Computed tomography criteria for discrimination of adrenal adenomas and adrenocortical carcinomas: analysis of the German ACC registry

Stephan Petersenn*, Paul-Ajoy Richter1,*, Thomas Broemel2, Christian O Ritter3, Timo Deutschbein4, Frank-Ulrich Beil1, Bruno Allolio4, Martin Fassnacht4,5 and for the German ACC Study Group

ENDOC Center for Endocrine Tumors, Altonaer Street 59, 20357 Hamburg, Germany, 1Department of Internal Medicine, University of Hamburg, Hamburg, Germany, 2Praenobis Hamburg, Hamburg, Germany, 3Department of Radiology and 4Endocrine and Diabetes Unit, Department of Internal Medicine I, University Hospital and 5Comprehensive Cancer Center Mainfranken, University of Würzburg, Würzburg, Germany

*(S Petersenn and P-A Richter contributed equally to this work)

Correspondence should be addressed to S Petersenn
Email stephan.petersenn@endoc-med.de

Abstract

Objective: Thresholds of 2–20 hounsfield units (HU) in unenhanced computed tomography (CT) are suggested to discriminate benign adrenal tumors (BATs) from malignant adrenal tumors. However, these studies included only low numbers of adrenocortical carcinomas (ACCs). This study defines a HU threshold by inclusion of a large cohort of ACCs.

Design: Retrospective, blinded, comparative analysis of CT scans from 51 patients with ACCs (30 females, median age 49 years) and 25 patients with BATs (12 females, median age 64 years) diagnosed during the period of 2005–2010 was performed.

Methods: Tumor density was evaluated in unenhanced CT by two blinded investigators.

Results: Median tumor size was 9 cm (range 2.0–20) for ACCs vs 4 cm (2.0–7.5) for BATs (P<0.0001). In ACCs, the median unenhanced HU value was 34 (range 14–74) in comparison with 5 (K13 to 40) in BATs (P<0.0001). ROC analysis revealed a HU of 21 as threshold with the best diagnostic accuracy (sensitivity 96%, specificity 80%, and AUC 0.89). However, two ACCs that were 5 and 6 cm in size would have been missed. Setting the threshold to 13.9 allowed for 100% sensitivity, but a lower specificity of 68%.

Conclusions: This first large study on ACCs confirmed that the vast majority of ACCs have unenhanced HU 21. However, to avoid misdiagnosing an ACC as benign, a threshold of 13 should be used.

Introduction

An incidentally detected adrenal lesion (incidentaloma) is a common finding during abdominal computed tomography (CT). Lesions of 1 cm or larger are reported to occur in up to 5% of patients. Most of these lesions in patients with no known malignancy will prove to be benign. Usually, conventional imaging tries to discriminate benign adrenal tumors (BATs) – mainly adrenocortical adenomas – from non-adenomas. This group mainly consists of adrenocortical carcinomas (ACCs) and adrenal metastases from other malignancies, and sometimes pheochromocytomas. A pooled analysis of multiple published studies by Boland et al. (1) suggested an attenuation threshold of 10 hounsfeld units (HU) for regular unenhanced CT to detect malignant lesions with a sensitivity of 98% and a specificity of 71%. However, all of the larger studies on this topic primarily focused on
adrenal metastases and included very few ACCs (1, 2, 3, 4, 5, 6, 7, 8, 9).

ACCs have an annual incidence of only one to two cases per million people, but are one of the most aggressive endocrine malignancies. ACC is more frequent in women than in men (1.5:1) and appears at any age, but the highest prevalence is in the fourth and fifth decades. Patients with functional tumors often present with hormonal symptoms (40–60%), in the majority with Cushing’s syndrome, while either patients with non-functional tumors suffer from symptoms of the mass effect or the tumor is, in ~15% of cases, detected incidentally (10, 11). During CT scans, ACCs are typically heterogeneous tumors with ill-defined margins, calcifications, and evidence of necrosis or hemorrhage. Most lesions are larger than 6 cm when detected (12, 13). There are only a small number of studies evaluating the radiological characteristics of ACCs. However, early detection of these rapidly growing tumors might be the only chance for cure in these patients and, therefore, reliable imaging criteria are needed to avoid missing the diagnosis of an ACC. Thus, the purpose of this study was to evaluate differences in radiological behaviors of ACCs and BATs, re-evaluating the performance of unenhanced CT scans in a large cohort by independent investigators.

Subjects and methods

Patient recruitment

The German ACC registry includes a large number of patients with proven ACC and was therefore utilized to study the radiological behavior of adrenocortical cancer in comparison with BATs. The ACC registry is approved by the Ethics Committee of the University of Würzburg. All patients provided written informed consent. This database was searched to identify patients with a localized ACC diagnosed between 2005 and 2010 with digitally available CT scans. ACC had been histopathologically proven and patients had undergone unenhanced abdominal CT imaging before any other treatment. Initial screening identified 167 patients with localized tumors, but no digital data of the CT were available for 62 patients and enhanced CT was performed only in 54 patients. Thus, finally 51 patients (21 males and 30 females) with a median age of 49 (range 17–79 years) were analyzed. The control group included 25 patients with BATs, with histological exclusion of malignancy or a follow-up of the last 6 months without any relevant tumor growth (12 males and 13 females with a median age of 64 (range 25–85 years)). This was an unbiased selection of patients with adrenal tumors recruited from the centers in Würzburg and Hamburg. The number of patients was limited by the strict entry criteria and the focus on CT, excluding patients with magnetic resonance imaging (MRI).

Image analysis

The CT images of all patients were independently analyzed using the MERLIN PACS Software (Phoenix-PACS) for morphological characteristics by two investigators blinded for the pathological diagnosis. Mismatches found were solved by consensus. The following morphological characteristics were documented: localization (left vs right adrenal gland vs bilateral disease), size (maximum axial diameter), shape (sustained (maintained) vs not sustained), margin definition (well defined vs ill defined), tumor homogeneity (homogeneous vs inhomogeneous), and unenhanced HU (Fig. 1 A and B). The region of interest (ROI) was carefully hand drawn over one-half to two-third of the mass, avoiding necrotic or hemorrhagic areas (14).

Statistical analysis

Results are expressed as median (range) unless otherwise mentioned. Data analysis was performed using GraphPad Prism 6.0 for Windows (GraphPad Software Inc., San Diego, CA, USA). The P value of <0.05 was considered statistically significant. The Mann–Whitney test and Fisher’s exact test were used, as appropriate. The thresholds for tumor size and HU values were calculated by ROC analysis, aiming for a sensitivity of at least 95% for ACCs. Inter-rater reliability was calculated by concordance correlation coefficient. Furthermore, CI for inter-rater reliability was calculated.

Results

Baseline characteristics and an overview of the main radiological findings are provided in Table 1.

Tumor size

The median diameter of ACC was 9 cm (range 2.0–20 cm). Forty-six ACCs (90%) had maximal diameters of more than 5 cm, while four (8%) had maximal diameters between 3 and 5 cm and one (2%) had a maximal diameter smaller than 3 cm. The median diameter of BATs was 4 cm (range 2.0–7.5 cm, Fig. 2A). ROC analysis aiming for at least 95% sensitivity to detect ACCs revealed a threshold of 3.9 cm (sensitivity 98%, specificity 40%, AUC 0.93, 95% CI 0.88–
One ACC would have been missed by this suggested threshold: it was a left-sided tumor with a well-defined margin, inhomogeneous appearance, and unenhanced HU of 34. Using a threshold of 5 or 6 cm resulted in an even lower sensitivity of 90 or 80%. By limiting the analysis to those tumors smaller than or equal to 6 cm (ten ACCs and 23 BATs) or 8 cm (18 ACCs and 25 BATs) in size, the threshold of 3.9 cm demonstrated a sensitivity of 90 and 94% and a specificity of 44 and 40% respectively.

Shape, margin, and homogeneity of the tumors

None of the analyzed tumors (ACCs and BATs) were sustained in shape. Twenty-eight ACCs (55%) had well-defined margins, whereas this was the case in 21 BATs (84%, $P<0.05$) leading to an odds ratio (OR) of 4.3 for ACCs with an ill-defined margin (sensitivity 45% and specificity 84%). All ACCs had an inhomogeneous appearance, while 14 of the BATs (56%) were homogeneous ($P<0.0001$). Thus, an inhomogeneous tumor structure (with a sensitivity of 100%) resulted in an OR of 130 for ACCs, but the specificity was rather low with 56%.

CT densitometry

First, inter-rater reliability was analyzed comparing results for all tumors, with a very high agreement (inter-rater reliability $P=0.96$, 95% CI 0.93–0.97). To understand the relevance of observer variation in more detail, we analyzed the mean differences in HU for unenhanced CT values between the two radiologists. As we found some association with absolute HU values, with a few outliers for high readings, we limited the analysis to the critical range of HU values between −30 and 30, resulting in a mean difference of 2.5 HU.

The median unenhanced HU value in the ROI was 34 (range 14–74 HU) in ACCs, while the median HU value in BATs was 5 (range −13 to 40 HU, Fig. 3A; $P<0.0001$). ROC analysis aiming for at least 95% sensitivity to detect ACCs revealed a threshold of 21 for unenhanced HU (sensitivity 96%, specificity 80%, AUC 0.89, 95% CI 0.79–0.98, $P<0.0001$; Fig. 3B). However, by using this threshold, two ACC patients would have been misclassified. These tumors were 5 and 6 cm in size and inhomogeneous, but with a well-defined margin and HU values of 14 and 20. Setting the threshold to 13.9 HU allowed for 100% sensitivity, but a lower specificity of 68%. Notably, applying the currently used threshold of 10 HU resulted in an even lower specificity of 58% (Table 2). Thereby, in ten out of 25 BATs, additional imaging would have been required to discriminate them from malignant tumors. On the other hand, unenhanced HU higher than 39.5 were highly suggestive of ACCs (sensitivity 31% and specificity 96%).

Again, we performed separate analyses for tumors smaller than or equal to 8 and 6 cm in size respectively, in which the specificity for the threshold of 13.9 fell as expected, but was still 68 and 65% respectively (Fig. 3C and D, Table 2). As revealed by AUC values and 95% CIs, both CT densitometry and tumor size demonstrated high accuracy for the correct diagnosis, without any significant difference between the two. However, with the emphasis on high sensitivity of at least 95%, CT densitometry clearly revealed the best accuracy.
The combination of CT densitometry (as the single best parameter) with morphological parameters such as shape, margin, and/or homogeneity resulted in a loss of accuracy. Even when combining CT densitometry with the second best parameter tumor size, accuracy dropped due to a loss of either sensitivity or specificity. In the best scenario, applying optimal thresholds of 21 HU and 5.8 cm instead of those with a sensitivity of at least 95% and requesting either of them to be positive for the diagnosis of ACCs, sensitivity increased slightly to 98%, but specificity dropped to 68%.

**Discussion**

CT densitometry has frequently been reported as an effective imaging technique to differentiate BATs from malignant lesions of the adrenal glands and a cut-off of <10 HU in unenhanced CT was proposed to exclude malignant tumors (1). However, this threshold was never investigated including a sufficiently large cohort of ACCs. In this study, we demonstrate that by using a threshold of 13 HU no ACC would have been missed.

While many recommendations on the management of adrenal incidentaloma still rely on the size of the adrenal mass (15, 16, 17), a recent guideline from the Italian Association of Clinical Endocrinologists emphasizes the importance of specific imaging criteria such as unenhanced HU (18). However, the size has certainly some predictive value, because most adenomas are smaller than 4 cm, whereas ACCs have a median size of 11 cm (19). Although the 3.9 cm cut-off of this study yields a higher sensitivity (98% for ACCs in comparison with the 5 cm cut-off), it involves a lower specificity (40%). Nevertheless, only one ACC patient would have been missed using a threshold of 4 cm, while five and ten patients would have been misdiagnosed when applying the 5 cm and the 6 cm cut-offs respectively. Although the number of patients with ENSAT stage I tumors (i.e., tumor size ≤5 cm) is only between 3 and 6% in most series

![Figure 2](http://dx.doi.org/10.1530/EJE-14-0916)
(20, 21, 22), it should be taken into consideration that the size of a tumor may depend on the time point of diagnosis during development of the disease. With more frequent use of radiological procedures, ACCs may be diagnosed at an earlier stage and with smaller size, making size a less reliable parameter to differentiate BATs and ACCs.

One of the first publications on unenhanced CT was the study by Lee et al. (6), which characterized 66 adrenal masses by three independent radiologists and classified 22 to be malignant, but did not differentiate the kind of malignancy or mention the HU for the different tumor entities. Boland et al. (1) summarized later data from ten publications detailing on only 22 ACCs of a total of 223 malignant tumors and 272 benign lesions. Subsequent studies mostly re-evaluated the suggested threshold of 10 HU in a series focusing on BATs in comparison with adrenal metastases. In addition, Hamrahian et al. (4) provided data for thresholds of 10 and 20 HU in 290 patients with adrenal masses including 15 ACCs, but focused on the comparison of adenomas/hyperplasia (156 patients) vs non-adenomas. The same applies to Ctvrtilik et al. (2), analyzing a cut-off of 23 HE. Zhang et al. (23) studied 41 ACCs, but did not determine an attenuation threshold. Table 3 provides a summary of all studies analyzing unenhanced CT that included more than three ACCs. All but one of these 64 ACCs had a HU of more than 20, but even the largest study (3) included only 17 ACCs. One of these 64 ACCs was reported as having an unenhanced HU of only 0.7, but was postoperatively carefully reviewed and described as a spongiocyte hyperplasia with a Weiss score of 5 (3).

In this study, we utilized the German ACC registry including its large series of histopathologically proven ACCs. The analysis indicated that the median attenuation of ACC is significantly higher than that of BAT. The vast majority of ACCs had unenhanced HU of more than 21, a threshold with a high sensitivity (96%) and reasonable specificity (80%) for ACCs. However, two ACCs 5 and 6 cm in diameter were missed by that threshold. Setting the threshold to 13.9 HU allowed for 100% sensitivity, but a lower specificity of 68%. The frequently used 10-HU threshold diagnosed ten out of 25 BATs (40%) as non-adenoma and would therefore lead to additional investigations. Owing to the higher incidence of adrenal adenomas compared with carcinomas, this also means that adrenal tumors with unenhanced CT values of more than 10 HU are frequently adenomas. Patients with HU values between 13.9 and 21 should be examined very carefully.

Table 2  | Threshold for HU to detect ACCs with related sensitivity, specificity, AUC, and significance level for all patients, patients with tumor sizes smaller than or equal to 8 cm, and patients with tumor sizes smaller than or equal to 6 cm respectively.

<table>
<thead>
<tr>
<th>Tumors (cm)</th>
<th>n</th>
<th>Threshold (HU)</th>
<th>Sensitivity (%)</th>
<th>Specificity (%)</th>
<th>AUC</th>
<th>Significance level</th>
</tr>
</thead>
<tbody>
<tr>
<td>All</td>
<td>76</td>
<td>21.0</td>
<td>96.1</td>
<td>80.0</td>
<td>0.89</td>
<td>P &lt; 0.0001</td>
</tr>
<tr>
<td></td>
<td></td>
<td>13.9</td>
<td>100</td>
<td>68.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>10.0</td>
<td>100</td>
<td>58.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>≤8</td>
<td>43</td>
<td>21.5</td>
<td>88.8</td>
<td>80.0</td>
<td></td>
<td>P &lt; 0.001</td>
</tr>
<tr>
<td></td>
<td></td>
<td>17.0</td>
<td>94.4</td>
<td>76.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>13.9</td>
<td>100</td>
<td>68.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>≤6</td>
<td>33</td>
<td>22.5</td>
<td>80.0</td>
<td>78.3</td>
<td></td>
<td>P &lt; 0.05</td>
</tr>
<tr>
<td></td>
<td></td>
<td>17.0</td>
<td>90.0</td>
<td>73.9</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>13.9</td>
<td>100</td>
<td>65.2</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 3  
Attenuation values of benign adrenal tumors and adrenocortical carcinomas in published studies that included at least three ACCs with HU data provided. PubMed was searched for the following keywords: incidentaloma (adrenal mass or adrenal tumor or adrenal cancer or ACC) and computed tomography (CT) and hounsfield units (HU).

<table>
<thead>
<tr>
<th>Author</th>
<th>BATs</th>
<th>ACCs</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n</td>
<td>Mean HU (± s.d.)</td>
</tr>
<tr>
<td>Chvrtlik et al. (2)</td>
<td>37</td>
<td>6 ± 13</td>
</tr>
<tr>
<td>Grousset et al. (3)</td>
<td>41</td>
<td>17</td>
</tr>
<tr>
<td>Remer et al. (7)</td>
<td>105</td>
<td>18 ± 14</td>
</tr>
<tr>
<td>Slattery et al. (34)</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Hamrahian et al. (4)</td>
<td>79</td>
<td>19</td>
</tr>
<tr>
<td>Szolar et al. (9)</td>
<td>24</td>
<td>8 ± 18</td>
</tr>
<tr>
<td>Rockall et al. (8)</td>
<td>20</td>
<td>11</td>
</tr>
</tbody>
</table>

*Threshold: 23 HU (sensitivity 89% and specificity 100%).
*Observer 1.
*Median HU.
*Threshold: 20 HU (sensitivity 58% and specificity 97%).

carefully. By limiting our analysis to tumors that were ≤6 or 8 cm in size (the group of tumors most difficult to determine the diagnosis before surgery), a threshold of 13.9 provided the highest sensitivity, whereas thresholds of ~21 HU offered a good compromise between sensitivity and specificity, again suggesting careful discussion with patients in-between and considering other criteria in addition. The mean difference between observers of 2.5 HU may be relevant to readings close to the thresholds. We therefore suggest readings by two independent observers for HU values in the critical range, as well as careful interpretation in the context of observer variation.

As already described by others, there are more criteria to differentiate incidentalomas, such as margin definition and tumor homogeneity. According to Boland et al. (13), cortical adenomas have no specific morphological features, and most are small, smooth, and heterogeneous with a well-defined margin when detected. On the other hand, the presence of ill-defined margins is often a sign of a more aggressive nature. ACCs are typically ill-defined, heterogeneous, and irregular lesions with calcifications in up to 30% of cases (5, 24, 25). In this study, 21 out of 25 BATs (84%) had a well-defined margin, and 28 out of 51 ACCs (55%) were also well defined. Zhang et al. (23) reported similar results, with 29 out of 41 ACCs (71%) being described as well defined. Furthermore, 11 out of 25 patients with BATs (44%) had an inhomogeneous tumor structure, whereas patients with ACCs were inhomogeneous in all cases. Nevertheless, the analysis of tumor margin and tumor homogeneity yielded statistically significant criteria to differentiate adrenal tumors in our series, and may therefore add to the interpretation of the CT in individual cases.

Obviously, there are alternative imaging tools such as chemical shift MRI (CS-MRI) or 2-[18F]-fluorodeoxyglucose positron emission tomography (FDG–PET) to characterize adrenal lesions. There exists some controversy about which test (CT or CS-MRI) might offer superior sensitivity and specificity for detecting lipid-rich lesions. CS-MRI might be superior when evaluating lipid-poor adenomas (26). In general, MRI is less standardized and more expensive than CT imaging, although it has the advantage of lacking radiation (14, 27, 28, 29). The most important disadvantage of CS-MRI is that, to the authors’ knowledge, there are no studies of CS-MRI criteria with a relevant number of ACCs (30). 18F-FDG–PET may offer additional information. Metabolically active malignant lesions trap 18F-FDG intracellularly, whereas most benign lesions fail to accumulate this radioisotope (3, 31, 32, 33). Thus, FDG–PET might be helpful in patients with undetermined adrenal mass by conventional imaging.

Our study has certain limitations. First, the number of analyzed BATs and ACCs is still relatively low. However, it included almost as many ACCs as the seven largest studies on adrenal imaging (describing at least three ACCs) together (Table 3). Secondly, the selection of patients for this study was not prospective or consecutive, but based on the availability of digital CT scans. We believe that this procedure is unlikely to harbor a relevant bias. Thirdly, only in a subset of 12 patients with BATs, the diagnosis was proved by histology. However, all BAT patients without surgery had at least one imaging in a time interval of at least 6 months. As virtually all ACCs grow very rapidly, the likelihood that we missed an ACC seems to be rather low. Fourthly, we did not analyze the value of delayed wash-out CT. However, this presumably very accurate method was...
performed in <10% of the patients preventing any reliable analysis. Finally, our study did not include other ‘non-adenomas’ like metastases from extra-adrenal malignancies or pheochromocytomas. However, the later are usually easy to diagnose with the measurement of plasma or urinary metanephrines and the results of imaging studies on metastases are already very strong not requiring additional data. In addition to the large number of ACCs, another strength of this study should be mentioned, namely the fact that all images were reviewed by two independent and blinded readers.

In conclusion, our study including the by far largest number of ACCs of any adrenal imaging study confirms that adrenal masses with a mean HU value below 21 are most probably of benign nature. However, to avoid misdiagnosing an ACC as a benign adenoma, a lower threshold of 13 should be used. For tumors with HU levels between 13 and 40, additional measures such as delayed wash-out CT, MRI with chemical shift and wash-out, or FDG–PET may be warranted to further exclude or suggest the malignancy of the lesion.

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
5 Korobkin M, Brodeur FJ, Yutzey GG, Francis IR, Quint LE, Dunnick NR & Kazerooni EA. Differentiation of adrenocortical adenomas from nonadenomas using CT attenuation values. AJR. American Journal of Roentgenology 1996 166 531–536. (doi:10.2214/ajr.166.3.8623622)


Received 25 October 2014
Revised 10 January 2015
Accepted 14 January 2015