Journal of Medical
Screening

Editorials

Spiral computed tomography screening

Lung cancer is one of the leading causes of cancer related death in both men and women in the world. This year, in the United States alone, an estimated 164 000 new cases of lung cancer will be identified and the disease will cause 156 000 deaths. Despite major efforts to improve cancer treatment over the past 20 years, progress has been modest. Nearly 90% of lung cancer is attributable to cigarette smoking. Avoiding uptake of the habit would be an effective means of primary lung cancer prevention. Smoking cessation lowers risk but the large population of former smokers remains at risk for lung cancer. Therefore, the development of an effective secondary prevention tool for lung cancer screening, if proven to decrease mortality from the disease, would be of great public health importance.

The strongest evidence regarding screening efficacy would come from a randomised clinical trial (RCT) with cancer specific mortality as the end point.[1] Screening for lung cancer with chest *x* ray (CXR) with and without sputum cytology was not shown to be effective in the three United States National Cancer Institute sponsored RCTs in the 1970s that addressed their utility.[2] The Mayo lung project (MLP) is the most illustrative of the trials. The MLP was a randomised screening trial of CXR and pooled sputum cytology every four months versus usual care, which consisted of a recommendation to have an annual CXR and sputum cytology (but without further efforts to achieve adherence to the recommendation). After an initial prevalence screen, 9211 high risk participants (male only) were randomised. There was an increase in the number of cancers detected in stage I and II in the screened arm (206 *v* 160). Interestingly, the actual number of cases detected at late stage was comparable. This led to a large improvement in five year survival in the screened group (31% *v* 13%). Nevertheless, the lung cancer death rate was virtually identical in both groups (3.2 *v* 3.0 per 1000 person years). Even after prolonged study follow up, lung cancer mortality was virtually the same in both study arms, despite substantially better survival time after diagnosis of lung cancer in the screened arm of the study. A recent update by Marcus *et al* at the United States National Cancer Institute with follow up to 25 years confirmed this.[3] Total mortality in the study arms was identical, demonstrating good distribution of risk factors as a consequence of the randomisation. This trial shows good evidence of overdiagnosis: tumours that do not come to attention in the absence of screening efforts. The MLP was designed to have a 90% power to detect a 50% reduction in lung cancer mortality. Because a more realistic reduction in mortality might have been missed, the United States National Cancer Institute is now conducting the large scale Prostate, Lung, Colorectal and Ovarian (PLCO) trial. Over 142 000 of the accrual goal of 148 000 participants have been enrolled.

The Early Lung Cancer Action Project (ELCAP) published their results of a baseline prevalence screen with spiral computed tomography (CT) which revealed an impressive relative increase in the proportion of lung cancer cases at early stages and an increased detection rate compared with CXR.[4] Two earlier studies had looked at the utility of spiral CT in Japanese groups with similar results.[5 6] None of these studies was controlled. In ELCAP, 1000 high risk, symptom free volunteers had a screening CXR and spiral CT. The *Lancet* reported on the results in individuals with one to six non-calcified nodules (NCNs). More diffuse abnormalities and four tumours diagnosed at a later stage were excluded from analysis. Spiral CT detected one to six NCNs in 233 (23%) participants. CXR detected one to six NCN in 68 (7%) participants. Twenty seven participants had a malignancy diagnosed and only seven of these were detectable on CXR. Twenty three (85%) of these malignancies were stage I (SEER distribution 22%). Twenty one of the malignant NCNs were adenocarcinomas, twice the expected distribution. Although this information is promising, as shown in MLP, an increased yield of stage I lung cancers does not necessarily translate into a mortality reduction. Non-randomised trials of screening modalities cannot readily account for lead time, length bias, or overdiagnosis bias. A large RCT with a mortality end point is required to assess the effectiveness of this technology. Before making public health recommendations, a strong knowledge of net benefit versus forms of early diagnosis and resulting treatment is required on both medical and ethical grounds.[7]

For these reasons the United States National Cancer Institute is planning to conduct a multicentre RCT of lung cancer screening with low dose spiral CT scan in current and former cigarette smokers. The primary end point will be lung cancer mortality. In preparation for such a trial, the NCI is planning a randomised feasibility study of several thousand participants designed to evaluate the willingness of subjects to be randomised, to determine the rate of abnormal examinations, and to estimate some of the operating characteristics of spiral CT in the screening setting. The feasibility study should be underway by late 2000 or early 2001.

BARNETT S KRAMER

National Cancer Institute,
Division of Cancer Prevention,
Bldg 31, 31 Center Drive,
Bethesda, MD 20892-2580, USA

1 Prorok PC, Kramer BS, Gohagan JK, eds. *Screening theory and study design: the basics in cancer screening*. New York: Marcel Dekker 1999:29–53.
2 Walpaw DR. *Early detection in lung cancer: case finding and screening in Medical Clinics of North America*. In: Cancer Screening and Diagnosis. 1996;**80**: 63–81.
3 Marcus P, Bergstralh EJ, Fagerstrom RM, *et al*. Lung cancer mortality in the Mayo Lung Project: the impact of extended follow-up. *J Natl Cancer Inst* 2000;**92**:1308–16.

4 Henschke C, McCauley D, Yankelevitz D, *et al*. Early lung cancer action project: overall design and findings from baseline screening. *Lancet* 1999;**354**:99–105.
5 Shusuke S, Shodayu T, Feng L, *et al*. Mass screening for lung cancer with mobile spiral computer tomography scanner. *Lancet* 1998;**351**: 1242–5.
6 Kaneko M, Eguchi K, Obmasu H, *et al*. Peripheral lung cancer: screening and detection with low-dose spiral CT versus radiography. *Radiology* 1996; **201**:798–802.
7 Malm H. Medical screening and the value of early detection: when unwarranted faith leads to unethical recommendations. Hastings Center Report, January–February 1999:26–37.

In search of better ways to transmit information about screening tests

Elsewhere in this issue (see page 123), Goyder *et al* express a view that the reporting of screening tests as "positive" or "negative" should be abandoned. Instead, reports should contain interpretations based on risks or probabilistic data. This recommendation is put forth as a first step towards remedying the problem that screened individuals may interpret the term "positive" to mean having the disorder and "negative" as being unaffected. The system of classifying screening test results as "positive" or "negative" thus becomes the bête noire to be dispatched before communication barriers can be torn down and replaced by an information system that will truly respond to individuals' needs.

Alternative terminology is proposed. As an illustration of this, the authors report interpretations of screening mammograms as: "normal", "further tests needed", "suspicious", and "malignant". The "normal" interpretation is synonymous with "negative". The other three interpretations call for subsequent action and thus can be classified as "positive". In this example, the words "positive" and "negative" are avoided but the concept remains. Health care providers and screened individuals readily understand the difference between "no action" and "further action needed" whether or not "positive" and "negative" are the specific words used.

The authors cite difficulties in choosing a single cut off for interpreting cholesterol measurements as one of the hazards of dichotomous screening test results. This example would more appropriately be applied to show why cholesterol should not be used as a screening test. It has long been recognised that cholesterol is not a suitable screening test for coronary artery disease, even though it is an important risk factor.[1] Cholesterol measurements in subjects who do not develop coronary artery disease overlap to such a great extent with those who do that no cut off can be defined that effectively segregates a high risk group which includes a significant proportion of cases. Cholesterol measurement can be helpful in documenting response to lifestyle changes or drugs, but its use in this context is not for screening purposes. The poor performance of cholesterol serves to reinforce the need for expert judgment in selecting which screening tests are to be recommended for introduction into practice. Wald and Cuckle have described a method for systematically analysing the performance of screening and diagnostic tests.[2] Guidelines such as these allow screening policy to be developed on a rational basis and also provide useful information of the type that Goyder *et al* would like to communicate more effectively.

Any screening test which is judged by accepted guidelines to be suitable for use in practice will have the capacity to segregate risk for the associated medical disorder sufficiently that a cut off can be determined for identifying individuals whose risk is high enough that further action is recommended. The rationale for setting cut offs is based on knowing comparative distributions of screening test measurements in affected and unaffected populations. With those distributions in hand, policy makers can determine a cut off (or range of possible cut offs) that achieves a satisfactory and logical balance between detection of the disorder and false positives. This key element needs to predate the introduction of any screening test into practice. If those responsible for providing interpretations of screening test results were to take the recommendation of Goyder *et al* literally and report only risks, the recipients of such reports would be left without expert guidance to help determine which screened individuals might need to consider further action.

Improving communication with people who are either candidates for screening or recipients of screening services is an important goal. Achieving that goal, however, is not an easy matter. Health care workers responsible for delivering screening services are faced with serious constraints on their time and resources. Any strategy for improving communication will need to work within these constraints. Whatever path is chosen, it would be unwise to abandon the concept of "positive" and "negative"; such an action would not be a substitute for ensuring that screening tests are properly explained and interpreted. The solution to the problem is education, not changing the terminology.

J E HADDOW

Foundation for Blood Research,
PO Box 190, Scarborough,
Maine 04074, USA

1 Wald NJ, Law M, Watt H, *et al*. Apolipoproteins and ischaemic heart disease: Implications for screening. *Lancet* 1994;**343**:75–9.
2 Wald NJ, Cuckle H. Reporting the assessment of screening and diagnostic tests. *Br J Obstet Gynaecol* 1989;**96**:389–96.