Deceiving Numbers: Survival Rates and Their Impact on Doctors' Risk Communication

Odette Wegwarth, PhD, Wolfgang Gaissmaier, PhD, Gerd Gigerenzer, PhD

Background. *Increased 5-y survival for screened patients is often inferred to mean that fewer patients die of cancer. However, due to several biases, the 5-y survival rate is a misleading metric for evaluating a screening's effectiveness. If physicians are not aware of these issues, informed screening counseling cannot take place.* ***Methods.*** *Two questionnaire versions (''group'' and ''time'') presented 4 conditions: 5-y survival (5Y), 5-y survival and annual disease-specific mortality (5YM), annual disease-specific mortality (M), and 5-y survival, annual disease-specific mortality, and incidence (5YMI). Questionnaire version ''time'' presented data as a comparison between 2 time points and version ''group'' as a comparison between a screened and an unscreened group. All data were based on statistics for the same cancer site (prostate). Outcome variables were the recommendation of screening, reasoning behind recommendation, judgment of the screening's effectiveness, and, if judged effective, a numerical estimate of how many fewer people out of 1000 would die if screened regularly. After randomized allocation, 65 German physicians in internal medicine and its subspecialities completed either of the 2 questionnaire versions.* ***Results.*** *Across both versions, 66% of the physicians recommended screening when presented with 5Y, but only 8% of the same physicians made the recommendation when presented with M (5YM: 31%; 5YMI: 55%). Also, 5Y made considerably more physicians (78%) judge the screening to be effective than any other condition (5YM: 31%; M: 5%; 5YMI: 49%) and led to the highest overestimations of benefit.* ***Conclusion.*** *A large number of physicians erroneously based their screening recommendation and judgment of screening's effectiveness on the 5-y survival rate. Results show that reporting disease-specific mortality rates can offer a simple solution to physicians' confusion about the real effect of screening.* ***Key words:*** *decision rules; risk communication or risk perception, shared decision making, health literacy, numeracy.* ***(Med Decis Making 2011;31:386–394)***

According to the concept of shared medical decision making, the technical knowledge of risks and benefits of medical interventions is held by the physician, who then shares this knowledge with patients to enable them to decide according to their preferences.[1] If physicians do not have this knowledge, which involves understanding health statistics, effective risk communication and shared decision making cannot take place. Numerous studies found that the format in which numeric health data are presented can generate divergent interpretations amongst physicians—a phenomenon that has been largely explained by the fact that some formats yield larger numbers than others and thus are more persuasive.[2–7] In the past, researchers mainly focused on the persuasiveness of relative risk reduction compared with absolute risk reduction and number needed to treat.[4,7,8] Another persuasive format, however, might be the 5-y survival rate.

The 5-y survival rate is probably the most common statistic used to report the progress in treating cancer. Improvements in 5-y survival are often considered an unambiguous sign of success: If patients who receive screening tests or examinations tend to live longer than those who do not, society's enormous investments in early detection and improved treatments must be paying off. However, although the 5-y survival rate is a valid measure for comparing cancer therapies in randomized trials, it is not valid for comparing differently diagnosed groups (e.g., survival before versus after the introduction of

Received 7 December 2009 from the Max Planck Institute for Human Development, Harding Center for Risk Literacy, Berlin, Germany. Revision accepted for publication 9 September 2010.

Address correspondence to Odette Wegwarth, Max Planck Institute for Human Development, Lentzeallee 94, 14195 Berlin, Germany; telephone: +49-30-82406-695; fax: +49-30-82406-394; e-mail: wegwarth@mpib-berlin.mpg.de.

DOI: 10.1177/0272989X10391469

screening; survival of unscreened versus screened people). In fact, changes in 5-y survival over time and groups were found to be completely unrelated to changes in cancer mortality for the 20 most common solid tumors ($r = .00$) in the United States.[3,9]

To understand why, it is helpful to look at how the 5-y survival statistic is calculated in the context of screening:

$$\text{Disease-specific 5-year survival rate} = \frac{\text{number of persons diagnosed with a specific cancer still alive 5 years after diagnosis}}{\text{number of persons diagnosed with a specific cancer in the study population}}.$$

In this calculation, the key term to notice is *diagnosed*, which appears in the numerator and denominator of the survival statistic of a specific cancer. Cancer can be diagnosed by either symptoms or screening. By definition, screening detects cancer before it causes symptoms. Because of this property, screening can bias 5-y survival rates in 2 ways: 1) by prolonging the period in which patients are known to have cancer and 2) by including people with nonprogressive cancer in the statistic. The first, called *lead-time bias*, accounts for the fact that screening may only reduce the time to diagnosis without increasing the time to death. This prolonged period of being diagnosed makes patients attending screening more likely to enter the 5-y survival statistic. Yet this may have no bearing on real prolonged or saved lives.[10] The second phenomenon, called *length-time bias* or *overdiagnosis bias*, concerns the detection of lesions that meet the pathological definition of cancer yet never become clinically significant due to their prolonged preclinical phase or their lack of propensity to progress.[10] The inclusion of nonprogressive and slowly progressive cancer inflates the 5-y survival statistic in 2 ways: First, it inflates the number of persons diagnosed (incidence) with a specific cancer in the study population (the denominator of the survival equation). Second, it inflates the number of diagnosed persons still alive 5 y after diagnosis (the numerator of the equation), because people with slowly progressive and nonprogressive cancer have a better prognosis than people with aggressive cancer and are thus more likely to survive the next 5 years.

Because 5-y survival rates do not allow reliable judgments on improved cancer control due to these biases, physicians are advised to use disease-specific mortality rates for a specific cancer.[11] This statistic does not depend on diagnostic procedures and therefore is not prone to screening-induced biases.

$$\text{Annual disease-specific mortality} = \frac{\text{number of persons who die from a specific cancer over 1 year}}{\text{number of persons in the study population}}.$$

Nonetheless, changes in survival rates over time[9] and across differently diagnosed groups[12] are still reported in the relevant medical journals. What do doctors conclude from these?

The present study is, to the best of our knowledge, the first to investigate how different measures used to report progress against cancer influence physicians' recommendations of screening and judgments of its effectiveness. A literature search of PubMed, MEDLINE, and ISI Web of Knowledge using layers of subject headings combining *5-year survival* as a constant and the subjects of *doctors' understanding*, *clinicians' understanding*, *physicians' understanding*, *risk communication*, *bias*, and *success against cancer* did not reveal any study in this vein. We assessed how 5-y survival rates, mortality rates, and 2 combinations of 5-y survival rates, mortality rates, and incidence—which were varied over time or group—would alter physicians' recommendations of screening as well as their judgments of its effectiveness. Furthermore, we investigated physicians' knowledge about lead-time bias and length-time bias. This study was exploratory in nature.

METHOD

Materials

We developed 2 versions of a survival survey: version ''group'' and version ''time.'' Although otherwise alike, version ''group'' always presented data as a comparison between a screened and an unscreened group and version ''time'' as a comparison between 1975 and 2004. Each version rested on a repeated-measures design and comprised 4 conditions: Condition 5Y provided data on 5-y survival, condition 5YM on 5-y survival and annual disease-specific mortality, condition M on annual disease-specific mortality, and condition 5YMI on 5-y survival, annual disease-specific mortality, and incidence (see Web Appendix 1).

To mask the fact that data shown in the 4 conditions and in both versions always referred to the potential effect of screening for the same cancer site

Table 1 Comparison of the Characteristics of the 2 Groups of Physicians

Characteristic	Version "Group" (n = 34)	Version "Time" (n = 31)	Total (N = 65)
Position, n (%)			
Junior physician	16 (48)	14 (45)	30 (46)
Senior physician	9 (26)	7 (23)	16 (25)
Head physician	9 (26)	10 (32)	19 (29)
Work environment, n (%)			
Private practice	4 (11)	4 (13)	8 (12)
Hospital	18 (54)	20 (64)	38 (59)
Research hospital	12 (35)	7 (23)	19 (29)
Years since graduation, range (M, SD)	1–35 (12.5, 9.4)	1–38 (15.2, 10.6)	1–38 (13.8, 10)
Age, range (mean, SD)	27–65 (40.2, 9.8)	27–66 (44.8, 10.1)	27–66 (42.4, 10.1)

(prostate), screenings and tumors were labeled with capital letters. The survey face sheet stated that all screenings mentioned in the following conditions were noninvasive, detected tumors for which several treatment options exist, and were first used routinely at the beginning of the 1990s. Each of the 4 conditions was preceded by an introduction of a 55-y-old healthy patient seeking advice from the doctor on whether to be screened for tumor A, B, or C and so forth.

Four outcomes were measured: recommendation of screening (yes, no, I can't decide), reasoning behind recommendation (open answer format), judgment of screening's effectiveness (yes, no), and, if judged effective, a numerical estimate of how many fewer people out of 1000 would die if they were regularly screened. After the 4 conditions, physicians were further asked if they knew what lead-time bias and length-time bias are and, if so, to explain each of these.

To investigate with as little bias as possible what doctors would conclude from the commonly reported 5-y survival statistic, condition 5Y was always first, followed by the other 3 conditions in the listed order. Combined conditions aimed at investigating whether more information fosters a detection of lead-time (condition 5YM) and length-time bias (condition 5YMI). Both survey versions were piloted by 6 physicians and refined in response.

Rationale for Data Source

Data on 5-y survival rates, disease-specific mortality rates, and incidence presented in the 4 conditions were drawn from the Surveillance, Epidemiology and End Result (SEER) program for prostate cancer and for the time points 1975 and 1999.[13] When our study was conducted in late 2008, no reliable data for a comparison between a screened and an unscreened group had yet been released (results from 2 randomized controlled trials on the effect of screening for prostate cancer were not released until March 2009[14,15]). Hence, numbers shown in version "group" were also orientated on the SEER database. To make conditions appear independent from each other, we marginally modified original SEER data from condition to condition in the 2 versions. Yet compared with the 20 most common cancer sites, the temporal increase of 5-y survival and incidence for prostate cancer ranked among the highest.[9] This raised concerns that results of our study might not be representative of other cancer sites. We thus decided to halve all numbers reported in version "time" when used in version "group." For instance, when an increase of 31 percentage points was reported for condition 5Y in version "time," an increase of 16 percentage points was reported for condition 5Y in version "group" (for details, see Web Appendix 2).

Sample and Procedure

Sixty-eight German physicians in internal medicine and its subspecialities participated in the study. Three physicians were excluded from analysis because they did not provide answers throughout all of the 4 conditions for recommendation of screening and judgment of screening's effectiveness, which was required for the within-subject comparisons. Of the final 65 physicians, 34 physicians completed the version "group" and 31 the version "time." A breakdown of the physicians' characteristics is shown in Table 1. Characteristics of respondents receiving the 2 versions did not differ with respect to the position, the work environment, years

since graduation, or age (all 95% confidence intervals included 0).

Responses were obtained through personal overtures by one author (n = 41) and through approaching physicians during further educational training (n = 24). The response rate was 98% for personal overtures and 65% for physicians approached during training.

When physicians agreed to participate, they were randomly allocated to one of the versions and were tested either at their work site or at a separate place within the facility where their training was being held. After giving informed consent, physicians were instructed to work individually, complete all questions following each of the conditions, and not to return to conditions that had already been completed. No payment was made for participation.

Analysis

All data were stored and analyzed with SPSS 16. If a response was lacking for any outcome, except recommendation of screening and judgment of screening's effectiveness, it was coded as a miss.

For the coding of reasoning behind recommendation (qualitative data), 2 independent raters first reviewed all reasons to establish categories that would cover these reasons. After the review, the raters discussed and consented on the categories. Next, each rater independently assigned physicians' responses to the categories. The interrater reliability was r = 0.91. Explanations of lead-time bias and length-time bias were coded as either correct or incorrect depending on whether they resembled the epidemiological definition.[10] Because the study was exploratory in nature, all outcomes were analyzed on a descriptive level.

RESULTS

Screening Recommendation

What would an informed pattern of recommendation look like? The 5-y survival statistic alone does not allow an unbiased judgment of the benefit of screening. Thus, if physicians are aware of this problem, they should chose either ''no'' or ''I can't decide'' in condition 5Y. The other 3 conditions provided information on disease-specific mortality, which allows an evaluation of the effect. In accordance with the SEER results, we presented mortality data that showed a minimal increase instead of a decrease for the screening group or the later time point. Thus, if physicians knew that they are advised to look at mortality data, one would expect them to choose ''no'' in these conditions. Therefore, an informed pattern of recommendation over the 4 conditions should be the following: I can't decide/no, no, no, no. One could argue, however, that physicians could have been knowledgeable about which statistic is best to look at but responded to the questions according to the practice of defensive medicine—a reaction to the unpredictabilities of the legal system in which a physician can be sued for doing too little but rarely for doing too much.[3,16] In such a case, the minimal increase in mortality could be viewed as negligible and thus, for defensive reasons, the screening as recommendable. In this case, a defensive yet informed pattern of recommendation over the 4 conditions would be the following: I can't decide/no, yes, yes, yes.

Figure 1 shows the actual pattern of recommendations for each physician across the 4 conditions—ordered by the number of ''yes'' choices—and for both versions. As can be seen, the 2 survey versions yielded roughly similar response patterns. No more than 4 physicians in version ''group'' (n = 34) and 1 physician in version ''time'' (n = 31) showed the informed pattern. The defensive yet informed pattern was not found with any of the physicians. Instead, in both versions, the patterns suggest that many physicians focused on the changes in 5-y survival. All conditions that included the 5-y survival statistic (5Y, 5YM, 5YMI) prompted considerably more ''yes'' choices than the condition presenting only disease-specific mortality (M). Forty-three of 65 physicians recommended screening when presented with only 5-y survival data. In contrast, only 5 of the same physicians recommended screening when presented with only disease-specific mortality data. The combined conditions, intended to foster physicians' insight by making them aware of lead-time bias and length-time bias, only partly attenuated the misleading effect of the included 5-y survival rate: Twenty of the physicians recommended screening when presented with condition 5YM, and 36 of them did so when later presented with condition 5YMI. One might conjecture that due to the repeated measurement design, physicians would show carryover effects. However, of all 65 physicians, only 6 always gave the same response throughout the 4 conditions.

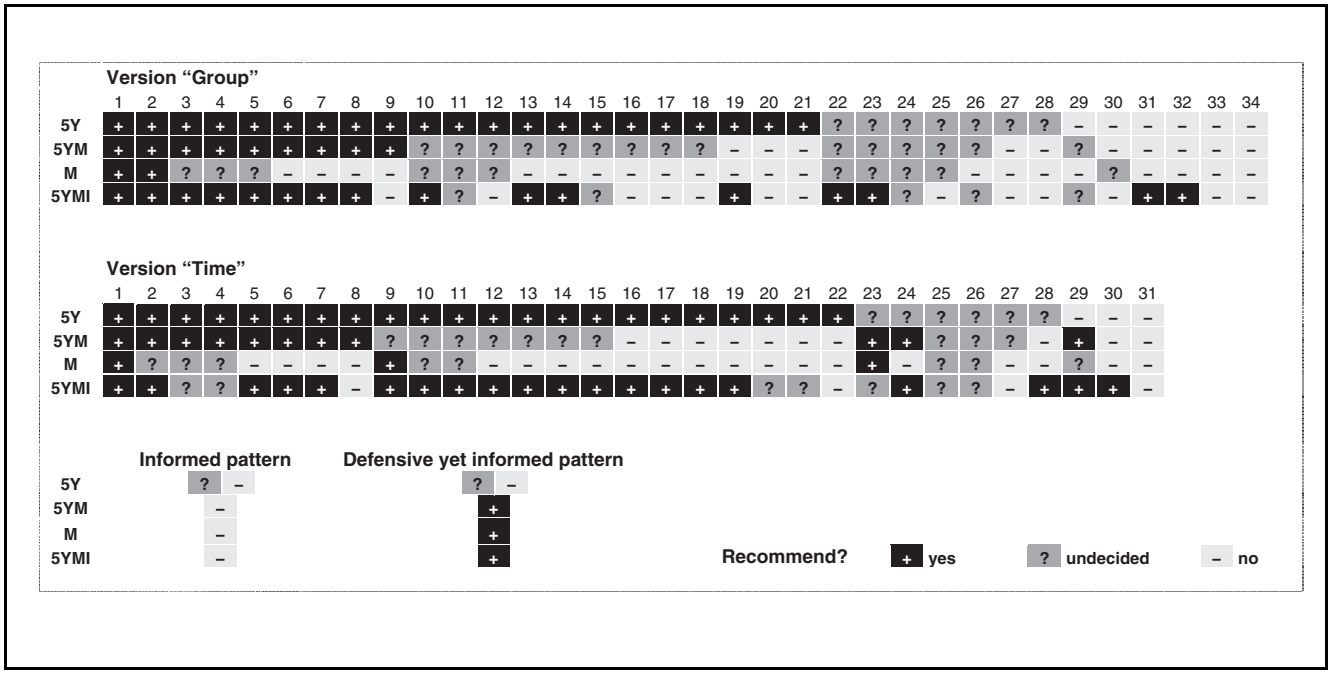

Figure 1 Screening recommendations of 34 physicians receiving information for a comparison between a screened and an unscreened group (version ''group'') and of 31 physicians receiving information for a comparison between the time points 1975 and 2004 (version ''time''). Conditions: 5Y = 5-y survival; 5YM = 5-y survival and annual disease-specific mortality; M = annual disease-specific mortality; and 5YMI = 5-y survival, annual disease-specific mortality, and incidence.

Reasons for Recommendation

What reasons did physicians give for their recommendations? For both versions, the most frequent reason for a recommendation in favor of the screening was the increase in the 5-y survival rate over time or groups. Often, the physicians described this increase as ''meaningful,'' ''clinically significant,'' or ''exemplifying the merits of early detection.'' For condition 5YMI only, another prominent reason was what we call the *incidence-mortality fallacy*. Here, a physician would infer from an increased 5-y survival rate, plus increased incidence and a stable mortality rate, that screening must be effective, not realizing that more screening inflated the incidence and thereby the 5-y survival rate. For instance, one physician wrote on his sheet, ''5-year survival better for screened group, mortality between groups equates, yet incidence is higher in screened group, thus fewer people die in the screened group.'' However, unless tumor biology (aggressiveness of the tumor) was suddenly to change, the number of people who developed the disease (incidence) would not be expected to influence the 5-y survival rate, that is, the prognosis of an individual case. Thus, if incidence and 5-y survival rates increase at the same time while mortality rates remain unchanged, increased incidence may reflect changes in clinical detection practice rather than changes in true occurrence.

Recommendations against screening were mainly triggered by physicians focusing on the mortality rate and the subsequent impression that the benefit is either negative or does not exist. For the occasions in which physicians could not decide whether to recommend screening, no prominent main reason was detected. Table 2 gives an overview of the reasoning for each condition.

Judgment of Screening's Effectiveness

Five-year survival rates affected the judgment of screening's effectiveness in the same way as for recommendation. Across versions, 51 of the 65 physicians judged the screening effective when presented with 5-y survival rates only. When presented with the next condition providing 5-y survival and mortality data together, 31 physicians changed their minds—now, 20 of 65 considered the screening effective. Shown the third condition, which

Table 2 Physicians' Reasons for Each of the Recommendation Options (Yes, No, I Can't Decide), by Condition

	5Y			5YM			M			5YMI		
	Yes	No	/	Yes	No	/	Yes	No	/	Yes	No	/
Significant increase in survival	43			18						16		
No benefit in survival					17	8		39	9		11	5
Incidence-mortality fallacy										19		
Therapy v. screening		4	7		2	5		1			3	
Data inconclusive/insufficient		4	4		1	3	1	1	2			3
Possible improved quality of life				1		1	3		8			3
Lead-time bias						3					2	
Poor therapy						2						
Missing value		1	2	1		3	1			1	1	1

Note: Numbers express the frequency of mention. A slash indicates ''I can't decide.'' Therapy v. screening indicates feeling unable to judge whether temporal changes of data are due to the success of screening or therapy. 5Y = 5-y survival; 5YM = 5-y survival and annual disease-specific mortality; M = annual disease-specific mortality; 5YMI = 5-y survival, annual disease-specific mortality, and incidence.

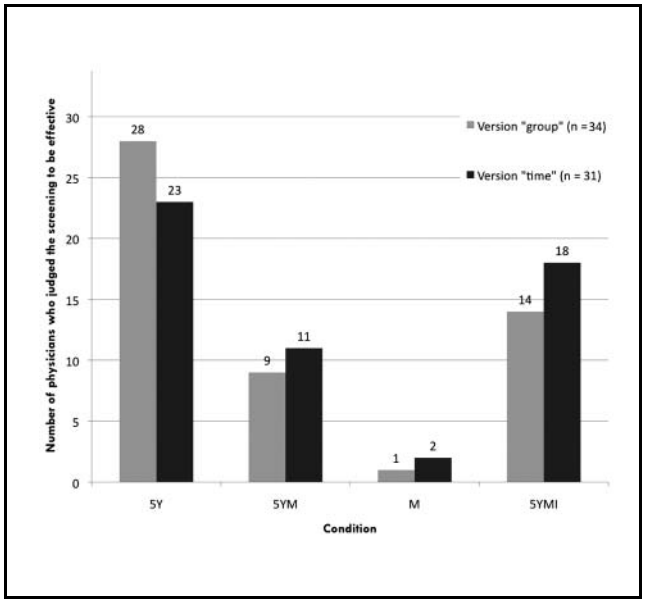

Figure 2 Number of physicians who assumed the screening to be effective, shown by condition and for the 2 versions, group: n = *34, and time:* n = *31. Conditions 5Y and 5YMI led the majority to assume the screening to be effective.*

presented only mortality data, the number of physicians who still judged the screening effective reduced to 3. When given information on 5-y survival, mortality, and incidence data together in the last condition, the number of physicians judging the screening effective rose again to 32. Figure 2 shows the influence of the different statistics on the physicians' judgments by version, which again are roughly comparable.

Numerical Estimate of Screening's Effectiveness

When a physician judged the screening to be effective, he or she was subsequently asked how many fewer people would die out of 1000 if they were regularly screened. The change in the disease-specific mortality rate over the 2 time points that we drew from the SEER results suggested that the answer is 0; no man has been saved from dying from prostate cancer since the introduction of screening. Thus, the best answer to the question of how many fewer would die is 0. These results were based on temporal changes, and one could argue that such changes cannot be exclusively attributed to the effect of screening. Results released later from 2 randomized controlled trials on the effects of prostate-specific antigen screening showed somewhat comparable results as the temporal data, however.[14,15]

As Table 3 shows, 0 was not the answer that the physicians arrived at when confronted with the 3 conditions that included 5-y survival rates. At the median level, within these conditions, physicians expected between 13 and 150 fewer deaths in 1000 screened people. Because numbers presented in the conditions of version ''group'' were only halves of those presented in version ''time,'' overestimations for the former were smaller. In both versions, condition 5Y led to the highest overestimations of the effectiveness and also to the largest variation among estimates. The combined 5YM and 5YMI conditions generated smaller variation and estimates; nevertheless, compared with the real reduction of cancer mortality, they still yielded high overestimations. Because 62 of 65 physicians in condition M had

Table 3 Physicians' Numerical Estimates of the Effectiveness of Screening by Condition and Version

	Median Estimate			
	Condition			
Version	**5Y**	**5YM**	**M**	**5YMI**
Group	30	13	—	14
Range	2–500	2–200	—	2–100
Total *n* (misses)	23 (5)	7 (2)	0 (1)	9 (5)
Time	150	25	.003	30
Range	2–980	0.01–250		1–500
Total *n* (misses)	21 (2)	7 (4)	1 (1)	13 (5)

Note: Estimates were based on the question of how many fewer people would die out of 1000 if they regularly attended the screening. Only physicians who indicated upfront that they judged the screening to be effective were asked to provide a quantitative estimate.

already correctly judged the screening to be ineffective in advance, only 3 physicians were asked for a numerical estimate. The 1 physician who finally gave an estimate would have been correct if we had reported a decrease instead of an increase in disease-specific mortality.

Knowledge of Lead-Time Bias and Length-Time Bias

After physicians had worked through the conditions, they were asked if they knew about the lead-time bias and the length-time bias and, if so, were asked to give an explanation of each. Fifty-four of the 65 physicians did not know what the lead-time bias was. Of the remaining 11 physicians who indicated they did know, only 2 explained the bias correctly. With only 1 exception, no physician knew what the length-time bias was. However, when asked to explain the bias, the 1 physician did not explain it correctly either.

DISCUSSION

Our exploratory study assessed how 5-y survival rates, mortality rates, and 2 combinations of 5-y survival rates, mortality rates, and incidence—varied over time or group—would alter physicians' recommendations of cancer screening and their judgment of its effectiveness. To recapitulate, across both survey versions, only 5 of 65 physicians showed informed recommendation patterns. More than two-thirds of the physicians erroneously based their screening recommendations on changes in 5-y survival rates over time or group. Furthermore, the majority judged the screening to be effective and overestimated this effectiveness by up to 150 fewer deaths in the screened group when presented with 5-y survival rates. Knowledge of the 5-y survival-related biases in the context of screening evaluation was scarce to nonexistent. Only a few physicians felt that they required information other than survival rates to make a recommendation and to decide whether the screening would be effective. Combined conditions, intended to foster a detection of lead-time bias (condition 5YM) and length-time bias (condition 5YMI), only partly attenuated the misleading effect of the included 5-y survival rates. Results further showed that disease-specific mortality enabled physicians to judge the effectiveness of the screening correctly. The 2 versions of the questionnaire yielded comparable results.

An open question concerns how physicians arrived at the numerical estimates of the effectiveness of screening. Across conditions, between 10% and 50% of the physicians seemed to have based their calculations on changes in 5-y survival rates over time or group. In a few further instances, physicians used the corresponding 5-y survival rate of 2004 or of the screened group, respectively. However, for most of the estimates, it was impossible to reconstruct how physicians made their calculations. A better understanding of how physicians based their estimates of effectiveness on the 5-y survival rate—alone and in combination with other statistics—would help to reveal the nature of the confusion in more detail.

Strengths and Limitations of the Study

Our study was based on a repeated-measures design. Such a design has the potential shortcoming of carryover effects that might prevent participants from changing their responses from one condition to

the other. However, across both versions, only 6 of the 65 physicians adhered to their initial recommendation, not changing it in the following conditions. Also, of the 51 physicians who judged the screening to be effective in condition 5Y, only 3 still did so in condition M, and many changed their judgment again when presented with condition 5YMI.

A limitation of the present study is that we used a convenience sample of 65 physicians in internal medicine. Thus, we do not know about the generalizability of our results. A study with a representative sample would be needed to investigate if findings hold true for a wider range of physicians. At the same time, the study involved physicians of various hierarchies from a variety of work environments. In addition, results are clear-cut, so that we assume a fair generalizability of our findings.

Another limitation might be seen in the fact that we told physicians within the introduction of each condition that the presented data were obtained from a randomized controlled trial (RCT; see the web appendices). Although RCTs produce the best available evidence on which to base medical actions, they do not necessarily make a wrong statistic in a wrong context right. At the beginning of our article, we made clear that 5-y survival rates are a valid measure for comparing cancer therapies in RCTs but not for comparing differently diagnosed groups (screening v. symptoms) in or outside of RCTs, due to lead-time bias and overdiagnosis bias. In addition, it might be argued that physicians used the data presented uncritically, overlooking the fact that the trial was on screening and not on therapies. However, we used words relating to screening (e.g., *screened group*, *unscreened group*) 4 times within each condition, but in none did we include any words relating to therapy or treatment.

Implications for Policy, Practice, and Medical Training

It is important to us that our work is not misconstrued: This article is not meant to suggest that there has been no real progress in cancer care. Instead, we want to highlight that survival rates have the potential to deceive physicians' perception of screening's effectiveness and therefore wrongly influence their recommendations. Our findings lead to the important issue of which metrics should be presented and in what context in the medical literature. Policies exist, such as CONSORT (http://www.consort statement.org/), that recommend the reporting of changes in mortality for screening evaluation, and an increasing number of medical journals are now subscribing to such policies. However, these policies appear not always to be enforced, and thus the reporting of survival rates in the context of screening is still common.[12] We believe that the implementation and enforcement of such policies merit serious consideration. Equally important, better training of physicians in understanding health statistics at medical schools may lessen their vulnerability to being misled. For clinicians who wish to correctly inform their patients about the true benefit of screening, it is essential for them to learn that improving 5-y survival rates over time and over differently diagnosed groups may not reflect a reduced disease burden and should not be taken as evidence of improved prevention, screening, or therapy. Improved survival rates may instead reflect more cases being diagnosed and unchanged mortality, as suggested, for instance, for prostate cancer.[14,15] In contrast, mortality is a clear-cut number that decreases with improvement in cancer control, be it through successful early detection or better treatment. A health system that expects physicians to correctly inform their patients about medical interventions should thus encourage the reporting of mortality rates in medical journals and medical training.

Contributors: All authors declare that they participated in conceptualizing the study, analyzing the data, and writing the article. All authors had full access to all of the data (including statistical reports and tables) in the study and can take responsibility for the integrity of the data and the accuracy of the data analysis.

Conflicts of interest: None declared.

Ethical approval: The Ethics Committee of the Max Planck Institute for Human Development approved the study, and all participants consented to participation at the beginning of the survey.

Funding/support: This study was funded by the Harding Center for Risk Literacy at the Max Planck Institute for Human Development (Germany). The authors declare independence from these funding agencies.

Role of the funding source: The funding source did not affect the study design, data collection, analysis and interpretation of the data, writing of the report, or the decision to submit the article for publication.

REFERENCES

1. Charles CA, Gafni A, Whelan T. Shared decision-making in the medical encounter: what does it mean? (or, It takes at least two to tango). Soc Sci Med. 1997;44:681–92.

2. Mühlhauser I, Kasper J, Meyer G. FEND: understanding of diabetes prevention studies: questionnaire survey of professionals in diabetes care. Diabetologia. 2006;49:1742–6.

3. Gigerenzer G, Gaissmaier W, Kurz-Milcke E, Schwartz LM, Woloshin S. Helping doctors and patients to make sense of health statistics. Psychol Sci Public Interest. 2007;8:53–96.

4. Covey J. A meta-analysis of the effects of presenting treatment benefits in different formats. Med Decis Making. 2007;27:638–54.

5. Ghosh AK, Ghosh K. Translating evidence-based information into effective risk communication: current challenges and opportunities. J Lab Clin Med. 2005;145:171–80.

6. Fahey T, Griffiths S, Peters TJ. Evidence based purchasing: understanding results of clinical trials and systematic reviews. Br Med J. 1995;311:1056–9.

7. Naylor CD, Chen E, Strauss B. Measured enthusiasm: does the method of reporting trial results alter perceptions of therapeutic effectiveness? Ann Intern Med. 1992;117:916–21.

8. Hembroff LA, Holmes-Rovner M, Wills CE. Treatment decision-making and the form of risk communication: results of a factorial survey. BMC Med Inform Decis Mak. 2004;4:20.

9. Welch HG, Schwartz LM, Woloshin S. Are increasing 5-year survival rates evidence of success against cancer? JAMA. 2000; 283:2975–8.

10. Gordis L. Epidemiology. 4th ed. Philadelphia: Saunders; 2008.

11. Extramural Committee to Assess Measures of Progress Against Cancer. Measurement of progress against cancer. J Natl Cancer Inst. 1990;82:825–35.

12. Henschke C, International Early Lung Cancer Action Program Investigators. Survival of patients with stage I lung cancer detected on CT screening. N Engl J Med. 2006;355:1763–71.

13. Ries LAG, Harkins D, Krapcho M, et al. SEER Cancer Statistics Review, 1975–2003 (National Cancer Institute). 2005 [updated 2005; cited March 26, 2008]; Available from: URL: SEER Web site: http://seer.cancer.gov/csr/1975_2003/

14. Andriole GL, Crawford ED, Grubb RL III, et al. Mortality results from a randomized prostate-cancer screening trial. N Engl J Med. 2009;360:1310–9.

15. Schröder FH, Hugosson J, Roobol MJ, et al. Screening and prostate-cancer mortality in a randomized European study. N Engl J Med. 2009;360:1320–8.

16. Steurer J, Held U, Schmidt M, Gigerenzer G, Tag B, Bachman LM. Legal concerns trigger PSA testing. J Eval Clin Pract. 2009;15: 390–2.