

REVIEW

Presentation of Benefits and Harms in US Cancer Screening and Prevention Guidelines: Systematic Review

Tanner J. Caverly, Rodney A. Hayward, Elyse Reamer, Brian J. Zikmund-Fisher, Daniel Connochie, Michele Heisler, Angela Fagerlin

Veterans Affairs Center for Clinical Management Research, Ann Arbor, MI (TJC, RAH, MH, AF); Department of Internal Medicine, University of Michigan Medical School (TJC, RAH, ER, BJZF, MH, AF), Center for Bioethics and Social Science in Medicine (TJC, BJZF, DC, AF), and Department of Health Behavior and Health Education (BJZF), University of Michigan, Ann Arbor, MI.

Correspondence to: Tanner J. Caverly, MD, MPH, Veterans Affairs Medical Center, 2215 Fuller Road, Ann Arbor, MI 48105 (e-mail: tcaverly@med.umich.edu).

Abstract

Background: Cancer prevention and screening guidelines are ideally suited to the task of providing high-quality benefit-harm information that informs clinical practice. We systematically examined how US guidelines present benefits and harms for recommended cancer prevention and screening interventions.

Methods: We included cancer screening and prevention recommendations from: 1) the United States Preventive Services Task Force, 2) the American Cancer Society, 3) the American College of Physicians, 4) the National Comprehensive Cancer Network, and 5) other US guidelines within the National Guidelines Clearinghouse. Searches took place November 20, 2013, and January 1, 2014, and updates were reviewed through July 1, 2015. Two coders used an abstraction form to code information about benefits and harms presented anywhere within a guideline document, including appendices. The primary outcome was each recommendation's benefit-harm "comparability" rating, based on how benefits and harms were presented. Recommendations presenting absolute effects for both benefits and harms received a "comparable" rating. Other recommendations received an incomplete rating or an asymmetric rating based on prespecified criteria.

Results: Fifty-five recommendations for using interventions to prevent or detect breast, prostate, colon, cervical, and lung cancer were identified among 32 guidelines. Thirty point nine percent ($n = 17$) received a comparable rating, 14.5% ($n = 8$) received an incomplete rating, and 54.5% ($n = 30$) received an asymmetric rating.

Conclusions: Sixty-nine percent of cancer prevention and screening recommendation statements either did not quantify benefits and harms or presented them in an asymmetric manner. Improved presentation of benefits and harms in guidelines would better ensure that clinicians and patients have access to the information required for making informed decisions.

When clinicians discuss cancer screening and prevention with patients, their perceptions about the magnitude of benefits and harms help shape their recommendations (1). Although it is useful for clinicians if guidelines are concise in their recommendations, estimates of the magnitude of both benefits and harms should be clearly delineated, along with information on the reliability and quality of the evidence that provides the basis

of those estimates. The provision of such information is essential for allowing clinicians to assess the guideline committee's conclusions and provides important information for patient consultations.

The manner in which the benefits and harms are presented can have a large influence on risk perceptions (2). Only absolute risk informs how likely an outcome is for an individual

Received: March 23, 2015; Revised: August 4, 2015; Accepted: December 22, 2015

Published by Oxford University Press 2016. This work is written by (a) US Government employee(s) and is in the public domain in the US.

(or group of individuals) (3). Unless the absolute risks of both benefits and harms are effectively communicated, one cannot determine whether the treatment has a clinically meaningful effect or whether the potential benefits outweigh the potential harms and burdens. Relative risk reductions are less informative for the care of individual patients. Moreover, patients, physicians, and policy makers perceive relative risk reductions as larger and more persuasive than the corresponding absolute risk reductions (2). Accordingly, a consensus exists about the importance of presenting absolute risk information in reports of randomized trials (4,5), in reports of observational studies (6), in clinical practice guidelines (7,8), and in patient decision aids (3,9). When the potential benefits and harms of medical interventions are not presented in a balanced fashion, testing and treatment decisions can be adversely affected (10). This may then lead to inappropriate physician recommendations and result in either underuse of needed and desired care or in overuse of unnecessary and unwanted care.

Arguably, clinical practice guidelines that review and rate the quality of evidence before making practice recommendations are best suited to providing this essential clinical information. The Institute of Medicine's standards for guideline development identifies that "a clear description of potential benefits and harms should be provided for each recommendation in a clinical practice guideline" (11). How these important clinical resources present benefits and harms has not been previously evaluated. In order to determine whether cancer screening and prevention guidelines present the benefit and harm information clinicians need to make decisions—in the format that best promotes clarity about the occurrence of important outcomes (ie, absolute effects)—we systematically examined how guidelines present benefit-harm information.

Methods

The rationale for expecting that potential benefits and harms be quantified in comparable terms (as absolute risks) is particularly strong in the context of recommendations for the use of cancer screening and prevention services in healthy populations. Evidentiary standards are high in this context (12). Thus, high-quality data on potential benefits and at least some of the important potential harms is reliably available for these services. Accordingly, this review focuses on how benefits and harms are presented for recommended cancer prevention and screening services ("positive recommendations").

Guideline Selection Strategy

We first identified all United States Preventive Services Task Force (USPSTF) guidelines related to cancer screening or prevention. We excluded USPSTF statements rated "I" for insufficient evidence and statements focusing solely on counseling or children/adolescent populations. We also retrieved the most recent guidelines from several other prominent national organizations making recommendations on the included USPSTF topics, including: the American Cancer Society (ACS), the American College of Physicians (ACP), and the National Comprehensive Cancer Network (NCCN). We next performed a search within the National Guidelines Clearinghouse (NGC)—a comprehensive database of evidence-based clinical practice guidelines maintained by the Agency for Healthcare Research and Quality—to identify all additional discrete US guidelines making recommendations on the same USPSTF topics. The search strategy is detailed in the [Supplementary Methods](#) (available online). As for

the other guidelines, we excluded guidelines within the NGC if: 1) the guideline merely duplicated or referred to other recommendation statements; 2) the guideline was not relevant to a specific USPSTF-covered topic area; 3) the focus was solely on behavioral counseling; or 4) the guideline stated there was not sufficient evidence to make a recommendation for or against a preventive service. All searches took place between November 20, 2013, and January 1, 2014. We reviewed any updates to the guidelines included in this review as of July 1, 2015.

Selection of Specific Recommendations

Two trained research assistants (ER and DC) reviewed the eligible guideline documents in their entirety, including abstracts and online appendices and tables, to abstract each specific recommendation within the document. A guideline document could contain a number of specific recommendations. Each of these specific recommendations identifies a particular action the guideline group recommends doing or not doing within a population. The PI of the research team (TJC) then reviewed each recommendation to ensure that only discrete statements about a specific intervention for a specified population were included. The research assistants were instructed not to include recommendations for populations designated as high-risk based on specific personal or family history risk factors (eg, screening for colorectal cancer in patients with inflammatory bowel disease). As stated above, the focus of this study was to understand how both the benefits and harms of a cancer prevention service are presented for recommended services. Accordingly, we included all discrete recommendations for the use of the preventive service in a particular population ("positive recommendations"). Those that recommended discussing the benefits and harms of the service with patients were also included as "positive recommendations." We excluded negative recommendations (discrete recommendations against a service). In contrast to positive recommendations for the use of a preventive service in a healthy population, negative recommendations were considered inappropriate for this review for two reasons: 1) it seemed unreasonable to require that harms are quantified for preventive services where there is adequate evidence suggesting very little or no benefit; and 2) there may be a good rationale for recommending against preventive services that does not require a clear understanding of how benefits compare with harms (eg, not an important enough population health problem).

Data Collection

Data Extraction and Management

The full text of each eligible guideline—along with all associated appendices and tables—was downloaded to an online storage site (Zotero) for abstraction (<https://www.zotero.org/>; Accessed January 10, 2016). Because guidelines often appear in multiple forms (eg, as a summary in the National Guidelines Clearinghouse, as a document on the developer's website, or as a published manuscript within a medical journal), efforts were made to find the most complete guideline document available. The goal of this study was to evaluate how benefits and harms are presented within guideline recommendations that are written to directly inform clinical practice. Thus, we did not assess the way benefits and harms were presented in any separate evidence reviews undertaken by the guideline developers.

The two research assistants used a standardized form to independently retrieve both quantitative and qualitative data from each specific recommendation and the guideline overall. Then, TJC reviewed the abstracted information to check

the accuracy of the coding. Differences in how the data was coded were highlighted and resolved by three members of the research team (ER, DC, and TJC), consulting the guideline document as needed. The following reflect the types of information abstracted in the review.

Information About Incidence of the Target Cancer

We included whether the incidence of the target cancer (over a specified timeframe) was presented. If presented, we identified whether this risk information was presented with or without a denominator. We also documented whether life-years or quality-adjusted life-years lost because of the target cancer were mentioned.

Information About Quality of Evidence

Using National Guidelines Clearinghouse documentation, details from the guideline documents, and, if applicable, any guideline development procedures documented on an official website, we identified: whether a systematic review was completed before the recommendation process, the methods used to analyze the quality and strength of the evidence, and whether there was an explicit rating scheme for the quality of evidence and strength of recommendation.

Information on Benefits and Harms

Only statements about an improvement in clinically important outcomes (ie, improvements in morbidity or mortality) were included as information about the benefits of the cancer screening and prevention services. Statements about the accuracy of screening, for instance, were not counted as a discussion of benefits. While screening test characteristics such as sensitivity and specificity are important, this information alone is not generally accepted as evidence to support a policy of screening (13). Any statements regarding negatively valued attributes of the screening test or preventive treatment were coded as harms (14). We then coded whether benefits and harms were quantified if mentioned, and, if quantified, whether the data were presented using relative risk or absolute risks. All of the following formats counted as a measure of absolute effect: number needed to invite/screen/treat/harm, natural frequencies (eg, x in 1000 or 1 in x), absolute percentages, and number of life-years or quality-adjusted life-years gained per population. The exact text within the guideline regarding all benefit and harm information was abstracted for qualitative analysis and error checking. To provide information about the accessibility of benefit-harm information, the location of this information within each guideline was recorded.

Analysis

We used descriptive statistics (eg, means, standard deviations, frequencies) to summarize the guideline data and how the risk/benefit information was presented.

Comparability Rating

Each guideline received a “comparability rating” based on how benefit and harm information was presented. This served as our primary outcome. Recommendations presenting absolute effect information for both benefits and harms received a “comparable” rating. “Asymmetric” ratings were given when the information was presented in an uneven fashion. This occurred for the following reasons: 1) because a recommendation for a service was made without mentioning what the potential benefit might

be, 2) because the benefit was mentioned but the possibility of any harms was not mentioned, 3) because the benefit was quantified while the harms were mentioned but not quantified, or 4) because the benefits were quantified in terms of a relative risk reduction while the harms were quantified in terms of absolute risk increase (mismatched framing) (15,16). “Incomplete” ratings were reserved for recommendations where: 1) benefits and harms were both mentioned but neither were quantified or 2) both harms and benefits were provided using relative effects only. Because presentation of benefits and harms is just one aspect of developing high-quality clinical practice guidelines, these comparability ratings cannot be used to make an overall assessment of the quality of a guideline.

The number of US guidelines available was not large enough to allow for extensive statistical evaluation of factors associated with a higher or lower comparability score. However, we tabulated the primary outcome by guideline type, guideline year, and target cancer to evaluate for any patterns that might emerge.

Results

Positive Recommendations Reviewed

Fifty-five positive recommendations were identified within 32 guideline documents that met inclusion criteria. These guidelines included a variety of different cancers and recommendations for screening and preventive services (Table 1).

Presentation of Incidence

The incidence of the target cancer over a specified timeframe (typically over one year) was mentioned in 50.9% (n = 28) of the recommendations. Only 10.9% (n = 6) presented the incidence as a proportion (ie, specified a denominator). Similarly, a mere 10.9% (n = 6) of the recommendations quantified the importance of the target cancer in terms of life-years lost because of the cancer.

Presentation of Quality of Evidence

Sixty-seven point three percent of the recommendations (n = 37) were informed by a prior systematic review. The majority of the recommendations included an explicit rating scheme for the

Table 1. Types of interventions with positive recommendations

Target cancer	Screening/preventive service
Breast cancer	Screening mammography Clinical breast exam Preventive medications (tamoxifen, raloxifene)
Prostate cancer	Prostate-specific antigen (PSA) Preventive medication (5-alpha reductase inhibitors)
Colorectal cancer	Fecal immunochemical test (FIT) Flexible sigmoidoscopy Colonoscopy
Cervical cancer	Computed tomography colonography Papanicolaou smear Human papillomavirus (HPV) testing HPV vaccination
Lung cancer	Low-dose computed tomography (CT) screening

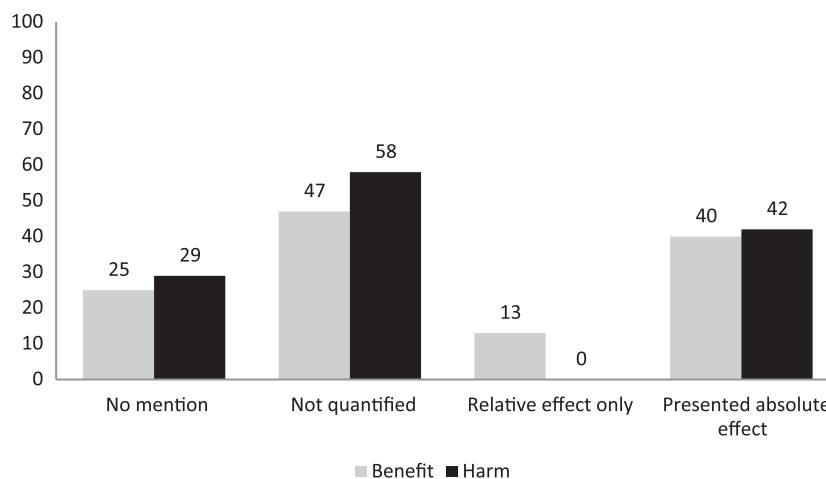


Figure 1. Proportion of 55 positive recommendations presenting benefit and harm information.

quality of the evidence (63.6%, $n = 35$) and for the strength of the recommendation (72.7%, $n = 40$).

Presentation of Benefits and Harms

The proportion of the time benefits and harms were mentioned and quantified, and the proportion quantified as relative and absolute effects are depicted in Figure 1. Overall, 25.4% ($n = 14$) of the recommendation statements failed to mention what the clinically important benefits of the cancer prevention might be while 29.1% ($n = 16$) failed to mention any potential harms.

There were no notable differences in the locations of the benefit and harm information. The majority of the time both benefits and harms were found in the text of the main guideline article if mentioned (71% and 64% of the recommendations, respectively).

Comparability Ratings

The proportion of positive recommendations with comparable, incomplete, and asymmetric comparability ratings is depicted in Figure 2. Among the 55 positive recommendations, 30.9% ($n = 17$) received a comparable rating—presenting absolute effects for both benefits and harms so that the trade-offs could be directly compared. 54.5% ($n = 30$) received an asymmetric rating and of these 30: 14 made a recommendation for a service without mentioning what the potential benefit might be; four mentioned the benefit but did not mention the possibility of any harms; eight quantified the benefit and mentioned but did not quantify the harms; and four quantified the benefits in terms of a relative risk reduction (thus showing a larger risk reduction) while quantifying the harms in terms of absolute risk increase (thus showing a small risk of harm), a practice known as mismatched framing (15,16). Finally, 14.5% ($n = 8$) received an incomplete rating. No cases were found in which both benefits and harms were reported as only relative risks, resulting in all incomplete ratings because of the fact that both benefits and harms were mentioned but not quantified.

The complete list of specific recommendations reviewed—grouped by comparability rating and cancer type—is presented in Supplementary Tables 1–3 (available online). Table 2 presents guideline recommendations on breast cancer screening with mammography as a representative example because different guidelines on this topic received asymmetric, incomplete, and

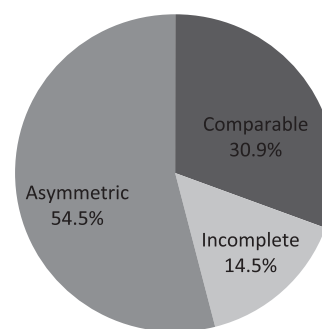


Figure 2. Proportion with asymmetric, incomplete, and comparable ratings among 55 positive recommendations for cancer screening & prevention services. Recommendations received a comparable rating if absolute effects were presented for both benefits and harms, an incomplete rating if the magnitude of benefits and harms were not able to be compared (both presented but neither quantified), and an asymmetric rating if the presentation was uneven.

comparable ratings. No clear patterns in ratings emerged based on the year the guideline was produced. Our power was limited to detect statistically significant differences across different guideline groups.

Discussion

In this systematic review of recommendations for cancer screening and prevention services, US guidelines typically did not provide information on benefits and harms so that they could be compared. In fact, 69% of positive recommendations either did not quantify benefits and harms or presented them in an uneven manner. Fewer than one in three of the positive recommendations received a comparable rating, presenting absolute effects for both benefits and harms so that the trade-offs could be directly compared. If patients and physicians have inaccurate perceptions about the magnitude of benefits and harms from cancer prevention and screening, it may be, at least in part, because the essential information is not readily accessible in important patient-care resources.

The USPSTF has a rigorous and highly transparent process for developing guidelines. This process includes developing an outcomes table that contains quantification of absolute benefits and absolute harms so that they can be compared (to aid determination of net benefit for Task Force members) and a

Table 2. Representative examples of asymmetric, incomplete, and comparable ratings for clinical practice guidelines on mammography screening for breast cancer*

Rating	Guideline title	Guideline developer	Year	Reason for rating	Benefits sample text	Harms sample text	Citation/link
Asymmetric	Diagnosis of breast cancer	Institute for Clinical Systems Improvement	2012	Benefits of screening mentioned but harms of screening not mentioned at all	“Regular mammographic screening has been shown to reduce mortality in breast cancer.”	Not mentioned	https://www.icsi.org/_asset/v9191q/DxBrdis.pdf Accessed November 21, 2013.
Incomplete	Position statement on screening mammography	American Society of Breast Surgeons	2011	Benefits and harms cannot be compared because neither quantified	“Patients with breast cancers detected in an earlier stage are more likely to be candidates for breast-conserving surgery, less likely to require chemotherapy, and will have an improved survival...”	“Some patients will undergo additional imaging or even biopsy for lesions that are subsequently found not to be malignant.”	https://www.breastsurgeons.org/statements/PDF_Statements/screening_mammography.pdf Accessed November 21, 2013.
Comparable	Screening mammography for women age 40 to 49 years	American College of Physicians	2007	Absolute effects presented for both benefits and harms	“This relative risk reduction corresponds to about 5.6 deaths prevented per 10 000 women screened...”	“...cumulative rates of false-positive mammograms of 38% after 10 mammograms”	Qaseem A, Snow V, Sherif K, et al. Screening mammography for women 40 to 49 years of age: a clinical practice guideline from the American College of Physicians. <i>Ann Intern Med.</i> 2007;146(7):511–515. http://www.ncbi.nlm.nih.gov/pubmed/17404353 Accessed January 20, 2014.

*Tables of all guidelines reviewed, organized by comparability rating and target cancer, are presented in [Supplementary Tables 1–3](#) (available online).

consistent format for presenting recommendations to the public (17). Perhaps as a result of this process, five out of six USPSTF guidelines reviewed presented benefit and harm information in a comparable fashion by our criteria. Nonetheless, the USPSTF often did not present absolute effect information in a way that aligns with current recommendations for optimal risk presentation, such as using a consistent denominator for all outcomes (eg, x in 1000) (3) or summarizing numeric outcome information in a consistent format within a summary table (7,8). Overall, the way benefits and harms were presented was quite variable across the guideline groups we reviewed. Presentation of benefits and harms may be an underappreciated task in the writing of current guidelines on cancer screening and prevention. In addition to guidelines, clinicians may also seek out online point-of-care resources to inform clinical decision-making. Thus, we also reviewed benefit-harm presentation in two popular point-of-care resources, UpToDate and DynaMed, to explore the extent to which these resources attend to the task of presenting benefits and harms. At the time of our review (01/2014), three out of seven UpToDate articles and five out of seven DynaMed articles reviewed (regarding similar cancer screening and prevention topics) received “comparable” ratings.

Presenting high-quality quantitative estimates for the magnitude of benefits and many of the important harms is possible for the cancer prevention services we reviewed. Thus, our findings led us to consider potential contributors to the lack of clarity in discussing benefits and harms. One reason may be a belief among guideline developers that recommendations should be as simple as possible. However, that reasoning does not justify omitting this information from the text and appendices. Another reason might be a fear among guideline developers that too much transparency about the magnitudes of benefits and harms may lead individual clinicians and patients to decide against utilizing a preventive service when guideline developers believe the service is valuable. However, this would run counter to informed consent and a guiding ethical principle of medicine—respect for the individual and their autonomy. Moreover, in cases where the magnitude of the benefits is large and the harms very small (such as with HPV vaccination), transparent benefit-harm presentation could increase uptake. Finally, there may be a belief that from a population perspective any cancer screening or prevention service with a statistically significant improvement in mortality is worthwhile. Such reasoning makes clear presentation of benefits and harms unnecessary. This reasoning also implies that clinical judgment about whether the mortality benefit outweighs the harms is unnecessary. Unfortunately, difficult judgments about net benefit cannot be resolved through statistical significance testing (18).

It is important to note the relatively recent shift in communication about the benefits and harms of cancer screening (19). Certain harms such as false positives, incidental findings, and overdiagnosis are increasingly recognized as important factors for decision-making in the cancer screening and prevention context. The National Cancer Institute recently issued its first “patient and physician guide” on lung cancer screening (<http://www.cancer.gov/newscenter/qa/2002/NLSTstudyGuidePatientsPhysicians> Accessed January 10, 2016.). This guide strives to present the benefits and harms of lung cancer screening in a format that allows for easy comparisons, facilitating informed judgment about the trade-offs. More broadly, the GRADE group now recommends that guideline developers present absolute effects for important benefits and harms within a standard “summary of findings” table (7,8). The guidelines reviewed in this study were largely written prior to these GRADE recommendations. It is possible

that future cancer screening and prevention guidelines may put more emphasis on the way benefits and harms are presented. Ideally, our study will motivate guideline developers to consistently include outcome summary tables similar to the ones the GRADE group recommends.

Our findings align with other work that has reviewed how benefit and harm information is presented in other domains. News stories about cancer screening tests often dramatize benefits while ignoring any potential for harms (20). An analysis of the content of 409 direct-to-consumer cancer center advertisements found that treatment benefits were mentioned in just 27% of the advertisements and only quantified in 2% (21). Harms were rarely mentioned (2% of the advertisements) and were never quantified. Furthermore, the presentation of benefits and harms in medical journals does not appear to be dramatically better. Among 359 articles in five high-impact medical journals, absolute risk information was presented in just 26 (7.2%) articles (22). An analysis of 119 systematic reviews in high-impact medical journals found that roughly 50% failed to present absolute risk information (15). Among randomized controlled trials of cancer screening, harms are rarely quantified (eg, only 4% of 57 trials quantified the chance of a false-positive finding) (23). It is reasonable to expect more transparent benefit-harm presentations in clinical practice recommendations. Indeed, an argument can be made that transparent presentation is even more important in the guideline document than in published evidence review reports that might support guideline development—because it is the guideline document that explicitly aims to inform clinical practice. Yet, benefit-harm presentation is often overlooked in current guidelines. Clarity about benefits and harms would promote more accurate perceptions about important outcomes (3,9) and support clinician and patient decision-making regarding whether an intervention is appropriate given the context and the patient’s preferences (3,24). A summary table of the absolute effects of the intervention could easily be included as part of standard recommendation statement templates.

Though our study represents a comprehensive review of US groups making patient care recommendations about cancer screening and prevention, there remain limitations. A “comparable” rating simply means that benefits and harms are presented in a format to allow potential comparison, while incomplete and asymmetric ratings mean that such a comparison was not possible. Thus, our study focuses on just one component of what is necessary for developing a high-quality guideline (11). The comparability ratings used in this study cannot be used as an indication of the overall quality of a guideline. A good comparability rating could be associated with low overall guideline quality and vice versa. Making a recommendation is complex and requires multiple judgments and a systematic and transparent process (11).

Prior to quantifying estimates of absolute effect for different outcomes, guideline groups must identify what the important outcomes are and determine the quality of evidence for each outcome. Determining estimates of absolute effect for each important outcome is a critical next step. The focus of our paper is how guidelines present these outcome estimates. Estimating the absolute effects of any intervention is necessary to determine whether the benefits outweigh the harms, costs, and other disadvantages. Nonetheless, estimating absolute effects for important outcomes may not be a straightforward task. In the context of screening tests, quantifying harms that occur downstream from the test itself may be particularly challenging. In some cases (eg, cervical cancer screening), instead of comparing no screening to any particular screening

interval, it may be more important to estimate the potential gains and risks of more frequent screening intervals vs less frequent intervals. If outcome rates vary considerably across groups, such as rates of harm for different age groups, current recommendations are to present the range of possible absolute effects across the groups or, alternatively, the absolute effects for each important subgroup (7). In rare instances where an intervention's benefit is known to be large and harms are felt to be "no more than small" (but difficult to quantify precisely), some might reasonably argue that quantifying the harms may be unnecessary. On the other hand, we believe that even in these cases it would still be good practice to estimate what "no more than small" means in quantitative terms. Commenting on how groups should move from outcome estimates to making a recommendation is beyond the scope of our paper. Nuanced judgments are required to move from outcome estimates to a concrete recommendation (25,26).

This paper focuses on how benefits and harms are presented to clinicians. An in-depth discussion of the extent to which clinicians should present numerical information to patients and how best to communicate this information to patients is beyond the scope of this paper. It is often argued that patients do not understand numbers and thus it might be better to avoid providing risk information as it might be too confusing. However, there are many articles that have provided advice on how to make complicated risk statistics easier for patients (and likely their providers) to understand (3,27).

We did not independently assess the quality of evidence supporting quantitative estimates. However, we did find that the majority of guidelines (67%) were supported by a systematic review and that most guidelines had explicit criteria for rating the quality and strength of evidence (64% and 73%, respectively). Also, ensuring that estimates were provided for all important outcomes was beyond the scope of this review. For instance, we counted the benefit-harm presentation as "comparable" if one clinically important benefit and one negative attribute of the intervention were quantified as absolute risks. We did not assess whether all substantive benefits and harms were quantified. Our rating method was chosen intentionally so that our review gave a best-case scenario regarding how important benefits and harms are presented in guidelines. Full transparency about benefits and harms would require that estimates for all important outcomes be provided (along with the degree of certainty in their magnitude), that all of this information is presented in an accessible location, and that any language describing benefits and harms is also balanced (7,8). If a guideline failed to receive a comparable rating using our method, then it would also fail more rigorous standards as well. We recommend the use of "summary of findings" tables, an approach proposed in a series of papers from the GRADE guidelines group, as the best method of summarizing and presenting outcome information. Unfortunately, none of the guidelines we reviewed summarized findings using this approach.

Of particular concern are guidelines that are asymmetric. Four of the 55 recommendation statements we reviewed presented the benefits of the intervention in terms of a relative risk reduction (larger number) while presenting the harms in terms of an absolute risk reduction (smaller number). While it is not likely this was done to intentionally mislead, this practice nonetheless mischaracterizes the trade-offs (28). Similarly, 22 of the 55 recommendation statements presented only the numerator regarding the incidence of the target cancer (ie, "In 2013, an estimated 232 340 women in the United States will be diagnosed with breast cancer and 39 620 women will die of the disease").

These large numbers do have some population health relevance. However, they can lead to a misperception about how likely an individual is to benefit from an intervention when the denominator is large (29). Using incidence rates is much more informative when proposing individual-level patient interventions (ie, "The age-adjusted annual incidence rate of cervical cancer is 6.6 cases per 100 000 women").

Although absolute effects are difficult to know precisely, firm recommendations should not be made without guideline developers and clinicians at least estimating how big the absolute benefits and harms are most likely to be. Groups recommending interventions for asymptomatic individuals should strive to clearly present absolute estimates for the chance of benefit and harm. The GRADE approach for creating summary of findings tables within guidelines—which includes how to present measures of absolute effect for both benefits and harms—could help standardize this task (7,8). Without access to transparent risk information on all clinically important outcomes, clinicians and policy makers cannot properly judge the trade-offs (10). Moreover, without such information clinicians cannot fully engage patients in true shared decision-making (3).

Funding

Dr. Caverly was supported by a VA Advanced Fellowship in Health Services Research & Development (HSR&D). This research was also supported in part by the Methods Core of the Michigan Center for Diabetes Translational Research (NIDDK P30DK092926) and also HX 13-001 (VA HSR&D Center of Innovation).

Notes

The views expressed in this article are those of the authors and do not necessarily reflect the position or policy of the Department of Veterans Affairs or the United States government. RAH receives royalties for authoring or co-authoring three topics in Up-To-Date, none of them on cancer screening. The authors have no other relevant financial disclosures.

The authors would like to thank Judith Smith, a health sciences informationist at the University of Michigan Medical School, for her expertise in helping with the systematic review process.

References

1. Eddy DM. Variations in physician practice: the role of uncertainty. *Health Aff (Millwood)*. 1984;3(2):74–89.
2. Akl EA, Oxman AD, Herrin J, et al. Using alternative statistical formats for presenting risks and risk reductions. *Cochrane Database Syst Rev*. 2011;(3):CD006776.
3. Trevena LJ, Zikmund-Fisher BJ, Edwards A, et al. Presenting quantitative information about decision outcomes: a risk communication primer for patient decision aid developers. *BMC Med Inform Decis Mak*. 2013;13(Suppl 2):S7.
4. Schulz KF, Altman DG, Moher D, for the CONSORT Group. CONSORT 2010 Statement: updated guidelines for reporting parallel group randomised trials. *BMJ*. 2010;340(mar23 1):c332–c332.
5. Ioannidis JPA, Evans SJW, Gøtzsche PC, et al. Better Reporting of Harms in Randomized Trials: An Extension of the CONSORT Statement. *Ann Intern Med*. 2004;141(10):781–788.
6. Von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP. Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *Prev Med*. 2007;45(4):247–251.
7. Guyatt GH, Oxman AD, Santesso N, et al. GRADE guidelines: 12. Preparing Summary of Findings tables—binary outcomes. *J Clin Epidemiol*. 2013;66(2):158–172.
8. Guyatt GH, Thorlund K, Oxman AD, et al. GRADE guidelines: 13. Preparing Summary of Findings tables and evidence profiles—continuous outcomes. *J Clin Epidemiol*. 2013;66(2):173–183.

9. Joseph-Williams N, Newcombe R, Politi M, et al. Toward Minimum Standards for Certifying Patient Decision Aids: A Modified Delphi Consensus Process. *Med Decis Making*. 2013. <http://mdm.sagepub.com/cgi/doi/10.1177/0272989X13501721>. Accessed January 10, 2016.
10. Tversky A, Kahneman D. The framing of decisions and the psychology of choice. *Science*. 1981;211(4481):453–458.
11. Institute of Medicine (U.S.), Committee on Standards for Developing Trustworthy Clinical Practice Guidelines, Graham R. *Clinical practice guidelines we can trust* [Internet]. Washington, D.C.: National Academies Press; 2011. <http://site.ebrary.com/id/10495436>. Accessed January 10, 2016.
12. Harris R, Sawaya GF, Moyer VA, Calonge N. Reconsidering the Criteria for Evaluating Proposed Screening Programs: Reflections From 4 Current and Former Members of the U.S. Preventive Services Task Force. *Epidemiol Rev*. 2011;33(1):20–35.
13. Barratt A, Irwig L, Glasziou P, et al. Users' guides to the medical literature: Xvii. how to use guidelines and recommendations about screening. *JAMA*. 1999;281(21):2029–2034.
14. Harris RP, Sheridan SL, Lewis CL, et al. The harms of screening: A proposed taxonomy and application to lung cancer screening. *JAMA Intern Med*. 2014;174(2):281–285.
15. Sedrakyan A, Shih C. Improving Depiction of Benefits and Harms. *Med Care*. 2007;45(Suppl 2):S23–S28.
16. Gigerenzer G, Gaissmaier W, Kurz-Milcke E, Schwartz LM, Woloshin S. Helping doctors and patients make sense of health statistics. *Psychol Sci Public Interest*. 2007;8(2):53.
17. Sawaya GF, Guirguis-Blake J, LeFevre M, Harris R, Petitti D. Update on the Methods of the U.S. Preventive Services Task Force: Estimating Certainty and Magnitude of Net Benefit. *Ann Intern Med*. 2007;147(12):871–875.
18. Feinstein AR. Invidious comparisons and unmet clinical challenges. *Am J Med*. 1992;92(2):117–120.
19. Woloshin S, Schwartz LM, Black WC, Kramer BS. Cancer Screening Campaigns—Getting Past Uninformative Persuasion. *N Engl J Med*. 2012;367(18):1677–1679.
20. Schwitzer G. A Guide to Reading Health Care News Stories. *JAMA Intern Med*. 2014. <http://archinte.jamanetwork.com/article.aspx?doi=10.1001/jamainternmed.2014.1359>. Accessed January 10, 2016.
21. Vater LB, Donohue JM, Arnold R, White DB, Chu E, Schenker Y. What Are Cancer Centers Advertising to the Public? A Content Analysis. *Ann Intern Med*. 2014;160(12):813–820.
22. Nuovo J, Melnikow J, Chang D. Reporting number needed to treat and absolute risk reduction in randomized controlled trials. *JAMA*. 2002;287(21):2813–2814.
23. Heleno B, Thomsen MF, Rodrigues DS, Jorgensen KJ, Brodersen J. Quantification of harms in cancer screening trials: literature review. *BMJ*. 2013;347(Sep 16 1):f5334–f5334.
24. Eddy DM. Comparing benefits and harms: The balance sheet. *JAMA*. 1990;263(18):2493–2505.
25. Andrews J, Guyatt G, Oxman AD, et al. GRADE guidelines: 14. Going from evidence to recommendations: the significance and presentation of recommendations. *J Clin Epidemiol*. 2013;66(7):719–725.
26. Andrews JC, Schünemann HJ, Oxman AD, et al. GRADE guidelines: 15. Going from evidence to recommendation—determinants of a recommendation's direction and strength. *J Clin Epidemiol*. 2013;66(7):726–735.
27. Fagerlin A, Zikmund-Fisher BJ, Ubel PA. Helping Patients Decide: Ten Steps to Better Risk Communication. *J Natl Cancer Inst*. 2011;103(19):1436–1443.
28. Naylor CD, Chen E, Strauss B. Measured Enthusiasm: Does the Method of Reporting Trial Results Alter Perceptions of Therapeutic Effectiveness? *Ann Intern Med*. 1992;117(11):916–921.
29. Huff D. *How to lie with statistics*. London: Penguin Books; 1991.