Performance Differences Between Conventional Smears and Liquid-Based Preparations of Thyroid Fine-Needle Aspiration Samples

Analysis of 47 076 Responses in the College of American Pathologists Interlaboratory Comparison Program in Non-Gynecologic Cytology

Andrew H. Fischer, MD; Amy C. Clayton, MD; Joel S. Bentz, MD; Patricia G. Wasserman, MD; Michael R. Henry, MD; Rhona J. Souers, MS; Ann T. Moriarty, MD

- Context.—Controversy exists about whether thyroid fine-needle aspirates (FNAs) should be processed with conventional smears or liquid-based preparations (LBPs).

Objective.—To compare the performance of conventional smears to LBPs for thyroid FNA slides circulated in the College of American Pathologists Interlaboratory Comparison Program in Non-Gynecologic Cytology.

Design.—Participant responses for thyroid FNA slides were compared with the reference diagnosis at the level of 3 general diagnostic categories: negative, suspicious, and malignant. For specific reference diagnoses of benign/goiter and papillary thyroid carcinoma, the participants’ specific diagnoses were analyzed and poorly performing slides were rereviewed.

Results.—The 47 076 thyroid FNA slide responses, between 2001 and 2009, included 44 478 responses (94%) for conventional smears and 2598 responses (6%) for LBPs. For the general reference category negative, participant responses were discrepant in 14.9% of conventional smears compared with 5.9% for LBPs (P < .001). The specific reference diagnosis of benign/goiter was misdiagnosed as a follicular neoplasm in 7.8% of conventional smears, compared with 1.3% of LBP. For the general reference category of malignant, participant responses were discrepant in 7.3% of conventional smears compared with 14.7% of LBPs (P < .001). The specific reference diagnosis of papillary thyroid carcinoma was misdiagnosed as benign/goiter in 7.2% of LBPs, compared with 4.8% of conventional smears (P < .001).

Conclusions.—LBPs performed worse than conventional smears for cases with a reference diagnosis of papillary thyroid carcinoma. However, LBPs performed better than conventional smears for cases with a benign reference diagnosis. Specific features in thyroid FNAs that may improve the diagnostic accuracy of LBPs and conventional smears are described.

experience. In this study, we compared the performance of LBP results with alcohol-fixed, Papanicolaou-stained smears and reviewed slides on which there was poor performance to identify potentially useful diagnostic features.

MATERIALS AND METHODS

The CAP NGC educational program circulates a large variety of cases to more than 10,482 participants and 2000 laboratories. The slides in the NGC program are donated from many laboratories to CAP with a proposed diagnosis. Slides are then reviewed by at least 3 members of the CAP Cytopathology Resource Committee composed of cytopathologists and cytotechnicians with experience diagnosing a broad range of sample types. Cases are rejected if one committee member feels that the case is of poor technical quality or is not a good example of the referring diagnosis. Poorly performing cases (defined as less than 70% concordance to the referring general diagnostic category) are reviewed by CAP at least annually and removed from circulation if they are felt, on rereview, to be poor examples of the diagnostic entities. In addition, if a participant questions the diagnosis, a case is rereviewed by members of the CAP Cytopathology Resource Committee with the potential to remove the case if it is felt to be a poor example of the diagnosis.

A search was performed for CAP NGC program responses in thyroid FNA cases between 2001 and 2009. The LBP category included SurePath (Becton, Dickinson and Company, Franklin Lakes, New Jersey), ThinPrep (Hologic, Inc, Bedford, Massachusetts), and alcohol-fixed cytopsins (all Papanicolaou stained). For cases prepared by direct smear, we included only alcohol-fixed, Papanicolaou-stained slides; we excluded air-dried Giemsa-stained slides to make the comparison equal for primary fixation and staining. Cases with fewer than 10 responses and fewer than 3 laboratory responses were excluded from the analysis; this helped to ensure a broad sampling of participants, with time for poorly performing cases to be excluded from the program. Slides with a reference diagnosis of unsatisfactory, parathyroid adenoma, and metastatic cancer not otherwise specified were also excluded from the analysis. With these exclusions, there were 47,076 total responses, with 44,478 responses for conventional tests (94.5%) and 2598 for LBPs (5.5%). Analyses were conducted at the level of total responses (including the laboratory response; n = 9,585; 20.4%), pathologist responses (n = 22,965; 48.8%), and cytotechnologist responses (n = 14,526; 30.9%). Comparison of performance was first made at the level of the General Diagnostic Category. The general diagnostic categories and the specific diagnostic possibilities in each category are as follows: negative (included goiter, Hashimoto thyroiditis, and parathyroid adenoma); suspicious for malignancy (included Hurthle cell neoplasm and follicular neoplasm); and positive for malignancy (included papillary thyroid carcinoma, medullary thyroid carcinoma, lymphoma, and anaplastic thyroid carcinoma). A discordant result was defined as a participant’s general diagnostic category not matching the reference general diagnostic category. For the Specific Reference Diagnoses of benign/goiter and papillary thyroid carcinoma, the participants’ specific diagnoses were subsequently analyzed. The 47,076 responses came from 837 slides of conventional smears (472 negative [56.4%], 108 suspicious [12.9%], and 257 positive [30.7%] reference diagnoses), and 49 slides of LBPs (13 negative [26.5%], 10 suspicious [20.4%], and 26 positive [53.1%] reference diagnoses). Slides from the 10 worst-performing cases were reexamined to try to help identify difficulties with the diagnosis.

Pearson $\chi^2$ statistics were used to test the null hypothesis that there was no association between performance and preparation type. All tests were run at the .01 significance level. The significance level is lower than the standard .05 because the $\chi^2$ test is sensitive to large sample sizes.

RESULTS

The $\chi^2$ analysis showed significant performance differences by preparation type for the general diagnostic categories of both negative and positive (Table 1). For slides with a reference general diagnostic category of “suspicous” (follicular neoplasm and Hurthle cell neoplasm), there was a high level of discordance that did not differ between conventional smears and LBPs (overall 37.1% versus 41.0 discordance, respectively; $P = .22$). Responses for slides with a negative general diagnosis were discordant in 14.9% of conventional smears, compared with 5.9% of LBPs ($P < .001$). Responses for cases with a positive general diagnosis were discordant in 7.3% of conventional smears, compared with 14.7% of LBPs ($P < .001$). Cytotechnologists had a higher discordant rate than did pathologists (in the same direction) for all general categories (Tables 2 and 3).

The performance differences for the cases with a negative general diagnostic category suggested that the
specific diagnosis of goiter was problematic for participants using conventional smears. To test that, we analyzed the specific diagnoses of participants for cases with a referring specific diagnosis of goiter. As shown in Table 4, only 69.2% of the participants’ diagnoses were goiter when diagnosing conventional smears, compared with 76.8% of responses for LBPs. Erroneous participant-specific diagnoses of follicular neoplasm were made in 7.8% of conventional smears, significantly higher than the 1.3% of LBPs in cases with the specific reference diagnosis of goiter (P < .001).

The performance differences for the cases with a positive referring general diagnostic category suggested that the specific diagnosis of papillary thyroid carcinoma was problematic for LBPs. We, therefore, analyzed the specific diagnoses of participants for cases with a reference diagnosis of papillary thyroid carcinoma. As shown in Table 5, 87.3% of participant diagnoses were papillary thyroid carcinoma when conventional smears were used, compared with 82.3% for LBPs. The most common erroneous response for both conventional smears and LBPs were benign goiters in 4.8% versus 7.2% of cases, respectively (P < .001).

In an effort to learn what diagnostic features may help participants to improve their recognition of a benign thyroid FNA on direct conventional smears, we reviewed 5 of the available conventional tests with the worst performance and a benign reference diagnosis. Cases were circulated to 5 members of the CAP Cytopathology Resource Committee for diagnostic tips. Figure 1 shows a direct smear of a reference benign goiter for which only 10 of 20 participant diagnoses (50%) were in the negative general diagnostic category. The slide appears to have been spray-fixed (because of the tufting of material into the small, circular areas on the slide). The specimen is quite cellular, but abundant colloid can be seen throughout. It is important to survey the overall amount of colloid because macrofollicular groups, by nature, tend to shed mostly colloid, leaving any microfollicular groups as relatively conspicuous components of some aspirates. Another important feature of a benign nodule (if papillary thyroid carci-
ma can be excluded) is the variation from group to group in the cytologic features. For example, if one compares the 2 follicular groups marked with thin arrows (Figure 1), there is a striking difference in nuclear size, chromatin compaction, and amount of cytoplasm. Figure 2 (from the same case) shows histiocytes bearing engulfed red blood cells. If papillary thyroid carcinoma and Hürthle cell neoplasms are excluded, the finding of macrophages is a very strong indicator of a benign follicular nodule.\(^5\) Figure 3 shows a different case of a direct smear reference example of a benign goiter in which 20 of 45 participants (44%) diagnosed a follicular neoplasm. This case and several other benign reference diagnoses cases with high discordance demonstrated marked nuclear pleomorphism. The variation suggests polyploidization, in which DNA is replicated more than 2 to 3 times without cell division (increasing by 4, 8, or 16 times) with a commensurate increase in cytoplasm.\(^6\) This type of atypia (benign endocrine atypia) strongly favors a benign diagnosis for follicular-type lesions; follicular neoplasms are generally much more monotonous. Figure 4 shows a direct smear of a reference diagnosis of benign goiter for which 22 of 60 participants (37%) gave the specific diagnosis of a follicular neoplasm. The aspirate is cellular, with some microfollicular groups. Abundant colloid was present in the background (not shown). Note that the occasional microfollicles in this case tend to be lined by cells with a very flattened or squamoid cytoplasm. The microfollicles in follicular neoplasms tend to be lined with more crowded cells that are cuboidal or even slightly columnar.\(^7\)

Five of the LBPs with the worst performance for a reference diagnosis of papillary thyroid carcinoma were also reviewed. Figure 5 shows a SurePath preparation of a case in which 42% (10 of 24) of participants diagnosed a benign goiter. This case and other poorly performing LBPs of papillary thyroid carcinoma demonstrated numerous macrofollicular groups, that is, broad, flat, or twisted 2-dimensional sheets of follicular cells. Papillary thyroid carcinoma (like benign goiters) is commonly macrofollicular. Like benign goiters, papillary thyroid carcinoma is also cystic in about 35% of cases and shows macrophages with hemosiderin (not shown in this Figure). A foreign-bodylike giant cell is present toward the lower part of the image, somewhat characteristic of papillary thyroid carcinoma. The combination of macrofollicular architecture and cystic changes still requires careful attention to the high-magnification nuclear features to exclude papillary thyroid carcinoma. At higher magnification (Figure 6), the characteristic LBP appearance of a papillary thyroid carcinoma is evident, with very fine chromatin texture, slightly oval-shaped nuclei, frequent eccentrically located nucleoli, and nuclear irregularity. Chromatin texture is dominated by fine lines in many cells (thick arrow). Such fine lines represent shallow folds of the nuclear lamina,\(^8\) similar to, but less well developed than, nuclear grooves. An intranuclear inclusion (long, thin arrow) is also present. Intranuclear cytoplasmic inclusions are highly suggestive of either papillary thyroid carcinoma or medullary thyroid carcinoma.

**COMMENT**

In this nationwide, educational, glass-slide survey of pathologists with varying backgrounds and varying exposure to different preparation types, we found evidence that papillary thyroid carcinoma was more easily recognized in conventional smears than it is with LBPs. Conversely, benign goiters appear to be more easily recognized with LBPs than it is with conventional smears.
The higher concordance for diagnosis of papillary thyroid carcinomas in our survey appears to support evidence in the literature (reviewed in Ljung) that LBPs may not permit as high a rate of definitive diagnosis of papillary thyroid carcinoma. Definitive, preoperative diagnosis of papillary thyroid carcinoma is important because it can allow a 1-step total thyroidectomy, potentially with lymph node sampling. A definitive diagnostic rate of papillary thyroid carcinoma should probably be monitored by laboratories, and a rate of greater than 60% is believed achievable with conventional smears. Difficulty in identifying intranuclear cytoplasmic inclusions in LBPs has been reported. In comparing cohorts with both conventional smears and LBPs to a cohort with only LBPs, Luu et al found a greater definitive diagnostic rate for papillary thyroid carcinoma in the former cohort.

In a previous review of papillary thyroid carcinoma cases in which respondents performed well compared with those in which they performed poorly in the CAP
NGC program, Renshaw et al\textsuperscript{12} noted the paucity of intranuclear cytoplasmic inclusions, the lack of pale chromatin, and the absence of nuclear enlargement in cases in which the performance was poor. Of note, the 4 ThinPrep FNA cases of papillary thyroid carcinomas in the Renshaw et al\textsuperscript{12} study from 2006 were cases with high performance and 100\% correct participant diagnoses. The good performance of participants reviewing the LBP slides in the previous study indicates that at least some cases of papillary thyroid carcinoma are accurately diagnosed with LBP by a broad group of pathologists. The current study averages the performance of the LBP and conventional smears for many individual cases, and on average, performance was worse with LBP than it was with conventional smears for the diagnosis of papillary thyroid carcinoma.

Review of the cases of papillary thyroid carcinomas with LBP preparations and poor performance in the present study suggests that participants may be unaware of the potential for macrofollicular architecture in some papillary thyroid carcinomas. Familiarity with the LBP appearance of papillary thyroid carcinoma is important, with attention to high-magnification features, including subtle nuclear lamina irregularity and fine chromatin texture with occasional intranuclear cytoplasmic inclusions.

The finding that LBP performs better than do conventional smears for benign goiters is not consistent with the suggestion some make that the key diagnostic features of watery colloid, larger-sized tissue fragments, or an inability to quantify the amount of colloid make LBP suboptimal for the diagnosis of goiters.\textsuperscript{2} The low diagnostic accuracy of benign lesions on direct conventional smears has not commonly been observed in other studies. Amrikachi et al\textsuperscript{13} found that only 3\% of cases diagnosed as follicular neoplasms on FNA smears proved to be benign adenomatous nodules. Saleh et al\textsuperscript{14} found that the rate of histologically confirmed benign colloid nodules was similar for LBP and conventional smears.

Our review of conventional smears with poor participant performance on reference benign goiter diagnoses suggested a few under-recognized diagnostic features. It is important to gauge the total amount of colloid in relation to the presence of any microfollicular groups because benign goiters can show a few microfollicular groups in the presence of a large amount of colloid (from predominantly macrofollicular groups). Histiocytes with hemosiderin favor a benign diagnosis (if papillary thyroid carcinoma and Hürthle cell neoplasms are excluded). “Endocrine atypia” (marked variation in nuclear size with a commensurate increase in cytoplasm) is a feature that favors a benign diagnosis. Variation from group to group in any cytologic feature favors a nonclonal benign goiter if papillary thyroid carcinoma can be excluded. Microfollicles bearing flattened rather than cuboidal cells may occur in benign nodules and do not have the ominous significance of microfollicles consisting of cuboidal or columnar cells.\textsuperscript{7}

The rate of correct responses is somewhat lower than that published in many large series on the accuracy of FNA test results. There are many potential reasons for this. First, participants have varying exposure to thyroid FNA slides, and they may not be familiar with the particular preparations sent for their review. Many participants may use a combination of LBP and conventional smears (including possibly air-dried preparations) as well as cell blocks in actual practice. There is likely an advantage to combining different preparation types in the diagnosis of thyroid FNA.\textsuperscript{3,7,11,15} Finally, this survey is an educational experience for participants, and in actual practice, a participant would likely defer a diagnosis or show problematic cases to others before rendering a potentially serious misdiagnosis.

References