Systematic Review and Meta-analysis of Diagnostic Accuracy of Percutaneous Renal Tumour Biopsy

Lorenzo Marconi, Saeed Dabestani, Thomas B. Lam, Fabian Hofmann, Fiona Stewart, John Norrie, Axel Bex, Karim Bensalah, Steven E. Canfield, Milan Hora, Markus A. Kuczyk, Axel S. Merseburger, Peter F.A. Mulders, Thomas Powles, Michael Staehler, Borje Ljungberg, Alessandro Volpe

Abstract

Context: The role of percutaneous renal tumour biopsy (RTB) remains controversial due to uncertainties regarding its diagnostic accuracy and safety.

Objective: We performed a systematic review and meta-analysis to determine the safety and accuracy of percutaneous RTB for the diagnosis of malignancy, histologic tumour subtype, and grade.

Evidence acquisition: Medline, Embase, and Cochrane Library were searched for studies providing data on diagnostic accuracy and complications of percutaneous core biopsy (CB) or fine-needle aspiration (FNA) of renal tumours. A meta-analysis was performed to obtain pooled estimates of sensitivity and specificity for diagnosis of malignancy. The Cohen kappa coefficient ($\kappa$) was estimated for the analysis of histotype/grade concordance between diagnosis on RTB and surgical specimen. Risk of bias assessment was performed (QUADAS-2).

Evidence synthesis: A total of 57 studies recruiting 5228 patients were included. The overall median diagnostic rate of RTB was 92%. The sensitivity and specificity of diagnostic CBs and FNAs were 99.1% and 99.7%, and 93.2% and 89.8%, respectively. A good ($\kappa = 0.683$) and a fair ($\kappa = 0.34$) agreement were observed between histologic subtype and Fuhrman grade on RTB and surgical specimen, respectively. A very low rate of Clavien $\geq 2$ complications was reported. Study limitations included selection and differential-verification bias.

Conclusions: RTB is safe and has a high diagnostic yield in experienced centres. Both CB and FNA have good accuracy for the diagnosis of malignancy and histologic subtype, with better performance for CB. The accuracy for Fuhrman grade is fair. Overall, the
Introduction

The management of renal tumours has evolved, with the increasing use of nonextirpative therapies for small renal masses (SRMs) in selected patients and the advent of effective targeted drugs for metastatic disease [1]. This has led to an increasing recognition of the importance of histologic characterisation of renal masses before treatment to tailor therapy based on tumour histology either in the localised or metastatic setting [2].

Percutaneous renal tumour biopsy (RTB) has been criticised due to concerns regarding its safety, diagnostic accuracy, and ability to distinguish tumour histologic subtypes and nuclear grade. Although fine-needle aspiration (FNA) and core biopsy (CB) have been used to sample renal tumours, the best technique has not been clearly defined [3]. Several recent studies have reported low complication rates and good diagnostic performance of RTB, but most studies were limited by small sample sizes, heterogeneous populations, different biopsy techniques, and lack of standardised definitions for diagnostic accuracy [4].

We performed a systematic review of the literature and meta-analysis to determine the diagnostic performance and safety of RTB in characterising malignancy, histologic subtype, and grade of renal tumours.

Selection of studies

Prospective or retrospective cohort studies providing data on accuracy for malignancy, tumour histotype and grade, and/or on complications of percutaneous CB or FNA of solid or cystic renal masses of any size in adult patients were included. Studies that fulfilled the following criteria were included for the evaluation of diagnostic accuracy for malignancy: (1) reference standard for tumour malignancy represented by pathology on surgical specimen of partial or radical nephrectomy performed after RTB, or clinical and radiologic follow-up of at least 12 mo showing presence or absence of tumour progression and/or onset of tumour-related symptoms; (2) availability of number of nondiagnostic biopsies; and (3) availability of number of diagnostic biopsies classified as true positives (TPs), false positives (FPs), false negatives (FNs), and true negatives (TNs) either as group totals or by case-by-case enumeration of diagnoses. Studies that did not provide data on all four elements of diagnostic accuracy were excluded.

Studies that provided data to assess concordance between tumour grade and/or histologic subtype between RTB and surgical pathology were included for the assessment of diagnostic accuracy for histologic subtype and/or grade.

Studies exclusively reporting complications of RTBs were also included. Complications were graded according to the Clavien-Dindo classification [8]. Studies on laparoscopic-assisted or ex vivo RTBs were excluded.

Data extraction

A data extraction form was developed a priori to collect information on study design, patient characteristics (age, gender, indication for RTB, comorbidities), tumour features (size, solid or cystic pattern), RTB characteristics (needle size, image guidance, number of cores, biopsy technique), reference standard (surgery performed, follow-up length and protocol), and outcome measures (accuracy and complications).

Quality assessment

Risk of bias (QUADAS-2 tool [9]) was assessed for studies included in the diagnostic accuracy meta-analysis and in the analysis of accuracy for tumour histotype and grade.
2.5. **Data analysis and statistical methods**

The rate of diagnostic biopsies (diagnostic yield) and nondiagnostic biopsies was assessed in all studies. When nondiagnostic cases were not reported by authors, the following definitions were used to define a nondiagnostic result: *normal renal tissue, extrarenal tissue, blood or necrosis only, inflammatory or fibrotic tissue only, insufficient material, or nonadequate tissue*. 

The accuracy for diagnosing malignancy was assessed on diagnostic RTBs. For each study, we built a $2 \times 2$ contingency table consisting of TP, FP, FN, and TN based on concordance between biopsy result and surgical pathology or clinical/radiologic follow-up. Meta-analysis was performed where appropriate on studies demonstrating homogeneity of population, outcome definition, and methods and timing of outcome measurement. The joint estimates of sensitivity and specificity and their 95% confidence intervals (CIs) were modelled using the *metandi* command in Stata v.13.1 (StataCorp, College Station, TX, USA). The summary receiver operating characteristic curve was plotted from this procedure. The *xmlreglogit* default was used, which fits a multilevel mixed-effects logistic regression using quadrature convergence. The *glmfit* (generalised linear latent and mixed model) formulation with spherical quadrature was used in rare instances with numerical convergence issues. The pooled estimates for sensitivity and specificity were based on bivariate analysis. For univariate analysis, forest plots for both sensitivity and specificity were generated using the *metan* command.

Sensitivity analysis was performed for studies with low risk of selection and flow bias and for studies reporting exclusively on FNA, CB, SRMs (<4 cm), or cystic masses. The Egger test was used to identify publication bias [10], and funnel plots were generated. The influence of potential publication bias was assessed using the “trim-and-fill” method [11].

Regarding the analysis of complications and accuracy for histologic subtype and grade, a narrative synthesis was provided using descriptive statistics. The Cohen kappa coefficient ($\kappa$) was calculated to define agreement between histotype and grade on biopsy and surgical specimen. To calculate agreement for histotype, a $7 \times 7$ contingency table was constructed with the following histologic subtypes: clear cell RCC, chromophobe RCC, papillary RCC, other malignant tumour, oncocytoma, angiomyolipoma, and other benign tumour. Regarding agreement for grade, a $4 \times 4$ contingency table was constructed with the four grade categories. The strength of agreement was considered poor for $\kappa < 0.2$, fair for $\kappa = 0.21–0.40$, moderate for $\kappa = 0.41–0.60$, good for $\kappa = 0.61–0.80$, and very good for $\kappa > 0.81$.

### 3. Evidence synthesis

#### 3.1. Quantity of evidence identified

A total of 1500 articles were identified by the literature search. Of these, 153 articles were selected for full-text screening and 57 (recruiting 5228 patients) were eligible for inclusion (Fig. 1). Thirty-three studies recruiting 2867 patients [12–44] were included in the meta-analysis of diagnostic accuracy for malignancy of FNA and CB. Nineteen studies were included in the analysis of accuracy for histotype and grade and 37 studies in the analysis of complications.

**Table 1** lists the main characteristics of the biopsied tumours in each study and technical details about the RTBs (needle size and type, image guidance, number of cores).

#### 3.2. Risk of bias and quality assessment

**Figure 2** shows the risk of bias assessment for studies included in the analyses of diagnostic accuracy for tumour malignancy, histotype, and grade. Overall, there was a high risk of bias across studies. Only five studies included in the meta-analysis were prospective [15,16,27,36,37]; the majority were retrospective case series including studies reporting all RTBs performed at a single institution and studies reporting only RTBs performed in patients who ultimately underwent surgery (ie, selection bias). There was only one prospective not fully paired comparative study of CB versus FNA [36]. The clinical setting of RTBs was heterogeneous within studies and across studies. There were concerns about differential-verification bias because pathology of the surgical specimen was available in all patients as a reference standard only in 11 of 33 studies included in the meta-analysis. Of these, only three high-quality papers reported on consecutive patients [15,36,37]. In the other 17 studies, pathology was available in a proportion of cases (median: 56%), and clinical and radiologic follow-up was used as reference standard for the remaining patients. Finally, the proportion of patients who had surgical pathology as the reference standard was not reported in five studies. A modest evidence of publication bias was found in the FNA studies (data not shown).

#### 3.3. Meta-analysis of accuracy of renal tumour biopsies for diagnosis of malignancy

The overall median rate of diagnostic RTBs was 92% (interquartile range [IQR]: 80.6–96.8%). **Figure 3** shows the forest plot of the 17 studies on CBs [14,16,18,20,21,26,29,30,33–38,40,42–44]. The rate of nondiagnostic biopsies was 0–22.6%. The estimates for sensitivity and specificity of diagnostic CBs based on bivariate analysis were 99.1% (95% CI, 96.4–99.8) and 99.7% (95% CI, 93.7–100), respectively. **Figure 4** shows the forest plot of the 18 studies on FNA [12,13,15,17,19,22–25,27,28,31,32,34,36,39,41,42]. The rate of nondiagnostic biopsies was 0–36%. The estimated sensitivity and specificity of diagnostic FNAs based on bivariate analysis were 93.2% (95% CI, 83–97.5) and 89.8% (95% CI, 78.6–95.4), respectively.

The sensitivity analysis for studies reporting RTBs of SRMs only ($n = 7$) [15,16,27,30,33,40,43] showed a sensitivity of 99.7% (95% CI, 81.5–100) and a specificity of 98.2% (95% CI, 83.3–99.8) (Fig. 5). The sensitivity analysis of
studies reporting RTBs of cystic masses only \((n = 4)\) \([13,21,25,41]\) showed a sensitivity of 83.6\% (95\% CI, 33.8–98.1) and a specificity of 98\% (95\% CI, 80.9–99.8) (Fig. 5). The estimates of sensitivity and specificity of studies with a low risk of selection and flow bias were 92.9\% (95\% CI, 79.6–97.8) and 84.3\% (95\% CI, 69.2–92.8).

### 3.4. Analysis of accuracy of renal tumour biopsies for tumour histotype

Five studies allowed the analysis of the agreement between tumour histotype on biopsy and surgical specimen with \(\kappa\) values \([14,28,37,45,46]\). Median \(\kappa\) value was 0.683 (IQR: 0.52–0.95), indicating a good degree of agreement.

Overall, 14 studies reported the concordance of tumour histotype between RTB and surgical pathology \([12,17,26,28,30,33,36,37,40,43,45–48]\) (Table 1). The median concordance rate was 90.3\% (IQR: 84–94.4\%). One study compared the concordance for histologic subtype of CB and FNA with final pathology, showing no significant difference (91\% vs 86\%, respectively; \(p = 0.45\)) \([36]\). The median concordance rate for the diagnosis of histotype in the six studies including SRMs only was 96\% (IQR: 90–100\%). Tumour subtype was not reported in studies including cystic masses only.

### 3.5. Analysis of accuracy of renal tumour biopsies for tumour grade

Seven studies allowed the analysis of the agreement between tumour grade on biopsy and surgical specimen with \(\kappa\) values \([14,15,24,26,37,45,49]\). Median \(\kappa\) value was 0.34 (0.13–0.52), which indicates a fair degree of agreement. In 17 studies, the authors reported the concordance rate between grade on RTB and surgical specimen \([12,14,15,24,26,30,33,36,37,40,43,45,47–51]\). Ten studies used the four-tier Fuhrman system, three studies used a two- or three-tier grading system (high vs low, or high vs...
Table 1 – Tumor characteristics, biopsy technical features and diagnostic results of the studies included in the analysis

<table>
<thead>
<tr>
<th>Study</th>
<th>n</th>
<th>Time period</th>
<th>Tumour characteristics</th>
<th>Technical characteristics</th>
<th>Diagnostic biopsies</th>
<th>Nondiagnostic biopsies, % (n)</th>
<th>Grade concordance</th>
<th>Histologic subtype concordance, % (n/n)</th>
<th>Complications, %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abel et al [12]</td>
<td>166</td>
<td>1991–2007</td>
<td>Solid or cystic, %</td>
<td>FNA or CB</td>
<td>22</td>
<td>9.2 (3.0–32.0)</td>
<td>100</td>
<td>Yes</td>
<td>5.4 (9)</td>
</tr>
<tr>
<td>Al Nazer and Mourad [49]</td>
<td>18</td>
<td>1990–1998</td>
<td>NR</td>
<td>FNA</td>
<td>20</td>
<td>20</td>
<td>100</td>
<td>No</td>
<td>7.7 (14/18) k = 0.65</td>
</tr>
<tr>
<td>Bishop et al [50]</td>
<td>33</td>
<td>1990–2010</td>
<td>FNA or CB</td>
<td>18</td>
<td>US or CT</td>
<td>9.2</td>
<td>100</td>
<td>No</td>
<td>NR</td>
</tr>
<tr>
<td>Blumenfeld et al [14]</td>
<td>18</td>
<td>1990–1998</td>
<td>NR</td>
<td>FNA</td>
<td>22</td>
<td>5.3</td>
<td>100</td>
<td>Yes</td>
<td>36 (9) (74/109)</td>
</tr>
<tr>
<td>Campbell et al [47]</td>
<td>25</td>
<td>1994–1996</td>
<td>Solid, 100</td>
<td>FNA</td>
<td>22</td>
<td>2.1</td>
<td>100</td>
<td>NA</td>
<td>NR</td>
</tr>
<tr>
<td>Chyhrai et al [16]</td>
<td>25</td>
<td>2004–2006</td>
<td>Solid, 100</td>
<td>CB</td>
<td>18</td>
<td>2.5</td>
<td>100</td>
<td>Yes</td>
<td>8 (2) (88/109)</td>
</tr>
<tr>
<td>Garcia-Solano et al [19]</td>
<td>31</td>
<td>2000–2006</td>
<td>Solid, 100</td>
<td>FNA</td>
<td>25</td>
<td>1.2</td>
<td>100</td>
<td>NR</td>
<td>NR</td>
</tr>
<tr>
<td>Halverson et al [47]</td>
<td>151</td>
<td>1999–2011</td>
<td>Solid, 100</td>
<td>FNA</td>
<td>22</td>
<td>1.5</td>
<td>100</td>
<td>NR</td>
<td>80 (8/10)</td>
</tr>
<tr>
<td>Hara et al [20]</td>
<td>33</td>
<td>1994–1999</td>
<td>Solid, 79; cystic, 21</td>
<td>CB</td>
<td>18</td>
<td>1.3</td>
<td>100</td>
<td>No</td>
<td>NR</td>
</tr>
<tr>
<td>Haubek et al [22]</td>
<td>169</td>
<td>1982–1988</td>
<td>Solid, % NR; cystic, %</td>
<td>FNA</td>
<td>23</td>
<td>1.5</td>
<td>100</td>
<td>5.3</td>
<td>NR</td>
</tr>
<tr>
<td>Juul et al [23]</td>
<td>301</td>
<td>1971–1983</td>
<td>Solid, 100</td>
<td>FNA</td>
<td>23</td>
<td>8.5</td>
<td>100</td>
<td>3.2</td>
<td>NR</td>
</tr>
<tr>
<td>Kelley et al [24]</td>
<td>43</td>
<td>1989–1993</td>
<td>Solid, 100</td>
<td>FNA</td>
<td>26</td>
<td>1.5</td>
<td>100</td>
<td>58 (7/12) k = 0.091</td>
<td>NR</td>
</tr>
<tr>
<td>Lebret et al [26]</td>
<td>119</td>
<td>1999–2005</td>
<td>Solid, 100</td>
<td>CB</td>
<td>18</td>
<td>0.9</td>
<td>100</td>
<td>10</td>
<td>NR</td>
</tr>
<tr>
<td>Li et al [27]</td>
<td>35</td>
<td>NR</td>
<td>Solid, 100</td>
<td>CB</td>
<td>18</td>
<td>1.5</td>
<td>100</td>
<td>1.9</td>
<td>NR</td>
</tr>
<tr>
<td>Masson et al [28]</td>
<td>87</td>
<td>NR</td>
<td>Solid, 100</td>
<td>CB</td>
<td>18</td>
<td>1.5</td>
<td>100</td>
<td>1.2</td>
<td>NR</td>
</tr>
<tr>
<td>Maturin et al [29]</td>
<td>152</td>
<td>1999–2005</td>
<td>Solid, 89; cystic, 11</td>
<td>CB</td>
<td>18</td>
<td>1.5</td>
<td>100</td>
<td>3.9</td>
<td>NR</td>
</tr>
<tr>
<td>Menouge et al [30]</td>
<td>250</td>
<td>1999–2009</td>
<td>Solid, 100</td>
<td>CB</td>
<td>18</td>
<td>1.5</td>
<td>100</td>
<td>20</td>
<td>NR</td>
</tr>
<tr>
<td>Mignon et al [31]</td>
<td>67</td>
<td>1985–1995</td>
<td>Solid, 100</td>
<td>CB</td>
<td>17</td>
<td>1.5</td>
<td>100</td>
<td>75 (46/61) k = 0.52</td>
<td>NR</td>
</tr>
<tr>
<td>Millet et al [45]</td>
<td>61</td>
<td>2006–2011</td>
<td>Solid, 100</td>
<td>CB</td>
<td>17</td>
<td>1.5</td>
<td>100</td>
<td>87.5 (21/24) k = 1</td>
<td>NR</td>
</tr>
<tr>
<td>Mondal [51]</td>
<td>24</td>
<td>2005–2009</td>
<td>Solid, 100</td>
<td>CB</td>
<td>18</td>
<td>1.5</td>
<td>100</td>
<td>87.5 (21/24) k = 1</td>
<td>NR</td>
</tr>
<tr>
<td>Reference</td>
<td>Year Range</td>
<td>Technique(s)</td>
<td>Size</td>
<td>Grade</td>
<td>Palpation-guided or US</td>
<td>穿刺术</td>
<td>CT</td>
<td>US</td>
<td>Percutaneous</td>
</tr>
<tr>
<td>-----------</td>
<td>------------</td>
<td>--------------</td>
<td>------</td>
<td>------</td>
<td>----------------------</td>
<td>-------</td>
<td>----</td>
<td>----</td>
<td>----------------</td>
</tr>
<tr>
<td>Mondal and Ghosh [32]</td>
<td>1984–1990</td>
<td>Solid, 100</td>
<td>FNA</td>
<td>21</td>
<td>NR</td>
<td>100</td>
<td>Yes</td>
<td>4.3 (4)</td>
<td>NR</td>
</tr>
<tr>
<td>Nicodero and Coklin [34]</td>
<td>1987–1991</td>
<td>NR</td>
<td>FNA</td>
<td>22</td>
<td>1–3</td>
<td>CT</td>
<td>26</td>
<td>Yes</td>
<td>NC</td>
</tr>
<tr>
<td>Renshaw et al [46]</td>
<td>1987–1995</td>
<td>NR</td>
<td>FNA</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
<td>100</td>
<td>No</td>
<td>–</td>
</tr>
<tr>
<td>Rybicki et al [35]</td>
<td>1990–2001</td>
<td>Solid, 86; cystic, 14</td>
<td>Solid, 100</td>
<td>CB vs FNA</td>
<td>18</td>
<td>2–3</td>
<td>CT</td>
<td>100</td>
<td>Yes</td>
</tr>
<tr>
<td>Schmidbauer et al [36]</td>
<td>2005–2007</td>
<td>4 (1.8) 3.9 [0.8–9]</td>
<td>Solid, 100</td>
<td>CB vs FNA</td>
<td>21</td>
<td>1–3</td>
<td>US</td>
<td>61</td>
<td>Yes</td>
</tr>
<tr>
<td>Solferini et al [37]</td>
<td>2001–2008</td>
<td>Solid, 3; cystic, 5 NR</td>
<td>CB</td>
<td>18</td>
<td>2</td>
<td>US</td>
<td>100</td>
<td>Yes</td>
<td>4.8 (2)</td>
</tr>
<tr>
<td>Somani et al [38]</td>
<td>1996–2006</td>
<td>Solid, 3; cystic, 5 NR</td>
<td>CB</td>
<td>16–18</td>
<td>2–4</td>
<td>US or CT</td>
<td>56</td>
<td>Yes</td>
<td>12.9 (9)</td>
</tr>
<tr>
<td>Strojan et al [39]</td>
<td>79 NR</td>
<td>28 (2.18) 3.4 [1–13]</td>
<td>Solid, 3; cystic, 5 NR</td>
<td>FNA</td>
<td>22</td>
<td>NR</td>
<td>US</td>
<td>100</td>
<td>Yes</td>
</tr>
<tr>
<td>Thullier et al [40]</td>
<td>1998–2006</td>
<td>Solid, 100</td>
<td>CB</td>
<td>18</td>
<td>NR</td>
<td>US or CT</td>
<td>56</td>
<td>Yes</td>
<td>22.6 (12)</td>
</tr>
<tr>
<td>Tod et al [41]</td>
<td>1988–1996</td>
<td>Cystic, 100</td>
<td>FNA</td>
<td>18, 22</td>
<td>1–4</td>
<td>US or CT</td>
<td>37</td>
<td>Yes</td>
<td>NC</td>
</tr>
<tr>
<td>Volpe et al [43]</td>
<td>1999–2007</td>
<td>Solid, 91; cystic, 9</td>
<td>CB</td>
<td>18</td>
<td>≥2</td>
<td>US or CT</td>
<td>24</td>
<td>Yes</td>
<td>10 (16)</td>
</tr>
<tr>
<td>Walton et al [48]</td>
<td>1999–2009</td>
<td>Solid, 3; cystic, 5 NR</td>
<td>CB</td>
<td>15–18</td>
<td>1–6</td>
<td>US or CT</td>
<td>42</td>
<td>No</td>
<td>–</td>
</tr>
<tr>
<td>Aribas et al [56]</td>
<td>1993–2010</td>
<td>Solid, 100</td>
<td>CB</td>
<td>18</td>
<td>1–3</td>
<td>US</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Elder et al [54]</td>
<td>25–31</td>
<td>Solid, 100</td>
<td>CB</td>
<td>18</td>
<td>NR</td>
<td>US or CT</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Iizumi et al [59]</td>
<td>1989–2009</td>
<td>8.3</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
<td>CT or US</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Johnson et al [60]</td>
<td>1993–1998</td>
<td>Solid, 100</td>
<td>FNA with or without CB</td>
<td>18, 20</td>
<td>NR</td>
<td>US</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Kroese et al [57]</td>
<td>2009–2010</td>
<td>Solid, cystic, 30.7</td>
<td>CB</td>
<td>18</td>
<td>1–4</td>
<td>CT</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Lemn [62]</td>
<td>1988–2003</td>
<td>2.5 [0.6–4]</td>
<td>Solid, 3; cystic, 5 NR</td>
<td>CB</td>
<td>18</td>
<td>US or CT</td>
<td>Both</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Study</td>
<td>n</td>
<td>Time period</td>
<td>Tumour characteristics</td>
<td>Technical characteristics</td>
<td>Diagnostic biopsies</td>
<td>Nondiagnostic biopsies, % (n)</td>
<td>Grade concordance</td>
<td>Histologic subtype concordance, % (n)</td>
<td>Complications, %</td>
</tr>
<tr>
<td>------------------</td>
<td>----</td>
<td>-------------</td>
<td>------------------------</td>
<td>---------------------------</td>
<td>---------------------</td>
<td>---------------------------------</td>
<td>-------------------</td>
<td>----------------------------------------</td>
<td>------------------</td>
</tr>
<tr>
<td>Li et al [58]</td>
<td>90</td>
<td>2004–2009</td>
<td>Solid, 100</td>
<td>FNA, CB</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NR</td>
</tr>
<tr>
<td>Nadel et al [64]</td>
<td>30</td>
<td>NR</td>
<td>4.6 [1–11]</td>
<td>FNA, 18, 1–3 CT</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NR</td>
</tr>
<tr>
<td>L. et al [68]</td>
<td>59</td>
<td>2004–2011</td>
<td>Solid, 100</td>
<td>CB, 18, 1–6 US</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NR</td>
</tr>
<tr>
<td>Pilotti et al [65]</td>
<td>132</td>
<td>NR</td>
<td>Solid, 100</td>
<td>FNA, 22, NR CT Fluoroscopy</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NR</td>
</tr>
<tr>
<td>Richter et al [55]</td>
<td>517</td>
<td>NR</td>
<td>Cystic, 60, solid, 40</td>
<td>FNA, CB, 18, (20–22) NR CT Fluoroscopy US or CT</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>No comp.</td>
<td></td>
</tr>
<tr>
<td>Tikkaakoski et al [66]</td>
<td>180</td>
<td>NR</td>
<td>Cystic, 100</td>
<td>FNA, &gt;20, NR US</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>Seeding: 0.6; pain: 0.6; haematuria: 0.6</td>
<td></td>
</tr>
<tr>
<td>Veltri et al [67]</td>
<td>145</td>
<td>NR</td>
<td>Solid, 100</td>
<td>FNA with or without CB, 18, 21,22 NR US or CT</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>Haematuria: 4; haematuria: 0.7; AV fistula: 0.7</td>
<td></td>
</tr>
</tbody>
</table>

AV = arteriovenous; CB = core biopsy; CT = computed tomography; FNA = fine-needle aspiration; k = Cohen kappa coefficient; MRI = magnetic resonance imaging; NA = not applicable; No comp. = no complications were observed; NR = not reported; SD = standard deviation; US = ultrasound.

Study included only in the analysis of complications of RTB.
intermediate vs low), and four studies used both the four-tier and two-tier grading system (Table 1). Most of the patients in these series (72% [range: 44.8–82%]) had low-grade RCC (Fuhrman I–II). Overall, the median concordance rate between grading on biopsy and surgical specimen was 62.5% (IQR: 52.1–72.1%) and 87% (IQR: 71–98%) using the four-tier and the two-tier Fuhrman grading system, respectively. One study compared the concordance of CB and FNA with final pathology, showing a higher concordance rate for CB (76% vs 28%; \( p < 0.05 \)) [36]. The six studies on FNA [12,15,24,49–51] reported concordance rates of 31.7–87.5% and 58–100% using the four-tier and the two-tier Fuhrman grading system, respectively. The 10 studies on CBs reported concordance rates of 43–93% [14,26,30,33,40,43,45,47,48]. The median concordance rate for tumour grade between RTB and surgical specimen in studies including SRM only \((n = 7)\) [15,30,33,40,43,45,47] was 66.7% (IQR: 60–69.8%) but increased to 86.5% (range: 80–93%) when a two-tier grading system was used. Accuracy for grading was not reported in studies including cystic masses only.

### 3.6. Analysis of complications of renal tumour biopsies

Ten [17–19,26,37,42,52–55] and five [23,33,56–58] studies reported the absence of any complications and major complications after RTB, respectively. In 22 studies, at least one complication was reported. The median overall complication rate across these studies was 8.1% (IQR: 2.7–11.1%). Only three Clavien grade ≥2 complications were reported.

Protocol-mandated computed tomography (CT) or ultrasound (US) imaging was routinely performed to assess the presence of postprocedural haematomas in 9 of the 37 series included in the analysis of complications.

Perirenal haematomas were observed in 18 studies. In 16 studies, Clavien 1 haematomas were reported in a median of 4.3% (IQR: 2.7–7.8%) of cases. Haematomas requiring blood transfusion (Clavien 2) occurred in three studies in a median of 0.7% of cases. Self-limiting haematuria (Clavien 1) was reported in 12 studies in a median of 3.15% (IQR: 1.1–8.6%) of patients [12,20,25,31,59–67]. One case of gross haematuria requiring admission for clot urinary retention and one of pseudoaneurysm treated with endovascular embolisation (Clavien 3a) were reported [29,63].

Lumbar pain after the procedure was reported in eight studies [16,20,30,59,60,64,66,68] with a median incidence of 3% (IQR: 1–4.8%). One case of septic shock after RTB in a pyelonephritic kidney was reported [64]. All other studies did not observe any inflammatory complications after RTB. Two studies reported one case of simple pneumothorax (Clavien 1) [36,63].

Only one case of seeding of a transitional cell carcinoma was reported in the studies in the analysis [66]. The presence of seeding was generally determined by clinical and radiologic follow-up. However, no evidence of seeding was found when the peritumoural and perirenal fat were routinely histologically assessed after tumour surgical removal [26,33].

### 3.7. Discussion

To our knowledge, this is the first systematic review including a meta-analysis of studies on the diagnostic performance of percutaneous RTB. The meta-analysis showed a high overall diagnostic rate for the procedure
(92%), with higher estimates for sensitivity and specificity for CB compared with FNA. The accuracy of RTB for the diagnosis of RCC histologic subtype was found to be good. A fair agreement between tumour grade at biopsy and on the final specimen was observed.

The use of percutaneous sampling of renal tumours historically was limited due to concerns about its safety, diagnostic yield, and accuracy, and for the perceived little impact of RTBs on clinical management [69]. However, the adoption of modern biopsy techniques and the growing expertise in performing biopsies, the progressively increased experience of pathologists in interpreting biopsy specimens, and the increased confidence of urologists in using biopsy results to support treatment decisions based on a better knowledge of the natural history of benign and malignant renal tumours have led to increasing indications of this procedure for the histologic characterisation of SRMs and metastatic primary renal tumours [3,70,71]. The current EAU urologic guidelines on RCC recommend that RTBs should be performed in patients in whom active surveillance is pursued and before ablative therapy and systemic therapy without previous pathology [1]. However, despite the wider indications and the encouraging diagnostic performance reported in experienced centres, the use of RTBs still remains limited outside academic centres and institutions with a special focus on urologic oncology [72,73].

Improving the quality of the evidence on RTB is crucial to better define the role of this procedure in the management of renal tumours. The current evidence base in this field is in fact limited by several factors. First, most studies on RTBs are retrospective, have relatively small sample sizes, and have heterogeneous populations. Second, the assessment of diagnostic accuracy of RTBs is hampered by the lack of surgical confirmation of the histology in a variable proportion of cases, by the use of different follow-up protocols to monitor the clinical behaviour of tumours that are not surgically removed after biopsy, by the adoption of
different definitions for biopsy success, and by the use of different biopsy techniques and protocols (CB vs FNA, CT vs US guidance, number and location of biopsies taken).

In the absence of large prospective multicentre studies using homogeneous biopsy techniques and standardised protocols for assessment of diagnostic outcomes, the present paper provides the best available evidence on diagnostic yield and accuracy of RTBs. The clinical question we assessed in the review was prioritised by an expert panel of clinicians (EAU RCC Guideline Panel), and the patient problem or population, intervention, comparison, and outcomes (PICO) elements were developed in conjunction with the panel. The search strategy and entire review process adhered to PRISMA guidelines and Cochrane review on diagnostic test accuracy principles. The strict methodological criteria used also ensured the inclusion in the meta-analysis of studies with lower risks of bias and a low level of clinical and methodological heterogeneity.

Our results show that percutaneous sampling of renal tumours harbours a diagnostic yield of 92% for the diagnosis of malignancy, indicating that a properly performed RTB can provide information for treatment decision making in most cases. However, nondiagnostic biopsies constituted a variable but non-negligible proportion of cases either for CB and FNA. This represents a matter of concern for clinicians; hence surgical exploration or repeat sampling is recommended when the biopsy of a radiologically suspicious renal mass is nondiagnostic [1]. Repeat biopsies have been reported to be diagnostic in a high proportion of cases (83–100%) in several series [38,44,63,74,75].

It should be acknowledged that a relevant proportion of nondiagnostic RTBs are in fact failed biopsies due to technical limitations (use of suboptimal technique, needle type, or image guidance) and/or to the intrinsic challenge of the procedure (ie, targeting a mass in an organ that moves with respiration).

The frequent inappropriate allocation of failed biopsies containing only normal renal parenchyma, blood, or fibrosis in the category of inaccurate biopsies has led to a potential underestimation of biopsy accuracy.

A fair evaluation of the diagnostic accuracy of informative RTBs is crucial for the definition of their utility, safety, and reliability in clinical practice when histologic information is used to make clinical decisions, such as supporting the choice of observation/active surveillance in the presence of a benign histology or a low-grade RCC.

In our meta-analysis, CB was shown to have excellent estimates for sensitivity (99.1%) and specificity (99.7%) in the assessment of tumour malignancy. Lower estimates for sensitivity (93.2%) and specificity (89.8%) for the diagnosis of malignancy were observed for FNA. Based on this data, CB may therefore be favoured over FNA for percutaneous sampling of renal tumours. However, because some authors suggest that the two techniques can provide complementary results and eventually increase diagnostic rates and accuracy [76], FNA may be performed in combination with...
CB in selected cases. A potential advantage of FNA is that it allows the intraprocedural assessment of the cytologic specimen, which can potentially confirm the appropriate location of the guiding cannula and increase the diagnostic yield of the subsequent CBs.

The subgroup analyses we performed showed excellent accuracy of RTB for diagnosing malignancy in SRMs, whereas the estimates for specificity and above all sensitivity of RTBs of cystic renal masses are significantly poorer compared with RTBs of solid masses (98% vs 83.6%, respectively). These findings, together with the potential risk of spreading of tumour cells resulting from cystic rupture during biopsy, support the current trend to limit the indications of RTBs for cystic renal lesions. Percutaneous sampling may still be indicated for Bosniak grade IV cysts, where clear enhancing solid nodules are visible within the lesion [1].

Our analysis also revealed a good degree of concordance between the diagnosis of RCC subtype on biopsy and on surgical specimens (90.3% in the overall population, improving to 96% in the analysis for SRMs; median $\kappa = 0.63$). The diagnosis of RCC subtype on RTBs is therefore reliable in most cases and can be safely used for treatment decision making. In fact, although an independent prognostic role for RCC histotype has not been clearly established, consensus of RCC subtype is useful for tailoring the best targeted therapy in systemic disease.

The assessment of tumour grade on RTBs is challenging [33,36,43,63]. According to our analysis, the degree of concordance of tumour grade on RTBs and surgical specimens is only fair (median $\kappa = 0.34$). The median reported concordance rate is 62.5%, improving to 87% when a simplified two-tier system (Fuhrman I–II = low grade; Fuhrman III–IV = high grade) is adopted. When only studies on SRMs are included in the analysis, the concordance rate using the four-tier system is slightly improved (66.7%); the concordance is similar using the two-tier system (86.5%).

Accuracy in the evaluation of tumour grade is important for clinical decision making but is limited by tumour heterogeneity and interobserver variability. Intratumoural grade heterogeneity has been reported in 5–25% of renal tumours [36,69]. Its potential impact on biopsy accuracy can be reduced by performing multiple biopsies in different areas of the tumour. Although the accuracy of RTBs for tumour grade is not optimal, Jeldres et al observed that models including other patient and tumour characteristics cannot reliably predict Fuhrman grade and therefore cannot substitute percutaneous biopsy for grade assessment [79].

Improving our ability to obtain samples that allow a reliable and accurate evaluation of tumour grade is clearly one of the main future goals of clinical research on RTBs. The classification of grading of renal tumours is evolving, and the recent International Society of Urological Pathology 2012 Consensus Conference accepted a new grading system with grades 1–3 of clear cell and papillary RCC based on nucleolar prominence, and grade 4 defined by the presence of extreme nuclear pleomorphism or sarcomatoid and/or rhabdoid differentiation [80]. Future studies will have to assess the accuracy of RTBs for this newly proposed grading system.

Finally, the present study indicates that percutaneous RTB is a safe procedure, with a very limited risk of significant (Clavien $\geq 2$) complications. Only one case of seeding was reported in the eligible studies, also when peritumoural and perirenal fat were assessed after surgical removal. Although another case of seeding was recently published [81], most of the few case reports date back to the late 1980s to early 1990s when different biopsy techniques were used [3]. Some authors suggest that the risk of this worrisome complication is minimal with the use of the coaxial technique [4], which should always be used in clinical practice. Other complications of RTB were mainly lumbar pain and haematomas/haematuria, which mostly resolved spontaneously without medical intervention. Although protocol-mandated CT or US imaging was not performed after biopsy in all included studies, the incidence of haematomas was relatively low (median: 5%), and blood transfusions were required on average in only 0.7% of cases.

This study has several limitations. Included studies may potentially be affected by selection bias and by the use of different reference standards (differential-verification bias) including variable clinical follow-up schedules for renal tumours that were not surgically removed. In spite of our efforts to standardise outcome definitions and measurements, clinical and methodological heterogeneity was inevitable, although it was minimised. The threshold for good quality studies for sensitivity analysis was based on only two domains of the QUADAS-2 risk of bias assessment tool (patient selection and flow/timing) because data were incomplete for all other domains. The duration of follow-up was also relatively short for most of the studies, and this confers uncertainty regarding biopsy-negative cases being TNs.

Nevertheless, this study represents the first meta-analysis of diagnostic accuracy of RTBs and is based on a robust methodology, with strict criteria for study selection that are rigorous, transparent, and reproducible.

Due to the lack of data in the literature, the study does not provide information on the learning curve needed by urologists and pathologists to take and interpret biopsy specimens, and on the optimal number of cores and biopsy pattern that should be performed to maximise diagnostic performance. Presently there is agreement that at least two good quality cores should be obtained in each case, but increasing the number of cores may increase the diagnostic yield of the procedure [33].

Further research is also needed to confirm whether biopsies of the peripheral part of the tumour should be preferred for larger lesions to avoid central necrosis or if both central and peripheral biopsies should be performed for SRMs, as previously suggested [82].
4. Conclusions

RTB has a high diagnostic yield and is associated with a very low risk of significant complications. Good estimates of sensitivity and specificity for the diagnosis of malignancy were found at meta-analysis either for CB and FNA, with better performance for CB. The accuracy of RTBs for the diagnosis of RCC subtype was found to be good; accuracy for nuclear grade was fair and improves using a simplified system (low vs high grade). However, the quality of the available evidence is moderate, and well-designed prospective cohort studies including consecutive patients and using valid reference standards to assess RTB performance are required to corroborate these findings and address knowledge gaps.

Author contributions: Alessandro Volpe had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

Study concept and design: Volpe, Marconi.

Acquisition of data: Volpe, Marconi, Stewart.

Analysis and interpretation of data: Volpe, Marconi, Norrie, Lam.

Drafting of the manuscript: Volpe, Marconi.

Critical revision of the manuscript for important intellectual content: Volpe, Marconi, Babestani, Lam, Hofmann, Stewart, Norrie, Bex, Bensalah, Canfield, Hora, Kuczyk, Merseburger, Mulders, Powles, Staehler, Ljungberg.

Statistical analysis: Norrie, Volpe, Marconi, Lam.

Obtaining funding: None.

Administrative, technical, or material support: None.

Supervision: Volpe, Marconi.

Other (specify): None.

Financial disclosures: Alessandro Volpe certifies that all conflicts of interest, including specific financial interests and relationships and affiliations relevant to the subject matter or materials discussed in the manuscript (eg, employment/affiliation, grants or funding, consultancies, honoraria, stock ownership or options, expert testimony, royalties, or patents filed, received, or pending), are the following: Thomas B. Lam is a company consultant for Pfizer, GSK, Astellas, and Ipsen. He receives company speaker honoraria from Pfizer, GSK, Astellas, and Ipsen. Karim Bensalah receives grants/research support from Pfizer and GSK, honoraria or consulted fees from Intuitive Surgical. Axel Bex receives company honoraria from Pfizer, is a trial participant for Pfizer Europe, participates on advisory boards for GSK and Novartis, is a company consultant for Pfizer and Novartis, and receives grants/research support from Pfizer. Steven E. Canfield receives company speaker honoraria from Amgen, Genomic Health, Algeta, and Bayer. Milan Hora receives company speaker honoraria from Covidien and Olympus. He participates in trials for Janssen, receives grants/research support from Ipsen, and receives company speaker honoraria from Janssen and Astellas. Markus A. Kuczyk is a shareholder at Bayer Healthcare, Astellas, Storz, Pfizer, and Wyeth, and Novartis. He is a company consultant for Karl Storz, Coloplast, Astra Zeneca, Astellas, and Storz, and receives company speaker honoraria from Pfizer, Astellas, Bayer, GSK, Pierre Fabre, and Janssen Cilag & Hexal. He is a trial participant for the ProteCt Study, Millennium Study C21004, and Millennium Study C21005 and at Astellas, Ipsen, and Janssen. He receives grants and research support for Pfizer and is a company consultant for Hexal AG. Axel S. Merseburger is a company consultant for Ipsen Pharma, Bayer, Astellas, Janssen Cilag, Novartis, and Pfizer. He receives company speaker honoraria from Ipsen Pharma, Wyeth, Astellas, Novartis, Pfizer, and SEP. He participates in trials for Astra Zeneca, Bayer, Pfizer, TEVA, Novartis, and Astellas. He receives grants/research support from Wyeth and participates in a company-sponsored speakers’ bureau for TEVA, Janssen, Pfizer, Astellas, Ferring, and Novartis. Peter F.A. Mulders receives company speaker honoraria from Astellas, Pfizer, Novartis, GSK AstraZeneca, and J&J. He participates in trials for Astellas, GSK, J&J, and Millennium. He receives grants/research support from Astellas, Wilex, GSK, and Pfizer. He participates in trials for J&J and Provenge. He receives grants/research support from Bayer. Thomas Powles is a company consultant for Novartis, Pfizer, and GSK. He receives company speaker honoraria from Novartis, Pfizer, and GSK. He participates in trials for GSK, Pfizer, and BMS. He receives grants/research support from GSK, Pfizer, and Novartis. He receives company speaker honoraria and participates in trials for Genentech. Michael Staehler is a company consultant for Pfizer, Novartis, GSK, Roche, and Astellas. He receives company speaker honoraria from Pfizer, Novartis, GSK, Roche, and Astellas. He participates in trials and receives fellowships and travel grants from Pfizer, Novartis, GSK, and Roche. He receives grants/research support from Pfizer, Novartis, GSK, Roche, Avo, and Bayer. He is a company consultant for Bayer and receives company speaker honoraria from Bayer and Avo. He participates in trials for Bayer, Avo, Wilex, and Immatics. Borje Ljungberg receives company speaker honoraria from GSK, Roche, Pfizer, and Novartis. He participates in trials for GSK, Medivation, Pfizer, and Janssen R&D. He is on the advisory board at Pfizer and GSK. The other authors have nothing to disclose.

Funding/Support and role of the sponsor: None.

References


