

Overdiagnosis in Cancer

H. Gilbert Welch, William C. Black

Manuscript received September 3, 2009; revised March 1, 2010; accepted March 5, 2010.

Correspondence to: H. Gilbert Welch, MD, MPH, Veterans Affairs Outcomes Group (111B), Department of Veterans Affairs Medical Center, White River Junction, VT 05009 (e-mail: h.gilbert.welch@dartmouth.edu).

This article summarizes the phenomenon of cancer overdiagnosis—the diagnosis of a “cancer” that would otherwise not go on to cause symptoms or death. We describe the two prerequisites for cancer overdiagnosis to occur: the existence of a silent disease reservoir and activities leading to its detection (particularly cancer screening). We estimated the magnitude of overdiagnosis from randomized trials: about 25% of mammographically detected breast cancers, 50% of chest x-ray and/or sputum-detected lung cancers, and 60% of prostate-specific antigen-detected prostate cancers. We also review data from observational studies and population-based cancer statistics suggesting overdiagnosis in computed tomography-detected lung cancer, neuroblastoma, thyroid cancer, melanoma, and kidney cancer. To address the problem, patients must be adequately informed of the nature and the magnitude of the trade-off involved with early cancer detection. Equally important, researchers need to work to develop better estimates of the magnitude of overdiagnosis and develop clinical strategies to help minimize it.

J Natl Cancer Inst 2010;102:605–613

Early detection has forced clinicians and researchers to contemplate a more expansive and, to many, counterintuitive definition of the word “cancer.” What most of us were taught in medical school is captured by the terse definition contained in the medical dictionary—“a neoplastic disease the natural course of which is fatal” (1). It was a simple definition that was largely accurate in an era when patients were diagnosed with cancer because they had signs and symptoms of the disease.

But that all changed after we became technologically able to advance the time of diagnosis and detect cancer early—before it produces signs and symptoms. Now it has become evident that the word “cancer” encompasses cellular abnormalities with widely variable natural courses: Some grow extremely rapidly, others do so more slowly, others stop growing completely, and some even regress. Clinicians are left with the realization that the word “cancer” is less a prediction about disease dynamics and more a pathological description made at a single point in time. Continued adherence to the dictionary definition of cancer, however, can lead to harm—including overuse of anticancer therapies.

Although not yet contained in medical dictionaries, recently, a new word has appeared in the medical literature to describe a side effect of our technological progress: “overdiagnosis.” This article is intended to summarize the phenomenon.

What Is Cancer Overdiagnosis?

Overdiagnosis is the term used when a condition is diagnosed that would otherwise not go on to cause symptoms or death. Cancer overdiagnosis may have one of two explanations: 1) The cancer never progresses (or, in fact, regresses) or 2) the cancer progresses slowly enough that the patient dies of other causes before the cancer becomes symptomatic. Note that this second explanation incorpo-

rates the interaction of three variables: the cancer size at detection, its growth rate, and the patient’s competing risks for mortality. Thus, even a rapidly growing cancer may still represent overdiagnosis if detected when it is very small or in a patient with limited life expectancy. Overdiagnosis should not be confused with false-positive results, that is, a positive test in an individual who is subsequently recognized not to have cancer. By contrast, an overdiagnosed patient has a tumor that fulfills the pathological criteria for cancer.

To understand overdiagnosis, one must first understand the heterogeneity of cancer progression, which can be diagrammed using arrows to represent different rates of cancer progression (Figure 1). The arrow labeled “fast” represents a fast-growing cancer, which is defined as one that quickly leads to symptoms and to death. The arrow labeled “slow” represents a slow-growing cancer, which is defined as one that leads to symptoms and death but only after many years. The arrow labeled “very slow” represents a cancer that never causes problems because the patient will die of some other cause before the cancer is large enough to produce symptoms. The most familiar clinical example is likely a small low-grade prostate cancer in an elderly male. The arrow labeled “nonprogressive” represents cellular abnormalities that meet the pathological definition of cancer but never grow to cause symptoms—alternatively, they may grow and then regress. Although the concept of nonprogressive cancers may seem implausible, basic scientists have begun to uncover biological mechanisms that halt the progression of cancer (2–4). Some cancers outgrow their blood supply (and are starved), others may be recognized by the host’s immune system or other defense mechanisms (and are successfully contained), and some are simply not that aggressive in the first place.

Overdiagnosis occurs when either nonprogressive cancers or very slow-growing cancers (more precisely, at a slow enough pace that individuals die from something else before the cancer ever

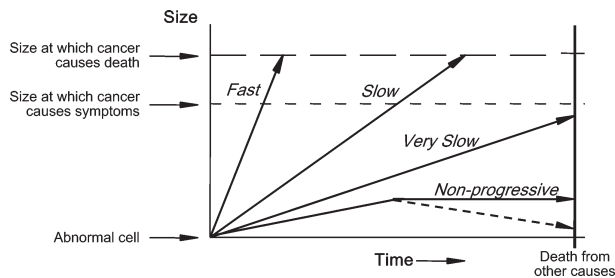


Figure 1. Heterogeneity of cancer progression. The arrow labeled “fast” represents a fast-growing cancer, one that quickly leads to symptoms and to death. The arrow labeled “slow” represents a slow-growing cancer, one that leads to symptoms and death but only after many years. The arrow labeled “very slow” represents a cancer that never causes problems because the patient will die of some other cause before the cancer is large enough to produce symptoms. The arrow labeled “nonprogressive” represents cellular abnormalities that meet the pathological definition of cancer but never grow to cause symptoms—Alternatively, they may grow and then regress (**dotted line**). (Figure 1 was previously supplied by the authors to Wikipedia.)

causes symptoms) are detected. These two forms of cancer have been collectively referred to as pseudodisease—literally false disease. Although we will not use the term subsequently, another definition of overdiagnosis is simply the detection of pseudodisease.

The conundrum in overdiagnosis is that clinicians can never know who is overdiagnosed at the time of cancer diagnosis. Instead, overdiagnosis can only be identified in an individual if that individual 1) is never treated and 2) goes on to die from some other cause. Because clinicians do not know which patients have been overdiagnosed at the time of diagnosis, we tend to treat all of them. Thus, overdiagnosis contributes to the problem of escalating health-care costs. But even where there no money involved, overdiagnosis would be a major concern: Although such patients cannot benefit from unnecessary treatment, they can be harmed.

Prerequisites for Overdiagnosis

The Existence of a Disease Reservoir

The first prerequisite for overdiagnosis is the existence a substantial number of subclinical cancers—in other words, a disease reservoir

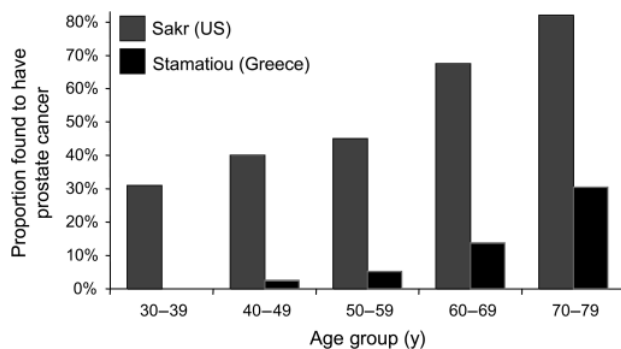


Figure 2. Prostate cancer reservoir in men dying from causes other than prostate cancer (and who were not known to have prostate cancer during life).

voir of detectable cancer. Inferences about the size of this disease reservoir come from the methodical inspection of tissues at autopsy in a series of individuals who died from causes other than cancer. This reservoir is most easily investigated in prostate and thyroid cancers because the glands are small enough to allow an exhaustive examination of thin sections of the entire organ. In addition, there have been multiple investigations of the reservoir in breast cancer.

Let us consider the data of two investigators who made age-specific estimates of the reservoir of prostate cancer from autopsies (Figure 2). Sakr et al. (5) examined the prostate glands of 525 American men who died in an accident; Stamatiou et al. (6) examined 212 Greek men who died of other causes and were not found to have palpable prostate cancer. Because additional estimates based on specimens obtained by radical cystectomy are similarly variable (7), it is clear that the reservoir of potentially detectable prostate cancer is highly age dependent and is probably in the range of 30%–70% in men older than 60 years.

Harach et al. (8) systematically examined the thyroid gland in 101 autopsies. They examined slices of thyroid tissue taken every 2.5 mm and found at least one papillary carcinoma in 36% of Finnish adults. Because many of the cancers were smaller than the width of the slices, they reasoned that they were missing some. Given the number of small cancers they did find and the number that they estimated they had missed (which was a function of size), Harach et al. concluded that the prevalence of histologically verifiable papillary carcinoma would be close to, if not equal to, 100% if one could look at thin enough slices of the gland.

Seven autopsy series have been directed at determining the disease reservoir of breast cancer (9). The four series that included age-specific data suggested that the proportion of middle-aged women who harbored undetected breast cancer ranged from 7% to 39%. Two explanations for this variability are possible that are germane to pathological estimates of the disease reservoir for any cancer. First, different series involve different pathologists, who may have different thresholds about whether to label a small abnormality as “cancer.” Second, different studies have different degrees of scrutiny, that is, some investigators did not look as hard as others. Among the seven series, for example, at one extreme, the investigators examined fewer than 10 slices per breast and at the other extreme, the investigators examined more than 200.

We have summarized the above data in the context of the lifetime risk of death or metastatic disease (Table 1). The lifetime risk of death or metastatic disease is perhaps the least ambiguous measure of the true disease burden for each cancer. The extent to which the disease reservoir exceeds this lifetime risk provides a crude estimate of the amount of overdiagnosis possible.

Activities Leading to Detection of the Disease Reservoir

But the existence of a disease reservoir of detectable cancer, by itself, will not lead to overdiagnosis. There must also be actions that tap it. Thus, the second prerequisite for overdiagnosis is activities leading to early cancer detection.

By far, the most obvious of these is cancer screening. The most familiar efforts involve cancer screening programs organized around a single test, such as mammography or prostate-specific antigen (PSA) testing. But cancer screening should be conceived

Table 1. Estimated size of the disease reservoir for three cancers, the lifetime risk of death or metastatic disease, and the probability of overdiagnosis where the entire disease reservoir detected

Cancer	Population	% With cancer (disease reservoir) (a)	Lifetime risk of death or metastatic disease* (b), %	Probability of overdiagnosis where entire disease reservoir detected† (c = [a - b]/a), %
Prostate	Men older than 60 y	30–70	4	87–94
Thyroid	Adults aged 50–70 y	36–100	0.1	99.7–99.9
Breast	Women aged 40–70 y	7–39	4	43–90

* The lifetime risk of death or metastatic disease was estimated by multiplying the lifetime risk of death reported by the Surveillance, Epidemiology, and End Results program (10) by 1.33, which more than accounts for the small proportion of patients diagnosed with metastatic disease who die from other causes (approximately 20%, 15%, and 10% of those with metastatic cancer of the prostate, thyroid, and breast cancer, respectively).

† This estimate is a lower-bound estimate because lethal and/or metastatic cancers do not always arise from prevalent cancers (those contained in the disease reservoir) but also from incident cancers (those not contained in the disease reservoir).

more broadly as any effort to detect cancer in those who have no symptoms of the disease. Thus, components of general periodic physical examination, such as searching for moles by closely inspecting the skin or seeking masses by palpating the neck, are also a form of screening.

Furthermore, interventions unrelated to screening can lead to early cancer detection. Pathological inspection of tissues removed in surgeries performed for reasons other than cancer may nonetheless find cancer. The most familiar example is prostate cancer detection following transurethral resection of the prostate for benign prostatic hyperplasia (11). However, the most important activity leading to unintended cancer detection undoubtedly involves the increased use of diagnostic imaging. Detailed imaging of the brain, thorax, abdomen, and pelvis intended to evaluate symptoms not suggestive of cancer nonetheless frequently detect abnormalities worrisome for cancer. Clinicians are familiar with this phenomenon, which is sometimes referred to as the detection of “incidentalomas.” For example, screening for colon cancer with computed tomography (CT) colonography detects extracolonic abnormalities in up to 50% of examinations (12).

The growth of early cancer detection activities is easiest to measure in organized cancer screening efforts, many of which did not exist two decades ago. Some of the growth is not simply in terms of the number of examinations but also in terms of the increasing sensitivity of the examination itself. It is very difficult to gauge the increase in screening physical examinations because these are not systematically recorded. However, increased use of

diagnostic imaging in general is well documented, particularly in the Medicare program (Figure 3) (13).

Evidence That Early Detection Has Led to Overdiagnosis

Randomized Trials of Screening

The strongest evidence for overdiagnosis comes from long-term follow-up after a randomized trial of screening. At the end of the trial, it is expected that the screening group will have a greater number of cancers detected than the control group, simply because screening advances the time of diagnosis and moves the detection of some cancers forward in time. If all of the excess of detected disease represents cancers that were destined to progress to clinical disease (ie, there is no overdiagnosis), the excess should disappear over time when both groups receive similar diagnostic scrutiny. In other words, the control group would be expected to “catch-up” to the screening group—because cancers appear clinically because of signs and symptoms. Although the duration of follow-up necessary to completely catch-up is equal to the lead time of the slowest growing cancer, a shorter interval may be sufficient to confirm overdiagnosis given the existence of competing mortality. A persistent excess in the screening group years after the trial is completed constitutes the best evidence that overdiagnosis has occurred.

Breast Cancer. Of the nine randomized trials of mammography, only one has reported long-term follow-up data on incident

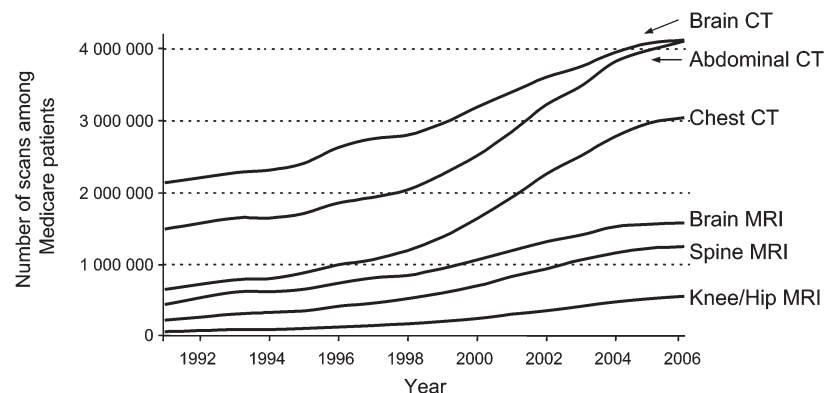


Figure 3. Trends in the number of various scans used in the Medicare population in the United States, 1991–2006. CT = computerized tomography; MRI = magnetic resonance imaging.

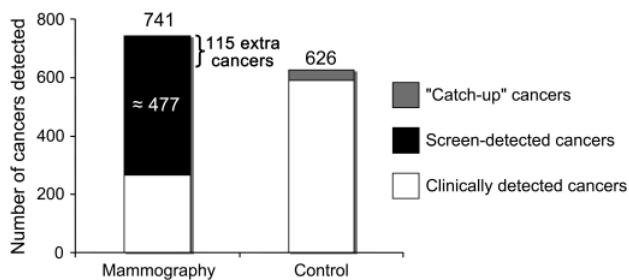


Figure 4. Number of breast cancers detected after 10 years in the Malmö randomized trial of mammography with the 35 additional “catch-up” cancers that appeared in the control group in the subsequent 15 years. “Extra cancers” refer to the difference between the mammography and control groups (after adding the catch-up cancers to the control group). They likely represent overdiagnosed cancers (see Supplementary Technical Appendix, available online).

cancers. The report on 15 years of extended follow-up after the end of Malmö mammographic screening trial provided evidence for breast cancer overdiagnosis (14). At the end of the 10-year trial, 741 breast cancers were detected in the screening group as compared with 591 in the control group. Over the subsequent 15 years, this difference of 150 cancers narrowed to 115, suggesting 35 catch-up cancers. The persistent excess of 115 cancers, however, suggests overdiagnosis.

The findings at the end of the trial, with the 35 catch-up cancers added, highlight a complexity in the estimation of overdiagnosis (Figure 4). One could say that 16% (115 in 741) of cancers detected in the screening group were overdiagnosis. Alternatively, one could restrict the denominator to screen-detected cancers because overdiagnosis can only occur in this subset (a clinically detected symptomatic cancer does not represent overdiagnosis). An earlier publication from the trial showed that 64.4% of cancers detected in the screened group were a consequence of screening, which suggested that about 477 were screen detected. Using this denominator, the risk that a mammographically detected cancer represents overdiagnosis is about 24% (115 in 477) (15).

Lung Cancer. Screening can result in overdiagnosis even among cancers that are traditionally viewed as the most rapidly growing

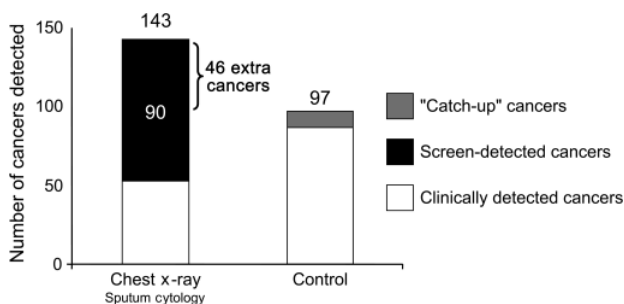


Figure 5. Number of lung cancers detected after 6 years in the Mayo clinic randomized trial of chest x-ray and sputum cytology screening with the 10 additional “catch-up” cancers that appeared in the control group in the subsequent 5 years. “Extra cancers” refer to the difference between the screening group and control group (after adding the catch-up cancers to the control group). They likely represent overdiagnosed cancers (see Supplementary Technical Appendix, available online).

and lethal. The Mayo trial of chest x-ray and sputum cytology screening (16) provided strong evidence for lung cancer overdiagnosis. At the end of the 6-year screening phase, 143 lung cancers were detected in the screening group as compared with 87 in the control group. In follow-up over the subsequent 5 years, 10 catch-up cancers appeared. Extended follow-up over the next 16 years identified no further catch-up cancers (17). Thus, the persistent excess of 46 cancers reflected overdiagnosis (Figure 5). The 46 extra cancers arose among the 90 screen-detected cases in the screening group. Using this denominator, the risk that a chest x-ray- and/or sputum cytology-detected cancer represents overdiagnosis is about 51% (46 in 90).

Prostate Cancer. Although there has been no long-term follow-up, the recently reported randomized trials of PSA screening for prostate cancer also provide some insight into overdiagnosis. The Prostate, Lung, Colorectal, and Ovarian Cancer (PLCO) trial (18) suffered from substantial contamination (ie, screening in the control group) and found no difference in prostate cancer mortality, nonetheless there was a 22% increase in prostate cancer detection in the screening group. It is not known whether this excess will ultimately be diminished by the appearance of catch-up cancers in the control group.

The European Randomized Study of Prostate Cancer (ERSPC) trial (19) used a lower PSA threshold for biopsy (3 vs 4 ng/mL) and a longer screening interval (every 4 years vs annually) and is believed to have had less contamination than the PLCO trial. It found that PSA screening was associated with a 20% reduction in prostate cancer mortality. There was a 70% increase in prostate cancer detection in the screening group—an extra 34 prostate cancers per 1000 men screened. This excess arose from the 58 screen-detected prostate cancers per 1000 men. If this excess represents overdiagnosis, the risk that a PSA-detected cancer represents overdiagnosis would be about 60% (0.034 in 0.058). However, it could be argued that there has been insufficient follow-up for catch-up cancers to become evident and that, therefore, this risk could be an overestimate.

A prior publication by the European group (20) suggested that the risk of overdiagnosis is, in fact, about this magnitude. The investigators estimated that 48% of all patients diagnosed in the screened group (which included both PSA- and clinically detected cancers) had been overdiagnosed (20). Application of that estimate to 82 per 1000 men diagnosed in the screening group during the trial would suggest that overdiagnosis had occurred in about 39 per 1000 men. Using this estimate, the risk that a PSA-detected cancer represents overdiagnosis is about 67% (0.039 in 0.058).

Observational Studies

Observational studies can also provide good evidence for overdiagnosis. In one striking example, investigators in Japan reported, after a first round of spiral CT screening (ie, prevalence screen), finding almost 10 times as much lung cancer as they had previously found in the same population using chest x-rays (21). At the completion of the 3-year screening program, lung cancer detection was virtually the same in smokers as that in never-smokers (22), producing a relative risk that approached 1:

$$\text{Relative Risk}_{\frac{\text{smokers}}{\text{never smokers}}} = \frac{\frac{\text{Lung Cancers Detected}_{\text{smokers}}}{\text{Number Screened}_{\text{smokers}}}}{\frac{\text{Lung Cancers Detected}_{\text{never smokers}}}{\text{Number Screened}_{\text{never smokers}}}}$$

$$= \frac{\frac{29}{31}}{\frac{11.5/1000}{10.5/1000}} = 1.1$$

Because a wealth of epidemiological investigation has demonstrated that the risk of smokers dying from lung cancer is at least 15 times higher than that of never-smokers (23), the Japanese data (the only large-scale CT screening study to include a similar proportion of smokers and nonsmokers) provide evidence that overdiagnosis can be a substantial problem with spiral CT screening.

Japanese investigators have also studied screening for a rare neuroendocrine cancer in children: neuroblastoma (24). Following the initiation of a national screening program, the number of children diagnosed with neuroblastoma more than doubled, and it went up almost fivefold in the group being screened—children younger than 1 year of age. Because some Japanese physicians were concerned about this trend (25), a group of pediatric oncologists decided to offer a “watchful waiting” strategy to the parents of infants with small cancers that were not obviously doing damage (26). Of the 17 couples offered the strategy, 11 accepted, and in each infant, the cancer regressed. Thus, these 11 cancers represented overdiagnosis. Subsequent studies of large-scale screening in Germany and Quebec found that screening detected about twice as many cancers as expected (suggesting overdiagnosis) but no change in neuroblastoma mortality (27,28).

Evidence That Overdiagnosis Is Happening in Populations

Although it is extremely difficult to assess when overdiagnosis has occurred in an individual, it is relatively easy to assess when overdiagnosis has occurred in a population. Rapidly rising rates of testing and disease diagnosis in the setting of stable death rates are suggestive of overdiagnosis. Let us now consider two hypothetical examples of rapid rises in the rate of diagnosis, one of which is suggestive of overdiagnosis and the other is not (Figure 6).

In the left panel of Figure 6, the rapid rise in cancer diagnosis is accompanied by a rapid rise in death from cancer. This pattern

suggests that the new diagnoses are life threatening and clinically important. This is the pattern that has been reported in esophageal adenocarcinoma (29).

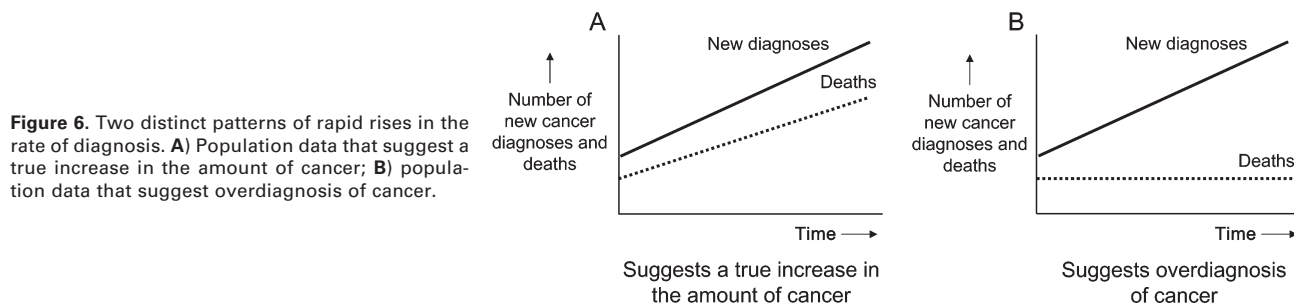
In the right panel of Figure 6, the rapid rise in cancer diagnosis is not accompanied by a rise in cancer death. This suggests that there is more diagnosis, but no change in the underlying amount of cancer destined to affect patients. It suggests overdiagnosis—the detection of very slow or nonprogressive cancers.

An alternative explanation is that there is a true increase in underlying amount of cancer destined to affect patients but that improvements in diagnosis and treatment coincidentally (and precisely) counterbalance the increase in new cancers—to leave cancer deaths unchanged. Although possible, this explanation is less likely. Not only is it not the most parsimonious explanation (it requires two assumptions instead of one) but also it requires that the rate of diagnosis and/or treatment improvement exactly match the increase in true disease burden (not too fast or mortality would fall, not too slow or mortality would rise).

The most credible population-based evidence for overdiagnosis comes from 30-year incidence and mortality data reported by Surveillance, Epidemiology, and End Results. For five cancers, the trends show increased rates of new diagnoses but not of deaths (Figure 7). In each case, increased screening activity or increased use of imaging tests capable of detecting incidentalomas is temporally associated with the increased rate of new diagnoses.

For thyroid cancer, the rate of diagnosis has more than doubled (from 4.9 per 100 000 to 10.6 per 100 000). Yet the rate of thyroid cancer death has been among the most stable of all cancers in the United States. The increase in new diagnosis has been confined to the histology with the most favorable prognosis (papillary thyroid cancer) and almost entirely consists of tumors less than 2 cm in diameter (30). The overdiagnosis of thyroid cancer likely reflects some combination of the increasing tendency of physicians to palpate the neck for masses (then refer for thyroid ultrasound) and incidental detection on ultrasounds and CT scans ordered for other reasons.

For melanoma, the rate of diagnosis has almost tripled (from 7.9 per 100 000 to 21.5 per 100 000). Again, the rate of death is generally stable (little change in the past 15 years). Although there may be an element of a true increase in clinically significant melanoma, these data suggest that most of the increase in diagnosis reflects overdiagnosis. The issue of overdiagnosis is well known to dermatologists (31–33). Because almost all the new diagnoses are localized (or in situ) melanomas and because their appearance



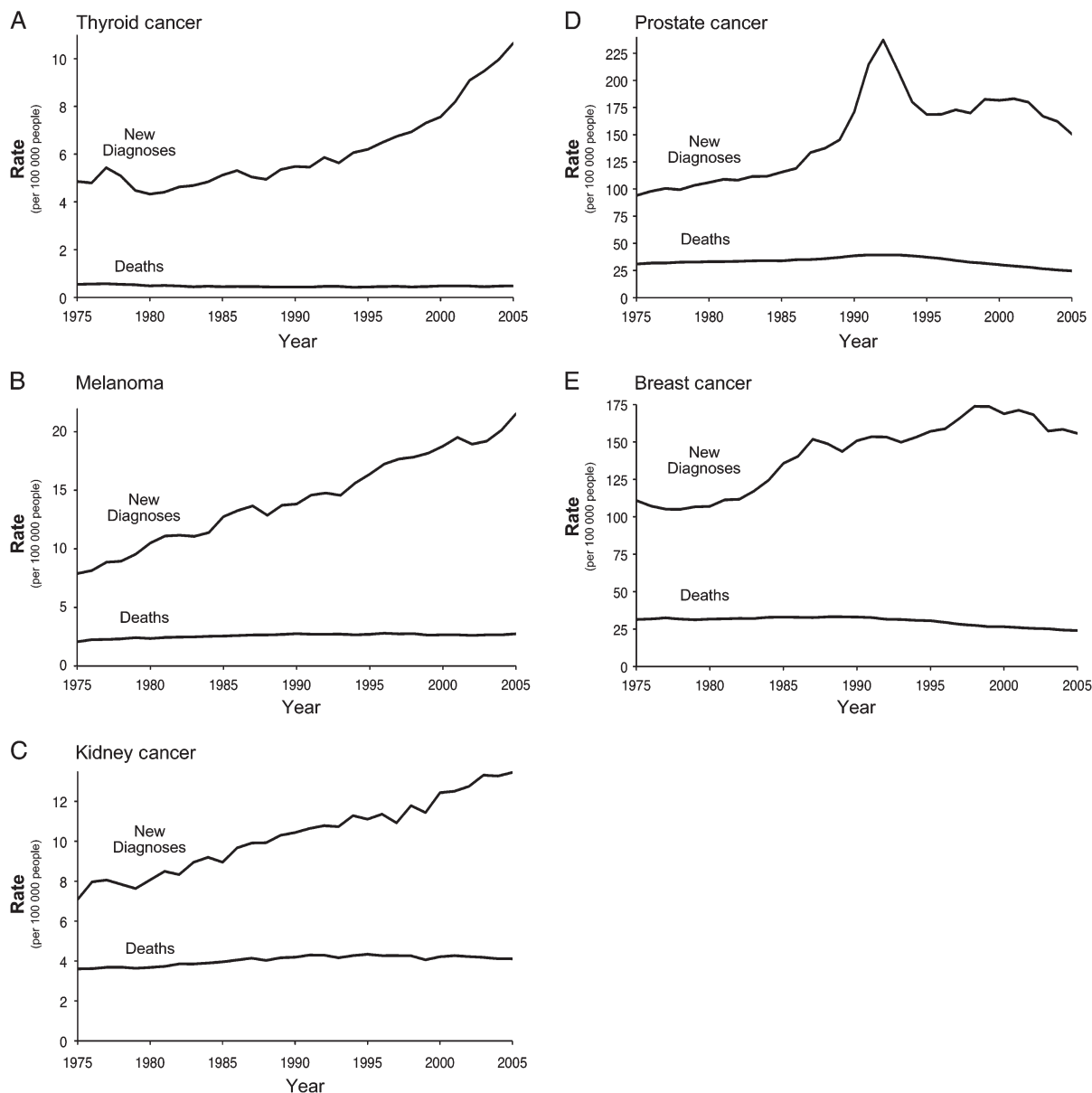


Figure 7. Rate of new diagnoses and death in five cancers in the Surveillance, Epidemiology, and End Results data from 1975 to 2005. **A)** Thyroid cancer. **B)** Melanoma. **C)** Kidney cancer. **D)** Prostate cancer. **E)** Breast cancer.

almost perfectly tracks the increase in population skin biopsy rates, overdiagnosis is likely the predominant explanation for the rise (34).

For cancers of the kidney and renal pelvis, rate of diagnosis has almost doubled over the past 30 years (from 7.1 per 100 000 to 13.4 per 100 000). However, the rate of death has been stable, with little change in the past 15 years. A recent investigation on the growth rate of 53 solid renal tumors, in which each tumor had at least two CT volumetric measurements 3 months apart before nephrectomy, demonstrated their variable natural history and the potential for overdiagnosis (35). Twenty-one (40%) had a volumetric doubling time of more than 2 years and seven (14%) regressed. Furthermore, slow-growing tumors were more common in the

elderly. Thus, it is likely that a substantial proportion of renal tumors represent overdiagnosis either because they do not grow at all or because their growth is too slow for the tumor to cause symptoms before the patient dies of other causes. Because there has been no systematic screening for these renal cancers, the increased rate of diagnosis is most likely because of incidental detection by the increasing use of abdominal ultrasound and CT.

Rising rates of diagnosis have occurred for both prostate and breast cancers. In both types of cancer, however, the story is more complex because the death rates for each are falling. In the past 15 years, prostate cancer mortality has fallen by about a third (from 38.6 per 100 000 to 24.6 per 100 000) and breast cancer mortality by about a quarter (from 33.1 per 100 000 to 24.0 per

100 000). This decrease reflects the combined effect of screening and improved therapy—and possibly, in the case of breast cancer, declining hormone replacement therapy use and women with new breast lumps presenting earlier for diagnostic mammography. But in both diseases, the combination of the data from randomized trials and from the population leaves little doubt that overdiagnosis is occurring. Thus, in these two diseases, we are left with the possibility that overdiagnosis because of early detection coexists with a mortality benefit from early detection. By contrast, in the case of the first three panels in Figure 7, it is difficult to identify a new and highly effective treatment capable of counterbalancing any true increase in incidence, resulting in unchanged mortality rates.

It is important to highlight those cancers for which there has been widespread screening yet little evidence that overdiagnosis is occurring in the population. There is little evidence of overdiagnosis of either cervical or colorectal cancer because the rate of diagnosis of both is falling (see Supplementary Technical Appendix, available online). If overdiagnosis is occurring as a consequence of screening for these two cancers, it is less cancer overdiagnosis and more overdiagnosis of the precursor lesions, for example, cervical dysplasia or adenomatous polyps.

Addressing the Problem

Overdiagnosis—along with the subsequent unneeded treatment with its attendant risks—is arguably the most important harm associated with early cancer detection. The impact of false-positive test results is largely transitory, but the impact of overdiagnosis can be life-long and affects patients' sense of well-being, their ability to get health insurance, their physical health, and even their life expectancy.

For clinicians and patients, overdiagnosis adds complexity to informed decision making: Whereas early detection may well help some, it undoubtedly hurts others. In general, there is no right answer for the resulting trade-off—between the potential to avert a cancer death and the risk of overdiagnosis. Instead, the particular situation and personal choice have to be considered. Often, the decision about whether or not to pursue early cancer detection involves a delicate balance between benefits and harms—different individuals, even in the same situation, might reasonably make different choices.

To address overdiagnosis, it is important to ensure that patients are adequately informed of the nature and the magnitude of the trade-off involved with early detection. This kind of discussion has been widely advocated as part of PSA screening but is nevertheless challenging for patients. They must first clearly understand the nature of the trade-off that although early diagnosis may offer the opportunity to reduce the risk of cancer death, it also can lead one to be diagnosed and treated for a “cancer” that is not destined to cause problems. Then, they must understand the magnitude of the trade-off. Each idea will be foreign and difficult, so they must be presented very clearly. We believe that this is best done through the construction of simple one-page balance sheets that frame the trade-off. We have provided one such example for screening mammography (Table 2).

The exercise of drafting a balance sheet highlights another important response: researchers need to work to develop reliable estimates of the magnitude of overdiagnosis. Consider the mammography example. In Malmö, there were 62 fewer breast cancer deaths and 115 women overdiagnosed (14)—a ratio of one breast cancer death avoided to two women overdiagnosed; yet others have argued that the ratio is 1 to 10 (37).

Admittedly, quantifying overdiagnosis is challenging. There are relatively few randomized trials of screening to start with, and even fewer will provide the needed long-term follow-up data. Nevertheless, even “best guess” estimates about the magnitude of overdiagnosis may play an important role in decision making. This effort will undoubtedly require modeling the natural history of the cancer, the impact of early diagnosis, and competing mortality. Although complex models may offer the highest degree of precision, their complexity can make it difficult for outsiders to review (or, in fact, even know) their structure and assumptions. Thus, we believe that there is an important place for more simple and transparent models in which all the assumptions, input values, and calculations are explicit and can be contained in a single spreadsheet.

A third response is to better understand patients' values regarding overdiagnosis. But researchers cannot understand patient values before patients understand the trade-off. Thus, efforts to determine preferences will need to be preceded by efforts to educate patients. Learning how sensitive patient preferences are to overdiagnosis (eg, whether changing the trade-off from 1:2 to 1:10 influences the decision to have mammography) will help inform us about how precise the estimates of overdiagnosis need to be.

A fourth response is to develop clinical strategies to help mitigate overdiagnosis. Overdiagnosis creates a powerful cycle of positive feedback for more overdiagnosis because an ever increasing proportion of the population knows someone—a friend, a family member, an acquaintance, or a celebrity—who “owes their life” to early cancer detection. Some have labeled this the popularity paradox of screening: The more overdiagnosis screening causes, the more people who feel they owe it their life and the more popular screening becomes (38). The problem is compounded by messages (in the media and elsewhere) about the dramatic improvements in survival statistics, which may not reflect reduced mortality, but instead be an artifact of overdiagnosis—diagnosing a lot of men

Table 2. Draft balance sheet for screening mammography in 50-year-old women*

Benefits	Harms
One woman will avoid a breast cancer death (36)	Between two and 10 women will be overdiagnosed and treated needlessly Between five and 15 women will be told that they have breast cancer earlier than they would otherwise yet have no effect on their prognosis Between 200 and 500 women will have at least one “false alarm” (50–200 will be biopsied)

* Among one thousand 50-year-old women undergoing annual mammography for 10 years. See Supplementary Technical Appendix (available online).

and women with cancer who were not destined to die from the disease (39).

It is possible that new insights from genomics will ultimately allow us to more accurately predict tumor behavior at the individual level. However, the field has not advanced to that point yet. We must explore other clinical strategies. One potential strategy to mitigate overdiagnosis is to raise the threshold to label a test as “abnormal” or the threshold to intervene. The diagnostic thresholds for common screening tests typically had their origins in arbitrary decisions (eg, PSA > 4 ng/mL). And the tendency over time has been for these thresholds to fall—either because we can see more (eg, microcalcifications on a mammogram) or because we learn that individuals below the threshold can still have cancer (leading some to argue for biopsies for PSA > 2.5 ng/mL).

The problem of overdiagnosis provides the motivation to investigate the other direction—testing higher diagnostic thresholds for labeling a screening test abnormal. One threshold to test is that of size—It may be better to simply ignore small abnormalities. This approach already has precedent in the use of size criteria to manage small pulmonary nodules (40) and adrenal lesions (41,42) incidentally detected on CT. There is an analogous threshold to test in laboratory values (such as a PSA)—that of magnitude—where it may be better to ignore what are now considered small elevations.

But most important may be to add an additional threshold that must be observed before labeling a screening test abnormal—that of growth. In spiral CT screening for lung cancer, demonstrating the growth of small lesions is now accepted as a prerequisite for biopsy—even among ardent screening proponents (43). Testing the effect of higher thresholds in randomized trials would offer the opportunity reduce not only overdiagnosis but also false-positive results.

Finally, there is much work to be done to incorporate the concept of overdiagnosis into the medical curriculum. Enthusiasm for new screening tests in the medical community is often based upon preliminary studies with inadequate study designs. Consequently, medical school curricula should incorporate formal coursework on how to evaluate screening tests and how to recognize overdiagnosis.

References

1. Dorland WAN. *Dorland's Illustrated Medical Dictionary*. 28th ed. Philadelphia, PA: W.B. Saunders Company; 1994.
2. Mooi WJ, Peeper DS. Oncogene-induced cell senescence—halting on the road to cancer. *N Engl J Med*. 2006;355(10):1037–1046.
3. Folkman J, Kalluri R. Cancer without disease. *Nature*. 2004;427(6977):787.
4. Serrano M. Cancer regression by senescence. *New Engl J Med*. 2007;356(19):1996–1997.
5. Sakr WA, Grignon DJ, Haas GP, Heilbrun LK, Pontes JE, Crissman JD. Age and racial distribution of prostatic intraepithelial neoplasia. *Eur Urol*. 1996;30(2):138–144.
6. Stamatou K, Alexizos A, Agapitos E, Sofras F. Incidence of impalpable carcinoma of the prostate and of non-malignant and precarcinomatous lesions in Greek male population: an autopsy study. *Prostate*. 2006;66(12):1319–1328.
7. Damiano R, Lorenzo GD, Cantiello F, et al. Clinicopathologic features of prostate adenocarcinoma incidentally discovered at the time of radical cystectomy: an evidence-based analysis. *Eur Urol*. 2007;52(3):648–657.

8. Harach HR, Franssila KO, Wasenius V. Occult papillary carcinoma of the thyroid: a “normal” finding in Finland. A systematic autopsy study. *Cancer*. 1985;56(3):531–538.
9. Welch HG, Black WC. Using autopsy series to estimate the disease “reservoir” for ductal carcinoma in situ of the breast: how much more breast cancer can we find? *Ann Intern Med*. 1997;127(11):1023–1028.
10. Ries LAG, Melbert D, Krapcho M, et al. *SEER Cancer Statistics Review, 1975–2005*. Bethesda, MD: National Cancer Institute; 2008. Based on November 2007 SEER data submission, posted to the SEER Web site. http://seer.cancer.gov/csr/1975_2005/. Accessed August 18, 2009.
11. Merrill RM, Feuer EJ, Warren JL, Schussler N, Stephenson RA. Role of transurethral resection of the prostate in population-based prostate cancer incidence rates. *Am J Epidemiol*. 1999;150(8):848–860.
12. Berland LL. Incidental extracolonic findings on CT colonography: the impending deluge and its implications. *J Am Coll Radiol*. 2009;6(1):14–20.
13. The Dartmouth Institute for Health Policy and Clinical Practice. *Dartmouth Atlas of Health Care*. Raleigh, NC: Lulu;2008.
14. Zackrisson S, Andersson I, Janzon L, Manjer J, Garne JP. Rate of overdiagnosis of breast cancer 15 years after end of Malmö mammographic screening trial: follow-up study. *BMJ*. 2006;332(7543):689–692.
15. Ramifications of screening for breast cancer 1 in 4 cancers detected by mammography are pseudocancers. *BMJ*. 2006;332:727. <http://www.bmj.com/cgi/content/full/332/7543/727>. Accessed March 15, 2009.
16. Fontana RS, Sanderson DR, Woolner LB, et al. Screening for lung cancer. A critique of the Mayo Lung Project. *Cancer*. 1991;67(4 suppl):1155–1164.
17. Marcus P, Bergstralh E, Zweig M, Harris A, Offord K, Fontana R. Extended lung cancer incidence follow-up in the Mayo Lung Project and overdiagnosis. *J Natl Cancer Inst*. 2006;98(11):748–756.
18. Andriole GL, Grubb RL, Buys SS, et al.; for the PLCO Project Team. Mortality results from a randomized prostate-cancer screening trial. *N Engl J Med*. 2009;360(13):1310–1319.
19. Schroder FH, Hugosson J, Roobol MJ, et al.; for the ERSPC Investigators. Screening and prostate-cancer mortality in a randomized European study. *N Engl J Med*. 2009;360(13):1320–1328.
20. Draisma G, Boer R, Otto SJ, et al. Lead times and overdiagnosis due to prostate-specific antigen screening: estimates from the European Randomized Study of Screening for Prostate Cancer. *J Natl Cancer Inst*. 2003;95(12):868–878.
21. Sone S, Takashima S, Li F, Yang Z, Honda T, Maruyama Y. Mass screening for lung cancer with mobile spiral computed tomography scanner. *Lancet*. 1998;351(9111):1242–1245.
22. Sone S, Li F, Yang Z, Honda T, Maruyama Y, Takashima S. Results of three-year mass screening programme for lung cancer using mobile low-dose spiral computed tomography scanner. *Brit J Cancer*. 2001;84(1):25–32.
23. Vineis P, Alavanja M, Buffler P, et al. Tobacco and cancer: recent epidemiological evidence. *J Natl Cancer Inst*. 2004;96(2):99–106.
24. Bessho F. Effects of mass screening on age specific incidence of neuroblastoma. *Int J Cancer*. 1996;67(4):520–522.
25. Bessho F. Where should neuroblastoma mass screening go? *Lancet*. 1996;348(9043):1672.
26. Yamamoto K, Hanada R, Kikuchi A, et al. Spontaneous regression of localized neuroblastoma detected by mass screening. *J Clin Oncol*. 1998;16(4):1265–1269.
27. Schilling FH, Spix C, Berthold F, et al. Neuroblastoma screening at one year of age. *N Engl J Med*. 2002;346(14):1047–1053.
28. Woods WG, Gao RN, Shuster JJ, et al. Screening of infants and mortality due to neuroblastoma. *N Engl J Med*. 2002;346(14):1041–1046.
29. Pohl H, Welch HG. The role of overdiagnosis and reclassification in the marked increase of esophageal adenocarcinoma incidence. *J Natl Cancer Inst*. 2005;97(2):142–146.
30. Davies L, Welch HG. The increasing incidence of thyroid cancer in the United States, 1973–2002. *JAMA*. 2006;295(18):2164–2167.
31. Swerlick RA, Chen S. The melanoma epidemic: more apparent than real? *Mayo Clin Proc*. 1997;72(6):559–564.
32. Dennis LK. Analysis of the melanoma epidemic, both apparent and real: data from the 1973 through 1994 surveillance, epidemiology, and end results program registry. *Arch Dermatol*. 1999;135(3):275–280.

33. Beddingfield FC III. The melanoma epidemic: res ipsa loquitur. *Oncologist*. 2003;8(5):459–465.
34. Welch HG, Woloshin S, Schwartz LM. Skin biopsy rates and incidence of melanoma: population based ecological study. *BMJ*. 2005;331(7515):481–484.
35. Zhang J, Kang SK, Wang L, Touijer A, Hricak H. Distribution of renal tumor growth rates determined by using serial volumetric CT measurements. *Radiology*. 2009;250(1):137–144.
36. US Preventive Service Task Force. *Effectiveness of Mammography in Reducing Breast Cancer Mortality*. Rockville, MD: Agency for Healthcare Research and Quality. <http://www.ahrq.gov/clinic/3rduspstf/breastcancer/bcscrnsu1.htm#results>.
37. Gotzsche PC, Hartling OJ, Nielsen M, Brodersen B, Jørgensen KJ. Breast screening: the facts—or maybe not. *BMJ*. 2009;338(7692):b86.
38. Raffle AE, Muir Gray JA. *Screening: Evidence and Practice*. New York, NY: Oxford University Press; 2007:68.
39. Welch HG, Schwartz LM, Woloshin S. Are increasing 5-year survival rates evidence of success against cancer? *JAMA*. 2000;283(22):2975–2978.
40. MacMahon H, Austin JH, Gamsu G, et al. Guidelines for management of small pulmonary nodules detected on CT scans: a statement from the Fleischner Society. *Radiology*. 2005;237(2):395–400.
41. Grumbach MM, Biller BM, Braunstein GD, et al. Management of the clinically inapparent adrenal mass (“incidentaloma”). *Ann Intern Med*. 2003;138(5):424–429.
42. Song JH, Chaudhry FS, Mayo-Smith WW. The incidental adrenal mass on CT: prevalence of adrenal disease in 1,049 consecutive adrenal masses in patients with no known malignancy. *Am J Roentgenol*. 2008;190(5):1163–1168.
43. International Early Lung Cancer Early Action Program Investigators. Survival of patients with stage I lung cancer detected on CT screening. *N Engl J Med*. 2006;355(17):1763–1771.

Funding

Department of Veterans Affairs Medical Center (H.G.W.).

Notes

The opinions in this manuscript are those of the authors and should not be interpreted as official positions of the Department of Veterans Affairs, the Department of Health and Human Services, or the US Government. The authors take sole responsibility for the study design, data collection and analysis, interpretation of the data, and the preparation of the manuscript.

Affiliations of authors: Veterans Affairs Outcomes Group, Department of Veterans Affairs Medical Center, White River Junction, VT (HGW); Department of Medicine (HGW) and Department of Community and Family Medicine (HGW, WCB), The Dartmouth Institute for Health Policy and Clinical Practice, Dartmouth Medical School, Hanover, NH; Department of Radiology, Dartmouth-Hitchcock Medical Center, Lebanon, NH (WCB).