ABSTRACT

Introduction: Patients with NSCLC may be treated with curative intent, yet they remain at high risk of both disease recurrence and second primary lung cancer (SPLC) and increased risk of early death. Guidelines provide recommendations for follow-up, but there is little consensus, and review of available evidence is necessary. The use of a systematic follow-up strategy for the detection of disease recurrence or SPLC after curative-intent treatment of NSCLC may increase the proportion of patients available for retreatment and increase the survival of patients with surveillance detection.

Methods: We performed a systematic review and meta-analysis of prospective studies on follow-up of NSCLC after curative-intent treatment to answer the following three questions: What is the effect of follow-up on detection of recurrence or SPLC? What is the effect of surveillance detection on curative-intent retreatment? What is the survival impact?

Results: Recurrence or SPLC was observed in 17.8% to 71% of patients.Scheduled imaging-detected recurrence in 60% to 100% of cases, yet neither computed tomography–based (OR = 2.31, 95% confidence interval [CI]: 0.27–19.49, p = 0.44) nor positron emission tomography–computed tomography–based follow-up (OR = 1.431, 95% CI: 0.92–2.22, p = 0.12) was statistically superior to standard follow-up strategies. Detection of disease recurrence/SPLC significantly increased the odds of curative-intent retreatment (OR = 4.31; 95% CI: 2.10–8.84, p < 0.0001). Curative-intent retreatment prolonged survival in reported studies.

Conclusions: The early detection of disease recurrence/SPLC may increase the likelihood of curative-intent retreatment and prolong survival. There is a clear need for prospective randomized controlled studies of follow-up to confirm effectiveness of available follow-up modalities.

Keywords: Non–small cell lung cancer; Surveillance; Follow-up; Retreatment; Survival

Introduction

Lung cancer is the most common cancer globally and is the leading cause of cancer mortality. Early detection and treatment provide the greatest likelihood of long-term survival. Despite curative-intent treatment, 5-year survival remains poor after surgery, ranging between...
30% and 75% (dependent on stage at resection)\(^2\) and 50% after stereotactic body radiotherapy (SBRT).\(^3\) These patients remain at high risk of disease recurrence and new primary disease after initial treatment.\(^4,5,6,7\) Follow-up after curative-intent treatment is therefore necessary to evaluate treatment response, to manage treatment-related complications, and for the detection of disease recurrence or second pulmonary lung cancers (SPLCs).\(^7\)

Accurate and early confirmation of disease recurrence may increase likelihood of change in medical treatment and guide and increase retreatment.\(^8,9,10\) The retreatment of local and locoregional recurrence of NSCLC has been studied and suggests acceptable patient tolerability and survival advantage after surgery or radiotherapy.\(^11–15\) Similarly, SPLC may be treated effectively and with high patient tolerability using both radiotherapy and surgical resection.\(^16–18\)

Clinical practice guidelines for posttreatment surveillance imaging of NSCLC have been formulated by a number of bodies despite a paucity of robust supportive evidence.\(^19–22\) Although there is imperfect consensus between guidelines, most statements support the use of computed tomography (CT) on a 6-monthly basis for 2 years after resection followed by annual scanning thereafter. Recent research has focused on subjective evaluations using patient-reported outcome measures (PROMs),\(^23,24\) etiopathologic biomarkers such as micro RNA and cell-free DNA,\(^25–27\) and functional imaging using positron emission tomography (PET)-CT,\(^28,29\) although the overall effectiveness of these measures remains to be confirmed.

We performed a systematic review and meta-analysis of existing literature to explore the effects of follow-up using surveillance modalities including clinical assessment, PROMs, biomarkers, and radiologic and nuclear medicine imaging. We sought to answer three questions. One, what is the effectiveness of surveillance in detection of disease recurrence and SPLC? Two, what is the impact of follow-up on retreatment? Three, what effect does retreatment of recurrence/SPLC have on survival?

### Materials and Methods

#### Protocol and Registration

The study protocol has been registered in the PROSPERO international prospective register of systematic reviews and is accessible at [http://www.crd.york.ac.uk/prospero/](http://www.crd.york.ac.uk/prospero/) (registration number: CRD42020201898).

#### Eligibility Criteria

For the purposes of this review, curative-intent treatment was defined as surgical resection with or without adjuvant therapy, stereotactic or radical radiotherapy, or chemoradiotherapy for patients with stage I to IIIA NSCLC. Follow-up was deemed to be any recurring health provider-patient interaction after treatment, intended to evaluate disease outcome and progression, and persisting at least 12 months after treatment. The control was deemed the respective standard of care. Primary outcomes were specified to be one or more of the following: (1) confirmed detection of new lung cancer (either locoregional or distal/metastatic recurrence or new metachronous lung cancer) using the follow-up method, including measures of diagnostic accuracy; (2) curative-intent retreatment of that newly diagnosed cancer as a direct result of detection on follow-up; and (3) survival after retreatment.

Studies with a one-off intervention (e.g., intervention for response assessment) and those on other lung cancers, such as small cell, carcinoid, thymoma, or mesothelioma, were excluded. Study design was limited to those with prospectively defined follow-up and further limited to English publications.

#### Search Strategy

A systematic search of MEDLINE (by means of PubMed), EMBASE, and the Cinahl was performed by AS and CC, with the final search performed on May 8, 2020 (EMBASE search strategy, Supplementary Table 1). This study was restricted to studies engaging prospectively collected data to provide protection from selection biases present in retrospective data and attrition owing to loss of follow-up and to provide higher tier evidence.

#### Data Collection Process

A total of 758 articles were retrieved from the search of databases in addition to hand-searching of reference lists in relevant articles. After removal of 123 duplicates, 632 titles and abstracts were independently evaluated. Of the remaining articles, 572 were deemed irrelevant, and full-text assessment was performed on the remaining 65 articles using the predefined inclusion and exclusion criteria. Subsequently, 52 articles were excluded owing to inappropriate study design, lack of comparator group, non–follow-up intervention, and inadequate length of follow-up (Fig. 1). Data were extracted independently by each assessor (CC and RS) using a standardized data collection form on Covidence (https://www.covidence.org/). Details regarding study identification, methods, patient population, interventions and outcomes were collected, and consensus between assessors was achieved by comparing the two forms and discussing any discrepancies.
Data Synthesis and Statistical Analysis

All analyses were conducted using the Review Manager (RevMan version 5.3 Computer program, Version 5.3. Copenhagen: The Nordic Cochrane Centre, The Cochrane Collaboration, 2014). We calculated the pooled OR with 95% confidence interval (CI) using a random-effects model owing to anticipated clinical trial heterogeneity. To address statistical heterogeneity, we performed chi-square to identify variation in effect estimates beyond chance by the finding of a chi-square p value less than 0.05 and the additional calculation of $I^2$ to describe the percentage of the variability in effect estimates owing to heterogeneity rather than sampling error consistent with Cochrane methods. We adopted standard thresholds for interpretation that an $I^2$ of 0% to 40% might not be important, 30% to 60% may represent moderate heterogeneity, 50% to 90% may represent substantial heterogeneity, and 75% to 100% may represent considerable heterogeneity.

Risk of Bias Assessment

Risk of bias was evaluated using the ROBINS-1 tool for nonrandomized studies of interventions\textsuperscript{30} by two reviewers (RS and MF), and discrepancies were resolved with consensus review (Fig. 2). This tool evaluates risk in relevant domains including confounding, selection, deviation from intended interventions, missing data, measurement of outcomes, and selection of the reported result. Risk of bias assessment results are displayed using the robvis visualization tool.\textsuperscript{31} The risk of bias was adjudged as low for two studies, moderate for seven studies, and serious for four studies with the main causes of risk of bias being deviation from intended intervention, incomplete reporting, missing data, and measurement and reporting biases.

Results

A total of 13 studies (5759 patients) were identified, in which a prospective surveillance strategy was identifiable for surveillance of NSCLC after curative-intent treatment. Surveillance modalities included history and physical examination (HPE), biochemistry, tumor markers, sputum cytology, chest radiograph (CXR), CT, PET-CT, magnetic resonance imaging (MRI), bone scintigraphy, and bronchoscopy (Table 1). In one study, there were no scheduled radiological follow-up
reported. There were no suitable surveillance studies available for analysis reliant on PROMs or biomarkers. Intended duration of follow-up was 2 years, 5 years, ongoing, or undefined.

Primary curative-intent treatment in the 13 studies included surgical resection (lobectomy, bilobectomy, or pneumonectomy), neoadjuvant chemotherapy followed by resection, resection with chemotherapy or radiotherapy, SBRT, and chemoradiotherapy. Confirmation of recurrences was variably reported and documented after biopsy and histologic confirmation (16 biopsies/121 recurrences), radiologic or PET suspicion, or radiologic or PET size progression before treatment or regression after retreatment.

<table>
<thead>
<tr>
<th>Study</th>
<th>D1</th>
<th>D2</th>
<th>D3</th>
<th>D4</th>
<th>D5</th>
<th>D6</th>
<th>D7</th>
<th>Overall</th>
</tr>
</thead>
<tbody>
<tr>
<td>Westeel 2000</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Chiu 2003</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Benamore 2007</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Subotic 2009</td>
<td>+</td>
<td>+</td>
<td>X</td>
<td>-</td>
<td>X</td>
<td>-</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Onishi 2011</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>X</td>
<td>-</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Choi 2011</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>Toba 2012</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>-</td>
<td>X</td>
</tr>
<tr>
<td>Hanna 2014</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>Lou 2014</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>Spratt 2016</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Gambazzi 2019</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>Conforti 2020</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>Toba 2020</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>X</td>
<td>X</td>
</tr>
</tbody>
</table>

Figure 2. Risk of bias assessment (green, low risk; yellow, moderate risk; red, serious risk).

Domains:
- D1: Bias due to confounding.
- D2: Bias due to selection of participants.
- D3: Bias in classification of interventions.
- D4: Bias due to deviations from Intended interventions.
- D5: Bias due to missing data.
- D6: Bias in measurement of outcomes.
- D7: Bias in selection of the reported result.

Judgement:
- Red: Serious
- Yellow: Moderate
- Green: Low
<table>
<thead>
<tr>
<th>Author Year</th>
<th>Country</th>
<th>Study Design</th>
<th>Mean Follow-Up (mo)</th>
<th>n</th>
<th>Age Mean</th>
<th>Sex (M/F)</th>
<th>Stage</th>
<th>Histology Ad/Sq/Other</th>
<th>Initial Treatment</th>
<th>Disease Reported</th>
<th>Follow-Up Modalities</th>
</tr>
</thead>
<tbody>
<tr>
<td>Westeel 2000</td>
<td>France</td>
<td>PC</td>
<td>131</td>
<td>192</td>
<td>60</td>
<td>177/15</td>
<td>I-IIIA</td>
<td></td>
<td>Surgical resection</td>
<td>Recurrence</td>
<td>SPLC</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>HPE and CXR 3-monthly for first 3 y and 6-monthly thereafter.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>CT and fiberoptic bronchoscopy 6-monthly for first 3 y and annually thereafter.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>From 8 y onward: annual CXR.</td>
</tr>
<tr>
<td>Chiu 2003</td>
<td>Republic of China</td>
<td>PC (part 2 only)</td>
<td>15</td>
<td>38</td>
<td>69</td>
<td>37/6</td>
<td>I-IIIA</td>
<td></td>
<td>Surgical resection</td>
<td>Recurrence</td>
<td>HPE, CXR, and LDCT 3-monthly for first 2 y and 6-monthly until end of fifth year. HPE included sputum cytology and CEA levels.</td>
</tr>
<tr>
<td>Benamore 2007</td>
<td>United States</td>
<td>RC (trial pts only)</td>
<td>77 (median)</td>
<td>40</td>
<td>60.5</td>
<td>25/15</td>
<td>II-IIIB</td>
<td></td>
<td>Any</td>
<td>Recurrence</td>
<td>SPLC</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Pts from two trials included:</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>INT 0161—CXR and biomarkers 3-monthly for first 2 y and 6-monthly thereafter.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>CT chest, upper abdo + brain 6-monthly for first 3 y.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>INT 0139—CXR and biomarkers 2-monthly for first year, 3-monthly for 2 y, and 6-monthly thereafter.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>CT chest, upper abdo + brain at 12, 18, and 24 mo and annually thereafter.</td>
</tr>
<tr>
<td>Subotic 2009</td>
<td>Serbia</td>
<td>PC</td>
<td>88</td>
<td>55</td>
<td>76/12</td>
<td>I-IIIA</td>
<td>18/67/3</td>
<td></td>
<td>Surgical resection</td>
<td>Recurrence</td>
<td>“Intensified follow up”:: monthly phone contact.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>HPE 1-monthly for first 3 mo, then 3-monthly until end of first year, 4-monthly for second and third years, and 6-monthly thereafter.</td>
</tr>
</tbody>
</table>

(continued)
<table>
<thead>
<tr>
<th>Author Year</th>
<th>Country</th>
<th>Study Design</th>
<th>Mean Follow-Up (mo)</th>
<th>n</th>
<th>Age</th>
<th>Sex (M/F)</th>
<th>Stage</th>
<th>Histology</th>
<th>Initial Treatment</th>
<th>Disease Reported</th>
<th>Follow-Up Modalities</th>
</tr>
</thead>
<tbody>
<tr>
<td>Onishi 2011</td>
<td>Japan</td>
<td>PC</td>
<td>—</td>
<td>121</td>
<td>71</td>
<td>80/41</td>
<td>I-IIIA</td>
<td>Ad/ Sq/Other</td>
<td>Surgical resection</td>
<td>Recurrence</td>
<td>PET-CT 6-monthly. (Continued)</td>
</tr>
<tr>
<td>Choi 2011</td>
<td>Korea</td>
<td>PC</td>
<td>—</td>
<td>358</td>
<td>62.9</td>
<td>239/119</td>
<td>I-IIIA</td>
<td>199/121/38</td>
<td>Surgical resection</td>
<td>Recurrence</td>
<td>HPE and CEA 3-monthly for first 2 y and biannually thereafter. PET-CT annually.</td>
</tr>
<tr>
<td>Toba 2012</td>
<td>Japan</td>
<td>RC</td>
<td>40.1</td>
<td>101</td>
<td>67.4</td>
<td>59/42</td>
<td>I-IIIA</td>
<td>76/20/5</td>
<td>Surgical resection</td>
<td>Recurrence</td>
<td>Chest CT or PET-CT alternately performed 6-monthly for first 3 y and PET-CT annually for next 2 y PET-CT performed at least once yearly.</td>
</tr>
<tr>
<td>Lou 2014</td>
<td>United States</td>
<td>RC</td>
<td>35</td>
<td>1640</td>
<td>67</td>
<td>695/945</td>
<td>I-IIIA</td>
<td>1222/292/127</td>
<td>Surgical resection</td>
<td>Recurrence</td>
<td>CT and HPE every 6-12 mo for first 2 y and annually thereafter. MnDCT and CXR at 3, 6, 12, 18, 24, 36, 48, and 60 mo after surgery.</td>
</tr>
<tr>
<td>Hanna 2014</td>
<td>Canada</td>
<td>PC</td>
<td>36</td>
<td>271</td>
<td>66.6</td>
<td>126</td>
<td>Any resection</td>
<td>Surgical resection</td>
<td>Recurrence</td>
<td>MnDCT and CXR at 3, 6, 12, 18, 24, 36, 48, and 60 mo after surgery.</td>
<td></td>
</tr>
<tr>
<td>Spratt 2016</td>
<td>United States</td>
<td>RC</td>
<td>23</td>
<td>366</td>
<td>77</td>
<td>170/196</td>
<td>I-II</td>
<td>257/99/10</td>
<td>SBRT (inoperable)</td>
<td>Recurrence</td>
<td>CT 3-monthly for first 2 y, 6-monthly for next 2 y, and annually thereafter.</td>
</tr>
<tr>
<td>Gambazzi 2019</td>
<td>Switzerland</td>
<td>RPG pilot</td>
<td>PET-CT 50 CE-CT 46</td>
<td>67</td>
<td>33/17</td>
<td>34/12</td>
<td>I-III</td>
<td>I-III</td>
<td>Any</td>
<td>Recurrence</td>
<td>2 arms: 1. PET-CT: (n = 50) 2. CE-CT: (n = 46) HPE + assigned imaging procedure 6-monthly for 2 y. (Continued)</td>
</tr>
</tbody>
</table>
Disease Recurrence and SPLC

Recurrence rates varied between 17.8% and 71% with median follow-up reported in just six studies ranging between 15 and 77 months. Rates of recurrence were lower, 11.1% to 22%, in earlier stage (stage I–II) disease,\(^{32,40}\) and higher, 52% to 72% in those with higher stage disease (stage IIIA).\(^{32,40,41}\)

Scheduled imaging-detected recurrence/SPLC in 60% to 100% of cases and symptomatic presentation lead to detection by unscheduled imaging in 0% to 40% of cases.\(^{35,39,42}\) Conforti et al.\(^{35}\) reported on 2261 patients observing a significantly higher likelihood of detection by scheduled follow-up compared with unscheduled detection for locoregional recurrences (88.4%, 95% CI: 84%–91%) and SPLC (93.2%; 95% CI: 84%–99%), but not for distant metastases (68.7%, 95% CI: 65%–73%, \(p < 0.0001\)). Lou et al.\(^{40}\) observed a higher detection rate of asymptomatic recurrences among patients with early stage by surveillance CT scans with 32% of detected recurrences in the early stage cohort (stage I–II) identified as a result of symptoms during unscheduled follow-up compared with 61% among patients in stage IIIA (\(p = 0.04\)).

Nine studies were available in which the proportions of disease recurrence or SPLC were detected by scheduled versus unscheduled radiological follow-up.\(^{13-36,39-43}\) The pooled estimate for odds of detection by scheduled follow-up provided an OR equal to 8.15 (95% CI: 3.5–18.9, \(p < 0.0001\)), although the estimate was associated with substantial trial heterogeneity, chi-square \(p\) value equal to 0.0001, and \(I^2\) equal to 94% (Fig. 3A).

Three studies were available in which it was possible to compare detection of disease recurrence or SPLC by CT versus standard follow-up within a scheduled follow-up strategy, and two further studies reported CT detection but comparator detection was unreported.\(^{45}\) Standard follow-up included 6 monthly biochemistry, HPE, and CXR in all studies, additional bronchoscopy in one study,\(^{42}\) and CT or MRI brain in one study.\(^{43}\) The pooled estimate provided an OR equal to 2.31 (95% CI: 0.27–19.49, \(p = 0.44\)). There was substantial trial heterogeneity, chi-square \(p\) value equal to 0.0003 and \(I^2\) equal to 88% (Fig. 3B).

The use of PET-CT in a scheduled follow-up strategy was not statistically different to standard follow-up including CT for the detection of recurrence or SPLC (Fig. 3C). Three studies were available in which it was possible to compare detection by PET-CT versus standard follow-up within a scheduled follow-up strategy; all studies included six monthly CT in the standard follow-up comparison and one study\(^{39}\) included bone scintigraphy and brain MRI.\(^{39}\) The pooled estimate provided an OR equal to 1.431 (95%
CI: 0.92–2.22, \( p = 0.12 \). There was no heterogeneity, chi-square \( p \) value equal to 0.48 and \( I^2 \) equal to 0\% (Fig. 3C).

Detection of SPLC was separately reported in four studies detected in 1.5\% to 7.5\% of patients\(^{35,38,42,43}\). Spratt et al.\(^{38}\) reported SPLC detection dynamics observing a median time to detection of 16.5 months (6.5–71.1 mo), with 32\% of these patients developing SPLCs after 2 years and a cumulative incidence at 18 months and 2 years of 3.3\% and 4.5\%, respectively, which continued to rise up to 6 years from the end of SBRT. The 2-year cumulative incidence rate for SPLC was 0\% for nonsmokers, 3.5\% for former tobacco smokers, and 15.1\% for current tobacco smokers (\( p = 0.005 \)), and 76\% of patients who developed SPLC were candidates for curative treatment.

### Diagnostic Accuracy

Report of diagnostic accuracy was available from four studies. Sensitivity for detection of recurrent disease by CXR was 21.2\%\(^{36}\) by CT 93\% to 94.2\%\(^{35,36}\) and by PET-CT 80.8\% to 94.4\%\(^{33,37,39}\) (Fig. 4). Specificity for detection of recurrent disease by CXR was 91.7\%\(^{36}\) by CT 72.0\% to 86.0\%\(^{33,36}\) and by PET-CT 62.0\% to 97.6\%\(^{33,37,39}\) (Fig. 4). Reporting of histologic confirmation of recurrence in diagnostic accuracy studies was incomplete with one study reporting histologic confirmation in just 13.2\%\(^{39}\).

### Risk of Recurrence

Conforti et al.\(^{35}\) undertook multivariate analysis to identify variables associated with increased risk of

---

**Figure 3.** (A) Forest plot of scheduled versus unscheduled detection of recurrence or second primary lung cancer. (B) Forest plot detection of recurrence or second primary lung cancer by scheduled CT versus standard scheduled radiology follow-up. (C) Forest plot detection of recurrence by scheduled PET-CT versus standard scheduled radiology follow-up. CI, confidence interval; CT, computed tomography; PET, positron emission tomography.
recurrence finding increased risk associated with positive lymph node status (hazard ratio [HR] = 2.00; 95% CI: 1.54–2.61), limited/minimal mediastinal lymph node sampling as compared with systematic mediastinal lymph node dissection (HR = 1.43, 95% CI: 1.10–1.86), whereas adjuvant chemotherapy was associated with protective effect (HR = 0.57; 95% CI: 0.43–0.76). Factors associated with increased risk of distant metastatic recurrence included T3/4 pathologic stage (HR = 1.30, 95% CI: 1.01–1.68) and positive lymph node status (HR = 1.76; 95% CI: 1.43–2.18) whereas there was some protective effect in never smokers (HR = 0.64; 95% CI: 0.43–0.88) and nonsquamous tumors (HR = 0.40; 95% CI: 0.33–0.49). Hanna et al. similarly found increased recurrence risk associated with stage III disease (2.96, 1.16–7.58) and bilobar compared with unilobar resection (6.81, 1.11–41.89).

Recurrence Site and Time

Hanna et al. detected 67.3% recurrence on minimal-dose CT within the first year of surveillance and 26.5% within the second year, and no asymptomatic patients were diagnosed in the fourth and fifth years. Spratt et al. detected disease recurrence or SPLC in 108 of 366 patients with stage I after SBRT in a cohort of patients in whom 41% had previously been treated for NSCLC. Of these, 84% were diagnosed within the first 2 years of treatment, 7.4% during year 3, and 8.3% after year 3. In a study of 2261 patients, Conforti et al. found that the hazards of recurrence were highest within 18 months of treatment, peaking between 6 and 12 months. Lou et al. observed a sustained recurrence risk, highest in the first 2 years after operation but remaining elevated through postoperative year 4. The rates of recurrence during the first 2 years were 7 to 10 recurrences per 100 person-years; they dropped to 2 per 100 person-years after year 4. Compared with patients with early stage, those with stage IIIA had a significantly higher rate of recurrence during the first 2 years after resection. After year 4, the rates of recurrence were similar between the two cohorts. Recurrence site was locoregional in 26.1% to 52.7% of cases, metastatic in 29.0% to 78.3%, and both thoracic and extrathoracic in 4% to 19%. Recurrence site was more likely to be metastatic for stage IIIA disease than stage I to II disease (85% versus 74%, p = 0.01).40

Retreatment

Four studies described retreatment as a consequence of detection method within a scheduled surveillance program, and three were available for meta-analysis (Fig. 5). The use of a scheduled follow-up...
strategy for disease detection after initial treatment was associated with a significantly increased odds of curative-intent retreatment when compared with disease recurrence detected by unscheduled investigation (OR = 4.31, 2.10–8.84, \( p < 0.0001 \)) with moderate level of statistical heterogeneity between studies (chi-square \( p = 0.04; I^2 = 69\% \)) (Fig. 5).

Benamore et al.\(^{31}\) reported retreatment with curative intent in 37% of recurrences which included seven patients treated for isolated brain metastases. Subotic et al.\(^{32}\) reported oncological retreatment for 54% of patients developing recurrence/SPLC. There was variable reporting of proportion receiving curative-intent treatment at 26% to 75%\(^{36,42,43}\) palliative intent at 25% to 47%\(^{43}\) and best supportive care at 22%.\(^{32}\) Hanna et al.\(^{36}\) reported follow-up of 271 patients finding 75% of asymptomatic patients detected by scheduled follow-up were suitable for curative-intent retreatment, yet no patients detected outside scheduled follow-up were suitable for curative-intent retreatment.

**Survival**

Hanna et al.\(^{36}\) reported that 37 of 49 patients (75.5%) with asymptomatic new or recurrent lung cancer underwent curative-intent retreatment (surgery or radiotherapy) with median survival of 69 months (range: 12–76 mo). Conversely, none of 14 patients with symptomatic tumor recurrences were able to undertake curative-intent retreatment and they patients had median survival of 15 months (range: 7–63 mo) from initial surgery. Asymptomatic patients unsuitable for curative-intent retreatment had significantly shorter median survival of 25 months (range: 6–48 mo) than those receiving curative-intent retreatment (\( p < 0.0001 \)).

Choi et al.\(^{44}\) reported on 111 disease recurrences among 358 patients and observed locoregional recurrence in 29 of 111 (26.1%) who had significantly longer median survival than those with distant metastases (4.2 ± 0.3 y versus 3.0 ± 0.2 y, mean ± SE, \( p = 0.008 \)). Patients with disease recurrence detected by CT had superior survival to those detected by HPE and patients with disease recurrence detected by PET-CT and CT had a not significant trend to longer survival than CT-detected recurrence alone (3.8 ± 0.2 y versus 3.3 ± 0.3 y, \( p = 0.179 \)).

Westeel et al.\(^{42}\) reported on 136 recurrence events among 196 patients with follow-up consisting of HPE, CXR every 3 months with CT chest and upper abdomen, and bronchoscopy scheduled 6 monthly for 3 years. Among asymptomatic patients, 3-year survival was 40% when the recurrence was exclusively thoracic, 45% when this thoracic recurrence was diagnosed by chest CT scan or fiberoptic bronchoscopy, and 70% if curative-intent retreatment was provided. In univariate regression analysis, detection of recurrence by a scheduled procedure, the absence of symptoms at the time of recurrence, thoracic recurrence site, and disease-free survival greater than 1 year from operation were associated with significantly enhanced overall survival (OS) from recurrence. Median survival was significantly longer in asymptomatic patients with the disease detected at recurrence diagnosis, scheduled examinations, recurrence greater than 1-year after resection, and for those with thoracic rather than extrathoracic recurrence. Similar results were found by Lou et al.\(^{40}\) reporting longer OS from the time of operation for patients whose recurrences were detected by surveillance CT scans than for those presenting with symptoms (\( p < 0.001 \)).

Conforti et al.\(^{35}\) reported OS for patients with disease recurrence detected by scheduled examinations compared with patients with recurrence detected by unscheduled examinations; they found a trend but no significant survival difference (OS HR = 0.81, 95% CI: 0.63–1.03, \( p \) value 0.08). OS rates at 3 years from diagnosis of recurrence were similarly poor in both groups (22.0% scheduled detection versus 21.8% unscheduled recurrence detection).

**Discussion**

Within this study, we set out to answer three questions. First, to estimate the effectiveness of surveillance in detection of disease recurrence and SPLC. We found that the use of a systematic follow-up strategy significantly increased the likelihood of detection of disease recurrence or SPLC OR equal to 8.15 (95% CI: 3.5–18.8, \( p < 0.0001 \)).

Four studies described the impact of scheduled screening on detection of recurrence/SPLC on curative-intent retreatment finding a significant increase in likelihood of provision of curative-intent retreatment (OR = 4.31, 2.10–8.84, \( p < 0.0001 \)). Five studies described survival outcomes suggesting poor survival after symptomatic presentation, trends to longer survival after scheduled surveillance, CT- and PET-detected recurrence, and longer survival for those able to undergo curative-intent retreatment.

Edelman et al.\(^{45}\) proposed principles for evaluating the use of testing in follow-up after curative-intent cancer treatment. These include the following: (1) the screening interval and duration of testing should be consistent with the maximal risk of recurrence and the natural history of the tumor; (2) risk of second malignancies should guide tests; (3) tests should be directed at the most likely sites of recurrence with high positive and negative predictive values, that is, test accuracy; (4)
therapy should be available that will result in cure, significant prolongation of life, or palliation of symptoms; and (5) initiation of earlier therapy should improve outcome.

In relation to screening interval and duration, this study found reports of disease recurrence or new SPLC in 18% to 71% of patients under surveillance with higher recurrence rates for stage IIIA (52% to 72%) than those in stage I to II (11% to 22%). Most of the recurrences were diagnosed within the first 2 years after treatment, but no clear conclusion regarding the screening interval can be drawn on the basis of the presented studies.

The findings of this study lend only limited support for the superiority of CT-based screening as preferred surveillance methodology for detection of disease recurrence and SPLC. The studies included in this analysis are published between 2000 and 2020, and there is a possibility that the diagnostic accuracy of CT imaging has evolved over this period which may influence detection rates and could be a contributing explanation for the heterogeneity observed between studies comparing CT with standard follow-up (Fig. 3A).

In relation to likely sites of recurrence, CT-based surveillance is currently embedded in available guidelines, including recommendations from the National Comprehensive Cancer Network, United States, suggesting chest CT every 6 months for 2 to 3 years after surgery, and every 3 to 6 months for 3 years after radiotherapy, and similarly in recent American Society of Clinical Oncology guidelines. Other guidelines are less prescriptive. The latest guidelines of the National Institute for Health and Care Excellence on lung cancer recommend to offer all people with lung cancer an initial specialist follow-up appointment within 6 weeks of completing treatment to discuss ongoing care with regular appointments after this rather than relying on the person requesting appointments when they experience symptoms and to offer protocol-driven follow-up.

Despite the brain being among the most common sites for recurrence and the fact that detection and treatment of small and symptomless solitary brain metastases may lead to long-term survival, monitoring the brain for recurrence has generally been neglected. In the current National Comprehensive Cancer Network guidelines, PET/CT and brain MRI are recommended after chest CT provides suspicion of relapse. Systematic use of PET/CT and brain MRI is discouraged in the American Society of Clinical Oncology guidelines, mainly on the basis of the lack of supporting evidence.

NSCLC is characterized by high levels of extrathoracic recurrence. The enhanced ability of PET-CT to detect extrathoracic recurrence addresses the third postulate of Edelman et al. and is used widely as standard of care in follow-up of cancers after curative-intent treatment including melanoma and as a second-step test in case of rising tumor markers, for example, in patients with ovarian and colorectal cancer. However, only a single randomized controlled trial was available which provided little compelling evidence of efficacy of PET-CT in follow-up. This trial may have been challenged by uncertain assumptions used for power calculations, differences in patient characteristics, varied exposure to PET-CT, and unusually high proportions of symptomatic recurrence at unscheduled time points. The study by Onishi et al. however suggests that PET-CT provides a diagnostic accuracy similar to that of multimodality surveillance imaging engaging brain MRI imaging using contrast media, contrast-enhanced whole-body CT, and bone scintigraphy. Thus, further evidence is needed to enable differentiation between the effectiveness of CT versus PET-CT for surveillance after curative treatment in patients with lung cancer. This also applies to newer technologies such as circulating tumor markers which are currently not recommended.

Key questions in evaluating follow-up after curative-intent treatment are whether systematic surveillance increases the detection of treatable disease and if this retreatment leads to important survival advantage (fourth and fifth principles of Edelmanns et al. discussed previously). This meta-analysis suggests that patients with recurrence detected during scheduled surveillance are more likely to be asymptomatic and more likely to be candidates for retreatment with curative intent suggesting important potential survival benefit.

Although SPLC are well recognized in surveillance follow-up, only four included studies clearly differentiate between disease recurrence and SPLC, despite the availability of accepted sorting strategies. In the absence of such distinction, studies may potentially incorrectly categorize all detected diseases as recurrence rather than recurrences and SPLC with the potential to disadvantage SPLC presentations.

Some of the challenges in evaluating the effectiveness of a follow-up strategy are first differentiating the effect of systematic follow-up independent of surveillance modality and second the modality-specific detection effects on disease recurrence and SPLC. Potential negative aspects of routine testing in follow-up may include patient anxiety, cost, and the costs and morbidity associated with investigation of incidental findings at routine scanning, and unnecessary intervention. Conversely, systematic follow-up may have unmeasured beneficial impacts in domains of quality of life, patient satisfaction, emergency presentation and readmission, communication, coordination, and continuity of care, which merit further investigation. None of
these aspects have been addressed in any of the studies presented here.

Limitations

There are significant limitations in the use of retrospective analyses of prospective datasets including nonprospective data specification, missing data, variable data interpretation, and the biases of inclusion, sampling, information, and nonrandomization, and there remains some uncertainty around the prospective specification of some data outcomes.

There was a degree of cohort heterogeneity between the included studies. The cohort described by Hanna et al.36 was predominantly stage I (79%), whereas the cohort described by Lou et al.40 consisted of stage III disease. Most studies however included mixed populations of stage I to III. Furthermore, Spratt et al.39 described follow-up in inoperable lung cancer patients treated with SBRT although 41% of the patients had previous lung resection or radiotherapy for NSCLC and therefore represented relapsed NSCLC rather than primary NSCLC. Westeel et al.42 permitted recruitment of patients undergoing curative-intent resection of synchronous lung and brain metastasis recurrence. There is also a degree of heterogeneity in trial design between included studies, both with regard to the type of intervention and scheduling. In one study, follow-up comprised routine brain MRI using contrast media, contrast-enhanced whole-body CT, and bone scintigraphy,39 elsewhere brain imaging only if clinically indicated,35 and in another CXR and CT thorax only.36 One study engaged CT or PET done 6 monthly on an alternating basis.37

The definition of symptomatic recurrence across studies is unclear. Intuitively, one might assume that the development of symptoms would precipitate unscheduled examination, and yet, one study reports symptomatic recurrence in both scheduled and unscheduled follow-up36 and a second study37 described the complete recurrence cohort as asymptomatic suggesting symptomatic recurrences detected by unscheduled investigation were excluded from the trial and therefore likely to favor screen-detected recurrence. Treatment-intent and retreatment modalities were also variably reported; in some studies, both were reported, and in other studies treatment-intent only was reported, and there was no clear differentiation between active anticancer treatment with palliative intent and best supportive care.

Conclusions

Patients undergoing curative-intent treatment for NSCLC remain at high risk of disease recurrence and SPLC, and early detection may increase likelihood of retreatment and survival. As yet, there is no clear evidence to confirm the most appropriate screening modality(ies), screening interval, or duration of screening. There remains a clear need to determine the qualitative and quantitative effectiveness of systematic follow-up after curative-intent treatment of NSCLC.

Supplementary Data

Note: To access the supplementary material accompanying this article, visit the online version of the Journal of Thoracic Oncology at www.jto.org and at https://doi.org/10.1016/j.jtho.2021.01.1622.

References


15. Ari


