



STATISTICS
by Thomas Lumley

Failing the screen test

Public screening for disease may sound good but the harm often outweighs the benefits.

Screening for disease is always a popular intervention. Early research on predicting Alzheimer's disease or cancer gets enthusiastic media coverage, and reluctance to implement universal screening is dismissed as short-sighted penny-pinching. The problem is that screening usually doesn't work very well because most people aren't sick.

In a sense, screening is the opposite of treatment. Treatment takes sick people and tries to make them healthy; screening takes healthy people and tries to diagnose them as sick. This innate harm of screening can only be outweighed by a benefit if there is an effective treatment, one that needs to be given early, and if there aren't too many people harmed by false diagnoses.

False positive diagnoses become more serious the rarer a condition is. Suppose you have a test that correctly classifies 99% of people. Think about using the test on a population of a million people. The number correctly diagnosed is at most the number who actually have the disease: if only 15,000 people have the disease, there can be at most 15,000 correct positive diagnoses; if only 500 people have

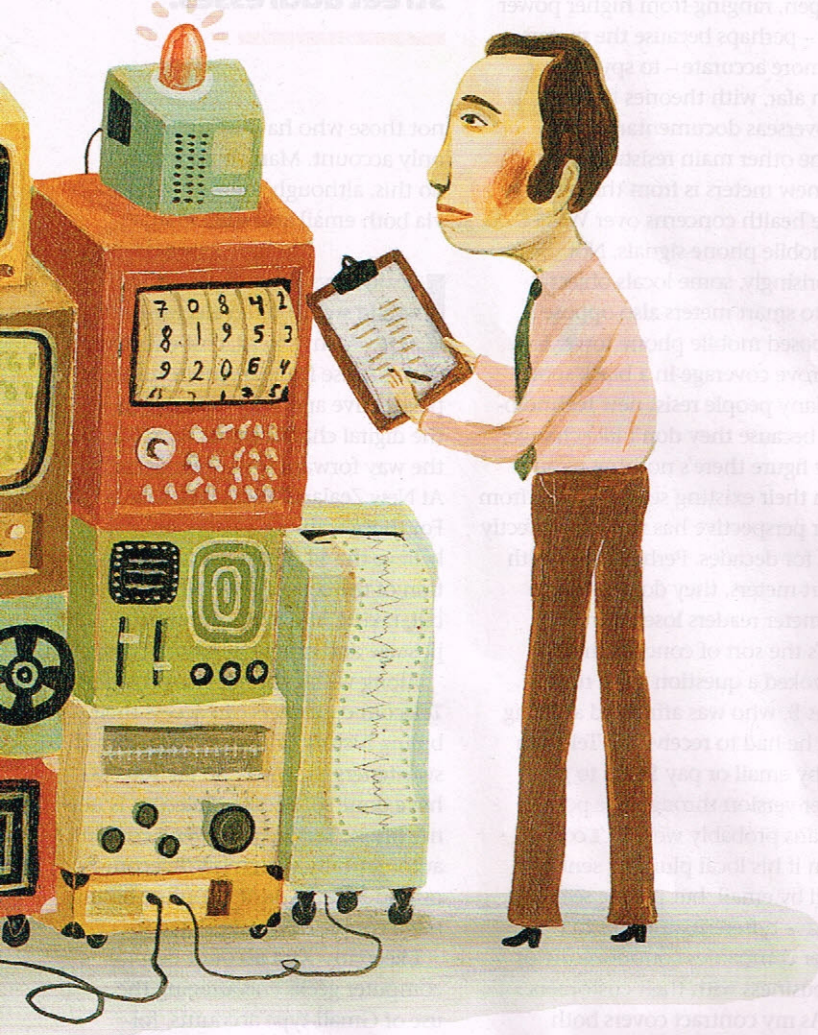
Screening is the opposite of treatment, as it takes healthy people and tries to diagnose them as sick.

the disease, at most there can be 500 correct positive diagnoses. Since the test is 1% inaccurate, about 10,000 people will be incorrectly diagnosed. If the disease is rare, the false positives will overwhelm the true positives and

most of those picked up by the screening will actually be healthy.

The first situation where false positives were studied carefully in medicine was in the 1940s with the Wassermann test for syphilis. It gave a false positive result in about 0.03% of people – 99.97% accuracy. When it was introduced, syphilis was much more common than 0.03% of the US population, so nearly all positive tests were correct. Penicillin rapidly made syphilis less common, but several US states required a Wassermann test before issuing a marriage licence. A Dr Garson pointed out in 1959 that the true prevalence of syphilis in South Dakota was only 0.03%, so false positive tests would be about as common as true positives. In technical language, the specificity of



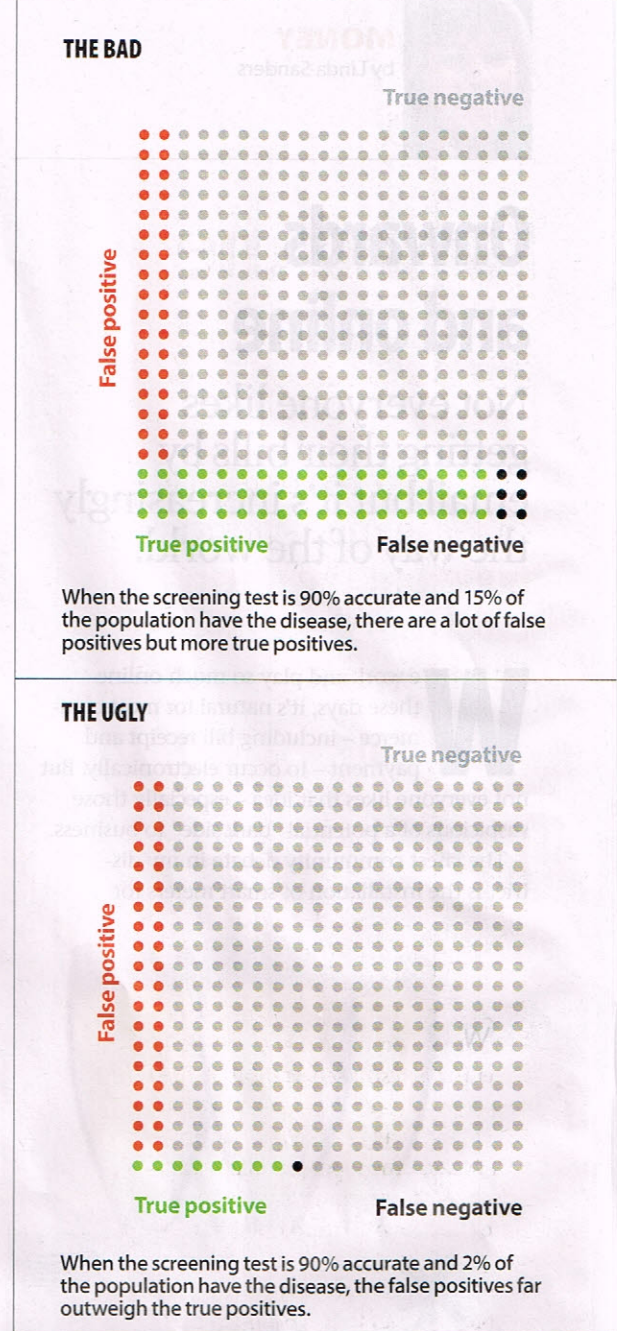


the test was 99.97%, but the positive predictive value was only 50%. That's low enough to make compulsory pre-marital screening distinctly problematic.

Most new screening technologies are not as accurate as the Wassermann test. A few months ago there were news stories about a potential test for Alzheimer's disease, which was reportedly 90% accurate in detecting the disease three years before it had a measurable effect on cognition. If we tested 1000 people aged 65-75, the US Alzheimer's

Association says that 150 of them would be expected to develop Alzheimer's within three years. The test would pick up 135 of the 150 true cases. Of the 850 who weren't going to develop the disease, about 85 would be diagnosed. Only 135 of the 220 positive diagnoses, 60%, would be accurate.

Is 60% good enough? In this case, no, not for public screening. The incorrect positive diagnoses would be crushing, and since there aren't yet any treatments that change disease progression, the benefit of the correct diagnoses would be relatively limited.



The test would be useful in selecting people for trials of new treatments. At the moment, treatments are tested on people who already have cognitive damage, but even a relatively inaccurate test would allow trials in healthy people at high risk, where there is the potential for more benefit.

In other situations a test of this accuracy can be useful even for screening. The Pap smear for cervical cancer is even less accurate, but it is very effective in preventing cervical cancer deaths, and the burden of a re-test or biopsy is less than the burden of a false diagnosis of Alzheimer's.

An ounce of prevention may be worth a pound of cure, but with population screening it is easy for an ounce of over-diagnosis to result in more than a pound of over-treatment. ■