

# A critical appraisal of overdiagnosis: estimates of its magnitude and implications for lung cancer screening

J M Reich

## See editorial, p298

Correspondence to:  
Dr J M Reich, Thoracic Oncology Program, Earl A Chiles Research Institute, Portland Providence Medical Center, 7400 SW Barnes Rd, A622, Portland, OR 97225, USA; Reichje@dnamail.com

Appendix C contains a glossary of less familiar terms.

Received 14 February 2007  
Accepted 30 April 2007

## ABSTRACT

**Background:** The magnitude of overdiagnosis is a critical and unresolved issue in lung cancer (LC) screening: (1) its contribution to the increase in survival constitutes specious evidence of benefit; (2) overdiagnosed individuals who undergo resection will experience a reduction in life expectancy, partially or completely offsetting the benefit received by others in whom earlier intervention proves curative.

**Method:** Critical analysis of studies in opposition and support of the view that LC screening imposes a substantial burden of overdiagnosis.

**Results:** Approximately 25%, possibly more, of radiographically (chest x ray) diagnosed LC appears to be overdiagnosed. Based on the observed tumour volume doubling time of low dose CT identified small malignant pulmonary nodules, CT will markedly augment lead time, increasing exposure to competing lethal morbidities, thereby increasing overdiagnosis.

**Conclusion:** To reduce all-cause mortality, CT screening will need to reduce LC mortality by an amount that exceeds the increase in mortality attributable to surgery and loss of pulmonary reserve in persons who are overdiagnosed or pathologically understaged (ie, with occult micrometastases). Presently, there is no evidence that CT screening will achieve any reduction in LC mortality.

Because of its frequency and lethality, identifiable high risk population and absence of effective non-surgical treatment, lung cancer (LC) is an ideal screening candidate. However, because controlled trials combining radiographic (chest x ray (CXR)) and cytological screening achieved no reduction in mortality, screening is not recommended by recognised professional societies (summarised by Bach and colleagues<sup>1</sup>). The development of low dose, spiral CT screening has generated a resurgence of interest because it is about fourfold more sensitive than CXR at identifying small peripheral LC, and because approximately 80% of CT identified LC are stage IA, which have a highly favourable surgical prognosis<sup>2</sup>: 5 year LC survival of surgically staged IA LC is about 70% while overall 5 year LC survival is approximately 15%.<sup>3</sup> CT screening advocates assert that improvement in LC survival, predicated on a predominance of stage IA at diagnosis, confirms its efficacy.<sup>4</sup> Others point out that LC survival is an invalid metric of efficacy because screening entrains biases—lead time, length biased sampling and overdiagnosis—that convey a spurious appearance of benefit,<sup>5</sup> and it fails to reflect related non-cancer deaths.

Non-cancer mortality is high in LC survivors. Brown and colleagues<sup>6</sup> used the US National

Cancer Institute sponsored Surveillance, Epidemiology and End Results (SEER) database<sup>7</sup> to compare their mortality to the values from the National Center for Health Statistics.<sup>8</sup> They reported that 10% died of non-LC, and that the non-cancer relative hazard of death was 2.73, nearly threefold that of colon and breast cancer (1.09 each). The age at death of 22 randomly selected persons who underwent successful (no recurrence) resection of stage IA non-small cell lung cancer (NSCLC) at Portland Providence Medical Center between 1980 and 1989 was 5 years less than the actuarial survival for healthy gender and age matched smokers (Reich JM, Asaph JW, unpublished data). The excess non-LC mortality most probably reflects the combined effects of cardiopulmonary morbidities attendant on cigarette smoking and surgical reduction of pulmonary reserve, which would be expected to foreshorten their course. The excess mortality will impact three categories of persons undergoing notionally curative resection: (1) those otherwise destined to die of LC in whom surgery proves curative; (2) those with occult micrometastatic disease (understaged), who account for an LC fatality of about 30% in stage IA NSCLC (fatality = 1–survival); (3) those in whom LC would not have become clinically evident in their lifetime (overdiagnosed). The benefit for persons in category 1 will be partially offset by this harm. Persons in categories 2 and 3 cannot benefit from surgery; in aggregate, it will diminish their longevity.

The crucial issue in LC screening therefore is whether earlier intervention in category 1 will achieve a mortality reduction that exceeds the mortality increase attributable to invasive procedures in persons with true and false positive tests, tissue over-interpretation (eg, classifying atypical adenomatous hyperplasia as cancer<sup>9</sup>), individuals pathologically understaged because of undetected micrometastases and overdiagnosed individuals. This analysis updates and augments Parkin's 1998 presentation to the American Cancer Society International Conference.<sup>10</sup>

It will: (1) examine the concept of overdiagnosis; (2) critically evaluate the evidence of its magnitude; and (3) estimate its magnitude and effect on survival and mortality in CT screening.

## CONCEPT OF OVERDIAGNOSIS

LC is generally conceived of as an inexorably progressive, highly lethal neoplasm. The mid-seventies Connecticut experience reflects this conception: the average annual prevalence (P) was 23/, and incidence (I), 46/100 000 person-years (p-y), giving a mean duration (D) = survival = P÷I of

0.5 years<sup>11</sup> (see glossary, appendix C). Screening programmes reveal a wide spectrum of aggressiveness. Peripheral adenocarcinomas and its variants (particularly bronchioloalveolar carcinoma—BAC), presenting as solitary pulmonary nodules (SPN), exhibit the slowest growth,<sup>12</sup> and are therefore most susceptible to being overtaken by competing lethal morbidities.

Overdiagnosis denotes detection of disease that would never have become clinically evident, a pathological finding lacking clinical import, detectable only by special means (screening or autopsy). Overdiagnosis, pseudodisease, clinically irrelevant cancer, iatrogenic pseudodisease and lanthanic disease are interchangeable terms. Overdiagnosed LC may be conceptualised as a clinical–pathological false positive diagnosis because it erroneously implies a lethal outcome.

Overdiagnosis<sup>13</sup> occurs under two circumstances: (a) some indolent cancers prove non-lethal (eg, prostate and thyroid cancers are frequent incidental autopsy findings); (b) the course of aggressive cancers may be overtaken by lethal competing morbidities. LC, due to a common cause (smoking), is frequently accompanied by chronic obstructive pulmonary disease (COPD) and coronary artery disease. Screening, by imposing length biased sampling and lead time bias, increases overdiagnosis by prolonging exposure to competing lethal morbidities.

The incidence of clinically relevant LC is numerically bounded: it is given by its annual mortality (M) + the proportion of cured LC that is not overdiagnosed; it can be approximated as  $M \div 0.85$ , assuming 15% survival and zero overdiagnosis. Diagnostic methods, in proportion to their sensitivity, can increase the aggregate number of diagnosed LC, thus exceeding this limit.<sup>13</sup> This spurious incidence increase is due to clinically irrelevant cases. If one subscribes to the belief that minute cancers arise frequently and are held in abeyance by immunological surveillance,<sup>14</sup> it follows that the number of diagnosed cancers (incidence) is limited only by screening frequency and tumour size resolvable by the method used.

## EFFECT AND QUANTIFICATION OF OVERDIAGNOSIS

Overdiagnosis combines a specious improvement in outcome, as measured by cause specific survival,<sup>15</sup> with a genuine increase in mortality. Definitive evidence for and quantification of overdiagnosis would require a study of untreated subjects with screen identified, histologically confirmed, pathologically staged LC. Decades later, a panel reviewing survival curves, pathology, medical records and autopsy findings would define LC and all-cause survival. Those dying with clinically silent LC would constitute the overdiagnosed cases. Such a study is, of course, ethically and practically unfeasible. A discussion of overdiagnosis quantification is inseparable from the measurements by which it is operationally inferred.

## STUDIES ADVANCED IN SUPPORT OF A LOW RATE OF OVERDIAGNOSIS

### Survival in unoperated persons

Flehinger and colleagues<sup>15</sup> reported on the pooled results of the National Cancer Institute sponsored Early Lung Cancer studies performed at the Mayo Clinic, Johns Hopkins and Memorial Sloan Kettering, identifying all cases of stage I (T1NOMO and T2NOMO) NSCLC, detected by symptoms, incidentally or by screening. They identified 45 clinically staged (c staged) T1 and T2 persons who either refused or were judged to have medical contraindications for surgery. Twenty-one (47%) were stage IA. Twenty (44%) received radiation therapy. Five of 45 (11%) not

undergoing surgery did not die of LC within the 5 year follow-up period (fig 1 in Flehinger and colleagues<sup>15</sup>).

Overdiagnosis was estimated by 5 year surgically untreated LC survival as 11%. This estimate is fallacious because non-LC deaths (overdiagnosed, by definition, if their LCs were clinically silent) were treated as withdrawals. Furthermore, the survival figure underestimates untreated stage I LC survival because the authors conflated screen identified patients with those identified by symptoms (rare in stage IA) and incidental radiographs, because somewhat more than half were stage IB and because all were c staged (ie, understaged).

Sobue and colleagues<sup>16</sup> reported all-cause survival of 42 persons with CXR detected c stage I LC diagnosed in 1976–1981, collected from 20 institutes comprising the Japanese Lung Cancer Screening Research Group, who either refused surgery or were judged not to be surgical candidates. In contrast with the predominance of prognostically favourable stage I adenocarcinoma<sup>12</sup> and its variants in the Early Lung Cancer Action Project (ELCAP) prevalence screening trial,<sup>2</sup> the authors reported a less favourable histological distribution: 25 (60%) were squamous cell and four (10%) were small cell. Radiation or chemotherapy, or a combination, was provided to 81%. All-cause 5 year survival was 14%. Medical record review showed that seven of 41 (17%) patients died of causes other than LC or its treatment in the 10 years of follow-up.

It is impossible to determine whether the 17% non-LC deaths in unresected individuals constitutes evidence of overdiagnosis or the (limited) efficacy of non-surgical therapy. It is likely that many were understaged (mediastinal lymph node involvement was assessed by linear tomography in almost all), which has the opposite effect of stage migration (ie, the outcome for the assigned stage would be less favourable (reverse Will Rogers effect)).

Henschke and colleagues<sup>17</sup> used the 1988–1994 SEER registry data to characterise the survival of resected and unresected persons with stage IA NSCLC ascertained by any means. The authors computed an 8 year LC death rate “which adjusted for deaths from all other causes” (the phrasing implies that these values were cause specific). For those undergoing lobectomy, 8 year LC survival ranged from 60% to 75%, depending on size (0.6–3.0 cm), and for those receiving no treatment, from 6% to 13%.

Reporting LC rather than all-cause survival can be misleading, because if deaths within 30 days of surgery are excluded, or if a disproportionate number of individuals allocated to surgical intervention die of comorbidities and are excluded from the analysis—treated as withdrawals (ie, censored data)—the consequence will be a spuriously favourable “5 year survival.” In effect, people who die from adverse effects of diagnosis or treatment are not considered LC deaths, even though LC was a directly contributing cause. Survival of unoperated persons does not quantify overdiagnosis; it identifies individuals who, because their neoplasms are indolent, are at risk of overdiagnosis. Censoring non-LC deaths in persons with clinically silent LC excludes those who are, by definition, overdiagnosed.

Unlike the Sobue and Mayo Lung Programme analyses, in which the cause of death was established by the investigator’s review of medical records, the authors relied on causes of death abstracted from the National Center for Health Statistics database of consolidated death certificates from each state’s Vital Statistics Office, which are less accurate.<sup>18 19</sup> For example, some states employ a hierarchical designation in which LC is assigned as the cause of death if it is listed as a prior diagnosis.

## Growth rate

Yankelevitz and colleagues<sup>20</sup> concluded that frequent overdiagnosis could be excluded in the Mayo Lung Project and Memorial Sloan Kettering studies because of the similarity of tumour volume doubling time (TVDT) of 87 cases of stage I LC to that of non-screen identified adenocarcinomas. The estimated median TVDT—101 days in the Mayo and 144 days in the Memorial study—was similar to the range of median TVDTs (61–269 days) reported for 97 non-screen identified lung adenocarcinomas compiled from several centres.<sup>21</sup>

The analysis assumes that all non-screen identified adenocarcinomas to which it compared the Mayo and Memorial screen identified adenocarcinomas proved or would have proven lethal. For example, a 1 cm SPN with a 144 day TVDT would require about 4 years to achieve a lethal size (see appendix A). Approximately one-third of healthy white male smokers in their late sixties would be expected to succumb to all causes within 4 years,<sup>22</sup> during which time a 1 cm LC might remain clinically silent.

In addition to methodological limitations in computing TVDT described in the paper, the authors make an implicit assumption that growth rate, because of its relationship to lethality, is the sole determinant of overdiagnosis. Other variables influencing overdiagnosis are: patient age and presence of competing lethal comorbidities, tumour size at diagnosis,<sup>12</sup> metastatic propensity and ability to circumvent immunological surveillance.<sup>14</sup>

## cDNA profile

Mulshine and Sullivan advanced the similarity between complementary DNA (cDNA) composition of screen identified and clinically identified LC<sup>23</sup>—implying similar biological potential—to support the view that overdiagnosis is “unlikely.”<sup>24</sup> Bianchi and colleagues<sup>23</sup> found a similar gene expression profile in 18 CT detected and 18 stage and cell type matched “symptom detected” cases. The accompanying editorial,<sup>25</sup> which addressed a concern that CT detected LC might be overdiagnosed if they were of limited biological potential, stated, “. . . quantitative real-time PCR and immunohistochemistry suggests that the aggressive biological potential of the CT detected cancers is similar to that of symptom-detected cases.”

Despite the limitation imposed by overmatching (see below), Bianchi and colleagues<sup>23</sup> reported that nine genes (see table 2 in Bianchi and colleagues<sup>23</sup>) involved in tumour growth were differentially expressed in the CT detected versus the “symptom detected” group. “All were expressed in the symptomatic tumours at a lower level than in the CT scan detected cases and loss of expression of five of these genes is known to be associated with tumour progression.” The study design may have been flawed by overmatching. It largely (11 of 18 cases) compared cDNA phenotypes in persons with screen versus “symptom detected” stage IA LC. Stage IA is virtually never “symptom detected.”<sup>112</sup> Ascertainment typically results from an incidental radiograph taken for unrelated reasons. In effect, they are comparable to screen detected cases and might therefore be expected to have a similar gene expression profile. A cDNA comparison with advanced LC of matched cell type would be more illuminating.

That the “aggressive biological potential” of CT detected LC can be inferred from its cDNA phenotype is open to question. “Often, the gene expression signatures obtained with the use of microarray analysis are difficult to interpret with respect to the biology of the underlying disease.”<sup>26</sup> The biological potential of LC is a function of several known variables—size, growth rate,

angiogenic and metastatic potential, ability to evade immunological surveillance, mitotic rate and mutation rate—among other factors.<sup>12</sup> Some of these variables are, presumably, subject to epigenetic influences. None are directly measurable with cDNA microarray phenotyping. Growth rate is the simplest measure of biological potential. CT identified LC presenting as SPN exhibit a vast difference—almost 35-fold—in TVDT (see below), from 46 days to 4 years. It is implausible that the biological potential of LCs possessing similar cDNA profiles and widely divergent TVDTs are similar.

## STUDIES ADVANCED IN SUPPORT OF A HIGH RATE OF OVERDIAGNOSIS

### Controlled trials

Two large randomised controlled trials allocated older male smokers to a screened versus an unscreened cohort following a prevalence CXR and cytological baseline screen to exclude persons with identifiable LC.

In the incidence portion of the Mayo Lung Programme,<sup>27</sup> 206 cases of LC were identified in the screened versus 160 in the unscreened cohort. The excess  $((206 - 160) \div 206 = 22\%)$  46 cases were characterised as “missing cases,” as a comparable number did not appear in the usual care cohort after extended follow-up.<sup>28</sup> Eddy postulated that they were overdiagnosed (ie, the number of LCs was approximately equal in both cohorts), the usual care arm counterparts with unidentified LC dying of competing morbidities.<sup>29</sup>

Kubik and colleagues<sup>30</sup> undertook a similar study in Czechoslovakia. Following a baseline screen, male smokers were randomly allocated to semi-annual CXR and cytological screening or to an unscreened control cohort. At 3 years and annually for 3 years thereafter, both cohorts were CXR screened. The LC cases in the screened versus the initially unscreened cohort, after prolonged follow-up, was  $(108 - 82) \div 108 = 24\%$ .

LC screening is predicated on the assumption that resection of curable LCs, premised progenitors of advanced LC, will achieve a commensurate, reciprocal reduction in the latter (stage shift). Neither trial demonstrated this essential effect: in the Mayo Lung Project, almost twice as many LCs (94, 46%) were completely resected in the intervention than in the usual care cohort (51, 32%). However, the number of advanced LCs identified during the entire study—112 in the intervention versus 109 in the usual care cohort—was not reduced. Similarly, in the Czech trial, nine persons identified in the intervention during the initial 3 year screen underwent “curative resection” versus eight in the control cohort.<sup>31</sup> At completion of the study, however, the number of resectable cases was similar in both cohorts: intervention 25/108 (23.1%); control 19/82 (23.2%).<sup>32</sup> Fontana and colleagues<sup>33</sup> summarised the three National Cancer Institute sponsored CXR and cytological screening studies, reporting that there were 303 advanced LC cases (stages III, IV) in the intervention groups compared with 304 in the controls.

Plausible explanations for failure of CXR screening to interdict advanced LC are:

- ▶ The majority of advanced LCs originate in proximal endobronchial sites, where radiographic screening lacks sensitivity.
- ▶ Advanced LCs that originate in the periphery of the lung, and are therefore potentially susceptible to early diagnosis, have such an aggressive natural history that early detection within practicable screening intervals is not achievable.
- ▶ Metastasis may occur in some cancers before they reach an actionable size. Each explanation for the absence of stage

shift implies that few of the excess screen identified, resectable LCs (principally stage I) were progenitors of advanced LC (ie, they were clinically irrelevant).

### Autopsy findings

McFarlane and colleagues<sup>34</sup> reported that the age adjusted incidence of previously unsuspected autopsy detected LC at Yale New Haven Medical Center was almost fourfold the community incidence rate in men and almost 15-fold higher in women. These “surprise” LCs neither caused nor contributed to the deaths of their patients.

In an updated review from the same site, Chan and colleagues<sup>35</sup> reported that one in six autopsy detected LCs were unrecognised before death; that 1% of men autopsied in 1973–1982 had previously unsuspected LC; and that most (70%) were resectable (TNM stages 0 to II) and were therefore presumably asymptomatic. Two other autopsy series reported similar findings.<sup>36–37</sup> These series demonstrate the existence of a large “reservoir” of CXR detectable, clinically irrelevant LC.

Chan’s analysis showed that the incidence of surprise LC was sixfold the estimated incidence of overdiagnosis—1.2/1000 p-y—in the Mayo study, indicating that the 22% estimate is conservative (appendix B). Autopsies are likely to underestimate the “reservoir” of CT identifiable malignant SPN: Dammas and colleagues<sup>38</sup> reported that, among 28 persons who had a CT scan demonstrating an SPN within 2 months of death, nine (22%) were not identified in the autopsy report.

### Histology and epidemiology:

If CT screening identified solely clinically relevant cancers, their histological and epidemiological features would approximate that of clinically identified cancer in the population.

Sone and colleagues<sup>39</sup> reported on a mass population CT screening of Japanese men and women over age 40 years (median 64), smokers and non-smokers. They identified 23 histopathologically staged LCs in the initial (prevalence) screen: 21 were IA and two were IB. Sixteen were BAC or well differentiated adenocarcinomas; three poorly differentiated adenocarcinoma; four squamous carcinoma. The proportion of cases in never smokers and females (each 0.44%) exceeded that in smokers and males (each 0.40%). The detection rate in the 3 year annual screening programme—406/100 000 p-y—was 11-fold the annual LC mortality rate—37.3/100 000 p-y—in this country region of Japan.

Yang and colleagues<sup>40</sup> in a review of LC at the Mayo Clinic in 1997–2003, found that BAC constituted 3.1% of all cases (Ping Yang, MD, PhD, written communication, 2005). Swensen and colleagues<sup>41</sup> reported that CT screening identified 11 cases of BAC (17% of screen identified LC), sixfold their clinical representation.

### Growth pattern:

Bach and colleagues<sup>42</sup> pointed out that “In the . . . ELCAP study, the number of cancers detected in the annual follow-up CT<sup>43</sup> is far less than that detected during the initial scan, even though the size of the detected lesions is similar. This difference in rates of detection is inconsistent with what one would expect if all the lesions grew at similar rates (ie, all were aggressive malignancies). Instead, that the size of the prevalent lesions detected at initial scan are consistent with those found at subsequent scans but that the rates of detection are substantially higher suggests that a meaningful proportion of cancers

detected during the initial scan would have behaved in an indolent manner”.

### QUANTIFICATION OF OVERDIAGNOSIS

The belief that overdiagnosis is unlikely entails its biologically implausible corollary: that a screen diagnosis of LC confers virtual immunity to death from all other causes. While the magnitude of CXR screening overdiagnosis cannot be determined with certainty, the per cent of “missing cases” in the Mayo and Czech studies, 22 and 24, respectively, furnish a lower bound. In the Mayo study, 95% of the intervention cohort received CXRs every 4 months; the control group received “usual care”—more than half of whom elected to have one or more CXRs in the study period. The computation of putative overdiagnosis assumes that none of the usual care cases was overdiagnosed. This is likely to be incorrect because some cases in the usual care cohort were screen identified. If the same percentage (22) were overdiagnosed in the controls who had one or more CXRs in the last 2 years of the trial,<sup>33</sup> there would be 125 clinically relevant LC. The percentage of overdiagnosed cases would then be  $(206 - 125) \div 206 = 39\%$ . No generally applicable figure, however, can be given for the magnitude of overdiagnosis, for it will be influenced by methodology (see below), smoking history, histological distribution, age, ethnicity and comorbidities of the screenees.

### MITIGATION OF OVERDIAGNOSIS

Overdiagnosis can be reduced by excluding individuals with clinically discernable, potentially lethal comorbidities (eg, COPD) from screening. However, this strategy will diminish screening efficiency because the incidence of LC is far higher in persons with impaired pulmonary function,<sup>44–45</sup> and a reduction in the absolute number of LCs will increase the proportion with false positive tests. For these reasons, Petty<sup>46</sup> advocated screening the high risk group, persons with identifiable COPD. Subset screening creates an irresolvable dilemma: the benefit of reducing overdiagnosis by screening healthy smokers is offset by its diminished efficiency. The greater efficiency conferred by screening persons with COPD (who are likely to have concurrent, latent, coronary artery disease) is offset by a reduction in benefit and an increase in overdiagnosis; both are due to their lower life expectancy. Furthermore, limited pulmonary reserve will increase their surgical morbidity and diminish their post-surgical quality of life.

### TUMOUR GROWTH KINETICS AND OVERDIAGNOSIS

Based on observations of tumour growth, both Weiss<sup>47</sup> and Geddes<sup>12</sup> calculated that the natural history of LC could be characterized by exponential growth, encompassing, at most, 40 tumour volume doublings (TVDT) (appendix A)

Aoki and colleagues<sup>48</sup> reported that, among resectable adenocarcinomas <3 cm in diameter, TVDT ranged from 42 to 1486 days, and half had a TVDT greater than 1 year. Hasegawa and colleagues<sup>49</sup> reported a mean TVDT of 452 days (1.24 years) in 61 malignant SPN identified by mass CT screening. Based on these TVDTs, approximately 10 years would be expected to elapse between screening diagnosis at 1 cm and death from LC (assuming constant TVDT); for 0.5 cm, the interval would be about 14 years.

Life expectancy (= median survival) in the USA at age 67 years—the median age of enrolment in the ELCAP trial<sup>2</sup>—in healthy male smokers is 16 years and in healthy female smokers 19 years.<sup>50</sup> Based on these TVDT, and assuming a linear inverse

relationship between age and life expectancy in this age group, one would predict that, after prescreening (to exclude individuals with significant comorbidities), 31% of males and 26% of females with a 1 cm SPN would die of all causes (of which LC would constitute a small per cent) had they not been screen identified.

### INFLUENCE OF CT IDENTIFICATION ON OVERDIAGNOSIS AND OUTCOME

Because CT augments lead time bias and introduces a favourable biological (phenotypic—see below) bias, it will increase overdiagnosis and hence both survival and mortality. Bach and colleagues<sup>51</sup> pooled the results of three one-armed CT screening trials and compared their outcomes with LC incidence and stage expectations modelled on SEER data. They reported 144 LC cases observed versus 44.5 expected. There was no diminution in advanced stage LC compared with the predicted values, implying that some of the 99.5 excess cases were overdiagnosed ( $99.5 \div 144 = 69\%$ ).

#### Lead time bias

With a mean TVDT of 1 year, 3 years are required for a tumour diameter to increase from 0.5 to 1 cm. For this reason, and because 1 cm tumours are frequently obscured in CXRs, CT, in comparison with CXR screening, will increase survival of stage I adenocarcinomas by approximately 3 years because of lead time bias. Under the dual assumptions that all-cause mortality remains constant and that earlier resection does not reduce the occurrence of occult metastases, 5 year stage IA LC fatality ( $\approx 30\%$ ) would be reduced by a factor of 3/5, increasing 5 year LC survival from  $\approx 70\%$  to  $\approx 90\%$ .

Non-LC mortality in the screened cohort of the Mayo trial<sup>27</sup> was 21.6/1000p-y = all-cause mortality, 24.8/- LC-mortality, 3.2/. The ratio of non-LC-mortality to LC-mortality was  $(21.6 \div 3.2) = 7:1$ . Based on the assumption that CT increases lead time by 3 years, the percentage of individuals diagnosed with clinically irrelevant stage I LC would increase to an unquantifiable but substantial amount: (1) the lead time estimate is conservative; (2) persons who develop LC have more severe obstructive airways disease (and coronary heart disease<sup>52</sup>) than the cohort from which they are drawn, and would therefore be expected to have a higher non-LC-mortality; (3) pulmonary resection is likely to foreshorten the course of lethal cardiopulmonary comorbidities.<sup>53</sup>

#### Biological bias

CT screening introduces a phenotypic bias because of the identification of an LC population differing in histological composition and biological potential from those identified either clinically or by CXR screening.<sup>41</sup> BACs have greater transradiancy because they retain the underlying air containing alveolar structure (lepidic growth). This “ground glass” appearance offers far less contrast with the surrounding lung.<sup>54</sup> For this reason, they are less detectable with CXR than are adeno- and squamous carcinomas, which, by contrast, have a solid density (hilic growth). Thus BACs are overrepresented in CT screenings, relative to both their proportion in CXR screenings and reflecting their slow growth, to the proportion of LC that become clinically evident.

Shimizu and colleagues<sup>54</sup> reported on 136 thin section CT evaluated air containing carcinomas under 20 mm in size, p-stage IA. One per cent showed pleural involvement; 2% vascular invasion; none showed lymphatic invasion. Relapse free 5 year

survival was 100%. Hasegawa and colleagues<sup>49</sup> reported on TVDT of CT identified small SPN according to their high resolution CT characteristics. The mean TVDT was 813 days among 19 with a pure ground glass appearance. The authors pointed out that a large number of these were CXR undetectable. Assuming no evolution to a more aggressive phenotype, a 1 cm BAC with an 813 TVDT would require 22 years to achieve a lethal size (10 cm).

#### Overdiagnosis and outcome

Although LC mortality is the most sensitive test of screening efficacy, outcome can best be gauged by complementing LC mortality with all-cause mortality.<sup>55</sup> All-cause mortality assesses the benefit conferred by earlier intervention with the offsetting harm done to persons with overdiagnosed and understaged LC and false positive tests.

All-cause mortality was higher in the screened than in the control cohort in both the Mayo and Czech trials.<sup>55</sup> Under the assumption that vital status has been ascertained in all participants after an adequate period of observation, which appears to be correct,<sup>28,30</sup> this suggests that CXR with cytological screening exerts a negative net effect. To offset the postulated adverse effect of increased overdiagnosis, CT screening would have to substantially reduce the number of cases with occult metastases and diminish the absolute number with advanced (unresectable) LC. Whether it will achieve the former by advancing the diagnosis by about 3 TVD is unknown. No reduction in the number of advanced LC was achieved in the CXR imaging studies<sup>27,30</sup>; none is evident in preliminary analyses of CT screening.<sup>41,51,56,57</sup>

#### EFFICACY VERSUS EFFECTIVENESS

Mortality (efficacy), surgical morbidity and overdiagnosis will be far lower in screening trials conducted in centres of excellence, which systematically exclude persons with identifiable significant comorbidities<sup>41</sup> than in population based settings (which gauge effectiveness).<sup>58</sup>

#### CONCLUSIONS

1. Although difficult to quantify, CXR with cytological LC screening results in a clinically important amount of overdiagnosis.
2. To improve outcome, CT screening, which will increase the proportion of overdiagnosed cases, will need to achieve a reduction in LC mortality sufficient to offset the immediate and long term adverse effects imposed by surgery in overdiagnosed, understaged and false positive cases.
3. Overdiagnosis will be greatest in series that confine their outcome assessment to CT identified stage I LC.

**Acknowledgements:** I am greatly indebted to Pamela Marcus, MS, PhD, epidemiologist, National Cancer Institute, for her critical reading of the manuscript and for her numerous incisive comments and contributions.

**Competing interests:** None.

#### REFERENCES

1. **Bach PB,** Niewoehner DE, Black WC. Screening for lung cancer: the guidelines. *Chest* 2003;**123**(Suppl1): 83S–8S.
2. **Henschke CI,** McCauley DI, Yankelevitz DF, *et al.* Early Lung Cancer Action Project: overall design and findings from baseline screening. *Lancet* 1999;**354**:99–105.
3. **Mountain CF.** Revisions in the international system for staging lung cancer. *Chest* 1997;**111**:1710–17.
4. **Miettinen OS,** Henschke CI. CT screening for lung cancer: coping with nihilistic recommendations. *Radiology* 2001;**221**:592–6.
5. **Reich JM.** Assessing the efficacy of lung cancer screening. *Radiology* 2006;**238**:398–401.

6. **Brown BW**, Brauner C, Minnotte MC. Noncancer deaths in white adult cancer patients. *J Natl Cancer Inst* 1993;**85**:979–87.
7. **NCI**. *Surveillance, Epidemiology, and End Results (SEER) program: SEER public use tape (1973–87)*. Bethesda, Maryland: Cancer Statistics Branch, Surveillance Program Division of Cancer Prevention and Control, NCI 1990.
8. **National Center for Health Statistics**. *Vital statistics of the United States, vol 2. Mortality*. Rockville, MD: National Center for Health Statistics, 1973–1987.
9. **Mori M**, Chiba R, Takahashi T. Atypical adenomatous hyperplasia of the lung and its differentiation from adenocarcinoma. *Cancer* 1993;**72**:2331–40.
10. **Parkin DM**, Moss SM. Lung cancer screening: improved survival but no reduction in deaths—the role of “overdiagnosis”. *Cancer Suppl* 2000;**89**:2369–76.
11. **US DHHS**. *Surveillance, epidemiology, and end results: Incidence and mortality data, 1973–77*. Bethesda, MD: National Cancer Institute, 1981, NIH publication No 81-2330.
12. **Geddes DM**. The natural history of lung cancer: a review based on rates of tumour growth. *Br J Dis Chest* 1979;**73**:1–17.
13. **Black WC**. Overdiagnosis: an underrecognized cause of confusion and harm in cancer screening. *J Natl Cancer Inst* 2000;**92**:1280–2.
14. **Dunn GP**, Bruce AT, Ikeda H, et al. Cancer immunoediting: from immunosurveillance to tumor escape. *Nat Immunol* 2002;**3**:991–8.
15. **Fleehinger BJ**, Kimmel M, Melamed MR. The effect of surgical treatment on survival from early lung cancer. *Chest* 1992;**101**:1013–18.
16. **Sobue T**, Suzuki T, Matsuda M, et al. Survival for clinical stage I lung cancer not surgically treated: comparison between screen-detected and symptom-detected cases. *Cancer* 1992;**69**:685–92.
17. **Henschke CI**, Wisnivesky JP, Yankelevitz D, et al. Small stage I cancers of the lung: genuineness and curability. *Lung Cancer* 2003;**39**:327–30.
18. **Feinstein AR**, Esdaile JM. Incidence, prevalence, and evidence: scientific problems in epidemiologic statistics for the occurrence of cancer *Am J Med* 1987;**82**:113–23.
19. **Kircher T**, Nelson J, Burdo H. Inaccuracy of death certificates in assigning cause of death. The autopsy as a measure of accuracy of the death certificate. *N Engl J Med* 1985;**313**:1263–9.
20. **Yankelevitz DF**, Kostis WJ, Henschke CI, et al. Overdiagnosis in chest radiographic screening for lung cancer: frequency. *Cancer* 2003;**97**:1272–5.
21. **Friberg S**, Mattson S. On the growth rates of human malignant tumors: implications for medical decision making. *J Surg Oncol* 1997;**65**:284–97.
22. <http://www.lifeexpectancy.com/help/1.0/Introduction.shtml> (accessed 23 November 2007).
23. **Bianchi F**, Hu J, Pelosi G, et al. Lung cancers detected by screening with spiral computed tomography have a malignant phenotype when analyzed by cDNA microarray. *Clin Cancer Res* 2004;**10**(pt 1):6023–8.
24. **Mulshine JL**, Sullivan DC. Lung cancer screening. *N Engl J Med* 2005;**352**:2714–20.
25. **Mulshine JL**, Weinstein JN. Is the gene expression pattern of lung cancer detected by screening with spiral computed tomography different from that of symptom-detected lung cancer? *Clin Cancer Res* 2004;**10**(pt 1):5973–4.
26. **Quackenbush J**. Microarray analysis and tumor classification. *N Engl J Med* 2006;**354**:2463–72.
27. **Fontana RS**, Sanderson DR, Woolner LB, et al. Lung cancer screening: The Mayo program. *J Occup Med* 1986;**28**:746–50.
28. **Marcus PM**, Bergstralh E, Fagerstrom RM, et al. Lung cancer mortality in the Mayo Lung Project: impact of extended followup. *J Natl Cancer Inst* 2000;**92**:1308–16.
29. **Eddy DM**. Screening for lung cancer. *Ann Intern Med* 1989;**111**:232–7.
30. **Kubik A**, Parkin DM, Khat M, et al. Lack of benefit from semi-annual screening for cancer of the lung: follow-up report of a randomized controlled trial on a population of high-risk males in Czechoslovakia. *Int J Cancer* 1990;**45**:26–33.
31. **Kubik A**, Polák J. Lung cancer detection: results of a randomized prospective study in Czechoslovakia. *Cancer* 1986;**57**:2427–37.
32. **Kubik A**, Parkin DM, Zatloukal P. Czech study on lung cancer screening: post-trial follow-up of lung cancer deaths up to year 15 since enrollment. *Cancer Suppl* 2000;**89**:2364–8.
33. **Fontana RS**, Sanderson DR, Woolner LB, et al. Screening for lung cancer. A critique of the Mayo Lung Project. *Cancer* 1991;**67**:1155–64.
34. **McFarlane MJ**, Feinstein AR, Wells CK, et al. The ‘epidemiologic necropsy’: unexpected detections, demographic selections, and changing rates of lung cancer. *JAMA* 1987;**258**:331–8.
35. **Chan CK**, Wells CK, McFarlane MJ, et al. More lung cancer but better survival. Implications of secular trends in “necropsy surprise” rates. *Chest* 1989;**96**:291–6.
36. **Kendre G**, Szende B, Lapis K, et al. Misdiagnosis of lung cancer in a 2000 consecutive autopsy study in Budapest. *Gen Diagn Pathol* 1996;**141**:169–78.
37. **Manser RL**, Dodd M, Byrnes G, et al. Incidental lung cancers identified at coronal autopsy: implications for overdiagnosis of lung cancer by screening. *Respir Med* 2005;**99**:501–7.
38. **Dammas S**, Patz EF Jr, Goodman PC. Identification of small lung nodules at autopsy: implications for lung cancer screening and overdiagnosis bias. *Lung Cancer* 2001;**33**:11–16.
39. **Sone S**, Li F, Honda T, et al. Results of three-year mass screening programme for lung cancer using mobile low-dose spiral computed tomography scanner. *Br J Cancer* 2001;**84**:25–32.
40. **Yang P**, Allen MS, Aubrey MC, et al. Clinical features of 5628 primary lung cancer patients. *Chest* 2005;**128**:452–62.
41. **Swensen SJ**, Jett JR, Hartman TE, et al. CT screening for lung cancer: five-year prospective experience. *Radiology* 2005;**235**:259–65.
42. **Bach PB**, Kelly MJ, Ramsey CT, et al. Screening for lung cancer: a review of the current literature. *Chest* 2003;**123**(Suppl 1):72S–82.
43. **Henschke CI**, Naidich DP, Yankelevitz DF, et al. Early lung cancer action project: initial findings on repeat screenings. *Cancer* 2001;**92**:153–9.
44. **Skillrud DM**, Offord DP, Miller RD. Higher risk of lung cancer in chronic obstructive pulmonary disease. *Ann Intern Med* 1986;**105**:503–7.
45. **Tockman MS**, Athonisen NR, Wright EC, et al. Airways obstruction and the risk for lung cancer. *Ann Intern Med* 1987;**106**:512–18.
46. **Petty TL**. Screening strategies for early detection of lung cancer: the time is now. *JAMA* 2000;**284**:1977–80.
47. **Weiss W**. A practical model of tumor growth. *J Respir Dis* 1980;**1**:41–9.
48. **Aoki T**, Nakata H, Watanabe H, et al. Evolution of peripheral lung adenocarcinomas: CT findings correlated with histology and tumor doubling time. *AJR Am J Roentgenol* 2000;**174**:763–8.
49. **Hasegawa M**, Sone S, Takashima S, et al. Growth rate of small lung cancers detected on mass CT screening. *Br J Radiol* 2000;**73**:1252–9.
50. Final report of the individual life insurance valuation mortality. <http://www.soa.org/research/individual-life/final-report-life-insurance-valuation.aspx> (accessed 10 December 2007).
51. **Bach PB**, Jett JR, Pastorino U, et al. Computed tomographic screening and lung cancer outcomes. *JAMA* 2007;**297**:953–61.
52. **Anthonisen NR**. Smoking, lung function, and mortality. *Thorax* 2000;**55**:729–30.
53. **Sin DD**, Wu L, Paul Man SF. The relationship between reduced lung function and cardiovascular mortality. *Chest* 2005;**127**:1952–9.
54. **Shimizu K**, Yamada K, Saito H, et al. Surgically curable peripheral lung carcinoma: correlation of thin-section CT findings with histologic prognostic factors and survival. *Chest* 2005;**127**:871–8.
55. **Black WC**, Haggstrom DA, Welch HG. All-cause mortality in randomized trials of cancer screening. *J Natl Cancer Inst* 2002;**94**:168–73.
56. **Patz EF**, Swensen SJ, Herndon II JE. Estimate of lung cancer mortality from low-dose spiral computed tomography screening trials: implication for current mass screening recommendations. *J Clin Oncol* 2004;**22**:2202–6.
57. **Gohagan JK**, Marcus PM, Fagerstrom RM, et al. Final results of the Lung Screening Study, a randomized feasibility study of spiral CT versus chest x-ray screening for lung cancer. *Lung Cancer* 2005;**47**:9–15.
58. **Handy JR Jr**, Asaph JW, Skokan L, et al. What happens to patients undergoing lung cancer surgery? *Chest* 2002;**122**:21–3.
59. **Reich JM**. Improved survival and higher mortality: the conundrum of lung cancer screening. *Chest* 2002;**122**:329–37.

## APPENDICES

## Appendix A

Tumour diameter can be computed from the exponential equation:

$$y = 0.001(1.26)^x$$

whose derivation is supplied in a previous paper,<sup>59</sup> where y is tumour diameter and x is the number of tumour volume doublings (TVD). The log form

$$x = \ln 1000y \div \ln 1.26$$

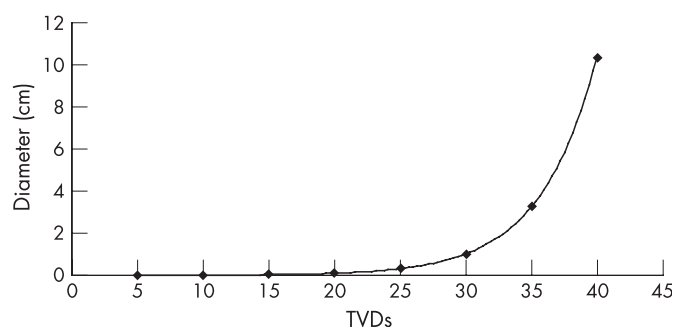
facilitates computation. The salient points agree with Geddes<sup>12</sup> comprehensive analysis (see table A1 and fig A1).

- (1) 27 TVDs are required to achieve a 0.5 cm diameter, the intervention threshold in US CT trials.
- (2) 30 TVDs are required to achieve a 1 cm diameter (0.5 g), the minimal threshold for CXR recognition, at which time the neoplasm has completed 75% of its natural history.
- (3) 35 TVDs are required to achieve a diameter of 3 cm, the upper limit of stage IA LC.
- (4) At 40 TVDs, tumour diameter is 10 cm (500 g), a lethal size.
- (5) An increase in diameter from 0.5 to 1 cm requires an eightfold increase in volume, which is accomplished by 3 TVDs.
- (6) At a mean TVDT of 1 year, 10 years are required to grow from a CXR threshold of 1 to 10 cm.

**Table A1** Tumour diameter based on the number of tumour volume doublings

No of TVDs	Tumour diameter (cm)
0	0.001
5	0.003
10	0.010
15	0.032
20	0.102
25	0.323
30	1.026
35	3.26
40	10.35

TVD, tumour volume doublings.



**Figure A1** Tumour diameter versus tumour volume doublings (TVDs).

## Appendix B

The LC incidence in the Mayo screened cohort was 5.5/1000 p-y. The proportion (46/206) of putatively overdiagnosed cases is 0.22. Overdiagnosis incidence is therefore  $0.22 \times 5.5/1000$  p-y = 1.2/1000 p-y or 16% of the incidence of clinically unsuspected LC discovered as a "surprise" (7.7/1000 p-y) in a concurrent US series of male autopsies.<sup>35</sup>

## Appendix C

### Glossary

Four biases introduced by screening:

- (1) Lead time bias refers to earlier detection of disease, in which intervention, as long as it is not harmful, will lengthen survival without necessarily improving longevity.
- (2) Length biased sampling refers to the screen identification of a more phenotypically favourable population of LCs due to the lengthier preclinical detectable phase of slower growing cancers. It is most in evidence in prevalence screens. Preclinical detectable phase refers to the asymptomatic period during which LC is detectable.
- (3) Overdiagnosis bias designates screen detection of cancers that would not otherwise have become symptomatic or clinically apparent before the individual died of other causes.
- (4) Biological (phenotypic) bias designates the histologically and prognostically favourable population of cases identified by screening versus clinically diagnosed lung cancer. It is most evident in CT screening, in which well differentiated adenocarcinomas and bronchioloalveolar carcinomas are disproportionately represented. These histological types are far slower growing and therefore prognostically more favourable than the population of clinically diagnosed

cancers, in which small cell, squamous cell and undifferentiated lung cancer are better represented.

- ▶ *Survival* designates the per cent of cases alive x years—conventionally, 5 years—after diagnosis. It may be given as a cause specific or an all-cause value.
- ▶ *Fatality* is the complement of survival ( $1 - \text{survival}$ ). Both survival and fatality may be reported as cause specific or all-cause values.
- ▶ *Mortality* is a measure of longevity. It is a rate, usually age adjusted, expressed as the number of deaths, over time, in a population or sample. In common with survival, mortality may be provided as a cause specific or an all-cause figure.  $\text{Non-LC mortality} = (\text{all cause mortality}) - (\text{LC-mortality})$ .
- ▶ *Prevalence* designates the number of cases at a given time in the sampled population.
- ▶ *Incidence* designates the number of new cases over time, usually annually, in the sampled population.
- ▶  $P = I \times D$
- ▶ Prevalence (P) can be computed as the product of incidence (I) and mean duration (D) of the condition. For example, the prevalence of stage IA LC would be the product of its incidence and the mean period of time it remains in stage IA.
- ▶ *Interval cancers* are those that typically present symptomatically between screenings. Their absence on earlier screenings may reflect either exceedingly rapid growth or their location within the mediastinum or proximal bronchi, sites in which radiographic imaging lacks sensitivity.
- ▶ *Stage IA classified as T1N0M0*, are peripheral tumours (T1)  $\leq 3$  cm, surrounded by lung or visceral pleura, with bronchoscopic evidence of spread no more proximal than the lobar bronchus. NO signifies no hilar or mediastinal metastases; MO indicates no discernable extranodal metastases. *Stage IB (T2N0M0)* is distinguished from IA by any of the following features: (1) T2 diameter  $> 3$  cm; (2) involvement of the main bronchus  $\geq 2$  cm from the main carina; (3) invasion of the visceral pleura; (4) obstructive pneumonitis or atelectasis not involving the entire lung. The prefixes "c" and "p" refer to clinical versus surgical–pathological staging.
- ▶ *Stage migration* ("Will Rogers effect") refers to the stage advance imposed by diagnostic methods that are more sensitive (eg, surgical–pathological versus radiographic staging). Stage migration improves the prognosis for a given stage.
- ▶ *Stage shift* refers either to a change in proportion of lower versus higher stages of cancer or to a reduction in the absolute number of advanced stage cases. Both a diminution in the proportion and a reduction in the absolute number of advanced stage cancers are necessary, although insufficient, indicators of screening efficacy. Lung cancer screening can be efficacious if, and only if, it achieves a reduction in the absolute number of advanced LC.
- ▶ *Hazard* is the rate of death within 1 year for persons alive at the beginning of the year. Non-cancer hazard is the rate of death from some other cause. Non-cancer relative hazard is the ratio: non-cancer hazard of a population with cancer  $\div$  non-cancer hazard in a matched population.