

# Understanding and communicating risk

## Measures of outcome and the magnitude of benefits and harms

Neil R. Bell MD SM CCFP FCFP James A. Dickinson MBBS PhD CCFP FRACGP Roland Grad MDCM MSc CCFP FCFP  
Harminder Singh MD MPH FRCPC Danielle Kasperavicius MPH Brett D. Thombs PhD

Understanding and communicating risk, such as the magnitude or effect size of the benefits and harms associated with preventive screening and other health care management interventions, is fundamental to evidence-based decision making by physicians and shared decision making with patients. Family physicians encounter several different measures of outcome and of magnitude or effect size that are used to describe benefits and harms. Some of these measures might be inappropriate to describe the benefits of screening, and others might be difficult for patients and physicians to understand.<sup>1,2</sup>

Effective communication of harms and benefits with patients can be challenging owing to limited basic numeracy and health literacy skills among many patients. Numeracy refers to an individual's ability to understand basic numerical concepts.<sup>2-4</sup> For example, patients might have difficulty understanding basic probability or converting between percentage and rate.<sup>4</sup> This issue has been described as "collective statistical illiteracy" and has been identified as an important barrier to patients' understanding of the benefits and harms associated with health decision making.<sup>2,3</sup> In Canada in 2014, 55% of Canadian adults were found to have inadequate numeracy skills.<sup>5</sup> *Health literacy* has been defined as "the ability to access, understand, evaluate and communicate information as a way to promote, maintain and improve health in a variety of settings across the life-course."<sup>6</sup> Similar to numeracy, a 2008 report from the Canadian Public Health Association found 55% or 11.7 million adults were estimated to have less than adequate health literacy skills.<sup>6</sup>

Physicians might also have challenges in understanding statistical measures. A recent survey of physicians from 8 countries, with various levels of training and specialty backgrounds, found that most had difficulty understanding commonly used measures of magnitude, such as relative risk (RR).<sup>7</sup> Physicians also had difficulty understanding basic concepts of risk and probability that are important in understanding and communicating the benefits and harms associated with preventive health care.<sup>2</sup> A 2012 study conducted in the United States found that most primary care practitioners had difficulty correctly interpreting the results of cancer screening studies and misinterpreted the measures used to describe the benefits of screening.<sup>8</sup> In spite of increased training in evidence-based medicine, statistics, and probabilistic reasoning in medical school curriculums, the level of physician statistical literacy does not appear to have increased during the past 40 years.<sup>9,10</sup>

This article will review and discuss the appropriateness, advantages, and disadvantages of commonly used measures of outcome (overall and disease-specific mortality, incidence or number of new cases, and 5- and 10-year survival) and magnitude or effect size (RR, absolute risk [AR], natural frequency, and number needed to screen [NNS]) that family physicians encounter when interpreting and communicating the harms and benefits of screening. The choice of outcome and magnitude or effect size measures and the format or frame used to present the information can result in different levels of understanding of the same probabilistic information by physicians and patients.<sup>1,2</sup>

### Key points

- ▶ Overall or disease-specific mortality demonstrated in a randomized clinical trial provides the highest-quality evidence for estimates of the benefits of preventive screening. Improved 5- or 10-year survival provides exaggerated estimates of the benefits of preventive screening owing to lead-time, length-time, and overdiagnosis bias. Increased disease detection (incidence or number of new cases) provides exaggerated estimates of the benefits of preventive screening owing to overdiagnosis.
- ▶ More than 50% of Canadians have inadequate numeracy and health literacy skills. Communicating the harms and benefits of preventive screening might be challenging with these patients.
- ▶ Patient and physician understanding of the magnitude or effect size of benefits and harms is improved when they are expressed as natural frequencies or in absolute terms such as absolute risk reduction with baseline risk. Visual displays of measures of magnitude or effect size such as 1000-person diagrams increase understanding for both physicians and patients.
- ▶ Proportionate measures of magnitude or effect size, such as relative risk, can lead patients and physicians to overestimate benefits; number needed to screen is less well understood by patients compared with other measures of magnitude or effect size, such as absolute risk reduction or natural frequencies.

In a previous article<sup>11</sup> in this Prevention in Practice series, we introduced a clinical case on screening for lung cancer with low-dose computed tomography (CT). Let's briefly review the case.

John is a 66-year-old man with a more than 30 pack-year history of smoking. John is a candidate for screening for lung cancer with low-dose CT based on recommendations from the Canadian Task Force on Preventive Health Care (CTFPHC). At a previous visit, John decided not to proceed with screening. After reading a news article on screening, John's wife asked him to reconsider his decision. At this visit, John and his wife are asking for additional information about screening for lung cancer. How could you best describe the potential benefits and harms of screening to John and his wife?

### Outcome measures encountered in preventive screening

Outcome measures of preventive screening that are frequently encountered by family physicians in information provided by government agencies, guideline developers, and advocacy groups on the benefits or harms of screening are outlined in **Table 1**. These measures include overall and disease-specific mortality, 5- or 10-year survival rates, and increased detection of disease (incidence or number of new cases).

**Overall and disease-specific mortality.** Reduced overall or disease-specific mortality in a randomized controlled trial provides the highest-quality evidence for estimates of the benefit of cancer screening.<sup>2,12-18</sup> Mortality rates in randomized controlled trials are unaffected by lead-time, length-time, or overdiagnosis bias. Mortality statistics are not affected by timing of disease diagnosis because all deaths that occur in the study population would be included in the overall or disease-specific mortality.<sup>2,12-18</sup> The main limitation of overall mortality is the requirement for very large sample sizes, while the limitations of disease-specific mortality relate to assignment of the cause of death and the potential difficulty in measuring harms related to screening.<sup>13</sup>

**Five- or 10-year survival rates.** Five- or 10-year survival rates are common survival statistics that consider patients *diagnosed* with a disease. Physicians will be aware of the frequent use of 5- and 10-year survival rates in cancer to describe the benefits of treatment and prognosis with different stages of disease.<sup>2,12,19</sup> However, physicians must be cautious in the use of 5- or 10-year survival rates to describe the benefits of preventive screening.<sup>2,14,15,20</sup> This occurs because of the increasing sensitivity of screening tests to identify smaller lesions and the heterogeneity of cancer progression where there is variability in the rates of cancer progression, with some cancers destined to fail to progress or grow so slowly that the patient will die of other causes.<sup>2,18,20</sup> Under these circumstances, 5- and 10-year survival can provide exaggerated estimates of the benefits of cancer screening.<sup>2,14,15</sup> Lead-time, length-time, and overdiagnosis bias describe how these factors could artificially increase 5- and 10-year survival rates.<sup>2,14,15</sup>

Lead-time bias is the earlier identification of patients with a disease owing to screening, but who have no change in the actual time of death owing to screening. Five- or 10-year survival rates are increased by early diagnosis because of the increased time from diagnosis to death in screened compared with unscreened patients. In both groups, there would be no change in the time of death.<sup>2,14,15</sup>

Length-time bias is the tendency for screening to identify more indolent or slower-growing cancers. Faster-growing, more aggressive tumours are more likely to become symptomatic and detected clinically rather than through screening; therefore, patients with screen-detected cancers have longer survival compared with those with clinically detected cancers. This apparent improvement in survival is incorrectly attributed to screening.<sup>2,14,15</sup>

Overdiagnosis is the detection of an "abnormality" or a "condition" that would ultimately not go on to cause symptoms or death.<sup>18,20</sup> In overdiagnosis bias, survival rates are increased because of the detection of patients with nonprogressive cancers (overdiagnosed patients) with screening. The overall number of patients who die of cancer would be unchanged but the 5- or 10-year

**Table 1. Outcome measures encountered in preventive screening**

MEASURE	HOW TO CALCULATE	ADVANTAGES AND DISADVANTAGES IN PATIENT RISK COMMUNICATION
Mortality (overall and disease specific)	The number of patients who died divided by the total number of patients in the study population. Mortality would be calculated separately for the control and intervention groups in randomized controlled trials	<ul style="list-style-type: none"> <li>Provides the highest-quality estimate of the benefits of cancer screening</li> <li>Results unaffected by lead-time, length-time, or overdiagnosis bias</li> </ul>
5- or 10-year survival rates (absolute rate)	The number of individuals who are alive at 5 or 10 years after the time of diagnosis of disease divided by the total number diagnosed with the disease	<ul style="list-style-type: none"> <li>Provides exaggerated estimates of the benefits of preventive screening owing to lead-time, length-time, and overdiagnosis bias</li> </ul>
Incidence (new cases)	The number of new events or cases that develop during a given time period in the total population at risk	<ul style="list-style-type: none"> <li>Provides exaggerated estimates of benefits of preventive screening owing to overdiagnosis</li> </ul>

survival rate is increased because of the inclusion of patients with nonprogressive or overdiagnosed cancer in the survival estimates of screened patients.<sup>2,15,18,20</sup>

**Incidence.** Incidence is the number of new cases or events that develop in a population at risk during a specified time interval. In screening, incidence can be increased owing to the detection of patients with disease that could progress to cause symptoms or death, or the detection of patients who are overdiagnosed.<sup>18</sup> Patients overdiagnosed and treated would be considered “survivors” of the disease, although if undiagnosed they would not have experienced symptoms or died of the disease.<sup>18</sup> Incidence provides exaggerated estimates of the benefits of preventive screening owing to overdiagnosis. Thyroid cancer provides an example of a disease where an increased incidence of disease was observed from screening.<sup>21-23</sup> In Canada<sup>21</sup> and other countries,<sup>22,23</sup> the increased detection or incidence of thyroid cancer

was found to have almost no effect on thyroid cancer mortality owing to overdiagnosis.

**Measures of magnitude or effect size encountered in preventive screening**

**Table 2** outlines the commonly used measures of magnitude or effect size that are used to describe outcome measures of screening and highlights the advantages and disadvantages of each measure. All examples in this table are taken from the National Lung Screening Trial.<sup>24-26</sup>

**Natural frequencies.** *Natural frequencies* can be defined as the number of persons with events juxtaposed with a baseline denominator of persons at risk.<sup>1,2,27</sup> In preventive screening, it is common to present the expected probabilities of outcomes in a population of 1000 persons undergoing screening compared with an equivalent population that is not being screened. Presentation of the results of screening trials in natural frequency

**Table 2. Measures of magnitude or effect size encountered in preventive screening**

MEASURE	ABBREVIATION	HOW TO CALCULATE	ADVANTAGES AND DISADVANTAGES IN PATIENT RISK COMMUNICATION	EXAMPLE* (REDUCTION IN LUNG CANCER MORTALITY)
Natural frequency	NF	Number of persons with events in a population	<ul style="list-style-type: none"> <li>• Highest levels of patient understanding and satisfaction</li> <li>• Denominator of 1000 people increases patient understanding of harms and benefits</li> <li>• Understanding increased when baseline risk is included</li> </ul>	13 of 1000 people died of lung cancer with screening; 16 of 1000 people died from lung cancer without screening. Thus, there were 3 of 1000 fewer deaths from lung cancer with screening
Absolute risk	AR	The number of events in the screened or control groups divided by the number of people in that group	<ul style="list-style-type: none"> <li>• Increases patient understanding of risk</li> <li>• Understanding increased when baseline risk is included</li> </ul>	AR in control group = 1.66% AR in screened group = 1.33%
Absolute risk reduction	ARR	Difference in the event rates between the screened and control arms of the study		ARR = 1.66% - 1.33% = 0.33%
Relative risk	RR	Ratio of the outcome measure (eg, overall mortality) in the screened group compared with the unscreened group	<ul style="list-style-type: none"> <li>• Can cause exaggerated perceived screening or treatment effects</li> </ul>	RR = 0.80
Relative risk reduction	RRR	The difference in event rates between the screened and control groups divided by the event rate in the control group	<ul style="list-style-type: none"> <li>• Can exaggerate the perceived treatment effect for both physicians and patients. Often presented as percentage without baseline risk</li> </ul>	RRR = 1.66 - 1.33/1.66 = 0.20 RRR = 20%
Number needed to screen	NNS	Reciprocal of the ARR	<ul style="list-style-type: none"> <li>• Decreased level of patient understanding compared with other measures of magnitude or effect size</li> </ul>	NNS = 308†

\*All measures describe the same reduction in lung cancer mortality.

†Differs slightly from 1/ARR in this example owing to rounding.

All examples are taken from the National Lung Screening Trial.<sup>24</sup>

Estimates were taken from the Canadian Task Force on Preventive Health Care systematic review and meta analysis on screening for lung cancer.<sup>25,26</sup>

formats has been found to improve the understanding of magnitude or effect size on benefits and harms by patients.<sup>1,2</sup> Patient understanding is increased with the inclusion of information on the baseline risk.<sup>1,2</sup>

**Absolute risk and absolute risk reduction (ARR).** The AR is the number of events in the screened or control groups divided by the number of people in that group. The ARR is the difference in the event rates between the control and treatment groups. Physicians and patients have greater understanding of risk differences when the results are presented as ARR compared with RR.<sup>1,2</sup>

**Relative risk and relative risk reduction (RRR).** The RR is the event rate in the screened group divided by the event rate in the control group. The RRR is the difference in event rates between the screened and control groups divided by the event rate in the control group. Presentation of risk in the form of RR or RRR can cause exaggerated perceived screening or treatment effects.<sup>1,2</sup> This is particularly problematic when the base rate is very low, in which case small changes in ARR can lead to a large change in RRR. This effect is shown in the example in **Table 2** of lung cancer screening.<sup>24-26</sup>

**Number needed to screen.** The NNS is the number of patients who must be screened in order to prevent 1 adverse event. The NNS is similar to the number needed to treat. The NNS is calculated as the reciprocal of the ARR. Although it has been advocated as a more easily understood measure of the benefits of an intervention, the presentation of results in this format resulted in lower levels of understanding by patients compared with other formats, such as natural frequencies.<sup>1</sup>

## Visual aids

The addition of visual aids to numerical information on risk improves the accuracy and comprehension of numerical data on risk by patients and physicians. Visual displays are better understood when they include both the “sick” and “healthy” populations. Commonly used formats include icon arrays, such as 1000-person diagrams and bar graphs. Examples of 1000-person diagrams developed by the CTFPHC on screening for breast, prostate, and lung cancer can be found on the CTFPHC website ([www.canadiantaskforce.ca](http://www.canadiantaskforce.ca)) and in previous articles in this series.<sup>28</sup>

## Framing of risk information

Framing is the expression of logically equivalent information (whether numerical or verbal) in different ways. Positive and negative frames refer to whether an outcome is described as a chance of survival (positive) or a chance of death (negative). Evidence suggests that positive framing is more effective than negative framing in persuading people to choose risky treatment options such as high-risk surgery.<sup>1,27,29</sup> Presenting information as

RR as opposed to ARR increased perceptions of treatment or screening benefit.<sup>1,2,29</sup>

## Bottom line

Overall or disease-specific mortality from randomized clinical trials provides the highest-quality evidence for the estimates of the benefits of preventive screening, while other measures of outcomes, such as 5- or 10-year survival rates or incidence, can result in exaggerated estimates of the benefits of screening. In presenting the magnitude or effect size of benefits and harms to patients, physicians should consider using measures of magnitude or effect size that are most effective in improving the understanding of patients. The most easily understood measure of magnitude or effect size is natural frequency supported by the use of knowledge translation tools, such as decision aids that feature 1000-person diagrams. Shared decision making and communication of the magnitude of benefits and harms might be challenging with some patients owing to inadequate health literacy and numeracy skills.

## Back to John

You take the opportunity to review the CTFPHC 1000-person diagram on lung cancer screening with John and his wife. This diagram is a visual representation of natural frequencies. After clarifying what the colours in the diagram indicate, John and his wife leave with a good understanding of the potential harms and benefits of screening and plan to discuss further before making a decision. John indicates that he will follow up if he wishes to proceed with the low-dose CT scan. 🌿

**Dr Bell** is Professor in the Department of Family Medicine at the University of Alberta in Edmonton. **Dr Dickinson** is Professor in the Department of Family Medicine and the Department of Community Health Sciences at the University of Calgary in Alberta. **Dr Grad** is Associate Professor in the Department of Family Medicine at McGill University and Senior Investigator at the Lady Davis Institute in Montreal, Que. **Dr Singh** is Associate Professor in the Department of Internal Medicine and the Department of Community Health Sciences at the University of Manitoba in Winnipeg and the Department of Hematology and Oncology for CancerCare Manitoba. **Ms Kasperavicius** is Research Coordinator for the Knowledge Translation Program at St Michael's Hospital in Toronto, Ont. **Dr Thombs** is Professor and William Dawson Scholar in the Faculty of Medicine at McGill University and Chair of the Canadian Task Force on Preventive Health Care.

### Competing interests

All authors have completed the International Committee of Medical Journal Editors' Unified Competing Interest form (available on request from the corresponding author). **Dr Singh** reports grants from Merck Canada, personal fees from Pendopharm, and personal fees from Ferring Canada, outside the submitted work. The other authors declare that they have no competing interests. For additional information on the conflict of interests of the Canadian Task Force on Preventive Health Care, please visit [www.canadiantaskforce.ca](http://www.canadiantaskforce.ca).

### Correspondence

**Dr Neil R. Bell**; e-mail [nrbell@telusplanet.net](mailto:nrbell@telusplanet.net)

### References

- Zipkin DA, Umscheid CA, Keating N, Allen E, Aung K, Beyth R, et al. Evidence-based risk communication: a systematic review. *Ann Intern Med* 2014;161(14):270-80.
- Gegerzenzer G, Gaissmaier W, Kurz-Milcke E, Schwarz L, Woloshin S. Helping doctors and patients make sense of health statistics. *Psychol Sci Public Interest* 2007;8(2):53-96.
- Peters E, Hibbard J, Slovic P, Dieckmann N. Numeracy skill and the communication, comprehension, and use of risk-benefit information. *Health Affairs* 2007;26(3):741-8.
- Schwartz LM, Woloshin S, Black WC, Welch GH. The role of numeracy in understanding the benefits of screening mammography. *Ann Intern Med* 1997;127(11):966-72.
- The Conference Board of Canada. *How Canada performs. Educational provincial rankings. Adults with inadequate numeracy skills*. Ottawa, ON: The Conference

- Board of Canada; 2014. Available from: [www.conferenceboard.ca/hcp/provincial/education/adlt-lownumn.aspx](http://www.conferenceboard.ca/hcp/provincial/education/adlt-lownumn.aspx). Accessed 2017 Dec 6.
6. Rootman I, Gordon-El-Bihbety D. *A vision for a health literate Canada: report of the Expert Panel on Health Literacy*. Ottawa, ON: Canadian Public Health Association; 2008.
  7. Johnston BC, Alonso-Coello P, Fredrich JO, Mustafa RA, Tikkinen KA, Neumann I, et al. Do clinicians understand the size of treatment effects? A randomized survey across 8 countries. *CMAJ* 2016;188(1):25-32.
  8. Wegwarth O, Schwartz LM, Woloshin S, Gaissmaier W, Gigerenzer G. Do physicians understand cancer screening statistics? A national survey of primary care physicians in the United States. *Ann Intern Med* 2012;156(5):340-49.
  9. Martyn C. Risky business: doctors' understanding of statistics. *BMJ* 2014;349:g5619.
  10. Bartlett G, Gagnon J. Physicians and knowledge translation of statistics: mind the gap. *CMAJ* 2016;188(1):11-12. Epub 2015 Nov 30.
  11. Grad R, Légaré F, Bell NR, Dickinson JA, Singh H, Moore AE, et al. Shared decision making in preventive health care. What it is; what it is not. *Can Fam Physician* 2017;63:682-84 (Eng), e377-80 (Fr).
  12. Measurement of progress against cancer. Extramural Committee to Assess Measures of Progress Against Cancer. *J Natl Cancer Inst* 1990;82(10):825-35.
  13. Kasperavicius D, Bell NR, Grad R, Dickinson JA, Moore AE, Singh H, et al. Response to using disease-specific mortality in discussion with patients [Letters]. *Can Fam Physician* 2017;63:828-30.
  14. Welch HG, Schwartz LM, Woloshin S. Are increasing 5-year survival rates evidence of success against cancer? *JAMA* 2000;283(22):2975-8.
  15. Welch HG, Woloshin S, Schwarz LM, Gordis L, Gøtzsche PC, Harris R, et al. Overstating the evidence for lung cancer screening: the International Early Lung Cancer Action Program (I-ELCAP) study. *Arch Intern Med* 2007;167(21):2289-95.
  16. Alibhai SM. Cancer screening: the importance of outcome measures. *Crit Rev Oncol Hematol* 2006;57(3):215-24. Epub 2005 Dec 20.
  17. Croswell JM, Ransohoff DF, Kramer BS. Principles of cancer screening: lessons from history and study design issues. *Semin Oncol* 2010;37(3):202-15.
  18. Welch HG, Black WC. Overdiagnosis in cancer. *J Natl Cancer Inst* 2010;102(9):605-13. Epub 2010 Apr 22.
  19. Mahoney J. Survival rates increase for Canadians with cancer: Statscan. *The Globe and Mail* 2010 Sep 12. Available from: [www.theglobeandmail.com/life/health-and-fitness/survival-rates-increase-for-canadians-with-cancer-statscan/article556422/](http://www.theglobeandmail.com/life/health-and-fitness/survival-rates-increase-for-canadians-with-cancer-statscan/article556422/). Accessed 2017 Dec 6.
  20. Moynihan R, Doust J, Henry D. Preventing overdiagnosis: how to stop harming the healthy. *BMJ* 2012;344:e3502.
  21. Topstad D, Dickinson JA. Thyroid cancer incidence in Canada: a national cancer registry analysis. *CMAJ Open* 2017;5(3):E612-6.
  22. Vaccarella S, Franceschi S, Bray F, Wild C, Plummer M, Dal Maso L. Worldwide thyroid-cancer epidemic? The increasing impact of overdiagnosis. *N Engl J Med* 2016;375(7):614-7.
  23. Ahn HS, Kim HJ, Welch HG. Korea's thyroid-cancer "epidemic"—screening and overdiagnosis. *N Engl J Med* 2014;371(19):1765-7.
  24. National Lung Screening Trial Research Team; Aberle DR, Adams AM, Berg CD, Black WC, Clapp JD, et al. Reduced lung-cancer mortality with low-dose computed tomographic screening. *N Engl J Med* 2011;365(5):395-409. Epub 2011 Jun 29.
  25. McMaster Evidence Review and Synthesis Centre Team. *Screening for lung cancer: systematic review and meta-analyses*. Calgary, AB: Canadian Task Force on Preventive Health Care; 2015. Available from: <https://canadiantaskforce.ca/wp-content/uploads/2016/03/lung-cancer-screening-systematic-reviewfinal-2.pdf>. Accessed 2017 Dec 13.
  26. Usman Ali M, Miller J, Peirson L, Fitzpatrick-Lewis D, Kenny M, Sherifali D, et al. Screening for lung cancer: a systematic review and meta-analysis. *Prev Med* 2016;89:301-14. Epub 2016 Apr 26.
  27. Gigerenzer G, Edwards A. Simple tools for understanding risks: from innumeracy to insight. *BMJ* 2003;327(7417):741-4.
  28. Moore AE, Straus SE, Kasperavicius D, Bell NR, Dickinson JA, Grad R, et al. Knowledge translation tools in preventive health care. *Can Fam Physician* 2017;63:853-8 (Eng), e466-72 (Fr).
  29. Moxey A, Dip G, O'Connell, McGettigan P, Henry D. Describing treatment effects to patients. How they are expressed makes a difference. *J Gen Intern Med* 2003;18(11):948-59.

### Suggested additional reading

1. Gigerenzer G, Gaissmaier W, Kurtz-Milcke E, Schwartz LM, Woloshin S. Helping doctors and patients make sense of health statistics. *Psychol Sci Public Interest* 2007;8(2):53-96.
2. Zipkin DA, Umscheid CA, Keating N, Allen E, Aung K, Beyth R, et al. Evidence-based risk communication: a systematic review. *Ann Intern Med* 2014;161(4):270-80.

This article is eligible for Mainpro+ certified Self-Learning credits. To earn credits, go to [www.cfp.ca](http://www.cfp.ca) and click on the Mainpro+ link.

Cet article se trouve aussi en français à la page 186.