

From ABCs to GRADE

Canadian Task Force on Preventive Health Care's new rating system for clinical practice guidelines

Neil Bell MD SM CCFP FCFP Sarah Connor Gorber PhD Marcello Tonelli MD SM FRCPC Kevin Pottie MD MCISc CCFP FCFP
Harminder Singh MD MPH FRCPC FACP Michel Joffres MD PhD Elizabeth Shaw MD CCFP FCFP

The Canadian Task Force on Preventive Health Care (CTFPHC) was reestablished in 2010 with a mandate to develop and disseminate clinical practice guidelines (CPGs) for primary and preventive care. The CTFPHC uses the GRADE (grading of recommendations, assessment, development, and evaluation) system to rate the quality of its evidence and the strength of its recommendation statements. The GRADE system provides a structured and transparent process for guideline development that begins at framing key questions and proceeds through the evaluation of evidence for benefits and harms, as well as incorporation of patient preferences and resource implications, to arrive at recommendations. This article outlines key concepts of the GRADE process to assist primary care practitioners in understanding the GRADE recommendations and discussing these recommendations with patients.

Background

Family physicians and other primary care health professionals often seek guidance from CPGs about how to better manage their patients. Family physicians are confronted with a bewildering array of CPGs developed by a large variety of government agencies and professional organizations. It is estimated that there are currently at least 2400 guidelines in the Agency for Healthcare Research and Quality's National Guideline Clearinghouse¹; more than 6400 guidelines in the database of the Guidelines International Network²; and more than 2700 in the Canadian Medical Association's CPG database.³ Each database includes multiple guidelines on the same topic, often with conflicting recommendations.^{4,5} Recently, there has also been increasing concern about the quality of CPGs owing to potential bias on the part of the guideline developers^{6,7} or the quality of the evidence used to develop the CPGs.^{5,7-9} For family physicians, these issues raise concerns about the validity of the recommendations and create confusion over which to apply in practice.

Family physicians are also confronted with a diverse range of systems used in CPGs to rate the quality of scientific evidence and the strength of recommendations. These different rating systems make it difficult for family physicians to understand and effectively communicate the benefits and harms of the practices recommended in CPGs to their patients. In 1979, the Canadian Task

Force on the Periodic Health Examination published one of the first systems to explicitly characterize the quality of evidence and strength of recommendations.¹⁰ This system ranked quality of evidence from I to III and classified the strength of recommendations from A to E. Although widely adopted because of its simplicity, this system did not provide a detailed quality assessment of the evidence for benefits and harms important to patients, nor did it explicitly consider the benefits versus possible harms in the strength of recommendations.¹¹ By 2002, at least 121 different systems had been developed that were used in publications, systematic reviews, and CPGs.¹² More recently, government and professional organizations tasked with the development of CPGs in Canada, the United States, Australia, and Europe have developed and implemented a variety of systems to evaluate the quality of evidence and rate the strength of recommendations for CPGs.¹³⁻¹⁵ Many of these systems use different letters, numbers, or symbols to communicate similar recommendations on specific health issues, and often the same letter or number has different meanings in the various systems.¹⁶

Why GRADE?

To overcome the problems related to the inconsistent rating of evidence and the confusion with different rating systems, an international group of health professionals, researchers, and guideline developers created the GRADE system in 2004.¹⁷ The GRADE system rates the quality of evidence and grades the strength of recommendations in systematic reviews, health technology assessments, and CPGs. The GRADE system is structured and transparent. It is designed for systematic reviews (eg, Cochrane systematic reviews) and guidelines that examine alternative management strategies or interventions, which might include no intervention or current best practice.¹⁸ The GRADE system also informs clinician and patient decision making in clinical practice settings and supports production of informed health policy. It is now used or endorsed by at least 70 different organizations throughout the world, including the World Health Organization, UpToDate, and the Cochrane Collaboration.¹⁹

The CTFPHC was reestablished with a mandate to develop and disseminate CPGs for primary and preventive care based on systematic analysis of scientific

evidence.²⁰ The CTFPHC guidelines address primary or secondary prevention of conditions with a substantial health burden; topics are selected based on literature review and input from practitioners and the public.

How do I interpret GRADE recommendations?

Many family physicians and primary care health professionals (who are the target audience of the CTFPHC guidelines) are potentially unfamiliar with GRADE processes and therefore might be unsure of how to interpret the potential benefits and harms of practices recommended by the CTFPHC. This article outlines key concepts of the GRADE process using examples from the recently published CTFPHC guidelines on breast cancer screening²¹ to assist primary care practitioners in understanding the GRADE recommendations and discussing these recommendations with patients.

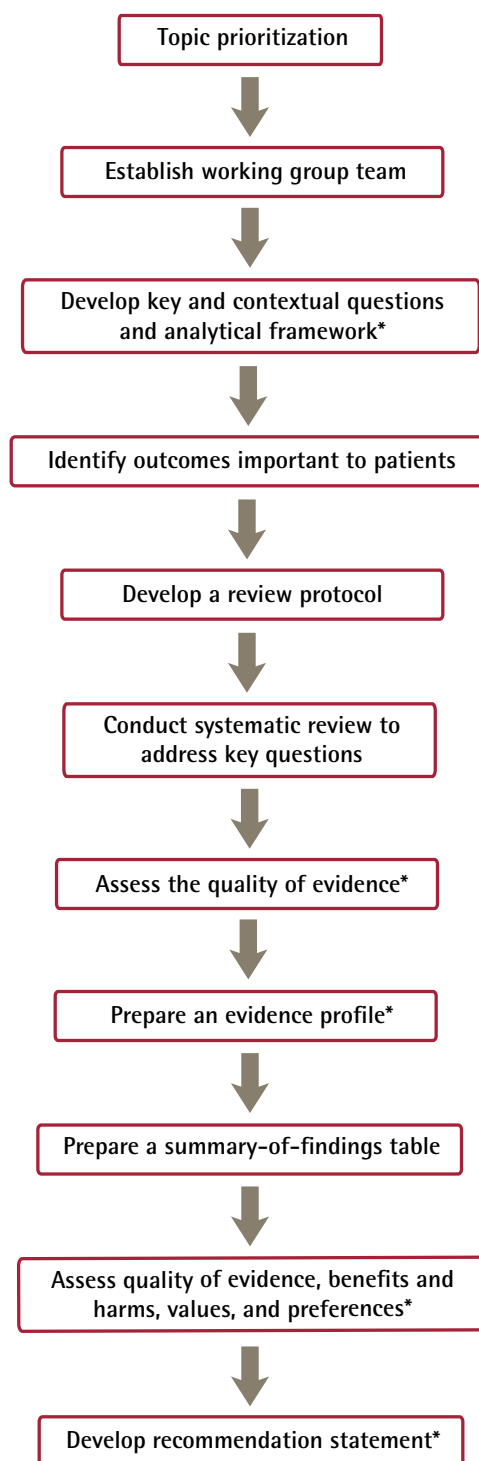
Although family physicians and other primary health care providers need not be aware of all the steps and processes involved in the development of CPGs using the GRADE methodology, consideration of several key elements in the GRADE guideline development process will ensure an overall understanding of the quality of evidence and strength of recommendations provided by this system. These elements include an understanding of the analytic framework and methods used in the literature review, the summaries of evidence tables, and the GRADE recommendations and how they can inform physician-patient decision making in clinical practice. More complete and detailed descriptions of the GRADE process for guideline developers and authors of systematic reviews have recently been published (www.gradeworkinggroup.org/publications/index.htm).

An overview of the CTFPHC guideline development process that highlights these key elements of GRADE is presented in **Figure 1**.

Does the guideline apply to my patient?

The importance of the analytic framework to practising family physicians and other primary health care practitioners is to provide an understanding of the patient populations to which the guideline recommendations would apply. This framework also identifies issues that were included or excluded from consideration in guideline development. The analytic framework and key questions provide the foundation for the literature review and guideline recommendation. This framework consists of a flow diagram with key questions and contextual questions. Key questions are those of main importance to clinicians and patients; they define the scope and focus of the evidence reviews. The contextual questions provide further information about how to interpret and apply the recommendations in our diverse Canadian settings; they also provide information about values and preferences, cost-effectiveness, and process and outcome

Figure 1. Steps in the CTFPHC guideline development process



CTFPHC—Canadian Task Force on Preventive Health Care.
*Highlighted steps are discussed in the paper.

indicators. Key questions are answered with a full systematic review, while for contextual questions a review of key studies and other systematic reviews is performed only for literature published in the past 5 years.

In the development of the analytic framework, guideline developers define the patient population, the intervention of interest, the comparator, and the outcome of interest. The process is also known as PICO (patient, intervention, comparator, outcome) and is now a widely accepted standard for development of guidelines and systematic reviews. An example of an analytic framework and key questions is shown in **Figure 2** and **Box 1**.

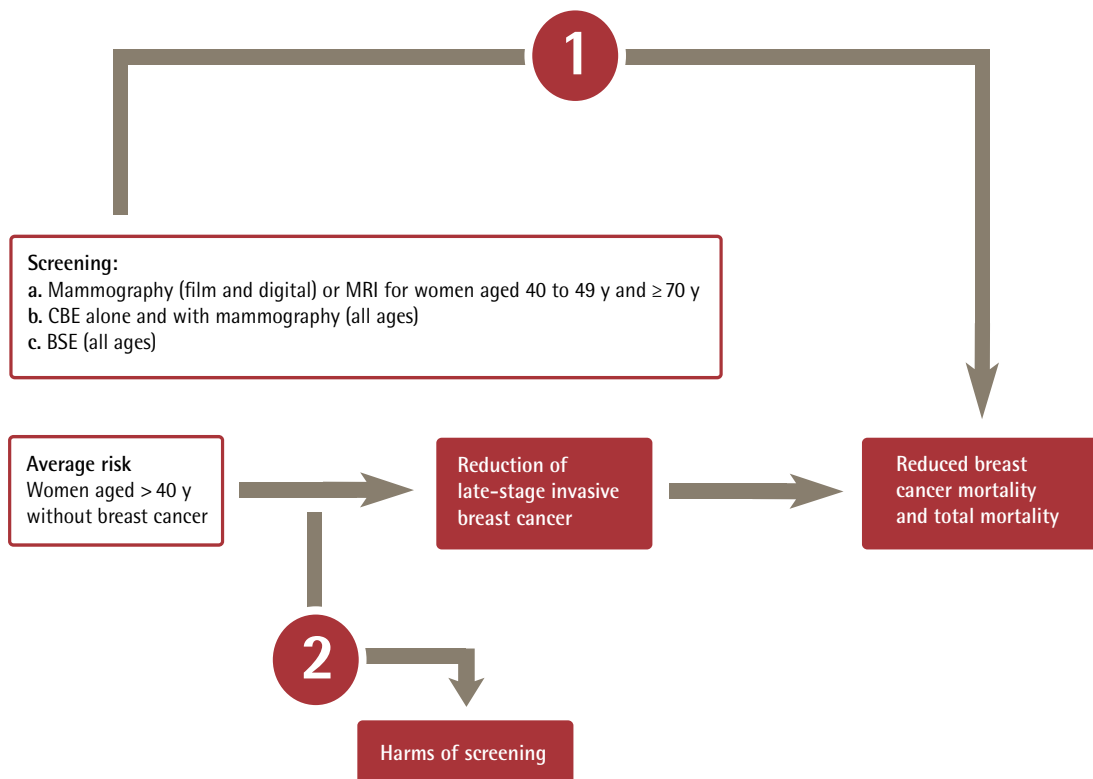
How good is the evidence?

In GRADE, the continuum of the quality of evidence is rated on a 4-point scale of high, moderate, low, or very low depending on the certainty that the results reflect the true effect of the intervention on the outcome (**Table 1**).²² Evidence is graded as high quality when the CTFPHC has high confidence that the true effect of an intervention or approach lies close to the estimate of effect, while lower-quality evidence indicates that the true effect might be substantially different from the

estimate of effect.²² The GRADE system considers several factors in determining the quality of the evidence. As a starting point, evidence of randomized controlled studies begins as high-quality evidence, while evidence from observational studies begins as low-quality evidence. Evidence can then be downgraded or upgraded depending on several factors. Evidence is downgraded based on consideration of 5 factors: risk of bias, inconsistency, indirectness, imprecision, and publication bias. Evidence can be upgraded based on 3 factors: large effect, dose response, and consideration of all possible confounders (**Tables 2 and 3**).²³⁻²⁹

For example, the evidence supporting the use of hormone replacement therapy for postmenopausal women in the early 1990s would have received a low-quality or low rating in the GRADE system because it was based on inconsistent observational studies.³⁰ Such a rating means further research could very likely have an important effect on the confidence in the estimate of effect and is likely to change the estimate. In fact, further research did show increased cardiovascular harms with hormone replacement, and this evidence would ultimately reverse the recommendation for hormone replacement therapy.

Figure 2. Example of an analytical framework using the breast cancer screening evidence review



BSE—breast self-examination, CBE—clinical breast examination, MRI—magnetic resonance imaging.

Box 1. Example of key questions using the breast cancer screening evidence review

Key questions

- 1a. Does screening with mammography (film and digital) or MRI decrease breast cancer mortality and all-cause mortality among women aged 40 to 49 y and ≥ 70 y?
- 1b. Does CBE screening decrease breast cancer mortality for women of all ages? Alone or with mammography?
- 1c. Does BSE practice decrease breast cancer mortality for women of all ages?
- 2a. What are the harms associated with screening with mammography (film and digital) and MRI?
- 2b. What are the harms associated with CBE?
- 2c. What are the harms associated with BSE?

Contextual questions

1. What are the patient values and preferences related to screening for breast cancer?
2. Are there subgroups of the Canadian population who have a higher burden of breast cancer or for whom it would be difficult to implement screening programs? Subgroup analysis that explores issues of burden of disease, screening rates, and special implementation issues includes the following:
 - aboriginal women,
 - women living in rural or remote areas, and
 - consideration of ethnicity
3. What is the optimal mammography screening frequency?
4. What is the cost-effectiveness of screening for breast cancer?

BSE—breast self-examination, CBE—clinical breast examination, MRI—magnetic resonance imaging.

Table 1. Interpretation of evidence levels used to GRADE CTFPHC recommendations

| EVIDENCE LEVEL | DEFINITION |
|-----------------|---|
| High | We judge evidence to be high quality when we are highly confident that the true effect lies close to that of the estimate of the effect. For example, evidence is judged to be high quality if all of the following apply: <ul style="list-style-type: none"> • there is a range of studies included in the analysis with no serious limitations; • there is little variation among studies; and • the summary estimate has a narrow CI |
| Moderate | We judge evidence to be moderate quality when we consider that the true effect is likely to be close to the estimate of the effect, but there is a possibility that it is substantially different. For example, evidence might be judged to be moderate quality if either of the following applies: <ul style="list-style-type: none"> • there are only a few studies and some have limitations but not serious flaws; or • there is some variation among studies or the CI of the summary estimate is wide |
| Low or very low | We judge evidence to be low or very low quality when the true effect might be substantially different from the estimate of the effect. For example, evidence might be judged as low quality if any of the following applies: <ul style="list-style-type: none"> • the studies have serious flaws; • there is important variation among studies; or • the CI of the summary estimate is very wide |

CTFPHC—Canadian Task Force on Preventive Health Care, GRADE—grading of recommendations, assessment, development, and evaluation.

Adapted from Balshem et al.²²

In summary, the GRADE system attempts to improve the estimate of the certainty of effects, thus providing clinicians and patients with more precise information on which to base their decisions.

The GRADE evidence tables

The GRADE system has developed specific approaches for the presentation of the results of the systematic literature reviews based on the analytic framework. The CTFPHC uses the GRADE evidence profile to present its results. The evidence profile table summarizes the size of the study population, the effect of the intervention, and the quality of the evidence. **Table 4** provides an example of an evidence profile developed for the CTFPHC guideline on screening for breast cancer in women aged 40 to 49 years.⁴

How does GRADE translate evidence into recommendations?

In GRADE, the assessment of the quality of evidence and the strength of recommendations are separate.

At present, GRADE recommendations are reported as either strong or weak. In addition to quality of evidence, GRADE also explicitly considers the balance between the benefits and harms, the values and preferences of patients, and the resource implications of an intervention in the determination of the strength of recommendations. While the quality of evidence and the balance between the benefits and harms are considered by the CTFPHC to be the most important elements, guideline developers might choose to place some or limited emphasis on resource implications and might have limited data on the values and preferences of patients for specific interventions.

Strong recommendations are more likely when there is a large difference between the benefits and harms and certainty around that difference, when there is greater certainty or similarity in values and preferences, and when the evidence quality is higher.³¹ Weak recommendations indicate that greater uncertainty exists (**Figure 3**).²⁹ Strong recommendations can be made even with low-quality evidence, assuming that the

Table 2. Factors that lead to decreasing evidence quality in the GRADE framework

| FACTORS | EXPLANATION | EXAMPLES |
|--|--|--|
| Risk of bias <ul style="list-style-type: none"> • Allocation concealment • Blinding • Loss to follow-up • Selective outcome reporting • Other limitations | <p>Whether those enrolling patients are aware of the group to which the next patient will be allocated</p> <p>Whether those involved in the study are aware of the study arm to which patients are allocated</p> <p>Whether many patients were lost during the trial, and whether the analysis was performed to take these losses into account</p> <p>When outcomes are reported based on the results and not based on a priori selection of important outcomes</p> <p>When there are other factors that could decrease the quality of the study</p> | An investigator assigns patients to the intervention group based on day of the week or date of birth, many patients drop out of the study, and the analysis is conducted to consider only those who adhered to the intervention ²³ |
| Inconsistency (variability in results) | If different studies provide widely different estimates of effect | There are 4 studies examining the effect of an intervention on an outcome; 2 studies show positive effects, 1 shows a negative effect, while the final study shows a null effect ²⁴ |
| Indirectness | The similarity between the question being addressed by the CTFPHC and the available evidence about the population, intervention, comparison groups, and outcomes | There are no data comparing drug A and B to each other, but there are data comparing drug A to a placebo and drug B to a placebo. There are no data on the effect of screening on cervical cancer mortality, but there are data on the surrogate outcome of invasive cervical cancer ²⁵ |
| Imprecision | Studies are imprecise if they have few patients or few events or very wide CIs around the estimates that lead to greater uncertainty about the findings | A group of studies has a wide CI and the clinical action that would be recommended changes depending on whether the upper boundary of the CI or the lower boundary of the CI represents the truth ²⁶ |
| Publication bias | An overestimate or underestimate in the effect that occurs when studies that find no effects are not published | Review authors fail to identify all relevant studies, negative studies are not published, or industry influence is suspected ²⁷ |

CTFPHC—Canadian Task Force on Preventive Health Care, GRADE—grading of recommendations, assessment, development, and evaluation.
Data from Guyatt et al.²³⁻²⁷

Table 3. Factors that lead to increasing evidence quality in the GRADE framework

| FACTORS | EXPLANATION | EXAMPLES |
|--|---|---|
| Large magnitude of effect and narrow CIs | When an intervention has a large effect on an outcome | Large effects are considered to be a risk ratio of at least 2 or a relative risk reduction of at least 50%. Having infants sleep on their backs has been associated with a 50% to 70% reduction in the risk of sudden infant death syndrome ²⁸ |
| Dose-response gradient | The presence of a dose-response relationship between intervention and outcome | When 50% of the population is immunized, it results in a 20% lower risk of disease; when 70% of the population is immunized, it results in a 40% lower risk; and when 90% of population is immunized, the risk is lowered by 80% ²⁹ |
| All plausible confounding | When inclusion of all plausible confounders or biases in observational studies that are unaccounted for in the analysis would result in an underestimate of an apparent treatment effect, or would increase the effect when no effect was found | Although an early study showed a positive association between vaccines and autism, further studies did not support this association, despite the fact that parents with autistic children might have been more likely to recall and report their vaccination experience than those whose children did not have autism ²⁸ |

GRADE—grading of recommendations, assessment, development, and evaluation.
Data from Guyatt et al.²⁸ and Santesso and Gauld.²⁹

Table 4. Evidence summary of benefits associated with screening mammography: *The content of the evidence profile table is provided in 13 standardized columns. The first column provides information about the number of studies and the study design used to determine the effectiveness of screening mammography for women in this age range (N=8 RCTs). Columns 2 to 7 provide an assessment of the quality of these studies. Footnotes provide further explanations as required. For instance, in column 3 (risk of bias) we indicate a serious concern about the potential risk of bias in the studies. This is based on the fact that only 3 of the 8 trials were considered truly randomized; in 5 of the trials it was not clear if investigators were blinded to the groups to which the patients were assigned or whether those enrolling patients were aware of which group patients were being assigned to. There were no other concerns about quality: results of all trials were consistent, the patients and the interventions were similar to the patients that we were studying, the samples sizes were large, the CIs were narrow, and there was no evidence of publication bias. Columns 8 to 11 in the table present the summary of our meta-analysis to determine the overall effectiveness of mammography screening in women aged 40–49 y. The number of deaths seen in the control and experimental groups are provided in columns 8 and 9. In columns 10 and 11, the estimates of the relative and absolute risk reductions that can be attributed to screening mammography are provided. Relative risk is used to compare risks between 2 different groups of people, often those who were exposed to an intervention and those who were not. Meta-analysis of mammography screening studies with women aged 40–49 y found a reduction of breast cancer risk of 15% (equivalent to an RR of 0.85) for women who were screened compared with women who were not screened. Absolute risk focuses on an individual's risk of getting a disease in a specific period of time and can be expressed as a percentage or a rate (eg, 10% or 1 in 10). In this example, this means that 474 fewer women per million (or 1 in approximately 2100) will die as a result of screening. Column 12 provides an overall rating of the quality. Column 13 highlights the importance of the results.*

| SUMMARY OF FINDINGS | | | | | | | | | | | | |
|---------------------|---|----------------------|---------------------------------------|---------------------------------------|-------------------------------------|----------------------|---|--------------------|------------------------------|---|----------|------------|
| NO. OF STUDIES | QUALITY ASSESSMENT FOR BREAST CANCER MORTALITY AT AGE 40–49 Y*W | | | | | | NO. OF BREAST CANCER DEATHS/ NO. OF PATIENTS (%) | | ESTIMATE OF MORTALITY EFFECT | | QUALITY | IMPORTANCE |
| | DESIGN | RISK OF BIAS | INCONSISTENCY | INDIRECTNESS | IMPRECISION | OTHER CONSIDERATIONS | SCREENING WITH MAMMOGRAPHY | CONTROL | RELATIVE (95% CI) | ABSOLUTE | | |
| 8 | Randomized trials [†] | Serious [‡] | No serious inconsistency [§] | No serious indirectness | No serious imprecision [¶] | None ^{††} | 448/152 300 (0.29) | 625/195 919 (0.32) | RR=0.85** (0.75 to 0.96) | 474 fewer per 1 000 000 (from 115 fewer to 792 fewer) NNS=2108 ^{†††} | Moderate | Critical |

NNS—number needed to screen, RCT—randomized controlled trial, RR—relative risk.

*The available data were based on women aged 39–49 y; however, the focus of the review was for those aged 40–49 y.

[†]Of the 8 studies, 5 were quasi-randomized and 3 were truly randomized.

[‡]Blinding and concealment were not clear for 5 studies, so only 3 trials were considered truly randomized.

[§]No heterogeneity exists; *P* value for testing heterogeneity is 0.48 and *I*²=0%.⁴

^{||}The question addressed is the same for the evidence regarding the population, intervention, comparator, and outcome.

[¶]Total sample size is large and the total number of events is > 300 (a threshold rule-of-thumb value).

^{††}Insufficient number of studies to assess publication bias.

**Estimates are based on a random-effects meta-analysis.

^{†††}NNS is the number of patients who would need to be invited to be screened about once every 2 years over a median of about 11 years to prevent 1 death from breast cancer.

balance between benefits and harms is clear and values and preferences are consistent, while weak recommendations can be made based on high-quality evidence. As an example, although only anecdotal evidence (low quality) suggests that parachutes are an effective intervention to reduce morbidity and mortality associated with jumping from an airplane,³² the recommendation to use a parachute would be classified as strong.

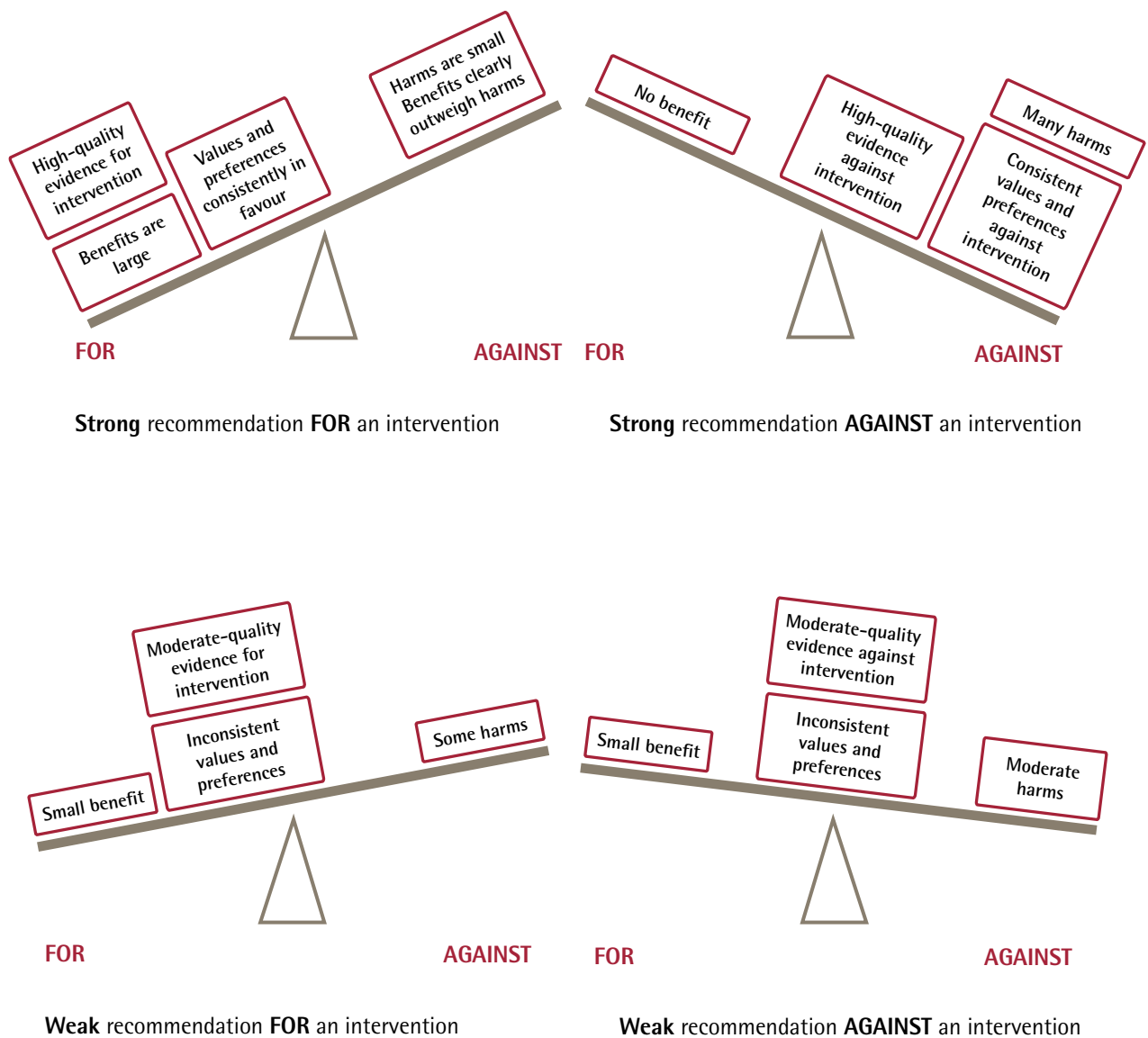
When the CTFPHC makes strong recommendations, clinicians can interpret this to mean that most individuals should receive the intervention in question, while for weak recommendations the focus shifts to helping patients make informed decisions, taking into account the benefits and harms, as well as their individual values and preferences. With weak recommendations, clinicians must recognize that different choices will be appropriate for different patients. For example, a weak recommendation against mammography in average-risk

women aged 40 to 49 years implies that (although most women of this age would not choose to be screened) regular screening could be appropriate in a 40- to 49-year-old woman who places a relatively higher value on preventing death from breast cancer and a relatively lower value on avoiding unnecessary tests and procedures. Similarly, a weak recommendation for mammography in average-risk women aged 50 to 74 years implies that screening would not necessarily be required or appropriate in a woman of this age who places a relatively lower value on preventing death from breast cancer and a relatively higher value on avoiding unnecessary tests and procedures.

How can GRADE recommendations inform patient-physician decision making?

Effective communication between physicians and patients is a key concept of family medicine and has

Figure 3. Balancing the benefits and harms to determine the strength of a recommendation



Based on a concept presented by Santesso and Gauld.²⁹


been associated with improved clinical outcomes for patients with a variety of health conditions.³³ Guideline developers are faced with the challenge of providing easily understood information on benefits and harms of recommendations to inform the discussion between physician and patient and assist in the decision-making process. Patients with the same condition or risk factors might have quite different values or preferences, life circumstances, or access to medical care. Understanding these differences along with the benefits and harms of the guideline recommendations

would support shared decision making by the patient and physician.

To support informed physician-patient decision making related to their guidelines, the CTFPHC has developed several tools in a collaborative manner with researchers and knowledge translation experts in Canada. These tools also undergo a process of internal and external peer review and user testing with patients and physicians to ensure that the scientific information is correct and presented in an easily understood format. Tools that have been developed and tested for

the CTFPHC guideline on screening with mammography include a video illustrating a doctor-patient interaction about screening, a list of frequently asked questions on breast cancer screening for patients, a flowchart to help women gauge whether screening is right for them, and decision aids that present risks and benefits in ways that patients can understand. An example of a tool developed to assist in decision making for screening with mammography for breast cancer for women between the ages of 40 and 49 is shown in **Figure 4**.^{*} These tools are available on the CTFPHC website (<http://canadiantaskforce.ca>).

Family physicians and GRADE

The GRADE system provides a rigorous approach to the development of CPGs that is increasingly being used by many professional and government organizations throughout the world. Family physicians need to be able to appreciate the benefits and harms and the certainty of evidence behind clinical recommendations. The use of the GRADE methodology by the developers of CPGs and systematic reviews can provide family physicians and other primary care health professionals with a guidepost of high quality for CPGs and systematic reviews. With the increasing use of GRADE, family physicians and other primary care health professionals should become familiar with the GRADE approach to assessment of the quality of evidence and the strength of recommendations so that they can effectively use CPGs and systematic reviews developed by this approach when making decisions with their patients. 

Dr Bell works in the Department of Family Medicine at the University of Alberta in Edmonton. **Dr Connor Gorber** works with the Public Health Agency of Canada in Ottawa, Ont. **Dr Tonelli** works in the Department of Medicine and in the Department of Public Health Sciences at the University of Alberta. **Dr Pottie** is Associate Professor in the Department of Family Medicine and in the Department of Epidemiology and Community Medicine, and works in the C.T. Lamont Primary Health Care Research Centre, all at the University of Ottawa. **Dr Singh** works in the Department of Internal Medicine and in the Department of Community Health Sciences at the University of Manitoba in Winnipeg and in the Department of Hematology and Oncology for CancerCare Manitoba. **Dr Joffres** works in the Faculty of Health Sciences at Simon Fraser University in Burnaby, BC. **Dr Shaw** works in the Department of Family Medicine at McMaster University in Hamilton, Ont.

Competing interests

None declared

Correspondence

Dr Sarah Connor Gorber, Public Health Agency of Canada, 785 Carling Ave, 514B4, Ottawa, ON K1A 0K9; e-mail sarah.connorgorber@phac-aspc.gc.ca

References

- Agency for Healthcare Research and Quality [website]. *National Guideline Clearinghouse*. Rockville, MD: Agency for Healthcare Research and Quality. Available from: www.guideline.gov. Accessed 2012 Feb 16.
- Guidelines International Network* [website]. Berlin, Ger: Guidelines International Network; 2013. Available from: www.g-i-n.net. Accessed 2012 Feb 16.
- Canadian Medical Association [website]. *CMA Infobase: clinical practice guidelines*. Ottawa, ON: Canadian Medical Association; 2013. Available from: www.cma.ca/clinicalresources/practiceguidelines. Accessed 2012 Jun 12.
- Han PK, Klabunde CN, Breen N, Yuan G, Grauman A, Davis WW, et al. Multiple clinical practice guidelines for breast and cervical cancer screening: perceptions of US primary care physicians. *Med Care* 2011;49(2):139-48.
- Brawley O, Byers T, Chen A, Pignone M, Ransohoff D, Schenk M, et al. New American Cancer Society process for creating trustworthy cancer screening guidelines. *JAMA* 2011;306(22):2495-9.
- Shaneyfelt TM, Centor RM. Reassessment of clinical practice guidelines: go gently into that good night. *JAMA* 2009;301(8):868-9.
- Kuehn BM. IOM set out "gold standard" practices for creating guidelines, systematic reviews. *JAMA* 2011;305(18):1846-8.
- Institute of Medicine. *Clinical practice guidelines we can trust*. Washington, DC: National Academies Press; 2011.
- Institute of Medicine. *Finding what works in health care: standards for systematic reviews*. Washington, DC: National Academies Press; 2011.
- Canadian Task Force on the Periodic Health Examination. The periodic health examination. *Can Med Assoc J* 1979;121(9):1193-254.
- Atkins D, Eccles M, Flottrop S, Guyatt GH, Henry D, Hill S, et al. Systems for grading the quality of evidence and the strength of the recommendations I: critical appraisal of existing approaches. The GRADE Working Group. *BMC Health Serv Res* 2004;4(1):38.
- Systems to rate the strength of scientific evidence*. Summary, evidence report/technology assessment: no. 47. AHRQ pub no. 02-E015. Rockville, MD: Agency for Healthcare Research and Quality; 2002.
- Harbour R, Miller J. A new system for grading recommendations in evidence based guidelines. *BMJ* 2001;323(7308):334-6.
- Dahm P, Djubegovic B. The Australian 'FORM' approach to guideline development: the quest for a perfect system. *BMC Med Res Methodol* 2011;11:17.
- US Preventive Services Task Force. *Grade definitions*. Rockville, MD: US Preventive Services Task Force; 2008. Available from: www.uspreventiveservicestaskforce.org/uspstf/grades.htm. Accessed 2012 Aug 30.
- Schünemann HJ, Best D, Vist G, Oxman AD; GRADE Working Group. Letters, numbers, symbols and words: how to communicate grades of evidence and recommendations. *CMAJ* 2003;169(7):677-80.
- GRADE Working Group. Grading the quality of evidence and strength of recommendations. *BMJ* 2004;328(7454):1490.
- Guyatt G, Oxman AD, Akl EA, Kunz R, Vist R, Brozek J, et al. GRADE guidelines: 1. Introduction—GRADE evidence profiles and summary finding tables. *J Clin Epidemiol* 2011;64(4):383-94.
- GRADE Working Group [website]. *Organizations that have endorsed or that are using GRADE*. GRADE Working Group; 2013. Available from: www.gradeworkinggroup.org/society/index.htm. Accessed 2013 Oct 31.
- Birtwhistle R, Pottie K, Shaw E, Dickinson JA, Brauer P, Fortin M, et al. Canadian Task Force on Preventive Health Care. We're back! *Can Fam Physician* 2012;58:13-5 (Eng), e1-4 (Fr).
- The Canadian Task Force on Preventive Health Care. Recommendations on screening for breast cancer in average-risk women aged 40-74 years. *CMAJ* 2011;183(17):1991-2001.
- Balshem H, Helfand M, Schunemann HJ, Oxman AD, Kunz R, Brozek J, et al. GRADE guidelines: 3. Rating the quality of evidence—introduction. *J Clin Epidemiol* 2011;64(4):401-6.
- Guyatt GH, Oxman AD, Vist G, Kunz R, Brozek J, Alonso-Coello P, et al. GRADE guidelines: 4. Rating the quality of evidence—risk of bias. *J Clin Epidemiol* 2011;64(4):407-15.
- Guyatt GH, Oxman AD, Kunz R, Woodcock J, Brozek J, Helfand M, et al. GRADE guidelines: 7. Rating the quality of evidence—inconsistency. *J Clin Epidemiol* 2011;64(12):1294-302.
- Guyatt GH, Oxman AD, Kunz R, Woodcock J, Brozek J, Helfand M, et al. GRADE guidelines: 8. Rating the quality of evidence—indirectness. *J Clin Epidemiol* 2011;64(12):1303-10.
- Guyatt GH, Oxman AD, Kunz R, Brozek J, Alonso-Coello P, Rind D, et al. GRADE guidelines: 6. Rating the quality of evidence—imprecision. *J Clin Epidemiol* 2011;64(12):1283-93.
- Guyatt GH, Oxman AD, Montori V, Vist G, Kunz R, Brozek J, et al. GRADE guidelines: 5. Rating the quality of evidence—publication bias. *J Clin Epidemiol* 2011;64(12):1277-82.
- Guyatt GH, Oxman AD, Sultan S, Glasziou P, Akl EA, Alonso-Coello P, et al. GRADE guidelines: 9. Rating up the quality of evidence. *J Clin Epidemiol* 2011;64(12):1311-6.
- Santesso N, Gauld M. *GRADE and guideline development*. Paper presented at: Canadian Task Force on Preventive Health Care Meeting; November 2011; Ottawa, ON.
- Guyatt GH, Oxman AD, Vist GE, Kunz R, Falck-Ytter Y, Alonso-Coello P, et al. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *BMJ* 2008;336(7650):924-6.
- Guyatt GH, Oxman AD, Kunz R, Falck-Ytter Y, Vist GE, Liberati A, et al. GRADE Working Group. Rating quality of evidence and strength of recommendations: going from evidence to recommendations. *BMJ* 2008;336(7652):1049-51.
- Smith GC, Pell JP. Parachute use to prevent death and major trauma related to gravitational challenge: systematic review of randomized controlled trials. *BMJ* 2003;327(7429):1459-61.
- Stewart MA. Effective physician-patient communication and health outcomes: a review. *CMAJ* 1995;152(9):1423-33.

^{*}Figure 4 is available at www.cfp.ca. Go to the full text of this article online, then click on **CFPlus** in the menu at the top right-hand side of the page.