




# BMJ Open Establishing thresholds for important benefits considering the harms of screening interventions

Lise Mørkved Helsingen <sup>1,2</sup> Linan Zeng,<sup>3,4</sup> Reed Alexander Siemieniuk <sup>3</sup>,  
Lyubov Lytvyn,<sup>3</sup> Per Olav Vandvik,<sup>5,6</sup> Thomas Agoritsas <sup>3,7</sup> Michael Bretthauer,<sup>1,2</sup>  
Gordon Guyatt<sup>3</sup>

**To cite:** Helsingen LM, Zeng L, Siemieniuk RA, *et al.* Establishing thresholds for important benefits considering the harms of screening interventions. *BMJ Open* 2020;**10**:e037854. doi:10.1136/bmjopen-2020-037854

► Prepublication history and additional material for this paper is available online. To view these files, please visit the journal online (<http://dx.doi.org/10.1136/bmjopen-2020-037854>).

Received 18 February 2020

Revised 10 October 2020

Accepted 20 October 2020



► <http://dx.doi.org/10.1136/bmjopen-2020-037854>



© Author(s) (or their employer(s)) 2020. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

For numbered affiliations see end of article.

## Correspondence to

Dr Lise Mørkved Helsingen; [lismhe@medisin.uio.no](mailto:lismhe@medisin.uio.no)

## ABSTRACT

**Context and objective** Standards for clinical practice guidelines require explicit statements regarding how values and preferences influence recommendations. However, no cancer screening guideline has addressed the key question of what magnitude of benefit people require to undergo screening, given its harms and burdens. This article describes the development of a new method for guideline developers to address this key question in the absence of high-quality evidence from published literature.

**Summary of method** The new method was developed and applied in the context of a recent BMJ Rapid Recommendation clinical practice guideline for colorectal cancer (CRC) screening. First, we presented the guideline panel with harms and burdens (derived from a systematic review) associated with the CRC screening tests under consideration. Second, each panel member completed surveys documenting their views of expected benefits on CRC incidence and mortality that people would require to accept the harms and burdens of screening. Third, the panel discussed results of the surveys and agreed on thresholds for benefits at which the majority of people would choose screening. During these three steps, the panel had no access to the actual benefits of the screening tests. In step four, the panel was presented with screening test benefits derived from a systematic review of clinical trials and microsimulation modelling. The thresholds derived through steps one to three were applied to these benefits, and directly informed the panel's recommendations.

**Conclusion** We present the development and application of a new, four-step method enabling incorporation of explicit and transparent judgements of values and preferences in a screening guideline. Guideline panels should establish their view regarding the magnitude of required benefit, given burdens and harms, before they review screening benefits and make their recommendations accordingly. Making informed screening decisions requires transparency in values and preferences judgements that our new method greatly facilitates.

## INTRODUCTION

The last two decades have seen an enormous increase in clinicians' reliance on clinical practice guidelines, including cancer screening. In parallel, standards for

guideline trustworthiness have been developed to ensure that guidelines are based on the best available evidence and are developed through a transparent process, including explicit statements of how the values and preferences of the target population influenced the recommendations.<sup>1-3</sup>

Over 110 guideline development organisations have adopted the Grading of Recommendations Assessment, Development and Evaluation (short GRADE) approach that provides a framework for judging the quality of evidence and going from evidence to clinical recommendations.<sup>4</sup> One of the key factors emphasised by GRADE when going from evidence to recommendations is the target population's values and preferences.<sup>5</sup>

GRADE guidance suggests that guideline panels recommend in favour of intervention when they believe that the majority of fully informed individuals would choose that intervention and against the intervention when they believe the majority would decline.<sup>4</sup> Both the GRADE approach and current standards for trustworthy guidelines are clear that guidelines should state explicitly how values and preferences judgements have influenced a recommendation, but do not provide detailed guidance of exactly how this should be achieved.<sup>1-3</sup>

Cancer screening guidelines have apparently suffered from this lack of guidance: they are both extremely variable in their articulation of underlying values and preferences and exhibit major deficiencies. A survey of 68 eligible cancer screening guidelines revealed that only 25 included a statement regarding the tradeoff between screening benefits versus harms and burdens (14 guidelines), or a statement of direction of the net effect (defined as benefits minus harms or burdens) (13 guidelines).<sup>6</sup>

Perhaps, as a result, cancer screening guidelines often vary in their recommendations, and often appear to ignore that screening programmes have limited uptake: some individuals are enthusiastic, while others are uninterested.<sup>7 8</sup> Thus, screening recommendations are likely to be preference-sensitive.<sup>9–11</sup> Panels addressing screening interventions should acknowledge this preference sensitivity. Implementing this GRADE principle should therefore involve specifying the magnitude of benefit required for people to accept the burdens and harms associated with screening. Despite this, not a single guideline panel addressing screening has to date been explicit and transparent in specifying their inferences regarding values and preferences.<sup>6</sup>

The presentation that follows presents a novel approach that, if applied, correct this serious limitation. The new method was developed in the context of a recent BMJ Rapid Recommendation guideline for colorectal cancer (CRC) screening,<sup>12</sup> which demonstrated the feasibility of using thresholds of required benefit to inform recommendations for clinical practice. The principles for establishment of benefit thresholds are described in [box 1](#).

### Guideline panel and tasks

Motivated by recent reports from randomised trials on long-term effects of sigmoidoscopy screening,<sup>13–15</sup> a multidisciplinary panel made recommendations for people aged 50–79 years over a timeframe of 15 years regarding four CRC screening options: faecal immunochemical testing (FIT) every year and every 2 years, a single sigmoidoscopy and a single colonoscopy—versus no screening.<sup>12</sup> In the BMJ Rapid Recommendations, the guideline panel takes an individual perspective, rather than a healthcare systems perspective. Cost-effectiveness and other contextual factors of relevance to healthcare systems—such as national screening programmes—were therefore not included in the process of moving from evidence to recommendations.<sup>16</sup>

The guideline panel included 22 people recruited according to the Institute of Medicine's standards for trustworthy guidelines and standards set by the BMJ Rapid Recommendations project, which takes a strict stance on conflicts of interest.<sup>1</sup> The panel consisted of members of the public with experience with CRC screening, clinicians, CRC screening experts and guideline methodologists. No panel member had financial conflicts of interest related to the guideline topic. Panel members with professional or intellectual conflicts were minimised, and those that existed were reported during the panel conferences and in the guideline paper.<sup>12</sup> The panel followed standards for trustworthy guidelines and used GRADE methods.

The panel defined the outcomes of interest and issued a systematic review and network meta-analysis (NMA) of clinical trials, and a complementary microsimulation study for current estimates for benefits, harms and burdens associated with screening.<sup>17 18</sup> A systematic review assessing peoples' values and preferences found limited evidence considering all four screening options

### Box 1 Principles for establishment of benefit thresholds

- ▶ The GRADE system for developing practice guidelines defines strong recommendations as those in which all or almost all fully informed individuals would make the same choice ('just do it').<sup>4</sup> GRADE distinguishes these from weak or conditional recommendations in which the majority of fully informed individuals would choose the suggested course of action, but a substantial minority would not ('it depends').<sup>4</sup>
- ▶ The GRADE definition of strong and weak recommendations defines the challenge guideline panels are facing when making screening recommendations: what would well-informed individuals choose when presented with the option of undergoing, or not undergoing, a screening intervention?
- ▶ A more specific framing of this question would picture individuals presented with the harms and burdens of a screening intervention and ask them, given those burdens and harms, what magnitude of benefit they would require to undergo screening.
- ▶ The term threshold refers to the smallest benefit individuals would require to make a choice to undergo screening, given the associated burdens and harms. If the benefit is lower than a person's threshold, that person would decline screening. If above, the person would undergo screening.
- ▶ According to GRADE, the distribution of peoples' thresholds should determine the direction and strength of recommendations. A narrow distribution of thresholds requiring small benefit would justify a recommendation for screening for all or almost all individuals. A narrow distribution of thresholds requiring large benefit would justify a recommendation against screening. A wide distribution of thresholds, with large variation in the importance people place on the benefits and harms associated with screening, would lead to weak recommendations for or against screening.
- ▶ In addition to values and preferences of required benefits, individual risk of developing and dying of cancer will likely influence an individual's screening decision: Two individuals who share the same threshold for screening benefit, for instance, a requirement that screening reduces lifetime risk of cancer by at least 1% (10 in 1000), may make different choices about screening if their risk of developing cancer is 20% or 2%. Given that relative risk reductions by screening are similar across disease risk,<sup>26–28</sup> an individual with a low risk of developing cancer might fall below the threshold and an individual with a high risk above the threshold. The former may decline screening, and the latter may undergo screening.
- ▶ Understanding these principles and acknowledging that values and preferences differ across individuals, guideline panels should make their best estimates of their target population's distribution of thresholds of required benefit, given the downsides of a particular screening intervention.

addressed by the guideline. The included studies showed large variability in preferences for different screening options across different study populations, when considering factors such as screening test and interval, required preparation before screening and reduction in CRC mortality risk.<sup>12</sup> Participants in the studies summarised were only presented with benefits from screening as relative risk reductions, and the review provided no information that directly could inform the threshold of absolute benefit required to undergo screening compared with no screening, or for choosing one screening option over

another.<sup>12</sup> Had that information been available, the exercise described in this paper would not have been necessary. Because, however, the studies provided no direct evidence regarding how much benefit people would require to be willing to undergo screening, the panel developed a new, transparent method for making estimates based on their experience.<sup>12</sup>

### New method for recommendation development

We developed and applied the following stepwise process for evidence presentation and development of recommendations:

1. *Presentation of burdens and harms of screening.* First, we assessed the harms and burdens associated with each screening method through our systematic review and complementary literature searches, and presented the results to the guideline panel. At this time, the panel had no access to the estimates for benefits of screening from the systematic review and NMA or the complementary microsimulation study.
2. *Panel surveys to establish thresholds for screening benefits.* Second, the panel addressed the issue of, given the harms and burdens information they had received, the magnitude of benefit people would require to undergo screening. For this purpose the panel completed three consecutive surveys documenting their views of the absolute reduction in CRC incidence or mortality the majority of people would require to tolerate the burdens and harms of screening.
3. *Consensus on thresholds for screening benefits.* Third, panelists came to a consensus regarding benefit thresholds through discussion of the survey results at video panel meetings, and still without knowledge of the benefits of screening from the systematic review of clinical trials and the microsimulation study.
4. *Recommendations based on benefit thresholds.* Finally, the panel got access to and reviewed the estimates of actual benefits of the four screening options from the systematic review of clinical trials and the microsimulation modelling.<sup>17 18</sup> The panel then formulated recommendations for screening based on the thresholds (derived in step 3) and the absolute benefits and harms.<sup>12</sup>

### Process and application

Figure 1 presents burdens and harms the panel reviewed in step 1 of the process.<sup>19</sup> The number of people under consideration is 1000, and the timeframe for the evidence is 15 years. All evidence was judged low certainty due to modelling.<sup>18</sup> The estimates include screening and additional investigation as a result of an initial positive screening test. For step 2 of the process, the panel responded to surveys considering thresholds of screening benefit that most fully informed people would require to undergo screening, given these associated burdens and harms. The number of panellists who responded to survey questions differed from 16 to 21. The three members of the public with CRC screening experience responded to all surveys. All panellists, regardless of survey response,

stated their views in panel meetings and agreed on the final recommendations. The methods co-chair of the guideline (LMH) administered the surveys and did not participate in voting.

### Survey 1

Table 1 shows examples of questions from the first survey, with votes of panel members who chose particular options. The survey focused on colonoscopy and on CRC mortality reductions (table 1A), with additional questions related to the other three screening options (table 1B). Possible thresholds presented to the panel were CRC mortality reductions of 1, 10, 20 or 30 per 1000. The threshold that proved most informative was 10 per 1000 (the value at which approximately half the panellists thought the majority of people would accept screening and half thought that the majority would decline). For the higher thresholds, panellists thought most patients would choose screening.

At a panel consensus videoconference, panellists agreed that the majority of individuals would accept screening for a reduction in CRC mortality of at least 10 in 1000 (1%). The panel also concluded that a reduction of 10 per 1000 (1%) in CRC incidence would be sufficient for a majority to choose CRC screening. The panel thought this would be true for any of the four screening options (FIT yearly, FIT every 2 years, sigmoidoscopy or colonoscopy).

### Survey 2

We then conducted a second survey focusing on possible choices of one screening option over another. Online supplemental appendix table 1 presents all results of the second survey.

Feedback from the first survey led us to clarify the response options as follows:

‘All or almost all’ means over 90% would choose this option.

‘Most’ means 75%–90% would choose this option.

‘Majority’ means 51%–74% would choose this option.

The panel agreed that an important difference of one invasive screening option over another would be 5 per 1000 for CRC mortality or incidence reduction; if results were similar (less than 5 per 1000 difference in CRC mortality or incidence reduction) almost all would choose FIT over either of the invasive tests (colonoscopy or sigmoidoscopy).

The results of the second survey and the ensuing panel discussion made it clear that most panel members felt people would perceive the burdens and harms of FIT as considerably less than those of sigmoidoscopy or colonoscopy. The panel, therefore, decided to revisit the FIT threshold decision from the first survey.

### Survey 3

In a third survey, we addressed limitations of the first survey. First, we used smaller gradients of benefit (we included thresholds of 1, 2, 3, 4 and so on to 10 per 1000); second, we avoided bias related to presentation by using a



## Among a 1000 people

	No screening	FIT every two years	FIT every year	Sigmoidoscopy	Colonoscopy
Procedure related mortality 15 years	0 per 1000 certainty →	No apparent difference ⊙⊙⊙⊙	No apparent difference ⊙⊙⊙⊙	No apparent difference ⊙⊙⊙⊙	No apparent difference ⊙⊙⊙⊙
Gastrointestinal perforation or bleeding 15 years	0 per 1000 certainty →	1 more ⊙⊙⊙⊙	1 more ⊙⊙⊙⊙	1 more ⊙⊙⊙⊙	2 more ⊙⊙⊙⊙
Other gastrointestinal adverse events 15 years	0 per 1000 certainty →	1 more ⊙⊙⊙⊙	1 more ⊙⊙⊙⊙	1 more ⊙⊙⊙⊙	2 more ⊙⊙⊙⊙
Cardiovascular adverse events 15 years	0 per 1000 certainty →	1 more ⊙⊙⊙⊙	1 more ⊙⊙⊙⊙	1 more ⊙⊙⊙⊙	1 more ⊙⊙⊙⊙
Screening tests needed 15 years	0 per 1000 certainty →	5134 more ⊙⊙⊙⊙	8715 more ⊙⊙⊙⊙	1000 more ⊙⊙⊙⊙	1000 more ⊙⊙⊙⊙
One or more colonoscopies needed 15 years	0 per 1000 certainty →	246 more ⊙⊙⊙⊙	347 more ⊙⊙⊙⊙	237 more ⊙⊙⊙⊙	1000 more ⊙⊙⊙⊙
Two or more colonoscopies needed 15 years	0 per 1000 certainty →	83 more ⊙⊙⊙⊙	101 more ⊙⊙⊙⊙	86 more ⊙⊙⊙⊙	105 more ⊙⊙⊙⊙

**Figure 1** Burdens and harms of colorectal cancer screening for a 3% (30 per 1000) 15-year risk of colorectal cancer. Multilayered presentation available at MAGIC website (<http://magicproject.org/190220dist/#/!sof/data-set/crc-30-per-1000>). **Definitions:** Gastrointestinal perforation and bleeding: perforations, gastrointestinal bleeding or transfusions requiring hospitalisation or an emergency department visit within 30 days after screening/work-up or surveillance colonoscopy. Other gastrointestinal adverse events: paralytic ileus, nausea, vomiting and dehydration or abdominal pain requiring emergency department visit or hospitalisation within 30 days after screening/work-up or surveillance colonoscopy. Cardiovascular adverse events: myocardial infarction or angina, arrhythmias, congestive heart failure, cardiac or respiratory arrest, syncope, hypotension or shock requiring hospitalisation or an emergency department visit within 30 days after screening, work-up or surveillance colonoscopy.

‘ping-pong’ approach (the order of presentation was 1, 10, 9, 2, 8, 3 and so on). The threshold at which approximately half the participants thought the majority would choose screening and half thought they would decline proved to be a mortality reduction and/or incidence reduction of 5 per 1000 (box 2). The panel, therefore, set thresholds for recommending in favour of screening of a reduction in either CRC mortality or incidence of 10 in 1000 for colonoscopy or sigmoidoscopy, and 5 in 1000 for FIT. Online supplemental appendix table 2 presents all results of the third survey.

### Application

Informed by the microsimulation model figure 2 presents the benefits in CRC mortality and incidence reduction achieved by each screening strategy for patients with 15-year CRC risk of 30 per 1000 (corresponding to a CRC

mortality of 9 per 1000). The results match closely with the thresholds the panel generated: for CRC incidence, reductions for sigmoidoscopy of 8 in 1000 and for colonoscopy of 10 in 1000; for mortality, reductions of 6 for yearly FIT and 5 for FIT every 2 years.<sup>18 19</sup> This led to the panel’s recommendation for screening for individuals with a baseline risk of 30 per 1000 over 15 years, and for no screening for those with a lower risk. Literature identified a wide variability in values and preferences among the target population, which were one of the reasons why the panel did not make any strong recommendations for or against screening.<sup>5 12</sup>

### DISCUSSION

In a guideline addressing CRC screening, we developed and demonstrated the feasibility of a method, through

**Table 1** (A) A single colonoscopy versus no screening. (B) A single colonoscopy versus a single sigmoidoscopy or faecal testing (annual or biennial for 15 years). Number of votes given for each response alternative.

(A)	(B)
<p>A patient who is screened with colonoscopy, has a 1 in 1000 (0.1%) lower risk of dying from colorectal cancer at 15 years. How would patients view such benefits?</p> <ul style="list-style-type: none"> <li>▶ Everyone would choose screening: 0</li> <li>▶ Most would choose screening: 1</li> <li>▶ A majority would choose screening: 3</li> <li>▶ A majority would decline screening: 30</li> <li>▶ Most would decline screening: 5</li> <li>▶ All would decline screening: 2</li> </ul> <p>A patient who is screened with colonoscopy, has a 10 in 1000 (1%) lower risk of dying from colorectal cancer at 15 years. How would patients view such benefits?</p> <ul style="list-style-type: none"> <li>▶ Everyone would choose screening: 0</li> <li>▶ Most would choose screening: 2</li> <li>▶ A majority would choose screening: 10</li> <li>▶ A majority would decline screening: 9</li> <li>▶ Most would decline screening: 0</li> <li>▶ All would decline screening: 0</li> </ul>	<p>Would patients be more or less reluctant to screen if the screening mode were sigmoidoscopy? Please state your opinion.</p> <ul style="list-style-type: none"> <li>▶ Much more reluctant: 4</li> <li>▶ A little more reluctant: 1</li> <li>▶ Equally reluctant (or enthusiastic): 9</li> <li>▶ A little less reluctant: 6</li> <li>▶ Much less reluctant: 1</li> </ul> <p>Would patients be more or less reluctant to screen if the screening mode were faecal testing? Please state your opinion</p> <ul style="list-style-type: none"> <li>▶ Much more reluctant: 0</li> <li>▶ A little more reluctant: 3</li> <li>▶ Equally reluctant (or enthusiastic): 2</li> <li>▶ A little less reluctant: 5</li> <li>▶ Much less reluctant: 11</li> </ul>

surveys and panel discussion, of eliciting panel members' views on the target population's thresholds of important benefit. We have shown how this method can be applied to guide a formal recommendation.<sup>12</sup>

### Strengths and limitations of our process

The method is potentially applicable to interventions with key benefit outcomes that can be weighed against a range of different burdens or harms. In cancer screening interventions, this benefit outcome is typically cancer mortality reduction, and in some instances incidence reduction. Our method is therefore potentially applicable to all cancer screening interventions.

The method may be less applicable when there is a large number of similarly important benefit outcomes. First, the ratings would require judgements for each of relevant benefit outcomes, thus increasing the burden on the panel. Second, and more problematic, would be addressing the permutations and combinations of benefit.

For instance, assume that three relevant benefit outcomes of an intervention are reducing death, stroke and myocardial infarction. The challenge would be

specifying simultaneously the benefits for all three outcomes. For instance, with a particular constellation of burdens and harms a 1% reduction in mortality, 2% in stroke and 3% in myocardial infarction may be sufficient to warrant implementation of an intervention. But it may also be possible that a 3% reduction in death, with no benefit in other outcomes might be sufficient. The possible permutations and combinations of benefit needed to outweigh harms could, even with three benefit outcomes, be very large.

A key element of the method is the panel's elicitation of the panel's inferences regarding peoples' values and preferences prior to knowing the evidence for absolute magnitude of benefit. This may not be possible, particularly when the panel is directly involved in creating the evidence summary. In this case, application of benefit thresholds to the evidence surprised even the content experts on the panel, perhaps testifying to the limited attention thus far given to the absolute magnitude of benefit people receive from screening. The approach we took ensured that views regarding values and preferences drove recommendations rather than the other way around.

This may be one reason our recommendations differed quite substantially from other CRC screening guidelines, which tend to give strong recommendations for screening for everyone from approximately 50–70 years of age.<sup>7,8</sup> There were several panel members with prior experience developing CRC screening recommendations, for whom the threshold approach led to recommendations that contradicted their prior views; one panel member was sufficiently uncomfortable and withdrew from the panel. This highlights the potential importance of one aspect of our methods: the ascertainment of the panel's estimation of thresholds before review of the benefits that screening actually achieves.

### Box 2 Screening with FIT every year versus no screening. Number of votes given for each response alternative.

A patient who is screened with FIT every year for 15 years, has a 5 in 1000 (0.5%) lower risk of dying from or being diagnosed with colorectal cancer at 15 years. Given the harms and burdens of screening, how would patients view such benefits?

- ▶ All or almost all would choose screening: 0.
- ▶ Most would choose screening: 4.
- ▶ The majority would choose screening: 5.
- ▶ The majority would decline screening: 4.
- ▶ Most would decline screening: 3.
- ▶ All or almost all would decline screening: 0.

Among a 1000 people					
	No screening	FIT every two years	FIT every year	Sigmoidoscopy	Colonoscopy
Colorectal cancer mortality 15 years	9 per 1000	5 fewer	6 fewer	5 fewer	6 fewer
	certainty →	⊕⊕○○	⊕⊕○○	⊕⊕○○	⊕⊕○○
Colorectal cancer incidence 15 years	30 per 1000	1 fewer	4 fewer	8 fewer	10 fewer
	certainty →	⊕⊕○○	⊕⊕○○	⊕⊕○○	⊕⊕○○

**Figure 2** Benefits of colorectal cancer screening for a 3% (30 per 1000) 15-year risk of colorectal cancer. Multilayered presentation available at MAGIC website <http://magicproject.org/190220dist/#!/sof/data-set/crc-30-per-1000>

We improved our methods in the course of developing the guideline. Particularly, our third survey was methodologically more sophisticated, including more explicit instructions, many more possible thresholds for consideration and safeguards against biases related to presentation, than was the first. The framing of the evidence may influence peoples' decisions.<sup>20</sup> One limitation of our approach was that we only used one method to present the benefits (presented as reductions in CRC mortality or incidence per 1000 people). Also, the first survey we performed in the panel had only a limited set of potential thresholds with large increments (1, 10, 20 or 30 prevented deaths or cancers). Had we used smaller increments the threshold for an important benefit may have been different. Those applying the approach in future may benefit from reviewing our experience.

Our framing of the evidence also had limitations in that it did not capture all issues relevant to screening. For example, different guidelines suggest repetition of sigmoidoscopy and colonoscopy even in the face of an initial negative test. Some simplification is likely necessary to make our approach feasible.

Even when guideline developers implement optimal methodology, making inferences regarding the distribution of decision thresholds is currently fraught with limitations. Sceptics may say there is no way of determining the validity of such a benefit threshold, and that the estimated threshold may be highly dependent on the composition of the panel, in the worst case by panel members' conflicts.<sup>21</sup> For the process to be valid, panels must understand that they are making their best estimation of the values and preferences of the guideline's target population, rather than their own values and preferences. These estimates must be rigorously grounded in the available evidence, including a systematic review of the published literature,<sup>22</sup> any information the panel has collected (such as a focus group study that can be quickly conducted),<sup>23</sup> the experience of clinician panel members in shared decision making and insights from the patients or members of the public serving on the panel.

Surveys of populations summarised in a systematic review could establish peoples' thresholds, and the higher quality the evidence regarding thresholds, the more likely that different panels would arrive at the same conclusion. As in this case, formal studies provided limited information, establishing only that values and preferences vary widely.

In the absence of high-quality evidence, the advantage of defining a threshold as described here is transparency. Potential users of the guideline can follow the complete process and get a full understanding of how the panel went from evidence to recommendations. If the individual does not agree with the defined threshold of required benefit, she can define for herself what the most suitable threshold should be, and act accordingly. Ideally, if this individual is a clinician helping patients with screening decisions, she will engage patients in shared decision-making.

### Challenges and future research

Panellists, understanding their charge, will first seek a systematic review of patients' values and preferences regarding benefit thresholds. As in this CRC screening recommendation, such a review will often discover only very limited information that is directly applicable.<sup>12</sup> Panellists may have to fall back on indirect evidence (for instance, value and preference studies from similar conditions), or their own personal experience and judgements. This experience may be particularly helpful when it comes from clinicians who have practiced shared decision-making with large numbers of patients.

These limitations are potentially addressable: Value and preference studies asking the directly relevant question (how much benefit people require to undergo screening) require only that granting agencies realise the importance of this area of investigation. Such studies will, however, face the obstacle of prior government and healthcare institution policies. Peoples' responses will be influenced by prior exposure to statements from



apparently trustworthy sources, presented without equivocation, that screening is in their best interest.<sup>24</sup>

Another issue is framing (for instance whether one presents the probability of dying vs the probability of surviving) which influences responses in a wide variety of situations.<sup>20</sup> Given their small sample size, whether this is true of guideline panels will require study of a large number of panels with efforts to standardise methodology.

Our BMJ Rapid Recommendations on CRC screening took an individual perspective, however, the method described here is potentially just as applicable to guidelines taking a public health perspective, that is, guidelines considering the introduction or removal of a screening programme. Whether taking a public health perspective and making recommendations for populations, or an individual perspective applicable in shared decision-making, it is necessary for guideline panels to estimate distributions of thresholds people would require for screening implementation. Taking the public health perspective, consider the lower age boundaries recommended for screening in current CRC guidelines: they include ages 40, 45, 50 and 60.<sup>7,8</sup> Each of these age thresholds implies a benefit people will require to undergo screening. Panels, particularly those taking a public health perspective, may consider many issues, including cost constraints, in their decisions. Nevertheless, the values and preferences of the public should surely be important. Thus, panels choosing 40 are likely to believe that people are ready to accept lower levels of benefit than panels choosing higher thresholds. Whatever age they choose, however, current guidelines never make the benefit threshold transparent.<sup>6</sup> A key consideration in choosing an age to start should be whether the majority of people older than that age, if well-informed, would choose screening.

Panels that take an individual approach to screening recommendations, as did ours, face an additional challenge. To apply our suggestion that individuals with a risk of CRC in the next 15 years of less than 30 in 1000 decline screening, and those with greater risk opt to screen, requires an estimate of individual risk. A number of risk calculators are available, but are limited in their predictive power.<sup>25</sup> With increasing availability of large datasets and new statistical methods, it is possible that the predictive performance of future risk models will improve. In any case, despite their limitations, risk models allow superior individualised shared decision-making: they improve on approaches that assume everyone—from a 50-year-old woman without other risk factors to a 70-year-old man with multiple risk factors—is at the same risk and therefore has the same to gain from screening.

Our discussion has, for two reasons, focused exclusively on application of our method in screening guidelines. First, this proposed specific method is most easily applicable when one has a single benefit outcome of primary interest—or perhaps two of similar importance. Such situations are common in screening when the relevant outcomes are reduction in cancer mortality, and in some instances cancer incidence. Such situations are much less

### Box 3 How guideline panels can establish thresholds for important benefits

#### Prerequisites

- ▶ The intervention has one key beneficial outcome that can be weighed against burdens and harms, or a highly selected number of key beneficial outcomes that bear similar importance for the recommendations.
- ▶ The panel follows standards for trustworthy guidelines, including appropriate management of conflicts of interest by the panel members.
- ▶ Systematic review of patient values and preferences does not provide high-quality evidence of the magnitude of benefit required to undergo the intervention(s) given its burdens and harms.
- ▶ The target population's threshold of the magnitude of benefit required to undergo the considered intervention(s) should be directly relevant for determining the direction and strength of recommendations.

#### 4-step implementation

Steps 1–3 should be performed *prior* to the panel seeing the best available benefit evidence

1. Present the panel with the evidence on burdens and harms for the intervention(s).
  - All estimates should be presented in absolute numbers.
2. Survey the panel on the magnitude of benefit required to undergo the interventions(s) given the burdens and harms from step 1.
  - Suggested thresholds of benefit should be presented on the same scale as burdens and harms, and should include the target population's most likely real benefits with small increments.
  - Terms used in the questions should be clearly specified (eg, 'almost all' = 90% or more).
  - Each panel member provides his/her best estimate of the values and preferences of the guideline's target population.
3. Discuss the survey votes in the panel and agree on threshold(s) of required benefit.
4. Present the panel with the best available evidence on benefits, and formulate recommendations based on the threshold(s) of required benefit defined in step 3.
  - Consider how recommendations are influenced by disease risk.
  - Remember that thresholds guide rather than dictate panel discussions, avoiding a too mechanistic approach as evidence of thresholds still will be limited.

For transparency, and for facilitating informed choices by guideline users, the guideline publication should present details of this process.

common outside of the screening and disease prevention context. Second, we have applied the approach only in the context of screening, and the impact of its use in other contexts remains speculative. Nevertheless, guideline developers may be wise to remain alert to contexts in which the approach might apply outside screening, and to determine its usefulness in such contexts. Examples might include antibiotics in self-limiting infections to follow-up after cancer treatment. Box 3 summarises our preliminary suggestions on how others might best implement this approach.

In our BMJ Rapid Recommendations on CRC screening, we only surveyed the guideline panel, which included lay persons eligible for screening. However, one could also have expanded the survey population

to target the general population (ie, the target population for cancer screening) and/or patient communities. Additional details would need to be provided to clearly communicate the intention of the survey and how results may be used, given that the general population may not have familiarity with practice guidelines and their development process.

## CONCLUSIONS

We have described the first application of a method to establish a threshold of required benefit to undergo screening, and the use of that threshold to inform a screening recommendation. The method will improve with subsequent implementation, most effectively by the conduct of values and preference studies providing empirical evidence of peoples' thresholds. In order to make informed choices about screening, guideline developers either address, despite the challenges, the minimum benefit people would require or abandon transparency in justifying their recommendations.

## Author affiliations

<sup>1</sup>Clinical Effectiveness Research, Department of Transplantation Medicine, Oslo University Hospital, Oslo, Norway

<sup>2</sup>Clinical Effectiveness Research, Faculty of Medicine, Institute of Health and Society, University of Oslo, Oslo, Norway

<sup>3</sup>Department of Health Research Methods, Evidence and Impact, McMaster University, Hamilton, Ontario, Canada

<sup>4</sup>West China Second University Hospital, Sichuan University, Chengdu, Sichuan, China

<sup>5</sup>Department of Medicine, Lovisenberg Diakonale Hospital, Oslo, Norway

<sup>6</sup>Institute of Health and Society, Faculty of Medicine, University of Oslo, Oslo, Norway

<sup>7</sup>Division of General Internal Medicine and Division of Clinical Epidemiology, University Hospitals of Geneva, Geneva, Switzerland

**Twitter** Thomas Agoritsas @ThomasAgoritsas

**Contributors** RAS, LL, POV, TA, MB, LMH and GG contributed to the conception and design of the work. LZ, RAS, LL, POV, TA, MB, LMH and GG contributed to the acquisition and interpretation of the data. LMH, MB and GG drafted the first version of the paper. LZ, RAS, LL, POV, TA, MB, LMH and GG authors revised the paper critically and approved the final version.

**Funding** LMH reports supported by a PhD grant from the Norwegian South Eastern Regional Health Authority during the conduct of this work (project number 2015051).

**Competing interests** All authors have completed the ICMJE form for disclosure of conflicts of interest. LMH, RAS, PV, TA, MB, LL and GG are authors of a guideline on colorectal cancer screening published in The BMJ 2019 (BMJ Rapid Recommendations), in which the described method was used. PV is the CEO of the non-profit MAGIC Evidence Ecosystem Foundation, who is responsible for the BMJ Rapid recommendations project, and GG and TA are board members of MAGIC. GG is the co-founder and co-chair of the GRADE Working Group, and PV and TA are members of the GRADE Working Group.

**Patient consent for publication** Not required.

**Provenance and peer review** Not commissioned; externally peer reviewed.

**Supplemental material** This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines,

terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

**Open access** This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

## ORCID iDs

Lise Mørkved Helsingen <http://orcid.org/0000-0002-2055-3978>

Reed Alexander Siemieniuk <http://orcid.org/0000-0002-3725-3031>

Thomas Agoritsas <http://orcid.org/0000-0002-6182-9969>

## REFERENCES

- 1 Institute of Medicine. *Clinical practice guidelines we can trust*. Washington, DC: The National Academies Press, 2011.
- 2 Qaseem A, Forland F, Macbeth F, *et al*. Guidelines international network: toward international standards for clinical practice guidelines. *Ann Intern Med* 2012;156:525–31.
- 3 Brouwers MC, Kho ME, Browman GP, *et al*. Agree II: advancing Guideline development, reporting and evaluation in health care. *CMAJ* 2010;182:E839–42.
- 4 Guyatt GH, Oxman AD, Vist GE, *et al*. Grade: an emerging consensus on rating quality of evidence and strength of recommendations. *BMJ* 2008;336:924–6.
- 5 Grade. Grade Handbook, 2013. Available: <https://gdtgradeproorg/app/handbook/handbookhtml> [Accessed 15 Jul 2020].
- 6 Zeng L, Helsing LM, Nampo FK, *et al*. How do cancer screening guidelines trade off benefits versus harms and burdens of screening? A systematic survey. *BMJ Open* 2020.
- 7 Bénard F, Barkun AN, Martel M, *et al*. Systematic review of colorectal cancer screening guidelines for average-risk adults: Summarizing the current global recommendations. *World J Gastroenterol* 2018;24:124–38.
- 8 Ebell MH, Thai TN, Royalty KJ. Cancer screening recommendations: an international comparison of high income countries. *Public Health Rev* 2018;39:7.
- 9 Vernooij RWM, Lytynen L, Pardo-Hernandez H, *et al*. Values and preferences of men for undergoing prostate-specific antigen screening for prostate cancer: a systematic review. *BMJ Open* 2018;8:e025470.
- 10 Mansfield C, Tangka FKL, Ekwueme DU, *et al*. Stated preference for cancer screening: a systematic review of the literature, 1990–2013. *Prev Chronic Dis* 2016;13:E27.
- 11 Sicsic J, Pelletier-Fleury N, Mounjid N. Women's benefits and harms trade-offs in breast cancer screening: results from a Discrete-Choice experiment. *Value Health* 2018;21:78–88.
- 12 Helsing LM, Vandvik PO, Jodal HC, *et al*. Colorectal cancer screening with faecal immunochemical testing, sigmoidoscopy or colonoscopy: a clinical practice guideline. *BMJ* 2019;367:i5515.
- 13 Atkin W, Wooldrage K, Parkin DM, *et al*. Long term effects of once-only flexible sigmoidoscopy screening after 17 years of follow-up: the UK flexible sigmoidoscopy screening randomised controlled trial. *Lancet* 2017;389:1299–311.
- 14 Holme Øyvind, Løberg M, Kalager M, *et al*. Long-Term effectiveness of sigmoidoscopy screening on colorectal cancer incidence and mortality in women and men: a randomized trial. *Ann Intern Med* 2018;168:775–82.
- 15 Miller EA, Pinsky PF, Schoen RE, *et al*. Effect of flexible sigmoidoscopy screening on colorectal cancer incidence and mortality: long-term follow-up of the randomised US PLCO cancer screening trial. *Lancet Gastroenterol Hepatol* 2019;4:101–10.
- 16 Alonso-Coello P, Oxman AD, Moher J, *et al*. Grade evidence to decision (ETD) frameworks: a systematic and transparent approach to making well informed healthcare choices. 2: clinical practice guidelines. *BMJ* 2016;353:i2089.
- 17 Jodal HC, Helsing LM, Anderson JC, *et al*. Colorectal cancer screening with faecal testing, sigmoidoscopy or colonoscopy: a systematic review and network meta-analysis. *BMJ Open* 2019;9:e032773.
- 18 Buskermolen M, Cenin DR, Helsing LM, *et al*. Colorectal cancer screening with faecal immunochemical testing, sigmoidoscopy



- or colonoscopy: a microsimulation modelling study. *BMJ* 2019;367:l5383.
- 19 MEEF. Evidence summaries for colorectal cancer screening, 2019. Available: <http://magicproject.org/190220dist/#/>
  - 20 Tversky A, Kahneman D. The framing of decisions and the psychology of choice. *Science* 1981;211:453–8.
  - 21 Norris SL, Burda BU, Holmer HK, *et al.* Author's specialty and conflicts of interest contribute to conflicting guidelines for screening mammography. *J Clin Epidemiol* 2012;65:725–33.
  - 22 MacLean S, Mulla S, Akl EA, *et al.* Patient values and preferences in decision making for antithrombotic therapy: a systematic review: antithrombotic therapy and prevention of thrombosis, 9th ED: American College of chest physicians evidence-based clinical practice guidelines. *Chest* 2012;141:e1S–23.
  - 23 Goodman SM, Miller AS, Turgunbaev M, *et al.* Clinical practice guidelines: incorporating input from a patient panel. *Arthritis Care Res* 2017;69:1125–30.
  - 24 Young B, Bedford L, Kendrick D, *et al.* Factors influencing the decision to attend screening for cancer in the UK: a meta-ethnography of qualitative research. *J Public Health* 2018;40:315–39.
  - 25 Usher-Smith JA, Harshfield A, Saunders CL, *et al.* External validation of risk prediction models for incident colorectal cancer using UK Biobank. *Br J Cancer* 2018;118:750–9.
  - 26 Deeks JJ. Issues in the selection of a summary statistic for meta-analysis of clinical trials with binary outcomes. *Stat Med* 2002;21:1575–600.
  - 27 Schmid CH, Lau J, McIntosh MW, *et al.* An empirical study of the effect of the control rate as a predictor of treatment efficacy in meta-analysis of clinical trials. *Stat Med* 1998;17:1923–42.
  - 28 Furukawa TA, Guyatt GH, Griffith LE. Can we individualize the 'number needed to treat'? An empirical study of summary effect measures in meta-analyses. *Int J Epidemiol* 2002;31:72–6.