Endoscopic ultrasound-guided fine needle aspiration of pancreatic lesions with 22 versus 25 Gauge needles: A meta-analysis

Antonio Facciorusso¹, Elisa Stasi², Marianna Di Maso¹, Gaetano Serviddio³, Mohammed Salah Ali Hussein¹,⁴ and Nicola Muscatiello¹

Abstract

Background: Robust data in favour of a clear superiority of 22 versus 25 Gauge needles for endoscopic ultrasound-guided fine needle aspiration are still lacking.

Objective: We aimed to compare the diagnostic sensitivity, specificity and safety of these two needles for endoscopic ultrasound-guided fine needle aspiration of solid pancreatic lesions.

Methods: A computerized bibliographic search was restricted to randomized controlled trials only. Pooled effects were calculated using a random-effects model and expressed in terms of risk ratio and 95% confidence interval.

Results: We analysed seven trials with 689 patients and 732 lesions (295 sampled with 22 Gauge needle, 309 with 25 Gauge needle, and 128 with both needles). A non-significant superiority of 25 Gauge in terms of pooled sensitivity (risk ratio: 0.93, 0.91–0.95 versus 0.89, 0.85–0.94 of 22 Gauge needle; \( p = 0.13 \)) and no difference in terms of specificity (1.00, 0.98–1.00 in both groups; \( p = 0.85 \)) were observed. Sample adequacy was similar between the two devices (risk ratio: 1.03, 0.99–1.06; \( p = 0.15 \)). Very few adverse events were observed and did not impact on patient outcomes.

Conclusion: Our meta-analysis reveals non-superiority of 25 Gauge over 22 Gauge; hence no definitive recommendations over the use of one particular device can be made.

Keywords

EUS, FNA, pancreas, sensitivity, specificity

Received: 4 September 2016; accepted: 24 October 2016

Introduction

Endoscopic ultrasound (EUS) represents a valuable and accurate diagnostic technique for the morphological characterization of pancreatic lesions; furthermore, EUS allows sampling of pancreatic tissue for cytopathological diagnosis by means of fine needle aspiration (FNA).

However, several technical and clinical features are known to influence the diagnostic performance of EUS-FNA, including location, size and tissue firmness of the lesion, experience of the endoscopist and availability of rapid on-site evaluation (ROSE) of EUS-FNA samples performed by a cytopathologist. On the other hand, whether specific procedural aspects such as use of a stylet, number of needle passes, or different needle calibres may have an impact on diagnostic accuracy and sample adequacy is still a matter of debate.

Among the most commonly adopted devices, 22 G and 25 G needles have proved effective and safe in clinical practice, and consequently most published studies have used these devices. Theoretically, smaller needles present several advantages, namely being associated with fewer bloody aspirates, and their higher flexibility and smaller calibre make needle passage and to-and-fro transversal movements easier through the pancreatic...
tissue. In fact, a meta-analysis published by Madhoun et al. in 2013 including both randomized controlled trials (RCTs) and retrospective studies concluded that 25 G needles are more sensitive than 22 G needles in the cytopathological diagnosis of pancreatic malignancies. However, serious concerns about the robustness of this finding should be raised, since it was not confirmed in the sensitivity analysis restricted to only prospective trials and after removing a single larger retrospective study.

Therefore, the aim of our meta-analysis is to provide an up-to-date overview on the comparison between 22 G and 25 G EUS-FNA for the diagnosis of pancreatic solid lesions. In order to maximize the reliability of the conclusions we decided to restrict our analysis to RCTs only.

Primary endpoints were pooled sensitivity and specificity. Secondary outcomes were diagnostic accuracy, sample adequacy and safety profile.

Methods
Inclusion and exclusion criteria
Only studies meeting the following criteria were included: (1) RCTs comparing EUS-FNA with 22 G and 25 G needle for pancreatic solid lesions; (2) full-text studies published in English; (3) articles reporting at least one of the following data: sensitivity (or data useful for its calculation), specificity (or data useful for its calculation), diagnostic accuracy, sample adequacy.

Search strategy
Figure 1 reports the search strategy followed in the meta-analysis.

Bibliographic research was conducted on PubMed, EMBASE, Cochrane Library and Google Scholar including all studies fulfilling inclusion criteria published until June 2016. Keywords used were “EUS”, “endosonography”, “endoscopic ultrasound”, “needle”, “FNA”, “pancreas”. Relevant reviews and meta-analyses on the use of EUS in pancreas solid lesions were examined for potential suitable studies. Authors of included studies were contacted to obtain full text or further information when needed.

Data extraction and management
Data extraction was conducted by two reviewers (AF and MdM) using a standardized approach (PRISMA Statement). Data on publication details (year of publication, name of first author and country), clinical characteristics (lesion size, location and number),
technical features (number of passes per lesion per needle, whether the same lesion was sampled with both needles or not, use of the stylet and ROSE availability), and diagnostic outcomes (sensitivity, specificity, sample adequacy, side effects) were extracted. Case reports and abstracts or studies with insufficient data were excluded.

The quality of the included studies was assessed by two authors independently (AF, MdM) according to the currently accepted criteria described elsewhere. Disagreements were resolved by discussion and following a third opinion (NM).

Statistical analysis

Chi-square and $I^2$ tests were used for across studies comparison of the percentage of variability attributable to heterogeneity beyond chance. A $p$-value < 0.10 for chi-square test and $I^2 < 25\%$ were interpreted as low-level heterogeneity.

As recommended by recent Cochrane guidelines, random-effects model with DerSimonian and Laird test was chosen a priori for all analyses (regardless of the level of heterogeneity), and then a fixed-effect model by means of Mantel–Haenszel test was performed as a sensitivity test. Summary estimates were expressed in terms of risk ratio (RR) and 95\% confidence interval (CI).

The pooled sensitivity and specificity of the two needles were compared using the bivariate approach, and a summary receiver operating characteristics (ROC) curve was built based on random-effect model.

Probability of publication bias was assessed using funnel plots and with Begg and Mazumdar’s test. Sensitivity analysis was further conducted applying the leave-one-out method and according to the quality of included studies.

All statistical analyses were conducted using RevMan version 5 from the Cochrane collaboration and OpenMeta[Analyst] software. For all calculations a two-tailed $p$-value of less than 0.05 was considered statistically significant.

Results

Literature search

As shown in Figure 1, 230 studies were initially identified. After preliminary exclusion of abstracts or papers not fulfilling inclusion criteria, 11 potentially relevant prospective articles were examined. Among these studies, four were excluded due to non-randomized design or because they compared different aspects of the EUS-FNA technique other than the calibre of needle used (such as the number of needle passes or availability of ROSE). Finally, seven RCTs with 689 patients and 732 lesions (295 sampled with 22 G EUS-FNA, 309 with 25 G EUS-FNA, and 128 with both needles) were included in the meta-analysis.

Characteristics of included studies

The main characteristics of the included studies are reported in Table 1.

The recruitment period ranged from 2007 to 2014. One RCT was conducted in Asia, two in the USA/Canada and the remaining four in Europe. All studies presented two well-balanced arms in terms of lesion features (number, location and size) and number of needle passes per lesion. The same lesions were sampled with both needles (in a random sequence) in two studies, while the study by Carrara et al. adopted a cross-over design with use of both needles in 18\% of lesions. All studies except those by Vilmann et al. and Lee et al. had their samples evaluated by a cytopathologist on site.

Quality was deemed high in four RCTs and moderate in three. More details on the methodological characteristics and quality of included articles are shown in Supplementary Table 1.

Sensitivity and specificity

As depicted in Figure 2(a), pooled sensitivity of the 22 G needle was 0.89 (95\% CI: 0.85–0.94), while sensitivity of the 25 G needle was 0.93 (0.91–0.95) (Figure 2(b)). On the other hand, specificity (pooled from six studies which reported specificity data) was the same in the two groups (1.00, 0.98–1.00 in both groups; Figure 3(a) and 3(b)). The bivariate generalized random-effect model showed a non-significant superiority of the 25 G needle in terms of sensitivity ($p=0.13$) and no difference in terms of specificity ($p=0.85$). In both cases, no evidence of heterogeneity was found ($I^2=12\%$ and 0\%, respectively).

The area under the ROC curve, which reflects the diagnostic accuracy of the procedures, was 0.98 for the 22 G and 0.99 for the 25 G needle (Figures 4(a) and 4(b)).

Sample adequacy

The forest plot of the comparison of sample adequacy is reported in Figure 5. RR for sample adequacy was very near to 1 with only a slight increase in favour of the 25 G needle (1.03, 0.99–1.06) which, however, did...
Table 1. Characteristics of the included studies.

<table>
<thead>
<tr>
<th>Study</th>
<th>Sample size</th>
<th>Study period</th>
<th>Country</th>
<th>Lesion size (mm)</th>
<th>Lesion number</th>
<th>Number of passes per lesion</th>
<th>Same lesion with both needles</th>
<th>Pancreatic head mass (%)</th>
<th>Stylet use</th>
<th>ROSE Quality</th>
</tr>
</thead>
<tbody>
<tr>
<td>Siddiqui 2009</td>
<td>22 G 64</td>
<td>2007–2008 USA</td>
<td>29.6</td>
<td>64</td>
<td>2.6 ± 1.2</td>
<td>No</td>
<td>66.8</td>
<td>Yes</td>
<td>Yes M</td>
<td></td>
</tr>
<tr>
<td>Camellini 2011</td>
<td>22 G 43</td>
<td>NA Italy</td>
<td>30.7</td>
<td>67</td>
<td>2.6 ± 1.2</td>
<td>No</td>
<td>72</td>
<td>NA</td>
<td>No M</td>
<td></td>
</tr>
<tr>
<td>Fabbri 2011</td>
<td>22 G 50</td>
<td>2007–2008 Italy</td>
<td>29 ± 0.7</td>
<td>50</td>
<td>2.8 ± 0.4</td>
<td>No</td>
<td>68</td>
<td>Yes</td>
<td>Yes H</td>
<td></td>
</tr>
<tr>
<td>Vilmann* 2013</td>
<td>22 G 28</td>
<td>2009–2010 Denmark Romania Germany</td>
<td>30.9 ± 14</td>
<td>28</td>
<td>2.8 ± 0.4</td>
<td>No</td>
<td>NA</td>
<td>No M</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lee 2013</td>
<td>22 G 94</td>
<td>2009–2010 Korea</td>
<td>33.2 ± 1.5</td>
<td>94</td>
<td>2.8 ± 1.2</td>
<td>No</td>
<td>32.9</td>
<td>Yes No H</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gimeno-Garcia*</td>
<td>22 G 78</td>
<td>2012 Canada NA</td>
<td>37.7 ± 1.9</td>
<td>94</td>
<td>3.1 ± 1.1</td>
<td>No</td>
<td>56.3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Carrara* 2016</td>
<td>22 G 46</td>
<td>2013–2014 Italy</td>
<td>38.4 ± 19.2</td>
<td>46</td>
<td>2.2 ± 1.1</td>
<td>Yes</td>
<td>34.1</td>
<td>NA Yes H</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Trials including either pancreatic and extra-pancreatic masses. Only pancreatic lesions were reported in the table and included in the analysis.
ROSE: Rapid On-Site Evaluation performed by a cytopathologist in the endoscopic room; M: Moderate; H: High; NA: Not Available

Figure 2. Sensitivity of 22 G (a) and 25 G (b) needles in individual study and pooled estimate. Pooled sensitivity of 22 G needles was 0.89 (95% CI: 0.85–0.94) while sensitivity of 25 G needles was 0.93 (0.91–0.95). The bivariate generalized random-effect model showed a non-significant superiority of 25 G (p = 0.13). No evidence of heterogeneity was found (I² = 12%).
Figure 3. Specificity of 22 G (a) and 25 G (b) needles in individual studies and pooled estimate. Pooled specificity of 22 G needle was 1.00 (95% CI: 0.98–1.00) in both groups. No evidence of heterogeneity was found (I² = 0%).

Figure 4. Weighted summary receiver operating characteristics (ROC) curve for studies involving the 22 G needle (a) and the 25 G needle (b). The area under the ROC curve was 0.99 for 22 G and 0.98 for 25 G needle.
not reach the significance threshold ($p=0.15$). No evidence of heterogeneity was observed ($I^2=0\%$, $\chi^2=4.92$, $df=6$; $p=0.55$) (Figure 5).

These results were confirmed in sensitivity and subgroup analysis. The funnel plot in Supplementary Figure 1 allows us to reject any concerns of publication bias, as confirmed with Begg and Mazumdar’s test ($p=0.64$).

### Adverse events

Since procedure-related adverse events were reported only in two studies,\textsuperscript{22,24} a meta-analysis of this endpoint could not be conducted.

Details on safety profile of the two procedures are reported in Supplementary Table 2. Among a total of 15 adverse events reported, 11 were experienced by patients in whom a 22 G needle was used (six cases of pancreatitis and five bleeding events) and four (one case of pancreatitis and three bleeding events) by those who underwent 25 G EUS-FNA. Of note, all reported adverse events were mild and did not impact on patient outcomes.

### Discussion

EUS-FNA has a pivotal role in the diagnostic algorithm of solid pancreatic lesions, but its diagnostic accuracy is strictly dependent on a series of tumour-related features (such as lesion size, number, histological type) and technical variables (such as needles adopted, number of passes or availability of an on-site pathologist in the endoscopic room for evaluation of sample accuracy).

Among the needles more frequently used, 22 G and 25 G have gained increasing popularity due to their manageability and safety.\textsuperscript{9,25,26} These same calibres are commonly adopted also for EUS-guided interventional procedures, as reported in previous studies.\textsuperscript{27,28}

Theoretically, larger needles (for instance 22 G or even 19 G) allow the collection of larger samples but may lead to an increased rate of complications. Moreover, they may cause some technical problems, mostly due to higher stiffness of the device, likelihood of bloody contamination or presence of cellular debris in the sample.

Due to these potential drawbacks of larger needles the 25 G needle has been successfully introduced into clinical practice, although clear and definitive evidence of its superiority over the 22 G needle is still lacking.

A previous meta-analysis published in 2013 found a significant superior sensitivity of 25 G (0.93 versus 0.85 of 22 G; $p=0.0003$) while specificity was similar (1 and 0.97 for 22 G and 25 G, respectively; $p=0.97$).\textsuperscript{9} However, the findings of this meta-analysis should be interpreted with caution as both RCTs and retrospective studies were included, and more of half of the whole patients’ population was derived from a single retrospective American study,\textsuperscript{29} which therefore had a significantly higher weight in the pooled analysis.\textsuperscript{9} In fact, when the sensitivity analysis was restricted to only prospective trials and after removing the aforementioned retrospective study,\textsuperscript{29} the statistical significance in favour of the 25 G needle was not present, thus raising serious concerns about the reliability of this finding. Moreover, several other RCTs have been published in the last few years in this field; therefore, the results of that meta-analysis clearly need to be updated.

In order to achieve robust and reliable conclusions, we decided to restrict our pooled analysis to only RCTs, thus including seven trials published between 2009 and 2016.

The first striking result, apparently in contrast with the findings of the aforementioned meta-analysis,\textsuperscript{9} is
the non-significant difference in terms of sensitivity between 22 G and 25 G EUS-FNA. In fact, pooled sensitivity for the 22 G needle was 0.89 (95% CI: 0.85–0.94), while sensitivity for the 25 G needle was 0.93 (0.91–0.95) \(p=0.13\); Figure 2). Therefore, although a non-significant trend of higher sensitivity in favour of 25 G was observed, a random-effect model failed to determine a clear superiority of one device over the other.

Since our analysis was restricted to RCTs, hence not influenced by any selection or outcome reporting bias, the theoretical advantages of 25 G needles are not found to be significant in clinical practice and the diagnostic outcomes of the two devices can therefore be considered comparable.

As expected, the specificity was similar in the two groups and the value of the area under the ROC curve, which reflects the diagnostic accuracy, was extremely high for both needles (0.98 for 22 G and 0.99 for 25 G).

Strictly related to sensitivity is sample adequacy, an endpoint which had been neglected in the previous meta-analysis. In this case, the potential advantages of larger needles (ability to collect larger tissue samples) are balanced by the easier use of 25 G needles through the pancreatic tissue; in fact, the RR for sample adequacy was close to 1 (1.03, 0.99–1.07; \(p=0.12\)).

Finally, our analysis confirmed the absolutely excellent safety profile of EUS-FNA either with 22 G and 25 G needle, since patients experienced some mild adverse events, which did not influence their clinical course, in only two studies.22,24

There are some limitations to our study. First of all, the low number of included studies and enrolled patients means particular caution is required in interpreting our findings. However, we deliberately decided to restrict inclusion criteria to only RCTs in order to provide more robust and reliable outcome estimates. Moreover, we think that our attempt to explore all the main outcomes (sensitivity, specificity, diagnostic accuracy, sample adequacy and safety profile) is to be commended, and represents a nearly unique analysis in this field. Second, there are other technical aspects such as use of stylet, ROSE availability, or number of passes, which may influence diagnostic accuracy of the procedure. Although a meta-regression analysis was unfortunately unfeasible due to the low number of included RCTs, no evidence of significant heterogeneity was found in our study and, as reported in Table 1, all RCTs were well balanced at baseline according to all these parameters.

In conclusion, despite these weaknesses, our meta-analysis reveals non-superiority of 25 G over 22 G, hence no definitive suggestions on the use of one particular device may be provided. Further RCTs are needed in order to confirm our results.

Acknowledgements
Antonio Facciorusso, MD designed the study and performed the statistical analysis; Marianna Di Maso, MD, Elisa Stasi MD, and Mohammed Salah Ali Hussein collected the data; Gaetano Serviddio, MD, PhD, and Nicola Muscatiello, MD revised the manuscript. All the authors approved the final draft submitted.

Declaration of conflicting interests
None declared.

Financial support
This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

References


