancer: An updated systematic review. *Evidence Syntheses*. 65(1) (2008).

30. 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*. 61(11), 1152-1559 (2012).

31. Barbey AK, Sloman SA. Base-rate respect: From ecological rationality to dual processes. *Behav Brain Sci*. 30(3), 241–297 (2007).

32. Fagerlin A, Zikmund-Fisher BJ, Ubel PA. How making a risk estimate can change the feel of that risk: shifting attitudes toward breast cancer risk in a general public survey. *Patient Educ Couns*. 57(3), 294–299 (2005).

33. Tversky A, Kahneman D. The framing of decisions and the psychology of choice. *Science*. 211(4481), 453–458 (1981).

34. Spiegelhalter D, Pearson M, Short I. Visualizing uncertainty about the future. *Science*. 333(6048), 1393–1400 (2011).

35. Power E, van Jaarsveld CHM, McCaffery K, Miles A, Atkin W, Wardle J. Understanding intentions and action in colorectal cancer screening. *Annals of Behav Med*. 35(3), 285–294 (2008).

36. Ryan M, Watson V. Comparing welfare estimates from payment card contingent valuation and discrete choice experiments. Health Econ. 18(4), 389–401 (2009).

37. Inadomi JM, Vijan S, Janz NK et al. Adherence to colorectal cancer screening. A randomized clinical trial of competing strategies. *Arch Intern Med*. 172(7), 575–582 (2012).

38. de Bekker-Grob EW, Ryan M, Gerard K. Discrete choice experiments in health economics: A review of the literature. *Health Econ*. 21(2), 145–72 (2010).

39. Sheen E. Supreme court review of the affordable care act: The future of health care reform and practice of gastroenterology. *Digestive Dis Sci*. 57(7), 1735–1741 (2012).

40. Sieg A. Colon capsule endoscopy compared with conventional colonoscopy for the detection of colorectal neoplasms. *Expert Rev Med Devices*. 8(2), 257–261 (2011).

Financial disclosure / acknowledgments

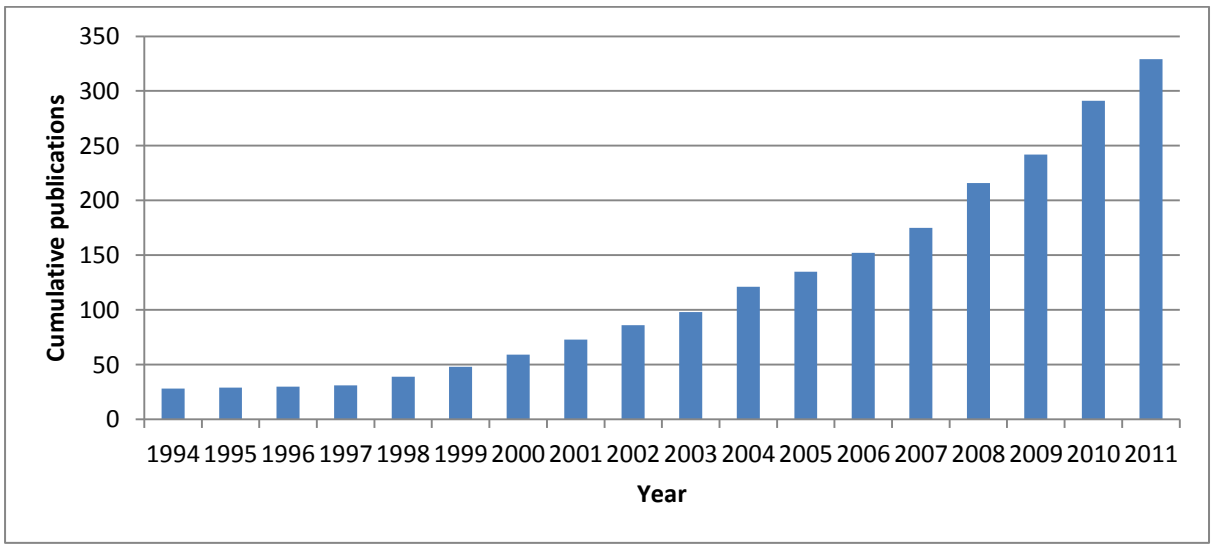
This article presents independent research commissioned by the National Institute for Health Research (NIHR) under its Programme Grants for Applied Research funding scheme (RP-PG-0407-10338). This work was undertaken at UCL and UCLH, who receive a proportion of funding from the NIHR Biomedical Research Centre funding scheme. The views expressed in this article are those of the authors and not necessarily those of the NHS, the NIHR or the Department of Health.

Figure 1. An example choice scenario, reproduced from van Dam et al. [20]

Choice options:	A	B	C
Preparation:	Enema, No fasting	Drinking of 0.75 liters of fluid, 12 hours fasting	None
Location:	At home	Hospital	None
Pain:	None	Mild pain	None
Risk of complications:	None	Small	None
The chance of dying from colon cancer decreases from:	3% to 1.8%	3% to 1.2%	3%
In the following 10 years you will undergo the test:	5x	2x	0x
Duration:	30 minutes	60 minutes	None

Suppose screening for colon cancer is introduced.
Which test do you prefer? *(Fill in: A, B or C)*

Colour Figure 2. Cumulative total of PubMed articles containing the term “conjoint analysis”.



Black & White Figure 2. Cumulative total of PubMed articles containing the term “conjoint analysis”.

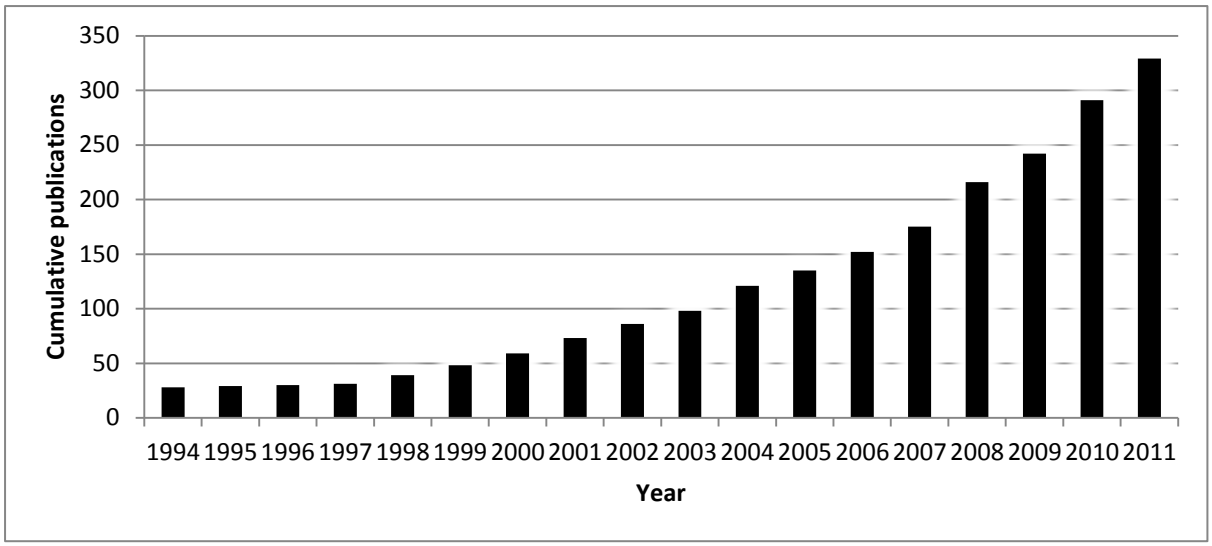


Table 1. Reproduced from Bridges et al. [17]

Table 1 – A checklist for conjoint analysis applications in health care.

1. Was a well-defined research question stated and is conjoint analysis an appropriate method for answering it?
   - 1.1 Were a well-defined research question and a testable hypothesis articulated?
   - 1.2 Was the study perspective described, and was the study placed in a particular decision-making or policy context?
   - 1.3 What is the rationale for using conjoint analysis to answer the research question?
2. Was the choice of attributes and levels supported by evidence?
   - 2.1 Was attribute identification supported by evidence (literature reviews, focus groups, or other scientific methods)?
   - 2.2 Was attribute selection justified and consistent with theory?
   - 2.3 Was level selection for each attribute justified by the evidence and consistent with the study perspective and hypothesis?
3. Was the construction of tasks appropriate?
   - 3.1 Was the number of attributes in each conjoint task justified (that is, full or partial profile)?
   - 3.2 Was the number of profiles in each conjoint task justified?
   - 3.3 Was (should) an opt-out or a status-quo alternative (be) included?
4. Was the choice of experimental design justified and evaluated?
   - 4.1 Was the choice of experimental design justified? Were alternative experimental designs considered?
   - 4.2 Were the properties of the experimental design evaluated?
   - 4.3 Was the number of conjoint tasks included in the data-collection instrument appropriate?
5. Were preferences elicited appropriately, given the research question?
   - 5.1 Was there sufficient motivation and explanation of conjoint tasks?
   - 5.2 Was an appropriate elicitation format (that is, rating, ranking, or choice) used? Did (should) the elicitation format allow for indifference?
   - 5.3 In addition to preference elicitation, did the conjoint tasks include other qualifying questions (for example, strength of preference, confidence in response, and other methods)?
6. Was the data collection instrument designed appropriately?
   - 6.1 Was appropriate respondent information collected (such as sociodemographic, attitudinal, health history or status, and treatment experience)?
   - 6.2 Were the attributes and levels defined, and was any contextual information provided?
   - 6.3 Was the level of burden of the data-collection instrument appropriate? Were respondents encouraged and motivated?
7. Was the data-collection plan appropriate?
   - 7.1 Was the sampling strategy justified (for example, sample size, stratification, and recruitment)?
   - 7.2 Was the mode of administration justified and appropriate (for example, face-to-face, pen-and-paper, web-based)?
   - 7.3 Were ethical considerations addressed (for example, recruitment, information and/or consent, compensation)?
8. Were statistical analyses and model estimations appropriate?
   - 8.1 Were respondent characteristics examined and tested?
   - 8.2 Was the quality of the responses examined (for example, rationality, validity, reliability)?
   - 8.3 Was model estimation conducted appropriately? Were issues of clustering and subgroups handled appropriately?
9. Were the results and conclusions valid?
   - 9.1 Did study results reflect testable hypotheses and account for statistical uncertainty?
   - 9.2 Were study conclusions supported by the evidence and compared with existing findings in the literature?
   - 9.3 Were study limitations and generalizability adequately discussed?
10. Was the study presentation clear, concise, and complete?
    - 10.1 Was study importance and research context adequately motivated?
    - 10.2 Were the study data-collection instrument and methods described?
    - 10.3 Were the study implications clearly stated and understandable to a wide audience?

Table 2. Search strategy

Retrieved articles contained these words:		
Conjoint analysis, discrete choice, discrete ranking **or** stated preference		
And		
Colonoscopy, CT colonography, barium enema, endoscopy, flexible sigmoidoscopy, faecal occult, faecal immunochemical **or** faecal DNA	**Or**	Bowel, colon, colorectal **or** rectal
And		
Cancer, neoplasm **or** malignancy		
And		
Screening		

Table 3. Description of conjoint analysis studies included in this review.

Study	Country	Aims	Study sample
Hol et al. 2010 [18]	Holland	"Determine individuals' preferences and to predict the uptake of CRC screening programmes with various screening tests, and the relative importance of different test characteristics for these preferences in an average-risk population [and] to identify differences in preference structures among subgroups in the population."	One group of screening-naïve individuals aged 50-74. One group of participants aged 50-74 who had been screened as part of a screening trial.
Nayaradou et al. 2010 [19]	France	"Our study aims at empirically identifying population preferences among the different characteristics of a mass CRC screening programme."	Members of the general population aged 50-74.
van Dam et al. 2010 [20]	Holland	"Determine how procedural characteristics of CRC screening tests determine preferences for participation, and how individuals weigh these against the expected health benefits from CRC screening."	One group of screening-naïve 50-74 year olds. One group of participants in a screening trial.
Hawley et al. 2008 [21]	United States	"The specific research objectives were (1) to describe preferences for CRC screening tests among African American (AA), Hispanic, and white primary care patients; and (2) to evaluate factors associated with preferences for guideline-recommended (eg, FOBT, SIG, COL, BE) and new technology (eg, FIT, V-COL) screening options in this diverse population."	African-American, Hispanic and White primary care patients aged 50-80 with no personal or family history of CRC.
Marshall et al. 2007 [22]	Canada	"The objective of the current study was to measure and quantify Canadian preferences for various CRC screening tests and for a "no-screening" option using utility-based methods."	Primary care patients aged 40-60 with no history of CRC, who were not institutionalised and understood sufficient English.
Salkeld et al. 2003 [24]	Australia	"The aim of the present study was to assess community decisions about CRC screening based on the trial evidence on harms and benefits."	Eligible CRC screening population aged 50-70 years who had completed an earlier rating exercise as part of the survey design.
Gyrd-Hansen & Søgaard 2001 [24]	Denmark	"This paper seeks to establish a representative utility function for cancer screening programmes which incorporates the utility and disutility associated with intangible effects [harms and benefits of screening]...The analysis is basically explorative, and seeks to establish whether public preferences exist for attributes associated with participation in cancer screening programmes."	Population aged 50.

Table 4. Summary results of conjoint analysis studies included in this review.

Study	Main findings
Hol et al.[18]	Significant predictors of preferences were type of screening test, screening interval and CRC mortality risk reduction. 5 or 10 yearly FS screening were both rated as preferable compared to 10 yearly colonoscopy or yearly FOBt.
Nayaradou et al. [19]	Mortality reduction, sensitivity, monetary cost and process were significant predictors of preferences.
van Dam et al. [20]	Pain, risk of complications, screening location, preparation, duration of procedure, screening interval and risk reduction of CRC-related death all influenced preferences. An extra 1% risk reduction in the CRC mortality was rated as equivalent to an additional 10 minutes of test duration, 5% was equivalent to a small risk of adverse effects, 10% to mild pain, 10% to an enema, 12% to oral preparation including fasting and 32% to preparation with additional fluid and fasting.
Hawley et al. [21]	37% of participants preferred colonoscopy, 31% preferred FOBt, 15% preferred barium enema and 9% preferred FS. When newer tests were included (CTC and FIT), these were rated more highly than existing tests. The most important attribute was what the test involved, followed by accuracy, frequency, discomfort and preparation.
Marshall et al. [22]	CTC was the most preferred test; the order of preferences for the remaining tests was colonoscopy, barium enema, flexible sigmoidoscopy and faecal DNA testing. 28.9% of participants consistently stated that they would prefer no screening. Attributes related to accuracy were more important than attributes related to process.
Salkeld et al. [23]	32% made decisions based only on the attribute of the number of CRC deaths prevented. 12% always chose no screening. Remaining participants would accept 853 false positives for one life saved from CRC death.
Gyrd-Hansen & Søgaard [24]	Increasing cost and number of screening tests over a lifetime had a negative impact on preferences. Increasing risk reduction of CRC-mortality had a positive impact on preferences. 34% preferred no screening.

Table 5. Summary of main subgroup differences between previously screened and unscreened participants (and between participants previously screened with different tests)

Hol et al. 2010 [18]	**Screening-naïve participants vs. previously screened participants:** The rationality test was less likely to be passed by screening-naïve participants compared to previously screened participants. All screening tests were preferred less by screening-naïve participants than previously screened participants. A longer screening interval was valued more highly by screening-naive participants than previously screened participants. **Participants with experience of endoscopy vs. participants without (including participants with experience of FOBT):** Preferences were more positive towards colonoscopy in participants with experience of FS or colonoscopy. Preferences were more positive towards FS and colonoscopy in participants with experience of FS than those with experience of FOBT.
van Dam et al. 2010 [20]	**Screening-naïve participants vs. CRC screening participants:** Risk-reduction of CRC death was valued more highly by screening-naïve participants than CRC screening participants. **Participants screened with flexible-sigmoidoscopy vs. participants screened with FOBt:** A hospital-based test was preferred by FS screening participants. A home-based test was preferred by FOBt screening participants. A longer screening interval and a greater risk-reduction were valued more highly by FS screening participants than FOBt screening participants. **Participants with previous endoscopy experience vs. participants without:** Pain was a more important attribute to participants with experience of endoscopy than those without. Hospital testing was positively associated with preferences for those with experience but a negative effect on participants without experience. Risk-reduction was valued less highly by participants without experience than those with experience.