How should we screen for breast cancer?

Using evidence to make medical decisions

Breast cancer is a killer. Screening for it is commonly seen as a life-saver, and moves to reduce screening programmes are understandably greeted with outrage. But, as **Howard Wainer** explains, evidence is a great help in deciding such issues.

Some 180 000 new cases per year of invasive breast cancer are diagnosed in women in the United States. About 40 000 of these women are expected to die from it. Breast cancer is second only to skin cancer as the most commonly diagnosed cancer, and second only to lung cancer in death rates. Among US women about one in four cancers are breast cancer and one in every eight US women can expect to be diagnosed with breast cancer. Figures for the United Kingdom are comparable, with about 45 000 new cases diagnosed each year.

However, some progress has been made in the battle against the horrors of breast cancer. Death rates have been declining over the past 20 years thanks to a combination of early detection and improved treatment. The first steps in early detection are self-examination and mammograms. The strategy is then to investigate any unusual lumps found from these relatively benign procedures with more invasive yet revealing methods – most particularly a biopsy.

The US Preventative Services Task Force periodically issues guidelines on cancer screening. In its 2002 guidelines it recommended that women begin having mammograms annually to screen for cancer when they reach the age of 40. Evidence available at that time strongly suggested that effective treatment, and hence women's survival, was crucially dependent on the early detection of tumours.

They made this recommendation despite the difficulty of interpreting the data. For example, suppose that a woman's expected survival is estimated to be only 5 years if a tumour is detected through self-examination. To be discovered in this way the tumour has to be of substantial size. Now suppose that the same woman had been having regular mammograms and that the tumour was discovered when it was still very tiny. In this instance, let us suppose that the average survival time is 10 years. We might conclude that the early detection has added 5 years to the woman's life. Surely such a result would support the use of annual mammograms. But how would our conclusion change if we learned that under usual circumstances it would take 5 years for a tumour to grow from a tiny size only detectable by a mammogram to a size large enough to be detectable by self-examination?

If these life expectancies were accurate we would certainly have to revise our estimate of the value of a mammogram. One conclusion might be that having a mammogram is not much fun, and it costs a few dollars, but it does no harm and might save a life.

In fact these numbers are not accurate. Women whose tumours were detected earlier lived longer, although how much longer is a difficult question to answer and subject to much legitimate debate. We shall return to this later.

Furthermore, a majority of tiny tumours never grow to become life-threatening. Instead they just sit there and do nothing. This raises an important issue: what is the appropriate action when a tiny tumour is discovered during a routine mammogram? The standard treatments – surgery, radiation, and chemotherapy – all have their own risks. Should the patient be subjected to these risks if the likelihood of what was discovered during the mammogram turning into something life-threatening is very small?

Thus we see that the decision to have a mammogram is not an unalloyed good. There are downsides that must be weighed against benefits before recommending any course of action. And such decisions need to use the most up-to-date evidence.

These issues are well known to the Preventative Services Task Force, which decided to review its recommendations in 2007. It convened a 16-member panel of experts appointed by the US Department of Health and Human Services and began its work gathering information. The Task Force contracted with the Evidence-Based Practice Center at the Oregon Health and Science University to collect all of the recent relevant work. This work was examined carefully, and at the end of its investigations it issued a report that modified its 2002 recommendations. It suggested a change in policy: that instead of all women beginning annual mammograms at 40, women not in any special-risk group (e.g. with no family history of breast cancer) should begin having biannual mammograms in their fifties.

The response to this change was fast and furious. It was claimed that the new guidelines were

politically motivated to save money. The defenders of the *status quo* were advocacy groups such as the American Cancer Society as well as professional groups that represented radiologists. And, of course, there was the usual chorus of politicians hoping to gain an advantage over those currently in office. The Obama administration immediately retracted the new guidelines and emphasised that it would not tolerate insurance companies using the new findings to disallow coverage of mammograms.

It is difficult to have a dispassionate discussion of this topic. The awful spectre of breast cancer hangs over half the population, and any attempt to diminish the availability of a practice that might reduce its terrible toll will be passionately opposed. Yet, let us try.

How effective are mammograms?

First let us consider how effective mammograms are. We are told that accuracy of mammograms varies from 80% to 90% depending on circumstances. That sounds pretty good. But what does such a probability mean? It is the probability of the mammogram being positive if you have cancer. Thus, if you have cancer, about 90% of the time the test will say so. But this probability is an answer to the wrong question, because we do not know if we have cancer. All we know is whether or not the test was positive. What we do wish to know is, if the test is positive, what is the probability that you have cancer? And, also, if the test is negative, what is the probability that we are cancer-free? Such probabilities are the characterisation of the test's accuracy that we should care about. These can be calculated using Bayes' theorem (with some base-rate data), but we can calculate them directly.

Let us calculate the answer to the question "if a mammogram is found to be positive, what is the probability that the patient has cancer?" We can estimate this probability from a fraction that has two parts. The numerator is the number of breast cancers found and the denominator is the number of positive mammograms. The denominator contains both the true and the false positives.

The numerator first: it is 180 000 cases.

The denominator has two parts: the true positives, 180 000, plus the false positives. How many of these are there? Each year there are 33.5 million mammograms given in the US. For the purposes of this discussion let us assume the more accurate figure, that when there is a cancer, 90% of the time it will be found, and when there is no cancer 90% of the time it will indicate no cancer. But this means that 10% of the time it will indicate a possible cancer when there is none. So, 10% of 33.5 million mammograms yield 3.35 million false positives. Thus the denominator of our fraction is 180 000 plus 3.35 million, or roughly 3.5 million positive mammograms.

Therefore the probability of someone with a positive mammogram having breast cancer is

$$\frac{180\,000}{3.5 \text{ million}}$$

or about 5%.

That means that 95% of the women who receive the horrible news that the mammogram has revealed something suspicious, and that they must return for further testing, are just fine. The mammogram lied to them. Is a test with this level of accuracy worth doing?

Without doubt, the answer to this question should be phrased in terms of human suffering and the length of time that lives are extended. I will get to that shortly. But to gain some understanding of at least one reason why the continuation of the old standard of annual mammograms has been so vehemently supported, it is revealing to look into the amount of money at stake.

Mammograms cost between $100 and $200 each. Let us use the smaller figure. So for 33.5 million mammograms the cost (conservatively) is $3.35 billion. Next, the 3.5 million positive mammograms are redone, adding $350 million to the total. This will yield about 350 000 positives that will require a biopsy. A biopsy's cost varies between about $1000 for a thin needle aspiration and $5000 for a fuller surgical procedure. Let us use the lower number. So to biopsy the 350 000 positive mammograms the cost would be at least $350 million. So far the cost is greater than $4 billion.

Now let us add on some other costs. Mammograms take at least 2–3 hours out of a working day, and biopsies at least 4 hours, often more. So, figuring conservatively, the two tests use up 75 million and 1.4 million hours of women's time, respectively. At even just $20 per hour, those 76 million hours represent an opportunity cost of $1.5 billion.

We have a conservative estimate of the annual costs of mammograms: $5.5 billion. In return for this cost we have the earlier detection of 180 000 breast cancers (actually 162 000, since 10% of the 180 000 that are eventually uncovered were missed in the initial mammogram). By dividing the costs of testing by the total number uncovered we can estimate the average cost of initial diagnosis:

$$\frac{\$5.5 \text{ billion}}{162\,000} = \$34\,000$$

Thus it costs at least $34 000 to diagnose one case of breast cancer. And for the unfortunate victims of the disease, this is only the start of the costs of treatment. With so much money at stake it would not be surprising that the *status quo* would be clung to by those whose livelihoods are closely tied to continuing into the future with the same rules we have used in the past.

This is just the monetary cost of detection. The human costs are far greater.

But how can we measure the emotional cost of the anxiety between the time one is told that there is an abnormality in the mammogram and the all-clear from the biopsy? We might also add in the complications that sometimes ensue after a biopsy (e.g. a staph infection) or the false positives from biopsies that can lead to aggressive treatment (e.g. surgery, radiation, chemotherapy) for a healthy woman. I will not try to place a dollar value on this aspect, but by any measure it is not a cost to be ignored when trying to value a mammogram.

It is surely more relevant to weigh the costs of detection against the likelihood of extending the patient's life. Here the picture gets more complicated. In 2009 Gøtzsche and Nielson[1] conducted a rigorous review of dozens of high quality studies of the value of mammography, encompassing more than half a million women. They concluded that "for every 2,000 women invited for screening throughout the ten years, one will have her life prolonged. In addition, 10 healthy women, who would not have been diagnosed if there had not been screening, will be diagnosed as breast cancer patients and will be treated unnecessarily. It is thus not clear whether screening does more good than harm."

Both aspects of this remarkable conclusion deserve our attention. First that only one out of every 2000 women screened will have her life extended. In the US, where 33.5 million women are annually screened, this means that 16,700 will have their lives extended because of the treatment triggered by the early detection afforded by a mammogram. This provides a new metric for our calculation of the value of a mammogram. Suppose we narrow the question: we ask "what is the detection cost for each life that is extended by that early detection?"

To calculate this we merely divide the total cost of screenings ($5.5 billion) in the US by the number of women whose lives were extended (16,700), yielding

$$\frac{\$5.5 \text{ billion}}{16\,700} = \$329\,000$$

A very large amount indeed, and it does not include any of the costs of treatment, nor the associated emotional costs. It also does not include the costs of both lives and treasure associated with Gøtzsche and Nielson's ominous second conclusion, that 10 healthy women per 2000 will be treated unnecessarily and hence suffer the

inevitable injury that modern cancer treatment entails. What sort of unnecessary treatment will they be subjected to? Is it just the relatively benign biopsy? Or is it something more? Gøtzsche and Nielson expand: "Finally, carcinoma in situ is much more likely to be detected with mammography and it is known that although less than half of the cases will progress to be invasive … these women will nevertheless be treated with surgery, drugs, and radiotherapy."

I have not tried to calculate the shortening of lives of the 167 000 healthy women who were incorrectly treated for breast cancer, but such a calculation would be crucial to balance the value of early screening.

The dilemma

Can we do anything to shift the balance? Yes, we must reduce the size of the denominator – the number of false positives. But so long as the ratio of healthy to sick people is so large, improving the accuracy of the test yields only limited help (going from 90% accurate to say 95% does not put much of a dent in it). The only real help would be reducing the number of mammograms administered that have only a tiny chance of revealing anything. But does this strategy, which surely reduces the number of false positives, also place women who do have cancer under unnecessary additional risk?

More light was shed on this very difficult triage decision in a September 2010 report on the results of a Norwegian study of the efficacy of mammograms[2]. This is the first study that examined the value of the early detection possible through mammograms when coupled with modern treatments such as hormonal therapy and other targeted drugs. In it the researchers compared the breast cancer death rates for women who had early detection with mammograms and those whose cancer was detected later after the tumour had grown enough to be noticed manually. They found that the difference in survival rates were small enough to be chalked up to chance.

The Norwegian result lends support to the growing concern over the widespread use of mammograms without regard to the often profoundly negative consequences for the many women on the receiving end of a false positive diagnosis. The price of false positives was worthwhile when successful treatment was dependent on early detection. But now modern treatments have seemingly thrown the earlier practice into a cocked hat.

But then, scarcely a week after the Norwegian study was published, the journal *Cancer* published another study, by Swedish researchers[3], that purported to show that women in their forties whose cancer was detected early by mammogram had a 26% lower death rate than women whose cancers were not detected early. This confusing result just added inertia to the *status quo*. This is unfortunate, because the Swedish study was marred by a methodological error that invalidates its results.

The Swedish study looked at the probability of death conditioned on the fact that the woman in question had a tumour. For the women who did not have a mammogram the numerator of this probability was the number of women who died of breast cancer and the denominator was the number of women who had a tumour that had grown large enough to have been discovered through self-examination. The women from the mammogram group had a different fraction. Its numerator was the same, the number of women who died from breast cancer. But the denominator was very different, and much more inclusive. It contained all women whose mammograms showed some sort of tumour. Most of those tumours were not destined to grow into anything, and surely some were false positives. Thus the value of the mammograms was much overstated. It is not clear how such a study could be corrected *post hoc*, but its flaws should have precluded its inclusion in the serious medical literature, where it could mislead those who are unfamiliar with the subtleties of self-selected samples.

An objective decision is only partially dependent on data for a procedure which generates more than $4 billion a year

My point is that these kinds of calculations emphasise that a screening device like a mammogram is not an unalloyed good. We must search for a balance to shape policy. I did not do the rough financial analysis in an attempt to try to weigh the value of a human life saved against some number of dollars. This is a difficult calculus for the money that is spent on unnecessary mammograms cannot be spent on other things that might have a greater positive impact on national health. No, my purpose was to suggest how hard it is to sharply limit a procedure that currently generates (in the US alone) more than $4 billion a year for its practitioners.

I note without further comment the reactions to the Norwegian study from three physicians whose commitment to the continued use of mammograms varies. Dr Barnett Kramer, associate director for disease prevention at the US National Institutes of Health, said: "This new study is very credible".

Dr Carol Lee, a radiologist at Memorial Sloan-Kettering Cancer Center and chair of the breast imaging commission of the American College of Radiology, said that the new study "affirmed that mammography saves lives."

Dr Laura Esserman, a professor of surgery and radiology at the University of California in San Francisco, said that it tells her "if you get the same treatment and the outcome is the same if you find the tumor earlier or later, then you don't make a difference when you find it early".

Conclusion

> "The problem is never how to get new, innovative thoughts into your mind, but how to get old ones out."
>
> Dee Hock

> "Old theories never die, just the people who believe in them."
>
> Einstein

Breast cancer is a horror. In earlier times, physicians had primitive weapons to combat it. The efficacy of those weapons was critically dependent on early detection. This being the case, the high number of false positives from mammograms could be tolerated. With modern treatments early detection is no longer crucial for success. Thus it is worthwhile to reconsider the value of pro forma mammograms. Making an objective decision is only partially dependent on data, for the size of the mammogram industry means that a substantial cutback in their use must overcome enormous momentum. This article is my attempt to explicate the issues, especially the statistical ones.

References

1. Gøtzsche, P. C. and Nielson, M. (2009) Screening for breast cancer with mammography: a review. *Cochrane Database of Systematic Reviews*, issue 4, article CD001877.
2. Kalager, M., Zelen, M., Langmark, F. and Adami, H. (2010) Effect of screening mammography on breast-cancer mortality in Norway. *New England Journal of Medicine*, **363**, 1203–1210.
3. Hellquist, B. N., Duffy, S. W., Abdsaleh, S., Björneld , L., Tabár, L., Viták , B., Zackrisson, S., Nyström, L. and Jonsson, H. (2010) Effectiveness of population-based service screening with mammography for women ages 40 to 49 years. *Cancer*, **116**. On-line in advance of publication, September 29, 2010.

Howard Wainer is Distinguished Research Scientist for the National Board of Medical Examiners and Adjunct Professor of Statistics at the Wharton School of the University of Pennsylvania.