Management of solitary pulmonary nodules: how do thoracic computed tomography and guided fine needle biopsy influence clinical decisions?

D R Baldwin, T Eaton, J Kolbe, T Christmas, D Milne, J Mercer, E Steele, J Garrett, M L Wilsher, A U Wells

Background: Computed tomography (CT) and fine needle guided biopsy (FNB) are often used in the assessment of patients with lung nodules. The influence of these techniques on clinical decision making has not been quantified, especially for small solitary pulmonary nodules (SPN) where the probability of malignancy is lower. A study was undertaken to determine the effect of CT and FNB derived information on clinical decision making in patients with a solitary pulmonary nodule <3 cm in diameter on initial chest radiography.

Methods: Clinical, physiological, and outcome data on 114 patients with an SPN <3 cm who had subsequent thoracic CT and FNB were extracted from the records of a specialist cardiorespiratory hospital in Auckland, New Zealand. Chest radiographs and CT scans were reported according to specified criteria by a thoracic radiologist. Computer generated summary sheets were used to present cases to each of six clinicians. Each case was presented three times: (1) with clinical data and chest radiograph only; (2) with the addition of the CT report; and (3) with all data including the result of the FNB. Clinicians were asked to specify their management on each occasion and to estimate the probability of the lesion being malignant. Reproducibility was assessed by re-evaluating 24 cases 1 month later.

Results: 33 (29%) nodules were benign, 35 (31%) nodules (malignant) were resected with negative node sampling, and 46 (40%) had a non-curative outcome (radiotherapy, incomplete resection, refused therapy). Intra-clinician decision making was consistent for all three levels of clinical data (median κ values 0.79–0.89). Agreement between clinicians on the need for surgery was lowest with chest radiography alone (κ=0.33), rose with CT information (κ=0.44), and increased further with the addition of the FNB data (κ=0.57). The proportion of successful decisions on surgical intervention (against the known outcome) increased with the addition of CT reports and further with FNB reports (p=0.006, Friedman’s test). The major benefit of the information added by CT and FNB reports was a reduction in unnecessary surgery, especially when the clinical perception of pre-test probability of malignancy was intermediate (31–70%). FNB data contributed most to the benefit (p<0.001). The addition of CT and FNB was cost efficient and can be applied specifically to patients with a low or intermediate probability of malignancy.

Conclusion: Both CT and FNB make cost effective contributions to the clinical management of SPN <3 cm in diameter by reducing unnecessary operations and increasing agreement between physicians on the need for surgery.
refused surgery (n=3). Thus, 114 patients were evaluable (63 men, median age 66, age range 37–83) and form the basis of this report. All patients underwent chest radiography and thoracic CT scanning prior to FNB. Clinical and investigative data were extracted from case records. Chest radiographs and CT scans were reported by a thoracic radiologist (DM), with the identification of features influencing the likelihood of malignancy and operability. The final clinical outcome was recorded after a minimum of 5 years of follow up.

Evaluation by clinicians
Data were entered onto a computer database from which summary sheets were generated for each patient, containing three levels of information: clinical data and chest radiograph report only (CO); the same data with the addition of the thoracic CT report (CT); and all data with the addition of the FNB report (FB). Six specialist respiratory physicians independently reviewed all CO forms. They were then given all CT forms after random reordering of the sequence of patients. Finally, all FB forms were reviewed, again after random reordering. For each form, clinicians estimated the probability of the SPN being malignant to the nearest 5%, and gave one of five management decisions: surgery, mediastinoscopy, non-surgical management of inoperable malignancy, observation, and nonsurgical management of a benign lesion. Reproducibility was evaluated by repeating the process for all six clinicians on 24 randomly selected patients 1 month later.

Analysis of data
Inter-observer and intra-observer agreement was measured using the kappa coefficient of agreement. Agreement between clinicians on the need for surgery (mediastinoscopy or resection), as opposed to non-surgical management, was evaluated. The probability of malignancy assigned by each clinician was categorised as <5%, 6–15%, 16–30%, 31–70%, 71–94%, and >95% (corresponding to clinically relevant categories of “very unlikely”, “unlikely”, “intermediate”, “probable,” and “virtually certain”). Weighted kappa values for all clinician pairs (n=15) were derived; differences in observer agreement between CO, CT, and FB observations were tested using Wilcoxon's rank sum test.

Decisions made by individual clinicians based on clinical summary sheets were evaluated against the actual management and outcome as follows:

A “successful decision” was defined as a match between:
(a) surgical intervention decision (mediastinoscopy or immediate surgery) and a surgical cure;
(b) conservative management (decision that surgery inappropriate) and either:
   • a benign outcome;
   • surgery with no cure;
   • palliative treatment of a malignant lesion.

An “unsuccessful decision” was defined as a match between:
(a) surgical intervention decision and either:
   • a benign outcome;
   • surgery with no cure;
   • palliative treatment of a malignant lesion;
(b) conservative management decision and a surgical cure.

Friedman’s test was employed to test the significance of trends across the three groups (CO, CT, and FB) for the ratio of successful to unsuccessful decisions. The McNemar’s test was used to analyse the separate effects of CT and FB on the ratio of successful to unsuccessful decisions. Further analyses were undertaken of the influence of CT scanning and FNB on the frequency of missed surgical cures and unnecessary operations.

The effect of CT scanning and FB on the successful to unsuccessful decision ratio was re-evaluated according to the perceived probability of malignancy before CT and FB. The mean probability of malignancy was calculated for all six clinicians, based on estimates to the nearest 5%, made from the CO forms. The probability of malignancy was subcategorised as low (<30%), intermediate (31–70%), and high (>70%).

Cost analysis
The cost per surgical cure was calculated according to five distinct clinical strategies aimed at curing malignant SPN:
• decisions based on clinical and chest radiographic data alone;
• decisions based on data with the addition of CT scans;
• decisions based on data with the addition of FNB;
• decisions based on clinical data alone for lesions with a high probability of malignancy based on chest radiography and clinical information with decisions based on FNB for low probability lesions;
• decisions based on CT data for lesions with a high probability of malignancy based on CT, chest radiography, and clinical information with decisions based on FNB for low probability lesions.

For these analyses the basic cost of a lobectomy was taken as the Health Related Group cost of £8363, CT UK National Health Service cost of £90, and FNB cost of £88 assuming a 5% admission rate for pneumothorax lasting 3 days.

RESULTS
Clinical features
In 63 patients (55%) there were no symptoms ascribable to the SPN; in the remaining 51 cases new or worsening symptoms at presentation consisted of cough (n=5), haemoptysis (n=12), exertional dyspnoea (n=9), and chest pain (n=19). Clinical features are shown in table 1. Based on levels of forced expiratory volume in 1 second (FEV1), few patients were clearly inoperable. Histological findings at needle biopsy or surgery and outcome are shown in table 2. Sixty eight patients (60%) had a potentially good outcome (successful resections, n=35; benign nodules, n=33).
Sensitivity and specificity of FNB
No useful sample was obtained in 16 patients (nine where the outcome was benign and seven where malignant). Where these samples were excluded there were three false negative samples and one false positive sample (sensitivity 96%, specificity 96%).

Variation in estimations of the probability of malignancy and management decisions
As shown in table 3, the reproducibility of clinical decisions and the estimation of the probability of malignancy (intra-observer agreement) were good for most clinicians, but did not increase with the addition of CT information or FNB data. The agreement between clinicians (inter-observer agreement) on the need for surgical intervention (resection surgery or mediastinoscopy) was lowest with chest radiography alone (κ=0.33), rose with the availability of CT information (κ=0.44), and increased further with the addition of FNB data (κ=0.57). Similarly, median weighted kappa values for the estimated probability of malignancy (for each possible paired combination of clinicians, 15 pairs in all) rose strikingly with increasing information (table 3).

Influence of CT and FNB on clinical decisions
The median ratio of “successful decisions” (management decision agrees with outcome) to “unsuccessful decisions” was lowest with clinical information alone (1.14, range 0.75–1.85), increased with the addition of CT information (1.51, range 1.15–2.45), and increased further with the addition of FNB data (2.09, range 1.92–3.56) (p=0.006, Friedman’s test). Table 4 shows the number of missed cures, unnecessary operations, and appropriate decisions to avoid surgery. The most important effect of additional information was seen when lesions were benign: with the addition of CT and FNB data clinicians were less likely to recommend surgery.

Effect of added information in relation to probability of malignancy
Table 5 shows the number of “successful decisions” in patients with low, intermediate, and high probabilities of malignancy (derived from the mean of estimates of the probability of malignancy, based on CO data). The greatest rise in “successful decisions” was seen in patients with an intermediate probability of malignancy. Decisions in the other probability groups become more favourable but not to the same degree, and the highest number of “unsuccessful decisions” was in patients with a high probability of malignancy. This was simply because there were more patients in that category. Unsuccessful decisions included surgery recommended for benign lesions, but there were also more patients who had unsuccessful surgical approaches for malignancy. On McNemar’s χ2 analysis, the greatest increase in “successful decisions” occurred with the addition of FNB data in patients with an intermediate probability of malignancy (p<0.001) and with the addition of CT data in patients with a low probability of malignancy (p=0.02).

Cost analysis
Table 6 shows that there was little difference in the cost per surgical cure for each of the five strategies. There were, however, important differences in the numbers of operations for benign lesions with all of the strategies employing FNB.

**DISCUSSION**
We have evaluated the effect of CT and FNB derived data on the appropriateness of management of a large cohort of patients with solitary pulmonary nodules using a novel scenario based approach. The greatest increase in numbers of successful management decisions occurred when the probability of malignancy was intermediate (although increases were seen for low and high probabilities), and was mainly attributable to the avoidance of surgery in benign SPN. Furthermore, the addition of CT and FNB data was cost effective and resulted in major increases in clinician agreement on the likelihood of malignancy and on management.

The management of the SPN is largely based on the perceived probability of malignancy which, in turn, is heavily influenced by patient age, smoking history (and other hazard factors) etc. The accurate application of Bayesian analysis requires knowledge of the overall prevalence of malignancy in the population under study. Thus, in routine practice, most clinicians estimate the likelihood of malignancy semi-quantitatively (for example, highly unlikely, intermediate probability, etc.). Table 1 shows agreement between clinicians on management was, at best, only moderately, increasing agreement on the probability of malignancy with additional data was mirrored by a striking reduction in variation in management decisions, although agreement between clinicians on management was, at best, only moderate to good.

From these linked observations two conclusions may be drawn. Firstly, a significant component of variation in management is likely to result from discrepancies between clinicians in estimating the probability of malignancy which can be minimised by the use of FNB, especially when the probability of malignancy is intermediate. Management decisions are more likely to be confident when the estimated probability of malignancy is very low or very high. Secondly, other factors also have an important influence on management decisions, given the significant variation in management in the present study, even with the availability of FNB data. The study has highlighted marked differences between decisions made by experienced respiratory specialists working in the same secondary/tertiary referral centre who meet together on a weekly basis to discuss individual cases of lung cancer and to agree on preferred management. Increasing agreement between clinicians with the availability of FNB will not only improve the appropriateness of clinical management, but may also reduce the risk of litigation.
The novelty of the use of this methodology in the present study lies in the definition of the added value of diagnostic tests, allowing the benefits of changes in physician perception (more cues, less unsuccessful surgery) to be set against the overall cost of management. In general, diagnostic tests are evaluated against a “gold standard” definition of a diagnosis. Pretest diagnostic probabilities, often based on a great deal of information including clinical assessment and ancillary investigations, are not taken into account. A test may be diagnostic without altering management and outcome, especially when a diagnosis is already highly probable. In the present study the restriction of FNB to patients with a low or intermediate probability of malignancy was not associated with a major increase in adverse outcome. The extra information gleaned from FNB in those with a high probability of malignancy had minimal practical impact. The true clinical utility of a diagnostic procedure depends on the information added by the test, and the resulting change in physician perception and management. Without this knowledge an evidence base for the utility and cost efficiency of diagnostic procedures is unattainable.

In most published series, patients with symptoms ascribable to a SPN and those with previous malignancy are not studied. However, the validity of excluding these subgroups can be questioned. The clinical reality of how best to manage SPN is not confined to asymptomatic patients or to those with no history of malignant disease. Thus, the present cohort is more representative of the range of patients encountered in routine clinical practice than some populations described in the medical literature. It is highly unlikely that the inclusion of these subgroups had a major influence on the main findings of the study. Both the presence of new symptoms and a history of malignancy influence the likelihood of malignancy and are thus taken into account in the management decisions reported in the study.

The “gold standard” against which study management decisions were designated as “successful” or “unsuccessful” was the actual outcome. For patients with benign SPNs, those with malignant SPNs undergoing curative resection, and those with malignant SPNs in whom disease was subsequently found to be incurable despite surgery, the definition of a “successful” management decision was straightforward. The “gold standard” was questionable only in the 35 patients who did not undergo resection of a malignant SPN. It can be argued that in some of these 35 cases optimal management was uncertain, even in retrospect, and that a surgical cure might have been missed. However, in all but five cases there was evidence of either inoperability (due to co-morbid conditions or poor lung function) or incurable disease, with the information gained from CT and FNB. Thus, the “gold standard” used for the purposes of this study was potentially inaccurate in less than 5% of the cohort.

Appropriate management is the goal of interpretation of clinical data. An unfavourable outcome can result from unnecessary surgery (surgery in benign disease or incurable malignant disease) or a missed cure (failure to resect curable malignant disease). Assigning relative importance to a missed cure and to unnecessary surgery is not entirely straightforward. Although a missed cure might appear to have more serious implications than unnecessary surgery at first sight, this is not necessarily the case if meticulous follow up is instituted and surgery is subsequently undertaken. Cummings et al argued that, in patients with a low probability of malignancy, the outcome differed little between immediate surgery and observation with a view to surgery (if a malignant growth pattern was observed at follow up). In support of this contention is the observation that there is no proven difference in survival between tumours of 1 cm diameter and those of 3 cm, a change in size which reflects a relatively small

### Table 4  Unsuccessful management decisions related to outcome for all clinicians

<table>
<thead>
<tr>
<th>Level of information</th>
<th>Missed cures, inoperable disease (n=35)</th>
<th>Surgery recommended, benign disease (n=46)</th>
<th>Surgery for benign lesion (n=32)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CO</td>
<td>5 (14%)</td>
<td>32 (70%)</td>
<td>18.5 (56%)</td>
</tr>
<tr>
<td>CT</td>
<td>3.5 (10%)</td>
<td>28.5 (62%)</td>
<td>13 (39%)</td>
</tr>
<tr>
<td>FN B</td>
<td>2.5 (7%)</td>
<td>28.5 (62%)</td>
<td>5 (15%)</td>
</tr>
<tr>
<td>Friedman p value</td>
<td>0.172</td>
<td>0.294</td>
<td>0.01</td>
</tr>
</tbody>
</table>

CO = clinical only; CT = CT report in addition; FN B = FNB report in addition. *Total possible cases where surgery could be recommended for incurable disease (1 surgery with nodes positive plus 35 malignant, palliative approach).
portion of the life of the tumour.22–26 When the probability of malignancy is not high, the avoidance of unnecessary surgery in benign disease and delayed surgery in malignant disease are equally important considerations. Thus, the information provided by FNB (resulting in a reduction in unnecessary surgery) is most valuable in patients with a low or intermediate probability of malignancy.

Reviews of management of SPN suggest that lesions of more than 3 cm are almost always malignant and recommend resection.22 This policy is thought to result in resection of relatively few benign lesions and those small cell carcinomas have a favourable outcome following complete resection.27 In the present study we have looked at clinician estimates of probability of malignancy and confirmed that CT and FNB are most useful (in absolute numbers) in the intermediate probability group. This is for lesions where FNB would be recommended. The small beneficial effect even in the high probability group would support the general view that FNB should be performed in lesions < 3 cm in diameter. This may, however, be flawed where there is a greater proportion of malignant nodules in the population, as is the case in the UK and New Zealand. However, we have shown that even with smaller nodules a strategy employing FNB in the low and intermediate probability group has little influence on cost or success of management. This strategy has the advantage of avoiding delay in those patients who have immediate surgery.

The cost analysis in this study has been deliberately limited to looking at the “attempt to cure” scenario. The value of FNB may have been underestimated because each cure is associated with less costly subsequent medical care. The major financial and clinical benefit of avoidance of unnecessary surgery was accounted for in our analysis.

In the present study we evaluated the routine investigation of pulmonary nodules, as performed at most centres. The use of formulae to quantify the probability of malignancy which might improve clinician agreement28–30 is not widespread and was not evaluated. The availability of specialist bronchoscopy techniques may influence the clinical approach and newer techniques such as ultrathin bronchoscopy31 require evaluation, but these methods establish a tissue diagnosis and therefore should have a similar impact to FNB on clinical decision making. Other approaches such as contrast enhancement measurements32 and positron emission tomography33–35 are increasingly applied at major academic institutions but are not generally available at smaller centres. These techniques may have a considerable additional influence on the level of agreement between clinicians and the accuracy of management. Less invasive surgical approaches may too have a role. The use of video assisted thoracoscopic biopsy may have less morbidity, especially for the more peripheral nodules, and provide 100% diagnostic accuracy.

Both CT and FNB make important contributions to the clinical management of SPN < 3 cm in diameter by reducing unnecessary operations and increasing agreement between physicians on the probability of malignancy and need for surgery. The addition of FNB is not associated with an overall increase in cost and may be applied to clinical strategies, reserving the investigation for nodules with a low or intermediate probability of malignancy.

Table 6 Cost analysis (median values)

<table>
<thead>
<tr>
<th>Strategy</th>
<th>Median cost per surgical cure (£)</th>
<th>Unsuccessful surgery</th>
<th>Surgery for benign lesions</th>
<th>Missed cure</th>
</tr>
</thead>
<tbody>
<tr>
<td>CXR alone all cases</td>
<td>22458</td>
<td>32</td>
<td>18.5</td>
<td>5</td>
</tr>
<tr>
<td>Addition of CT all cases</td>
<td>20431</td>
<td>28.5</td>
<td>13</td>
<td>3.5</td>
</tr>
<tr>
<td>Addition of FNA all cases</td>
<td>19582</td>
<td>28.5</td>
<td>5</td>
<td>2.5</td>
</tr>
<tr>
<td>Addition of FNA if lesion not 19258</td>
<td>30.5</td>
<td>5</td>
<td>1.5</td>
<td></td>
</tr>
<tr>
<td>Addition of FNA if lesion not 18982</td>
<td>28</td>
<td>7</td>
<td>3.5</td>
<td></td>
</tr>
</tbody>
</table>

CXR=chest radiograph; FNA=Fine needle aspiration.

Authors’ affiliations
D R Baldwin, Department of Respiratory Medicine, City Hospital, Nottingham NG5 1PB, UK
T Eaton, J Kolbe, T Christmas, D Milne, J Mercier, E Steele, J Garrett, M L Wilsher, A U Wells, Respiratory Services, Green Lane Hospital, Auckland, New Zealand

Funding source: none.

Conflict of interest: none.

REFERENCES