REVIEW

Recommended evaluation of adrenal incidentalomas is costly, has high false-positive rates and confers a risk of fatal cancer that is similar to the risk of the adrenal lesion becoming malignant; time for a rethink?

T J Cawood, P J Hunt, D O’Shea¹, D Cole and S Soule
Department of Endocrinology, Christchurch Hospital, Private Bag 4710, Christchurch, New Zealand and ¹Department of Endocrinology, St Vincent’s University Hospital, Dublin, Ireland

(Correspondence should be addressed to T J Cawood; Email: tom.cawood@cdhb.govt.nz)

Abstract

Objective: To assess the performance of current clinical recommendations for the evaluation of an adrenal incidentaloma.

Design and methods: Literature review. Electronic databases (Pubmed, Ovid and citation searches from key articles) from 1980 to 2008 were searched. Eligible studies were those deemed most applicable to the clinical scenario of a patient referred to an endocrinologist for assessment of an incidentally detected adrenal mass. Surgical series, histopathological series and oncological series were reviewed and most were excluded.

Results: The prevalence of functional and malignant lesions presenting as adrenal incidentaloma was similar to that quoted in most reviews, other than a lower incidence of adrenal carcinoma (1.9 vs 4.7%) and metastases (0.7 vs 2.3%). The development of functionality or malignancy during follow-up was rare (<1% becoming functional and 0.2% becoming malignant). During follow-up, false-positive rates of the recommended investigations are typically 50 times greater than true positive rates. The average recommended computed tomography (CT) scan follow-up exposes each patient to 23 mSv of ionising radiation, equating to a 1 in 430 to 2170 chance of causing fatal cancer. This is similar to the chance of developing adrenal malignancy during 3-year follow-up of adrenal incidentaloma.

Conclusion: Current recommendations for evaluation of adrenal incidentaloma are likely to result in significant costs, both financial and emotional, due to high false-positive rates. The dose of radiation involved in currently recommended CT scan follow-up confers a risk of fatal cancer that is similar to the risk of the adrenal becoming malignant. This argues for a review of current guidelines.

Introduction

The scenario of a patient being referred to an endocrinologist for assessment of an adrenal incidentaloma is becoming increasingly common, with improvements in imaging, increasing numbers of radiological investigations and an aging western world population. Current clinical management of such patients varies across different centres and countries, as there is a relative lack of prospective outcome data on which to base clinical recommendations, with practice generally led by consensus (1) and expert opinion (2). Current guidelines recommend various endocrine investigations and radiological imaging at baseline. If the lesion is thought to be benign and non-functional, it is suggested that these tests are repeated at between 6 and 12 monthly intervals for between 2 and 4 years, which typically would include between one and three follow-up computed tomography (CT) scans (1–3).

While such an approach aims to avoid missing an important functional or malignant lesion, because of the low prevalence of such conditions, investigations need to have very high sensitivity and specificity in order to perform adequately in this environment. The consequence of low specificity is a high false-positive rate, which can result in significant financial and emotional cost, with the potential for further tests and/or unnecessary adrenalectomy. Furthermore, there has been little attention paid to the possible harmful consequences of these investigations, in particular the exposure to ionising radiation from repeated CT scanning.

In order to assess the performance of the current clinical recommendations, we first reviewed the literature to find the most representative data on the prevalence of the various different pathologies that are likely to present as an adrenal incidentaloma. This involved excluding those studies that may bias towards
malignant or functional lesions, such as surgical, histopathological and oncology series.

We then reviewed the performance of the various investigations commonly used for assessment of adrenal incidentalomas, and applied the sensitivity and specificity data to the likely prevalence of the different pathologies at both initial presentation and during follow-up. Finally, we considered the potential induction of fatal cancer as a result of exposure to ionising radiation during repeated CT scanning.

Design and methods

A literature review was undertaken that was intended to include studies that were most applicable to the clinical scenario of a patient referred to an endocrinologist for assessment of an incidentally detected adrenal mass. Electronic databases (Pubmed and Ovid) from 1980 to 2008 were searched, and additional citations from key articles were also included. Pre-determined exclusion criteria were surgical series, histopathological series and oncological series, non-English studies (where inclusion criteria could not be accurately determined), studies where a proportion of the patients from larger multicentre studies may have been reported elsewhere, studies where the study population was not > 20 and studies where the study population was neither a random sample nor a consecutive series.

Using the search terms ‘adrenal’ + ‘incidentaloma’ and ‘adrenal’ + ‘ incidental’, from 1980 to 2008, 828 studies were found. After selecting those articles pertaining to the presentation or follow-up of adrenal incidentaloma, and only those written in English and where n > 20, there were 68 articles remaining. To these articles were added a further 42 studies that were obtained from citation searches from two leading reviews (4, 5), giving a total of 110 articles. A number of these were then excluded for a variety of reasons including review articles (28), surgical series (21), studies where a proportion of the patients from larger multicentre studies may have been reported elsewhere (15) and studies including oncology patients (7) (Fig. 1). A small number of additional exclusion criteria were applied where methodological issues were such that these studies were not deemed most applicable to the clinical scenario of adrenal incidentaloma. Examples of such exclusions are given below.

Prevalence of malignant and functional lesions in adrenal incidentaloma

The prevalence of malignant and functional lesions in true adrenal incidentaloma is likely to have been overestimated in the literature for a variety of reasons. The consensus definition of a true adrenal incidentaloma is an adrenal mass (> 1 cm diameter) detected incidentally during imaging performed for extra-adrenal complaints (therefore excluding those with severe or paroxysmal hypertension, hypokalaemia, clinical signs of hypercortisolism or hyperandrogenism). Also, those patients with a history of previous or current malignancy, who represent a group of patients with an increased risk of harbouring an adrenal metastasis, should arguably be referred to an oncology service and excluded from consideration of true incidentalomas. The operating characteristics of investigation algorithms intended to reasonably exclude a functional or malignant lesion in those with a low pre-test probability of malignancy may not be adequate in this patient group.

A number of studies should be excluded from consideration if the aim is to determine the likely prevalence of malignant and functional lesions in those presenting to endocrinologists as a true incidentaloma. Such studies include reports where some or all of the patients are from surgical series (6–12). These lesions have been deemed worthy of consideration for surgical removal, usually due to being larger or more suspicious of malignancy. These data are likely to overestimate the prevalence of malignancy compared with those referred to endocrinology with true adrenal incidentaloma (13). Also, a number of papers have been misrepresented in literature reviews (4), where the percentage of malignant lesions is reported as a percentage of those operated on, rather than as a percentage of those initially assessed with an incidentaloma, again exaggerating the prevalence of malignant lesions (e.g. 13% of those operated on versus 6% of the total (14), 14% of those operated upon versus 3% of total (15) and 12% in those operated upon versus 2.6% of total (16)). Other sources of potential bias include studies involving patients with a history, current diagnosis or suspicion of malignancy (16–20), as well as studies where the adrenal lesion is not incidental, but was detected during investigations based on symptoms, signs, hormone levels and imaging studies (21). Additional problems include studies where those with non-functional adrenal masses are excluded (8), studies where the primary imaging modality was ultrasound (and so smaller lesions that are more likely to be benign may be missed) (22) and studies that excluded patients with masses smaller than 2 cm in diameter (23). Further problems arise when including studies of selected patients rather than random or consecutive series (24), including studies of small size which are consequently more influenced by random variation (e.g. where a single case can alter reported prevalence by 5% (8, 17, 25)), and also studies where a proportion of the patients from larger multicentre studies may have previously been reported elsewhere in smaller reports involving less centres, so potentially over-representing any random biases (26–34). More recent papers, not yet established in review literature, also have potential sources of bias such as excluding lesions that appear
benign (Hounsfield units (HU) < 10) or patients with a specific CT diagnosis such as haematoma and myelolipoma (35).

In short, the most relevant studies for the scenario of true adrenal incidentaloma report on consecutive or unselected patients and exclude patients with a history of previous or current malignancy. In addition, where a larger multicentre report includes patients that may already have been described in smaller single-centre studies, only the larger report should be considered. Using these criteria and others detailed above, we consider the studies in Table 1 to be the most relevant (30, 34, 36–42), plus one study where we have excluded the oncology patients (19).

From these studies, the expected prevalence of primary adrenal cancer in true adrenal incidentaloma is less than half that quoted in recent extensive literature reviews (2, 4, 5) (mean 1.9%, median 1.4%, compared to mean 4.4–4.7%, (Table 2)). Similarly, when compared with earlier inclusive reviews, these selected studies find the prevalence of metastasis is less than half (mean 0.7%, median 0.2%, compared to mean 2.3%), the prevalence of phaeochromocytoma is two thirds (mean 3.1% compared to 5.2%), the prevalence of aldosteronoma is half (mean 0.6% compared to 1.1%), while the prevalence of subclinical Cushing’s syndrome is similar (mean 6.4%, median 6.0%, compared to mean 6.6%).

Therefore, by including surgical series and the other above-listed biases, the current literature leads to a significant overestimate of the chance of a true adrenal incidentaloma being either malignant or...
functional, with the exception of subclinical Cushing’s syndrome. With this lower pre-test probability of disease, the investigation algorithms based on the higher prevalence rates may be suboptimal for managing true adrenal incidentaloma. Furthermore, as CT technology has improved in the last two decades, the ability to detect smaller lesions is improving, and so the proportion of smaller, more likely benign lesions that are detected is likely to be increasing. This is supported by the negative correlation ($R = -0.4$, $R^2 = 0.16$) between the year of study and mean tumour size in the studies listed in Table 1. Therefore, even the lower prevalence estimates derived from the studies in Table 1 may be an overestimate when considering the population of patients now referred with adrenal incidentaloma, which may include an increasing proportion of patients with smaller adrenal masses.

There are no data giving an accurate figure for the prevalence of adenomas among adrenal incidentalomas. This is relevant because the performance of CT scanning in assessing adrenal lesions is partly related to its ability to identify lipid-rich lesions including adenoma by incorporating HU into diagnostic algorithms. Adenomas may be non-functional, but also include lesions causing subclinical Cushing’s syndrome and hyperaldosteronism (HA). Adenomas may be under-represented among surgical series as the majority of small, benign, non-functional lesions will be adenomas and hence less likely to be surgically excised. Estimates of the prevalence of adrenal adenomas among adrenal incidentalomas vary. One study suggests a prevalence of 52% (4), although this review does include a number of surgical series. The real prevalence of adenomas is likely to be nearer 80%, this being the total of apparently non-functioning adenomas (71.2%), plus lesions causing subclinical Cushing’s syndrome (7.9%) and lesions causing HA (1.2%) (5).

### Table 1

The prevalence of likely pathological entities in those patients presenting with true adrenal incidentaloma.

<table>
<thead>
<tr>
<th>Age (years)</th>
<th>Size of mass (cm)</th>
<th>Benign</th>
<th>Malignant</th>
<th>Primary adrenal carcinoma (%)</th>
<th>Metastases (%)</th>
<th>Non-functional (%)</th>
<th>Subclinical Cushing's (%)</th>
<th>Phaeo (%)</th>
<th>Aldosteronoma (%)</th>
<th>Total (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>Median</td>
<td>Boys</td>
<td>Girls</td>
<td>Mean or median</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>342</td>
<td>58</td>
<td>1.2</td>
<td>0.3</td>
<td>98.6</td>
<td>0.3</td>
<td>96.9</td>
<td>95.9</td>
<td>1.2</td>
<td>0.3</td>
<td>95.9</td>
</tr>
<tr>
<td>1991 USA</td>
<td>1992 Germany</td>
<td>Yes</td>
<td>Yes</td>
<td>1.2</td>
<td>0.3</td>
<td>98.6</td>
<td>95.9</td>
<td>1.2</td>
<td>0.3</td>
<td>95.9</td>
</tr>
<tr>
<td>1996 France</td>
<td>1997 Italy</td>
<td>Yes</td>
<td>Partially</td>
<td>1.2</td>
<td>0.3</td>
<td>98.6</td>
<td>95.9</td>
<td>1.2</td>
<td>0.3</td>
<td>95.9</td>
</tr>
<tr>
<td>1997 Greece</td>
<td>2000 Turkey</td>
<td>Yes</td>
<td>Yes</td>
<td>1.2</td>
<td>0.3</td>
<td>98.6</td>
<td>95.9</td>
<td>1.2</td>
<td>0.3</td>
<td>95.9</td>
</tr>
<tr>
<td>2007 Greece</td>
<td>2010 Italy</td>
<td>Yes</td>
<td>Yes</td>
<td>1.2</td>
<td>0.3</td>
<td>98.6</td>
<td>95.9</td>
<td>1.2</td>
<td>0.3</td>
<td>95.9</td>
</tr>
<tr>
<td>2007 Greece</td>
<td>2010 Italy</td>
<td>Yes</td>
<td>Yes</td>
<td>1.2</td>
<td>0.3</td>
<td>98.6</td>
<td>95.9</td>
<td>1.2</td>
<td>0.3</td>
<td>95.9</td>
</tr>
<tr>
<td>2007 Greece</td>
<td>2010 Italy</td>
<td>Yes</td>
<td>Yes</td>
<td>1.2</td>
<td>0.3</td>
<td>98.6</td>
<td>95.9</td>
<td>1.2</td>
<td>0.3</td>
<td>95.9</td>
</tr>
</tbody>
</table>

| | | | | Mean or median |
| | | | | 0.4 |
| | | | | 0.16 |

Follow-up of adrenal incidentalomas

Studies addressing the follow-up of adrenal incidentaloma are relatively low in number and have a considerable degree of heterogeneity. The inclusion of adrenal incidentaloma studies where the adrenalectomy rate is 100% would help to provide more accurate histological data and reduce the chance of underdiagnosis, but such a study would be unethical, with the general experience being that relatively few patients develop malignancy or clinically apparent functional lesions. Mindful of these and other limitations, we consider the studies included in Table 3 (19, 23, 33, 35, 36, 39, 41, 43–49) to be those most relevant when considering the natural history of adrenal incidentalomas that are initially considered benign and non-functioning.
For any test proposed for follow-up of adrenal incidentalomas, the following should be considered: i) what is the sensitivity and specificity of the test, particularly in the low-prevalence situation of adrenal incidentaloma follow-up, ii) what is the consequence of false-negative and false-positive results, iii) are there any risks associated with the test and iv) what are the financial costs of the test?

Tests for subclinical Cushing’s syndrome

Subclinical Cushing’s syndrome has not been adequately defined, and the natural history of this syndrome is unknown (4). The definitions of subclinical Cushing’s syndrome vary widely in the literature. No current test or definition accurately identifies those at risk of complications (such as glucose intolerance, obesity, osteoporosis and adrenal failure following adrenalectomy). Current definitions of subclinical Cushing’s syndrome include autonomous cortisol secretion in patients who do not have the typical signs and symptoms of hypercortisolism (2). Others use definitions that stipulate the absence of clinical signs of cortisol excess plus at least two biochemical abnormalities of the hypothalamic–pituitary–adrenal (HPA) axis (30, 50). The main drawback of these definitions is that the more tests that are performed looking for HPA abnormalities, the higher the rate of diagnosis of subclinical Cushing’s syndrome. Furthermore, if the disease is defined by laboratory abnormalities, rather than by some clinical correlate, it becomes difficult to define false positives, as all positives are by definition indicative of the diagnosis.

The absence of an agreed definition of subclinical Cushing’s syndrome or gold standard test for diagnosing subclinical Cushing’s syndrome makes the determination of specificity and sensitivity of tests used to diagnose subclinical Cushing’s syndrome problematic. The NIH position statement (2) and others including the Endocrine Society (51) recommend using the 1 mg dexamethasone (DXT) as the initial test for subclinical Cushing’s syndrome. Applying the commonly used definition of subclinical Cushing’s syndrome (the absence of clinical signs of cortisol excess, plus two or more abnormalities of basal or dynamic test of the HPA axis), the largest study is from Mantero et al. (30) comprising 946 adrenal incidentalomas, and 854 were characterised as non-hypersecreting and 92 as subclinical Cushing’s syndrome. Tests used to evaluate the HPA axis included urinary-free cortisol (UFC), plasma ACTH, serum DHEA sulphate (DHEAS), serum 17-hydroxyprogesterone (17-OHP), 24-h cortisol rhythm, 1 mg DXT, 100 µg CRH test and 250 µg ACTH test, although not all tests were performed on all patients.

The results (Table 4) indicate that ~15% of lesions classified as non-hypersecreting demonstrate a single abnormal test of the HPA axis, a significantly lower
Table 3: Studies detailing the follow-up of adrenal incidentalomas thought to be benign and non-functioning after initial diagnostic work-up.

<table>
<thead>
<tr>
<th>First author</th>
<th>Year</th>
<th>Number of patients (n&gt;20)</th>
<th>Mean age (years)</th>
<th>Size of masses (mean or median)</th>
<th>Follow-up (years)</th>
<th>Increased in size (%)</th>
<th>Unchanged in size (%)</th>
<th>Decreased in size (%)</th>
<th>Became malignant (%)</th>
<th>Became functional (%)</th>
<th>Developed adrenal Ca (%)</th>
<th>Developed metastases (%)</th>
<th>Developed Cushing’s syndrome (%)</th>
<th>Developed subclinical Cushing’s syndrome (%)</th>
<th>Developed phaeochromocytoma (%)</th>
<th>Developed aldosteronoma (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Song (35)</td>
<td>2007</td>
<td>71</td>
<td>NG</td>
<td>NG</td>
<td>2.7</td>
<td>0.0</td>
<td>NG</td>
<td>NG</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>100</td>
<td>NG</td>
<td>NG</td>
<td>NG</td>
<td>NG</td>
</tr>
<tr>
<td>Song (35)</td>
<td>2007</td>
<td>209</td>
<td>NG</td>
<td>NG</td>
<td>2.5</td>
<td>4.0</td>
<td>NG</td>
<td>NG</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>100</td>
<td>NG</td>
<td>NG</td>
<td>NG</td>
<td>1</td>
</tr>
<tr>
<td>Favia (33)</td>
<td>2000</td>
<td>90</td>
<td>NG</td>
<td>1.8</td>
<td>2.0</td>
<td>7.0</td>
<td>NG</td>
<td>NG</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>100</td>
<td>NG</td>
<td>NG</td>
<td>NG</td>
<td>NG</td>
</tr>
<tr>
<td>Barzon (43)</td>
<td>1999</td>
<td>75</td>
<td>56</td>
<td>2.5</td>
<td>4.0</td>
<td>16.0</td>
<td>81.0</td>
<td>2.7</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>100</td>
<td>8</td>
<td>2.7</td>
<td>4*</td>
<td>1.3</td>
</tr>
<tr>
<td>Follow-up</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bulow (44)</td>
<td>2006</td>
<td>229</td>
<td>64</td>
<td>2.5</td>
<td>2.1</td>
<td>7.4</td>
<td>87.4</td>
<td>5.2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>100</td>
<td>2.6</td>
<td>1.3</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Toveston (19)</td>
<td>2007</td>
<td>88</td>
<td>NG</td>
<td>2.6</td>
<td>2.0</td>
<td>12.5</td>
<td>87.5</td>
<td>0.0</td>
<td>1.1</td>
<td>0</td>
<td>1.1</td>
<td>98.9</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Barry (33)</td>
<td>1998</td>
<td>231</td>
<td>64</td>
<td>2.0</td>
<td>7.0</td>
<td>4.0</td>
<td>96.0</td>
<td>0.0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>100</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Libe (45)</td>
<td>2002</td>
<td>64</td>
<td>61</td>
<td>2.5</td>
<td>2.1</td>
<td>20.0</td>
<td>0.0</td>
<td>1.6</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>98.4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Siren (46)</td>
<td>2000</td>
<td>27</td>
<td>59</td>
<td>2.5</td>
<td>7.1</td>
<td>25.0</td>
<td>31.0</td>
<td>44.0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>100</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Rossi (36)</td>
<td>2000</td>
<td>32</td>
<td>NG</td>
<td>2.8</td>
<td>15.6</td>
<td>84.6</td>
<td>0.0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>100</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Bastounis (41)</td>
<td>1997</td>
<td>60</td>
<td>NG</td>
<td>3.2</td>
<td>3.6</td>
<td>3.7</td>
<td>97.3</td>
<td>0.0</td>
<td>0</td>
<td>0</td>
<td>100</td>
<td>NG</td>
<td>0</td>
<td>NG</td>
<td>NG</td>
<td></td>
</tr>
<tr>
<td>Grossrubatscher (47)</td>
<td>2001</td>
<td>53</td>
<td>NG</td>
<td>2.5</td>
<td>2.0</td>
<td>41.5</td>
<td>47.2</td>
<td>11.3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>100</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Emral (39)</td>
<td>2003</td>
<td>60</td>
<td>NG</td>
<td>2.0</td>
<td>0.0</td>
<td>0.0</td>
<td>NG</td>
<td>NG</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>100</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Mantero (49)</td>
<td>2000</td>
<td>53</td>
<td>NG</td>
<td>1</td>
<td>26.4</td>
<td>NG</td>
<td>NG</td>
<td>NG</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>100</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Bencsik (48)</td>
<td>1995</td>
<td>27</td>
<td>NG</td>
<td>&lt;3</td>
<td>1.8</td>
<td>3.7</td>
<td>NG</td>
<td>NG</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>100</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>95% CI</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8.0–21.3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>8.0–21.3</td>
<td>45.6–90.2</td>
<td>–2.4–16.4</td>
<td>0.0–0.4</td>
<td>0.0–0.0</td>
<td>0.0–0.2</td>
<td>99.6–99.8</td>
<td>0.0–0.7</td>
<td>–0.4–1.0</td>
<td>0.0–0.4</td>
<td></td>
</tr>
<tr>
<td>Median</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>8.0</td>
<td>0.0</td>
<td>0.0</td>
<td>100.0</td>
<td>0.0</td>
<td>0.0</td>
<td>100.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td></td>
</tr>
</tbody>
</table>

NG, not given; CI, confidence interval.

*These patients had any abnormal baseline hormonal test repeated, plus up to eight hormonal assessments in total, including 0800 h cortisol, 1800 h cortisol, morning ACTH, DHEAS, 17-OHP, 1 mg DXT and in some cases CRH test. The paper by Song is included three times, as these subgroups of patients were followed up either radiologically (71 patients), by histology or imaging (209 patients), or by clinical information (41 patients).

bRepresents one case of renal cell cancer metastasis in a patient with a known history of renal cell carcinoma; arguably this study should have been excluded from this analysis.

**Represents one case of adrenal non-Hodgkin’s lymphoma.
Table 4 Hormonal evaluation of non-hypersecreting adrenal incidentalomas and lesions characterised as causing subclinical Cushing’s syndrome, from Mantero (30).

<table>
<thead>
<tr>
<th>Test</th>
<th>Non-hypersecreting (%)</th>
<th>Subclinical Cushing’s syndrome (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low morning ACTH levels</td>
<td>15</td>
<td>79</td>
</tr>
<tr>
<td>Above normal UFC</td>
<td>11</td>
<td>75</td>
</tr>
<tr>
<td>Abnormal circadian rhythm of plasma cortisol</td>
<td>17</td>
<td>43</td>
</tr>
<tr>
<td>Blunted ACTH response to CRH</td>
<td>17</td>
<td>55</td>
</tr>
<tr>
<td>Cortisol not adequately suppressed by 1 mg dexamethasone</td>
<td>10</td>
<td>73</td>
</tr>
</tbody>
</table>

Available studies are confounded by heterogeneity, examining different patient populations, with different pre-test probabilities, and using total or fractionated metanephrines, with different laboratory cut-offs. Despite these limitations, there is a reasonable degree of consistency with urinary metanephrines having median sensitivity of 95% (77% (56), 97.1% (57), 90% (58), 100% (59), 100% (60), 94.7% (61) and 95% (62)) and specificity of 95% (93% (56), 91.1% (57), 98% (58), 94% (59), 99.1% (60), 95.3% (61) and 98% (62)). Plasma metanephrines have slightly higher sensitivity of 98% (97% (58), 99% (56), 98% (63), 100% (64) and 96% (65)), but lower specificity of 89% (85% (58), 89% (56), 92% (63), 96.7% (64) and 80% (65)).

Tests for HA

The recommended initial screening test for HA in hypertensive patients with adrenal incidentalomas is the ratio of ambulatory morning plasma aldosterone concentration to plasma renin activity (Aldos/PRA) (2, 4, 13). Normokalaemic HA occurs at a frequency of 7–38%, hence screening for HA should not be limited to those who are hypokalaemic (4). The ability of the Aldos/PRA to detect primary HA in resistant hypertension varies, depending on numerous factors including patient selection criteria, medication use and laboratory cut-offs. Typical expected performance of the Aldos/PRA is in the region of sensitivity 90% (78% (66), 89% (67), 92% (68), 90% (69), 96% (70) and 100% (71)) and specificity 90% (83% (66), 84% (71), 85% (70), 91% (69), 96% (67) and 100% (68)). There are less data on the performance of Aldos/PRA in the setting of adrenal incidentaloma, although studies that have retrospectively applied certain cut-offs have achieved sensitivity of 100% and specificity of 95, 96 and 99% (72–74). However, these figures may significantly overestimate the true performance of the Aldos/PRA ratio as in the quoted series the prevalence of HA was remarkably high at 14% (73) and 12% (10 of the 50 incidentaloma patients) (74), when compared with the expected prevalence of 1% in all adrenal incidentaloma patients and 4% in hypertensive adrenal incidentaloma patients (72). Of note, the true sensitivity of the Aldos/PRA as a screening test in adrenal incidentaloma remains unclear, as without additional definitive tests such as saline suppression in patients with a normal Aldos/PRA ratio, the sensitivity of the ratio for the detection of HA in the hypertensive patient with an incidentaloma remains conjectural.

Taken together, the sensitivity and specificity of the Aldos/PRA for detecting HA in patients with hypertension and an adrenal incidentaloma are likely to be in the region of 90–100%, depending on the cut-offs used and patient group studied, with a figure of 95% being arguably a reasonable estimate.

Proportion than those classified as subclinical Cushing’s syndrome. However, in the light of the current definition of subclinical Cushing’s syndrome (requiring two or more abnormal tests of the HPA axis), it is relevant to note that even the use of two out of the three most frequently abnormal HPA tests would still fail to capture a significant proportion of patients otherwise labelled as subclinical Cushing’s syndrome had more tests been performed. For example, an abnormal 1 mg DXT test plus a low ACTH only identified 55% of those in the subclinical Cushing’s syndrome group, and an abnormal 1 mg DXT plus above-normal UFC only identified 50% of the subclinical Cushing’s syndrome group. Therefore, a single 1 mg DXT will label 10% of the non-hypersecretors as subclinical Cushing’s syndrome (10% false positive, specificity 90%) and will fail to label 27% of the subclinical Cushing’s syndrome group (27% false negative, sensitivity 73%). The cut-off for morning cortisol used in this study was 5 μg/dl (138 nmol/l), and more stringent cut-offs would increase the false-positive rate.

This 73% sensitivity of the 1 mg DXT in subclinical Cushing’s syndrome is lower than that found in other studies (92% (45), 100% (40), 86% (52) and 83% (53)), although these other studies were much smaller (6–12 patients, compared with 92 (30)). When taken together, the 1 mg DXT was abnormal in 76% (110/131) (50) of those diagnosed with subclinical Cushing’s syndrome. The 90% specificity is in agreement with other studies of subclinical Cushing’s syndrome (91% (2, 54)).

**Tests for phaeochromocytoma**

Recommendations for clinical practice from the First International Symposium on Pheochromocytoma are that measurement of plasma or urinary fractionated metanephrines should be the first-line test for diagnosis, and reference intervals should favour sensitivity over specificity (55).
Imaging as a test for malignancy

The ability of radiological imaging to detect malignant adrenal incidentalomas varies depending upon the criteria used. CT is generally better able to differentiate between adenomas and non-adenomatous lesions, compared with its ability to diagnose or exclude specific pathological entities.

When selecting CT criteria that discriminate between adenomatous and other lesions, the majority of studies have determined thresholds for parameters such as HU in a retrospective manner. These thresholds are likely to perform less well when tested prospectively. Only a minority of studies (75) have prospectively evaluated the performance of HU thresholds obtained from CT adrenals, and derived a sensitivity of 71% and specificity of 100% for diagnosing adrenal adenomas. Studies vary in their patient populations, inclusion and exclusion criteria, threshold levels and/or combination of thresholds (e.g. pre-contrast CT + delayed contrast wash-out thresholds). There are also a variety of methods by which the malignant or benign status of the lesions has been determined, and many of the studies are too small to have any identified adrenocortical carcinomas (76, 77), and hence the ability of the given thresholds to identify these lesions is unknown.

With these caveats in mind, the most studied imaging modality for diagnosis of adrenal adenomas is CT, with or without delayed contrast washout. The typical performance of CT is sensitivity of 64% (78), 66.7% in lesions over 2 cm (79), 71% (75), 72% (76), 92% using 10 min contrast washout (80), 96% (81) and 96% using 3 min contrast washout (82). Using combined parameters, CT performance for identifying adenomas can range from 42% (using non-contrast CT HU ≤ 20 plus size ≤ 4 cm (83)) to 100% (when lesions with pre-contrast HU < 0 are excluded (considered benign) and those with pre-contrast HU > 43 are excluded (considered suspicious of malignancy) and when cysts, myelolipomas and phaeochromocytomas are excluded, and using an absolute percentage contrast washout threshold at 10 min of 52% (77)). Collating the data from these varied studies gives an approximate sensitivity of 75% (mean sensitivity 78% and median sensitivity 72%). This 75% figure is consistent with other estimates (75% (79) and 68–89% (4)).

Typical specificity of CT for diagnosing adrenal adenomas is 85.7% (79) in lesions over 2 cm, 92% using delayed enhanced CT (82), 95% (78), 95% (76), 95% using delayed enhanced CT (80), 96% using delayed CT (81), 100% (75), 98% using various exclusions as detailed above (77) and 100% using non-contrast CT HU ≤ 20 plus size ≤ 4 cm (83). Collating the data from these varied studies gives an approximate specificity of 95% (mean and median specificity of 95%). This 95% figure is consistent with other estimates (93–100% (4)).

Table 5: The performance of recommended tests for the initial evaluation of 1000 adrenal incidentalomas.

<table>
<thead>
<tr>
<th>Test</th>
<th>Sensitivity (%)</th>
<th>Specificity (%)</th>
<th>Prevalence (%)</th>
<th>Test positive</th>
<th>False negative</th>
<th>True negative</th>
<th>False positive</th>
<th>True positive</th>
<th>Test negative</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 mg DXT</td>
<td>73</td>
<td>90</td>
<td>90</td>
<td>141</td>
<td>610</td>
<td>859</td>
<td>842</td>
<td>17</td>
<td>49</td>
</tr>
<tr>
<td>Urinary metanephrines</td>
<td>95</td>
<td>95</td>
<td>95</td>
<td>47</td>
<td>922</td>
<td>293</td>
<td>982</td>
<td>2</td>
<td>49</td>
</tr>
<tr>
<td>Plasma metanephrines</td>
<td>98</td>
<td>89</td>
<td>98</td>
<td>137</td>
<td>863</td>
<td>89</td>
<td>944</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Renin/aldosterone</td>
<td>95</td>
<td>95</td>
<td>95</td>
<td>95</td>
<td>944</td>
<td>95</td>
<td>944</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Alendronate</td>
<td>80</td>
<td>0.6</td>
<td>80</td>
<td>5.7</td>
<td>944</td>
<td>90</td>
<td>697</td>
<td>14</td>
<td>14</td>
</tr>
<tr>
<td>CT</td>
<td>75</td>
<td>75</td>
<td>75</td>
<td>75</td>
<td>67</td>
<td>38</td>
<td>67</td>
<td>14</td>
<td>14</td>
</tr>
<tr>
<td>Adrenal cancer</td>
<td>1.9</td>
<td>1.9</td>
<td>1.9</td>
<td>1.9</td>
<td>1.9</td>
<td>1.9</td>
<td>1.9</td>
<td>1.9</td>
<td>1.9</td>
</tr>
</tbody>
</table>
Studies addressing the detection of adrenocortical carcinoma, which have found a 4 cm cut-off, had 90% sensitivity, even though 76% of lesions greater than 4 cm were benign (27, 84). Another study found that in tumours > 3 cm, CT had a sensitivity of 75% and specificity of 67% for detecting adrenocortical carcinoma (85). Hence, the ability of CT to correctly diagnose this low-prevalence pathology is considerably lower compared to the ability of CT to differentiate between adenomatous and other non-adenomatous lesions.

The performance of the recommended investigations for the initial evaluation and during the follow-up of adrenal incidentalomas initially thought to be benign and non-functional is given in Tables 5 and 6.

Risk of cancer due to imaging-associated radiation exposure

The most commonly used imaging modality in assessing adrenal incidentaloma is CT. X-ray imaging techniques, including CT, involve ionising radiation, which carries an associated risk of inducing malignancy. The magnitude of that risk has been estimated from studies where populations have been exposed to a known dose of ionising radiation, and the subsequent increase in cancer-associated deaths has been ascertained. This has been studied in depth by the United Nations Scientific Committee on the Effects of Atomic Radiation who publish regular reports, updated by the latest mortality data. The latest report assessing lifetime mortality risk for total cancer associated with ionising radiation was published in 2006 (86), and includes data from a cohort of 86 611 survivors of atomic bombings in Japan. Although some of this cohort received large doses of radiation (up to 4000 mSv), the cohort is fundamentally a moderate dose cohort, with the mean dose in the exposed group being 200 mSv, with > 50% of the exposed individuals in the cohort having doses < 50 mSv (86, 87). Other studies have looked at fractionated or chronic low-dose exposure, including the IARC 15-country nuclear worker study (88), the Techa River contamination with liquid radioactive waste (89, 90) and the Semipalatinsk study of the effects of local fallout from Soviet atmospheric nuclear weapons testing in Kazakhstan (91). There are concerns about bias in these four studies, which found substantially elevated risks for solid cancers compared with the Japanese atomic bomb cohort. These studies have low statistical precision, with wide confidence intervals for the risk estimates, which at least in the Techa River and nuclear worker study overlapped those from the Japanese cohort (86).

The effective dose of radiation associated with an abdominal CT scan is typically 10 mSv (92) (http://www.icrp.org/docs/Rad_for_GP_for_web.pdf; http://www.fda.gov/cdrh/ct/risks.html) (range 3.9 (93)–30 mSv (94)), which is equivalent to 3.3 years of
natural background radiation exposure. The effective dose of 10 mSv for CT abdomen compares to 2 mSv for CT head, 7 mSv for barium enema, 1.3 mSv for lumbar spine plain X-ray and 0.02 mSv for chest X-ray. Targeted CT of the adrenals may involve a lower radiation dose, with a scan of the upper abdomen delivering ~5 mSv (95). However, adrenal protocols involving delayed contrast washout measurement require scanning the adrenal region twice, and hence the total effective dose may be similar to a whole abdomen scan, but is delivered to the upper abdominal organs, which would consequently receive a higher dose than during a whole abdominal scan.

There is direct epidemiological evidence from human populations demonstrating that acute exposure to ionising radiation at doses in the 10–50 mSv range (i.e. the organ dose range typically delivered by two or three CT scans) increases the risk of some cancers (87). The corresponding CT-related risks can thus be directly assessed without the need to extrapolate measured risks to lower doses (92). Risk estimates for radiation-induced mortality for solid cancers and leukaemia, with a test dose of 10 mSv, are ~4.8% per Sv. For a CT scan of the abdomen, the estimated associated lifetime absolute risk of cancer-related death as a direct consequence of the radiation dose received during the scan (10 mSv) is 0.048% (86). An abdominal CT scan is therefore estimated to cause one cancer-related death for every 1000 (http://www.nap.edu/catalog/11340.html) – 2000 (http://www.icrp.org/docs/Rad_for_GP_for_web.pdf) abdominal CT scans. More recent estimates based on the lifetime risk of fatal cancer as a result of a single abdominal CT scan in those aged over 30 years (as in the vast majority of adrenal incidentaloma patients) are lower, in the region of one death per 5000 scans (92). The precise magnitude of the fatal cancer risk associated with CT scanning remains a matter of some debate, but it is generally accepted that although the risk is relatively small, it is not zero, and hence should be considered before ordering CT imaging, especially where the expected diagnostic yield is low.

**Economic cost**

In our institution, the cost of blood tests to assess functionality (routine biochemistry, cortisol as part of 1 mg DXT, urinary catecholamines and metanephrines, renin and aldosterone) is ~NZ$ 175 (US$ 120). Additional tests such as 24-h UFC, ACTH, DHEAS, 17-OHP and serum catecholamines and metanephrines add an additional NZ$ 290 (US$ 200). An abdominal CT with contrast costs NZ$ 720 (US$ 500). Therefore, for baseline assessment, investigations cost between NZ$ 895 (US$ 620) and NZ$ 1185 (US$ 820), depending on how many blood and urine tests are ordered.

**Implications of current clinical recommendations**

There are a number of consensus or clinical guidance publications with arguably the more influential publications being those by Young (2), the UpToDate review of which Young is the first author (3), and the NIH consensus statement (1). The recommendations are summarised in Table 7. The routine application of the recommended strategy for follow-up of adrenal incidentalomas is likely to be costly in terms of financial expense, radiation exposure and risk of inducing fatal cancer. Furthermore, this approach carries a high risk of false-positive diagnoses and the associated emotional burden to each patient, while

---

### Table 7 Summary of current clinical recommendations (1–3).

<table>
<thead>
<tr>
<th>Publication</th>
<th>Hormonal tests</th>
<th>Frequency</th>
<th>Duration (years)</th>
<th>Imaging</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>NIH Consensus statement 2002 (1)</td>
<td>1 mg DXT, plasma-free metanephrines, K and renin/aldo in those with hypertension</td>
<td>Annual</td>
<td>4</td>
<td>Monitor those &lt;4 cm, use additional criteria in those 4–6 cm</td>
<td>Two CTs, at least 6 months apart, no data to support continued imaging if no increase in size CT at 6, 12 and 24 months</td>
</tr>
<tr>
<td>Young, NEJM 2007 (2)</td>
<td>1 mg DXT, urinary metanephrines and catecholamines, K and renin/aldo in those with hypertension</td>
<td>Annual</td>
<td>4</td>
<td>Monitor those &lt;4 cm</td>
<td>CT at 6, 12 and 24 months</td>
</tr>
<tr>
<td>UpToDate, Feb 2008 (3)</td>
<td>1 mg DXT, urinary metanephrines and catecholamines, K and renin/aldo in those with hypertension</td>
<td>Annual</td>
<td>4</td>
<td>Monitor those &lt;4 cm</td>
<td>CT at 6, 12 and 24 months</td>
</tr>
<tr>
<td>Combined strategy</td>
<td>1 mg DXT, urinary (or plasma) metanephrines and catecholamines, K and renin/aldo in those with hypertension</td>
<td>Annual</td>
<td>4</td>
<td>Monitor those &lt;4 cm</td>
<td>One to three follow-up CTs (median 3, average 2.3)</td>
</tr>
</tbody>
</table>

www.eje-online.org
only offering a small chance of a true-positive diagnosis (often of debatable clinical benefit, such as subclinical Cushing’s syndrome or an adrenal metastasis). At initial evaluation, the false-positive rate for functional or malignant lesions is typically 5 times greater than the true-positive rate, and during follow-up of adrenal incidentalomas initially considered to be benign and non-functional, the false-positive rate is typically 50 times greater than the true-positive rate (Tables 5 and 6).

The cost to the patient of repeated clinic visits both financially, and also in terms of inconvenience and psychological distress, is impossible to quantify but is likely to be significant. The financial costs to the health-care system are approximately US$ 1630 for the tests alone (based on performing 2.3 CT scans, plus annual routine biochemistry, cortisol as part of 1 mg DXT, urinary catecholamines and metanephrines, renin and aldosterone for 4 years), without allowing for additional costs such as hospital overheads, consultation time and further evaluation of positive results. There should also be consideration of the costs associated with unnecessary adrenalectomy.

The CT imaging recommended exposes the average patient to 23 mSv of ionising radiation, and so exposes the individual to a chance of inducing fatal cancer of between 1 in 430 and 1 in 2170 (based on a risk of death of between 1 in 1000 and 1 in 5000, per single abdominal CT scan of 10 mSv). This is comparable with the risk of detecting an adrenal malignancy during 3 year follow-up of between 0 and 1 in 500 (based on median incidence of 0%, mean incidence 0.2%; Table 3).

It is important to comment in more detail on the two cancers detected during follow-up in the papers included in this review. The first was a renal carcinoma metastasis in a patient with a known history of renal carcinoma (19), and since patients with known malignancy are at increased risk of developing adrenal metastases, one could reasonably argue that this study should have been excluded from this analysis. Moreover, there is no proven benefit of adrenalectomy in patients found to have adrenal metastases during the investigation of adrenal incidentalomas (1). The other malignancy detected during follow-up showed a mass enlargement after 6 months and the patient was eventually diagnosed with non-Hodgkin’s lymphoma (45).

Current practice of interval CT imaging, consisting of two to three follow-up scans, is therefore almost as likely to cause fatal cancer as it is to find cancer (half of which are likely to already be disseminated), and so on a policy level is difficult to justify. In addition, because of the low prevalence of malignancy or endocrine functionality developing during follow-up of an adrenal incidentaloma initially thought to be benign and non-functional, the performance of the available diagnostic tests is poor, with many more false positives than true positives (Table 8).

### Conclusion

The precise figures presented are clearly open to criticism, partly due to the heterogeneity of the studies that have, and have not, been included in this review. However, such debate would distract from the broad conclusions that can reasonably be drawn from the data in the literature, which can be summarised as follows:

- The prevalence of endocrine functionality in adrenal incidentalomas is low (<10%).
- The prevalence of malignancy in adrenal incidentalomas is lower (<5%).
- The majority of cases of endocrine functionality are due to subclinical Cushing’s syndrome, the clinical importance of which is debatable.
- The risk of an adrenal incidentaloma, initially thought to be benign and non-functional, subsequently developing clinically important endocrine functionality or malignancy over 2–4 years is low, in the region of <1%, but does vary depending upon the definition of, and extent of testing for, subclinical Cushing’s syndrome.
Due to the low prevalence of these conditions, especially during follow-up, and the less-than-ideal performance of the tests used to diagnose endocrine functionality and malignancy, the chances of false-positive tests are many times higher than the chance of a true-positive result.

The chance of detecting malignancy during CT follow-up of adrenal incidentalomas initially thought to be benign and non-functional is broadly similar to the chance of inducing fatal cancer from the ionising radiation delivered by these CT scans.

In the light of these data, current practice is arguably difficult to defend on a number of levels including the standards of Hippocrates’ ‘Primum non nocere’ (first do no harm). Acceptable results, but with considerably less expense, both financial and emotional due to false-positive test results, could be obtained by taking reasonable steps to exclude functionality and malignancy at presentation (such as those recommended by the NIH Consensus statement and leading reviews summarised in Table 7 (1–3)), but not embarking on either biochemical or imaging follow-up in those thought to be benign and non-functioning. For those in a ‘grey zone’ – for example 4–6 cm lesions or other suspicious imaging characteristics, a single imaging follow-up study 3–6 months later could be undertaken. This follow-up imaging could be with MRI to reduce radiation exposure, although keeping with the same imaging modality has technical advantages for interval comparison. Based on the data presented, the default position after initial evaluation should be to either excise the lesion or discharge from follow-up. The small number of patients who subsequently develop clinical signs of hormone excess or malignancy should be investigated appropriately.

Further follow-up studies of patients with adrenal incidentalomas are required to define the optimal investigation algorithm that balances the false-positive and negative rates of endocrine test protocols as well as the financial and emotional costs of investigation. We would also argue that the data linking radiation exposure with future cancer risk must also be considered when formulating a management algorithm. Based on the available data reviewed in this paper, our view is that current clinical recommendations for the assessment and management of adrenal incidentalomas have yet to find the optimal balance.

Declaration of interest

The authors declare that there is no conflict of interest that could be perceived as prejudicing the impartiality of the research reported.

Funding

This research did not receive any specific grant from any funding agency in the public, commercial or not-for-profit sector.

References

Evaluation of adrenal incidentalomas

525


Received 14 April 2009
Accepted 5 May 2009