Can liquid-based preparation substitute for conventional smear in thyroid fine-needle aspiration?  
A systematic review based on meta-analysis

Yosep Chong¹, Soon-Jin Ji², Chang Suk Kang¹ and Eun Jung Lee¹

¹Department of Hospital Pathology, College of Medicine, The Catholic University of Korea, Seoul, Republic of Korea
²Medical Library, The Catholic University of Korea, Seoul, Republic of Korea

Abstract

Objective: Conventional smear (CS) using fine-needle aspiration cytology (FNAC) has been established as the test of choice for diagnosing thyroid lesions, despite low sample adequacy and inter-individual variations. Although a liquid-based preparation (LBP) technique has been recently applied to overcome these limitations, its clinical utility and its accuracy over CS are controversial. This study aimed to determine the true sensitivity and specificity of LBP in thyroid FNAC by meta-analysis.

Design: Systematic review with meta-analysis.

Methods: We searched major electronic databases (MEDLINE, EMBASE, Cochrane library, Google Scholar) with queries of ‘thyroid’, ‘LBP’ and ‘liquid-based cytology’. Original articles including cytohistologic correlation data comparing the accuracy of any LBP technique, such as ThinPrep, SurePath and Liqui-Prep, with CS were included for qualitative meta-analysis and preparation of synthesized reporter-operating curves (sROC).

Results: A total of 372 studies were screened and 51 original articles were eligible for full-text review; finally, 24 studies were chosen for the meta-analysis. Average sample inadequacy was significantly lower in two mainstream LBP methods (ThinPrep and SurePath) than CS. Specificity and sensitivity by sROC were similar or slightly superior for LBP vs CS. Various cytomorphologic changes by each method have been reported.

Conclusions: Although a learning curve is essential for adapting to the cytomorphologic features of the LBP technique, our results support the use of two mainstream LPBs alone in thyroid FNAC that LBP will increase the sample adequacy and reduce the workload with similar accuracy. More data and further evaluation are needed for the other LBP methods.

Introduction

Conventional smear (CS) using fine-needle aspiration cytology (FNAC) has been well established during the last few decades as the diagnostic test of choice for making initial diagnosis and treatment plans for thyroid lesions (1). It has been widely accepted as a primary diagnostic tool owing to its simplicity, safety, possibility of repetition and cost-effectiveness. The major technical limitations in this test are those that occur during the smearing procedure:
first, blood-obscuring background generated by abundant vasculature of thyroid lesions; second, poor cellularity due to extensive fibrosis or the cystic nature of the lesion itself; and third, person-to-person variation in the smearing technique often leading to dry artifacts. These problems result in a number of inadequate samples for making a proper diagnosis and cause a decrease in overall efficacy. In fact, many studies have reported a quite high rate of inadequate samples in thyroid FNAC using CS, up to 50.5% (2). Therefore, such limitations derived from the nature of the thyroid lesions and the essential parts of the procedure that involve various medical personnel have been a major hindrance to overcome.

Liquid-based preparation (LBP), or thin-layer preparation, was first introduced for gynecologic cervical smears that have similar limitations. Two major systems approved by the United States Food and Drug Administration, ThinPrep (Hologic, Marlborough, MA, USA) and SurePath (BD Diagnostics-TriPath Imaging, Burlington, NC, USA), are basically designed to reduce such variations and artifacts and are intended to produce representative, standardized smears by an automated process. Both techniques consist of collection of the aspirates in specially developed liquid fixative; followed by removal of cell debris, red blood cells and inflammatory cells; homogenization by vortexing and finally a sampling and slide-producing step either by vacuum application or a sedimentation method.

Over 25 years of wide spread use, the diagnostic utility of LBP in gynecologic samples has been relatively well-verified, clarifying its strong and weak points. It provides standardized slides of homogenous cellular smears with well-preserved cell morphology resulting in clearer visualization, shorter interpretation time and more reproducible results among various cytotecnicians and pathologists. In particular, dispersing cell clusters into single cells during the homogenizing step in LBP is an important strong point for gynecologic samples in which cell overlapping is a major hindrance for accurate interpretation.

In terms of application to thyroid FNAC samples, however, where the shapes of cell clusters and the nature of the background are valuable for accurate diagnosis, conflicting results on the diagnostic utility of LBP have been reported. These might be attributable to the diversity of the subject population, subtle differences in detailed procedures and the infancy of application of this technology in this specialized field.

However, many studies have been designed and conducted under pressure to some extent to favor a certain LBP product, potentially leading to biased results, and this should not be neglected. For example, many studies applied different interpretation criteria for sensitivity and specificity that were favorable for their preferred conclusion. Although many investigators now agree that application of LBP to thyroid FNACs is acceptable, to what extent we can trust the results of LBP, whether it is okay to use LBP alone or whether LBP should be applied in combination with CS are aspects that are not clear.

In this study, we performed a systematic review and meta-analysis of the comparative studies of LBP and CS, mainly on ThinPrep and SurePath methods, conducted in thyroid FNACs to draw less biased results and a statistically convincing conclusion on the diagnostic accuracy and utility of LBP in thyroid FNACs.

**Subjects and methods**

**Subjects**

This study was approved by the Institutional Review Board of The Catholic University of Korea, College of Medicine (SC15SIS0004). We searched major electronic databases from January 1, 1990 to January 5, 2015 for relevant articles published in medical journals with abstracts written in English. Included databases were MEDLINE (PubMed), Cochrane Library, EMBASE and Google Scholar. There were 7 LBP methods included in the search: ThinPrep, SurePath (also known as AutoCyte PREP), Liqui-PREP (LGM-International, FL, USA), CellPrepPlus (Biodyne, Seongnam, Korea), Cell & Tech (Cell & Tech Bio, Seoul, Korea), EasyPrep (YD Diagnostics Corp., Seoul, Korea) and HuroPath (formerly known as E-Prep, CelltraZone, Seoul, Korea). The queries used were: ‘(‘Thyroid gland’ (MeSH Term) OR ‘Thyroid gland’ (Text Word)) AND (‘liquid-based preparation’ (Text Word) OR ‘liquid-based cytology’ (Text Word) OR ‘ThinPrep’ (Text Word) OR ‘SurePath’ (Text Word) OR ‘AutoCyte PREP’ (Text Word) OR ‘Liqui-Prep’ (Text Word) OR ‘CellPrepPlus’ (Text Word) OR ‘Cell & Tech’ (Text Word) OR ‘EasyPrep’ (Text Word) OR ‘E-Prep’ (Text Word))’ for PubMed, ‘thyroid gland’/exp AND (‘liquid-based preparation’/exp OR ‘liquid based cytology’ OR ‘thinprep’/exp OR ‘surepath’/exp OR ‘autocyte prep’/exp OR ‘liqui-prep’/exp OR ‘cellprepplus’/exp OR ‘cell & tech’/exp OR ‘easyprep’/exp OR ‘e-prep’/exp) for Embase, ‘allintitle: thyroid AND (‘thinprep’ OR ‘surepath’ OR ‘autocyte prep’ OR ‘liqui-prep’ OR ‘cellpreppro’ OR ‘cell & tech’ OR ‘easyprep’ OR ‘e-prep’ OR ‘liquid based preparation’ OR ‘liquid based
cytology” for Google Scholar. All similar possible word variations were also searched. The attained records were retrieved and managed with EndNote X7.2.1 (Bld 10136, Thomson Reuters, New York, NY, USA).

### Study selection, reviewing and data retrieving

The process of study selection and reviewing is depicted in Fig. 1. After the initial search, any duplicates were removed from the results. Then, the title and abstracts of the records were screened by two independent reviewers (Y Chong and E J Lee). Case reports, letters, reviews, conference proceedings and posters were excluded. Any original studies with cytohistological correlation data were included for full-text reviewing and only a subset of the studies with eligible data was used for quantitative analysis. In addition, the cited references in each study were manually searched and reviewed to identify any additional relevant studies.

To apply the same standardized criteria for determining sensitivity and specificity, data from each study were retrieved and properly treated for quantitative analysis. Based on the treatment guidelines for thyroid lesions after FNAC, follicular neoplasms (FN) or Hurthle cell neoplasms (HCN) were considered to require surgical resection. Thus, we applied a tentative definition of positive, false positive, negative, and false negative solely based on the need for surgical resection. For example, FNAC results of FN or HCN with a histologic diagnosis of nodular hyperplasia (NH) or other benign lesions were tentatively considered as false-positive results, while the cases with histologic diagnosis of papillary carcinoma (PTC) were tentatively considered as positive results although the cytomorphologic features of FN/HCN and PTC are different. Likewise, FNAC results of atypia of undetermined significance (AUS) or benign follicular nodule with histologic diagnosis of FN or HCN were tentatively considered as false negative while the cases with NH or thyroiditis were tentatively considered as negative results. We hypothesized that the proportion of tentatively categorized false-positive or false-negative cases may be similar regardless of CS or LBP. Detailed interpretation criteria are shown in Table 1. Similar possible terminological variations among studies were recategorized into the relevant subcategories.

### Quality assessment of diagnostic accuracy studies

To assess the quality of the studies included in the meta-analysis, we incorporated the revised quality assessment of diagnostic accuracy studies, QUADAS-2, developed by Whiting et al. (3). QUADAS-2 consists of four key domains: patient selection, index test, reference standard and flow and timing. A few signaling questions relevant to the risk of bias and the applicability in the index test domain were designed and added to the quality assessment as follows: Was a standardized classification system used? Was any risk of bias according to individual variation in the obtaining technique avoided? Was any risk of bias during data transformation and utilization avoided as much as possible?

Two independent reviewers (Y Chong and EJ Lee) reviewed the included studies using the modified QUADAS-2 and made a judgment on the risk of bias and the applicability of each domain. The results were tabulated and summarized after a discussion about the studies with discrepant assessments. The final meta-analysis was performed after the exclusion of any studies.
Table 1  Detailed tentative criteria for interpretation of FNAC results compared with histologic diagnosis.

<table>
<thead>
<tr>
<th>Interpretation</th>
<th>FNAC results of CS or LBP</th>
<th>Histologic diagnosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Positive</td>
<td>Malignancy</td>
<td>PTC</td>
</tr>
<tr>
<td></td>
<td>Suspicious for malignancy</td>
<td>Other types of primary thyroidal malignancy including lymphoma</td>
</tr>
<tr>
<td></td>
<td>FN (suspicious for)</td>
<td>FA*/FC</td>
</tr>
<tr>
<td></td>
<td>HCN (suspicious for)</td>
<td>HCA*/HCC</td>
</tr>
<tr>
<td>False positive</td>
<td>Malignancy</td>
<td>Nodular hyperplasia/colloid cyst</td>
</tr>
<tr>
<td></td>
<td>Suspicious for malignancy</td>
<td></td>
</tr>
<tr>
<td></td>
<td>FN (suspicious for)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>HCN (suspicious for)</td>
<td></td>
</tr>
<tr>
<td>Negative</td>
<td>Benign follicular nodule</td>
<td>Nodular hyperplasia/colloid cyst</td>
</tr>
<tr>
<td>False negative</td>
<td>AUS/FLUS*</td>
<td>Thyroiditis (Hashimoto's/lymphocytic/granulomatous)</td>
</tr>
</tbody>
</table>

AUS, atypia of undetermined significance; FA, follicular adenoma; FC, follicular carcinoma; FLUS, follicular lesion of undetermined significance; HCA, Hurthle cell adenoma; HCC, Hurthle cell carcinoma; PTC, papillary thyroid carcinoma.

with a high risk of bias or concerns about bias in any of the domains.

Data extraction and analysis

The weighted average of sample inadequacy was calculated for each LBP method and CS. To determine the statistical significance, the weighted average difference was calculated in the studies that evaluated both LBP and CS using the same sample. A P value of less than 0.05 was defined as statistically significant. The reported sample inadequacy was categorized by the year of publication and compared by the mode of sampling method.

Results

Study selection, reviewing and data retrieving

The inclusion/exclusion process during the screening and selection steps is summarized in Fig. 1. A total of 372 papers were identified by the database search (81 in MEDLINE, 177 in EMBASE, 42 in Cochrane library and 72 in Google Scholar). After excluding 125 duplicates, a total of 247 records were screened by titles and abstracts. After 81 records were removed, 166 studies were subjected to further evaluation. Ninety-seven records of case reports, letters, reviews and conference proceedings were excluded, and only 51 studies were eligible for full-text reviewing (31 on TP, 12 on SP, 4 on Liqui-Prep, 1 on CellprepPlus, 1 on Cell & Tech, none on EasyPrep or E-prep). Among these, only 24 studies were eligible for data retrieval and qualitative synthesis (21 on TP, 6 on SP, 1 on Liqui-Prep, 1 on CellprepPlus, 1 on Cell & Tech, none on EasyPrep or E-prep).

Sample inadequacy of LBP

The average sample inadequacy using all of the case data from the studies using either LBP alone, or LBP and CS or LBP and a combined method, are summarized in Table 2 (4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20). The average sample inadequacy in TP studies (17 studies, 24,880 cases) was 16.4% in the cases processed by TP, 33.4% by CS and 7.7% by a combined method (Table 2A). Average sample inadequacy in the SP studies (4 studies, 1706 cases) was 18.9% in the cases processed by SP, 12.6% by CS and 4.0% by a combined method (Table 2B) (21, 22, 23, 24). Only one study each was included for calculating average sample inadequacy for Liqui-PREP (25), CellprepPlus (26) and the Cell and Tech method (27) (Table 2C, D and E).

Sample inadequacy using the case data of the comparative studies of LBP and CS is summarized in Table 3 (4, 5, 7, 10, 12, 13, 18). Average sample inadequacy in TP studies (7 studies, 14,251 cases) was significantly lower in TP (24.0%) than CS (33.4%) (P<0.01). Likewise, average sample inadequacy in SP studies (2 studies, 730 cases) was significantly lower in SP (7.1%) than CS (13.2%) (P<0.02), although the number of included studies was limited.

Sample inadequacy of TP studies gradually decreased from the first dates of publication to the more recent ones and the inadequacy trend line of TP was slightly lower than that of CS (Fig. 2A). The accumulated inadequacy
of TP studies more clearly demonstrates this finding (Fig. 2B). Sample inadequacy was significantly lower by TP than CS in samples collected by double sampling or syringe rinsing or directly collected to vial from the different patients (during the different periods) (Fig. 2C). However, sample inadequacy was significantly higher in TP than CS in the consultation slide or the samples collected by splitting (Fig. 2C).

### Sensitivity and specificity of LBP

A coupled forest plot of sensitivity and specificity of 17 TP studies is shown in Fig. 3A (4, 5, 7, 8, 9, 11, 12, 14, 15, 16, 17, 18, 19, 20, 28, 29, 30, 31). There was no obvious relationship between sensitivity and specificity among 17 studies using TP. Sensitivity and specificity of CS showed more homogeneous results than those of TP because TP

### Table 2

Reported sample adequacy (weighted average) of the studies using a LBP and a CS/combined method.

<table>
<thead>
<tr>
<th>Year</th>
<th>Authors</th>
<th>Total cases</th>
<th>Inadequacy (case number/%)</th>
<th>Sampling method</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>TP</td>
<td>CS</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2000</td>
<td>Scurry et al. (4)</td>
<td>728</td>
<td>401</td>
<td>40.9</td>
</tr>
<tr>
<td>2001</td>
<td>Afify et al. (5)</td>
<td>209</td>
<td>82</td>
<td>11.0</td>
</tr>
<tr>
<td>2004</td>
<td>Tulecke et al. (8)</td>
<td>115</td>
<td>115</td>
<td>6.1</td>
</tr>
<tr>
<td>2006</td>
<td>Fadda et al. (9)</td>
<td>1525</td>
<td>1525</td>
<td>11.3</td>
</tr>
<tr>
<td>2007</td>
<td>Michael et al. (11)</td>
<td>218</td>
<td>218</td>
<td>33.0</td>
</tr>
<tr>
<td>2008</td>
<td>Cavaliere et al. (12)</td>
<td>7750</td>
<td>7750</td>
<td>32.3</td>
</tr>
<tr>
<td>2008</td>
<td>Saleh et al. (13)</td>
<td>290</td>
<td>145</td>
<td>24.1</td>
</tr>
<tr>
<td>2008</td>
<td>Stamataki et al. (14)</td>
<td>252</td>
<td>252</td>
<td>4.0</td>
</tr>
<tr>
<td>2010</td>
<td>Ardito et al. (15)</td>
<td>353</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>2011</td>
<td>Gheri et al. (16)</td>
<td>2518</td>
<td>2518</td>
<td>9.4</td>
</tr>
<tr>
<td>2011</td>
<td>Luu et al. (17)</td>
<td>4101</td>
<td>2101</td>
<td>8.2</td>
</tr>
<tr>
<td>2012</td>
<td>Chang et al. (18)</td>
<td>4290</td>
<td>2523</td>
<td>12.6</td>
</tr>
<tr>
<td>2013</td>
<td>Mastorakis et al. (19)</td>
<td>1000</td>
<td>1000</td>
<td>4.1</td>
</tr>
</tbody>
</table>
data in some studies were limited by sample type and data quality (Fig. 3A). A forest plot of the combined TP and CS studies showed heterogeneity because of the limited number of included studies and the limited data quality of the included studies (Fig. 3C) (Q value). The sROC using these data showed similar curves between TP and CS; more precisely, a slightly higher curve for TP, showing a significant difference in specificity and sensitivity between TP and CS in thyroid FNAC (Fig. 3D). Although the sROC of the combined TP and CS was lower than the others, it should be interpreted with caution because it was based on data derived from a limited number of studies.

A forest plot of TP after exclusion of studies of limited quality showed more homogenous results (Fig. 4A) (4, 5, 7, 9, 11, 12, 14, 15, 18, 19, 20, 31). The heterogeneity of studies of TP and CS was similar (based on the Q value). Only one study was included in this analysis for combined TP and CS (Fig. 4C). The sROC derived from these studies showed similar but slightly higher curves for TP than the previous sROC (Fig. 4D).

For SP, a forest plot drawn from SP and CS is depicted in Fig. 5A and B, and it showed moderate heterogeneity (Q value) (21, 22, 24, 32, 33). A sROC using 5 SP studies and 3 CS studies showed curves with similar sensitivity and specificity (Fig. 5C).

### Quality assessment of diagnostic accuracy studies

Quality assessments of the included studies are summarized in Table 4 and Supplementary Fig. 1 (see section on supplementary data) given at the end of this article. Most studies in the final meta-analysis had a low risk of bias or concern in each domain according to risk of bias or applicability, which represents a relatively high level of credibility in the results of the meta-analysis. Three studies showed an uncertain risk of bias or concern in one domain among the TP studies. Only one study among SP studies showed uncertain concern in the applicability during the patient selection.

### Morphologic characteristics of LBP

Morphologic parameters could not be compared using meta-analysis because each study applied a customized strategy that could not be easily standardized for meta-analysis. Major morphologic characteristics of each LBP compared to CS described by a few key studies are summarized in Table 5 (8, 18, 27, 28, 29, 34, 35).

### Discussion

This study demonstrated that the clinical utility of LBP is almost the same or marginally better than that of CS for thyroid FNAC in terms of sample adequacy, sensitivity and specificity.

### Sample inadequacy of LBP

The sample adequacy was significantly superior for two mainstream LBP methods (ThinPrep and SurePath) than CS for most sampling methods (Table 3). More data on

---

**Table 3** Difference in sample inadequacy between LBP and CS (weighted average difference).

<table>
<thead>
<tr>
<th>Year</th>
<th>Authors</th>
<th>Total cases</th>
<th>Inadequacy (case number/%)</th>
<th>Sampling method</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>TP</td>
<td>CS</td>
</tr>
<tr>
<td>(A)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2000</td>
<td>Scurry et al. (4)</td>
<td>728</td>
<td>401</td>
<td>40.9</td>
</tr>
<tr>
<td>2001</td>
<td>Afify et al. (5)</td>
<td>209</td>
<td>82</td>
<td>11.0</td>
</tr>
<tr>
<td>2003</td>
<td>Cochand-Priollet et al. (7)</td>
<td>240</td>
<td>120</td>
<td>22.5</td>
</tr>
<tr>
<td>2006</td>
<td>Malie et al. (10)</td>
<td>744</td>
<td>459</td>
<td>3.9</td>
</tr>
<tr>
<td>2008</td>
<td>Cavaliere et al. (12)</td>
<td>7750</td>
<td>3875</td>
<td>32.3</td>
</tr>
<tr>
<td>2008</td>
<td>Saleh et al. (13)</td>
<td>290</td>
<td>145</td>
<td>24.1</td>
</tr>
<tr>
<td>2012</td>
<td>Chang et al. (18)</td>
<td>4290</td>
<td>2523</td>
<td>12.6</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>14,251</td>
<td>7605</td>
<td>24.0</td>
</tr>
</tbody>
</table>

(A) TP vs CS, (B) SP vs CS.

CS, conventional smear; LBP, liquid-based preparation; SP, SurePath; TP, ThinPrep.
sample adequacy are needed for the newly developed techniques such as Liqui-PREP, CellprepPlus and the Cell and Tech method (Table 2). We can see clearly that the sample adequacy is getting better over time after the introduction of LBP for thyroid FNAC (Fig. 2A and B). This must be due to a learning curve of the new technology. This trend suggests that the learning curve has reached a stage of maturity for the new technology.

As expected, sample adequacy was better for LBP (TP) than CS using sampling methods where relatively equal amounts of sample content might be distributed to LBP and CS, such as double sampling, direct to vial (different samples), and syringe rinsing methods. Sample adequacy using consultation slides and sample splitting methods showed better adequacy for CS than TP. This can be explained by the fact that most consultation slides contain samples from the primary screening laboratory that might contain generally lower cellularity. Sample splitting is a limited method inevitably producing disproportionate samples used for LBP and CS, as many prior studies have shown. Combined LBP and CS methods showed much lower rates of sample inadequacy than LBP or CS alone for both TP and SP studies, although the difference was not statistically significant (Table 2).

### Sensitivity and specificity of LBP

The sensitivity and specificity of LBP over CS using sROC showed similar or slightly better results for TP than CS (Fig. 3). The results were clearer after the exclusion of studies with poor or limited data quality (Fig. 4). The results were similar for SP studies as well (Fig. 5). For SP studies, there is a generalizability limitation because the included studies are all from either Belgium or South Korea.

On a side note to the results of meta-analysis, there were two studies that deal with cytohistological correlation data of other LBPs, one for CellprepPlus and one for EasyPrep (26, 36). The CellprepPlus study compared only 20 cases of CellprepPlus and CS, but it showed 100% sensitivity and specificity for CS and 71.4% sensitivity and 86.7% negative predictive value for CellprepPlus for the histologic correlation (data not shown) (26). The EasyPrep study compared only 28 and 26 cases of SP and EasyPrep and histologic diagnosis and showed 100% sensitivity and specificity for SP and 95.5% sensitivity and 80% negative predictive value for EasyPrep (36).

### Morphologic characteristics of LBP

The morphologic parameters were not feasible for meta-analytic comparison because it is generally thought that the standards of morphologic parameters vary widely among various sampling methods, LBP methods, study designs and investigators. However, we can summarize the general morphologic changes of LBP in thyroid FNAC compared to CS based on the consistent findings of most studies. In LBP, the nuclear size is smaller, the nuclear-to-cytoplasmic ratio is bigger, nucleoli are more prominent and nuclear membrane irregularity and...
Figure 3
Coupled forest plots of the studies using TP (A), CS (B) and combined TP and CS (C), and a corresponding sROC (D). CS, conventional smear; sROC, synthesized reporter operating curve; TP, ThinPrep.
Figure 4
Coupled forest plots after exclusion of limited studies using TP (A), CS (B) and combined TP and CS (C), and a corresponding sROC (D). CS, conventional smear; sROC, synthesized reporter operating curve; TP, ThinPrep.
nuclear grooves become more obvious. The cytoplasm is scantier in LBP. These changes are probably due to a lack of the smearing effect that can be a potential cause of dry or degenerative artifacts in CS. However, intranuclear pseudoinclusions are less evident on LBP than CS owing to the ingredient changes in the fixative solutions of LBP. Increases of 3-dimensional clusters and decreases of large papillae by fragmentation are other important features of LBP compared to CS. These can be understood to be a result of the homogenizing step of LBP. Therefore, LBP gives better cytomorphologic visibility for follicular clusters of the benign lesions while it loses the important papillary structure of PTCs. For follicular lesions, macrofollicular architecture, Hurthle cell changes and the presence of colloid (tissue-paper-like material) and macrophages serve as important features to suspect benign follicular lesions, which is similar in CS (8).

Figure 5
Coupled forest plots after exclusion of limited studies using SP (A), CS (B) and a corresponding sROC (C). CS, conventional smear; SP, SurePath; sROC, synthesized reporter operating curve.
Large oncocytic cells with a discohesive pattern are more common in malignant thyroid lesions. The most important and specific features for papillary carcinoma are the hobnail features in papillary carcinoma are often associated with cytoplasmic vacuole, background macrophages and lymph node metastasis. Incompletely lysed red blood cells are more prominent in EasyPrep.
Quality assessment of diagnostic accuracy studies

QUADAS-2 assessment of the included studies revealed that only a few studies had an uncertain risk of bias or concern in one domain, which means that the results of this meta-analysis are trustworthy. Furthermore, the results of the sROC curve before and after the exclusion of the studies with limited quality were consistent. From the beginning, we hypothesized that the simplified, tentative categorization of the cytopathological diagnosis into four groups according to the consequent surgical treatment plan (Table 1), true and false positive, or true and false negative, might not influence the results of the meta-analysis mathematically. However, there should be caution when applying the results of this study under some circumstances.

Conclusion

Based on the results of this study, we conclude that it is reasonable for LBP to be substituted for CS of FNAC of thyroid lesions for the following reasons. First, the sample adequacy is statistically superior for LBP than CS. Second, the sensitivity and specificity of LBP was similar or slightly superior to that of CS. Third, although an educational period is essential and there are pros and cons of cytomorphologic features using LBP in thyroid FNAC, it does not seem to greatly affect the accuracy of the diagnosis itself. Therefore, it is okay to trust the results of any of the two major LBPs (TP and SP) for thyroid FNAC, even when it is performed alone without additional CS. However, additional data and further evaluation are needed for the other LBPs to confirm their results.

Supplementary data
This is linked to the online version of the paper at http://dx.doi.org/10.1530/EC-17-0165.

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 work was supported by the research grant from Institute of Clinical Medicine Research in the Catholic University of Korea, Yeouido St. Mary's Hospital.

Author contribution statement
Y C designed the study, participated in screening, selection, and reviewing the references, data analysis and wrote the draft. S J J designed the query and search the databases. C S K reviewed the data analysis and revised the manuscript critically. E J L designed the study and participated in screening, selection, and reviewing the reference, data analysis and reviewed the final manuscript.

References

2 Rossi E, Morassi F, Santusano G, Zannoni G & Fadda G. Thyroid fine needle aspiration cytology processed by ThinPrep: an additional slide decreased the number of inadequate results. *Cytopathology* 2010 **21** 97-102. (doi:10.1111/j.1365-2309.2009.00659.x)
14 Stamatakis M, Anninos D, Brontzos E, Georgoulakis J, Panayiotides J, Christoni Z, Peros G & Karakitsos P. The role of liquid-based cytology


17 Luu MH, Fischer AH, Pisharodi L & Owens CL. Improved preoperative definitive diagnosis of papillary thyroid carcinoma in FNAs prepared with both thimprep and conventional smears compared with FNAs prepared with thimprep alone. Cancer Cytopathology 2011 119 68–73. (doi:10.1002/cncr.20124)


23 Sidipopoulos N, Dumont JI, Golden AC, Quinlisk FL, Gonzalez JL & Padmanabhan V. Quality improvement by standardization of procurement and processing of thyroid fine-needle aspirates in the absence of on-site cytological evaluation. Thyroid 2009 19 1049–1052. (doi:10.1089/thy.2009.0163)


