Performance of ThinPrep liquid-based cervical cytology in comparison with conventionally prepared Papanicolaou smears: a quantitative survey

Ovadia Abulafia, M.D., John C. Pezzullo, Ph.D., and David M. Sherer, M.D.

Department of Obstetrics and Gynecology, State University of New York at Brooklyn, Downstate Medical Center, 450 Clarkson Avenue, Box 24, Brooklyn, NY 11203, USA

Department of Pharmacology and Biostatistics, Georgetown University Medical Center, Washington, DC 20057

Received 16 October 2002

Abstract

Objective. The goal of this study was to evaluate the performance of ThinPrep, a liquid-based cytology preparation technique, in comparison with conventionally prepared Papanicolaou smears in detecting cervical pathology.

Methods. Forty-seven English-language articles published between January 1990 and September 2002 were identified through Medline and manual searches. After elimination of 5 nonprimary articles, 10 unpaired studies, 5 descriptive articles with insufficient or no data, and 3 articles with data that could not be recast into a consistent format, there remained 24 usable articles. Seventeen articles contained data comparing ThinPrep with conventional cytology for 35,172 patients; and 10 articles compared cytology with histology or other gold standard diagnoses for 21,752 patients. Three of these articles contained both types of comparisons. The standard cytology classification into negative, atypical, low-grade (LGSIL) and high-grade (HGSIL) squamous intraepithelial lesions, and carcinoma was applied; other categorization schemes were recoded as necessary. Concordance estimates, based on five-way and dichotomous (normal/abnormal) classifications, were obtained from the 17 studies containing paired cytology data. Sensitivity and specificity rates were obtained from the 10 studies with paired cytology and histology data.

Results. The two methods tend to agree in 89 and 92% of cases based on the five-level and dichotomous classifications, respectively. ThinPrep was reported as normal in 93.5% of cases of normal conventional smears. The remaining 6.5% of ThinPrep slides were classified as follows: atypical, 4.55%; LGSIL, 1.56%; HGSIL, 0.36%; invasive cancer, 0.007%. Sensitivity rates, relative to histology, were 68% (conventional) and 76% (ThinPrep), and specificity rates were 79% (conventional) and 86% (ThinPrep).

Conclusion. ThinPrep tends to be more sensitive and specific than conventional smears in detecting cervical dysplasia. The increased sensitivity of ThinPrep results in increased cytologic diagnosis of cervical atypia, LGSIL, HGSIL, and invasive cervical carcinoma.

© 2003 Elsevier Science (USA). All rights reserved.

Introduction

The Papanicolaou (Pap) smear has been the cornerstone of screening for cervical neoplasm for the last 50 years [1]. As a result of widespread application of this screening tool in industrialized countries, a significant reduction in the incidence of invasive cervical carcinoma has occurred during this period [2]. Papanicolaou smear has a reported sensitivity of 80% and specificity of 99% [3]. However, false-negative rates as high as 50% have been described [4]. False-negative smears have been attributed to sampling (preparation) errors, presence of obscuring material (blood or mucus), screening, and interpretation errors [5].

ThinPrep (CYTYC Corp., Boxborough, MA) involves a new technique for the preparation of cervical cytology specimens [6]. In this technique, the cytology sample obtained from the transition zone of the uterine cervix is the same as for the conventionally prepared Papanicolaou smear. However, to decrease the presence of obscuring material (blood and mucus) the cervical cytologic sample is dispersed in a liquid suspension and subsequently centrifuged and passed through a filter. The cervical cytology sample assessed is
obtained from the suspension resulting in a monolayer preparation with well-preserved cellular morphology and absence of air-drying, cell crowding, and obscuring cells, with minimal cell overlap.

Numerous studies have compared the performance of cervical slides prepared using the ThinPrep technology with conventionally prepared cervical slides [6–52]. The wide variety of subject populations and sampling schemes among these studies precluded the application of formal meta-analysis to only a limited number of studies. The objective of our study was therefore to perform a “quantitative survey” of as many of the existing studies as possible, evaluating the performance of the ThinPrep liquid-based cervical cytology preparation system in comparison with conventionally prepared Papanicolaou smears. Our specific goals are twofold: (1) to assess the consistency of the two methods, and (2) to assess the accuracy (sensitivity specificity) of each method with respect to a reliable (“gold standard”) diagnosis.

Methods

Data sources

Searches of computerized databases including Medline, and PubMed (National Library of Medicine, Bethesda, MD, USA) were conducted for literature published between January 1990 and September 2002 in the English language and containing the terms “ThinPrep,” “Papanicolaou smear,” “liquid-based cytology,” and “cervical dysplasia” [6–52]. Cross-referencing was also used. These searches revealed 47 relevant articles. Full-text copies of all articles were retrieved; these were reviewed by one person.

Study selection

Of the 47 articles identified, 5 were found to be descriptions or reviews of other published work. Another 10 articles described unpaired studies, and were unsuitable for our planned analyses. Another 5 articles were primarily descriptive, with insufficient or no tabulated data, and there were three articles whose data could not be recast into a format consistent with the others. After elimination of these, there remained 24 usable articles. Seventeen of these articles contained paired data comparing ThinPrep with conventional cytology on a total of 35,172 patients; and 10 articles contained paired data comparing cytology with histology or with some other “gold standard” (or approximation thereto) such as a clinical diagnosis from subsequent biopsy or surgery, or a consensus of senior cytologists or pathologists, with formal diagnosis on a total of 21,752 patients. Three of these articles contained data suitable for both kinds of analyses.

Analyses

Data regarding the agreement between ThinPrep and conventionally prepared Papanicolaou smears, and between histology findings following both preparation methods were extracted and placed into common, consistent tabular structure. Standard cytology classification into negative, atypical, low-grade (LGSIL) and high-grade (HGSIL) squamous intraepithelial lesions, and carcinoma was used throughout. Results were also dichotomized into normal and abnormal smears. For the 17 paired studies in which each sample was evaluated by both methods, the data were expressed as a 5 × 5 cross-tabulation of ThinPrep × conventionally prepared smear results. The numbers were normalized by expressing each cell as a percentage of the total number of samples.

Several methods for combining these 17 concordance tables into an overall weighted “consensus” table were explored: “patient-level averaging,” in which the patient counts in each cell were summed across the studies; “study-level averaging,” in which the normalized percentages were averaged across studies; and a “study-level median” approach in which, for each cell, the median of the percentages across all studies was taken. Each method had advantages and disadvantages, and the three methods produced consensus tables that were in reasonably good agreement with each other. The decision was made to use patient-level averaging, realizing that under this approach larger studies would exert more influence on the consensus table than smaller studies. The heterogeneity of the studies we analyzed precludes the meaningful estimation of confidence intervals around the overall average rates; we have therefore chosen not to display such intervals.

A measure of overall concordance is the fraction of specimens for which the specific results (expressed as one of the five categories defined above) were the same for the two methods. This measure of concordance was calculated for each study, along with exact 95% confidence interval (95% CI) based on binomial distribution.

Another measure of overall agreement was the fraction of specimens for which the dichotomous results, negative or positive (atypical LGSIL, HGSIL, or carcinoma), were different for the two methods. Exact 95% confidence intervals were also calculated for this measure.

To adjust for concordant results that might arise simply from chance (even if there were no agreement between the two methods), Cohen’s κ was calculated for the overall five-way and dichotomous cross-tabulations.

Since cervical cytology is considered a screening test and not a definitive diagnosis, sensitivity and specificity of both modalities could be calculated only if correlation between the cervical smear and histology diagnosis were available. For the 10 studies containing the required paired cytology and histology data, the results were cross-tabulated, and sensitivity and specificity rates were calculated and averaged across studies. A few of the 10 studies had information
suitable for only one rate (sensitivity or specificity) or for only one modality (conventional or ThinPrep). For analytical purposes, atypia was included with the negatives. We designed so that the positive data pertained to dysplasia and changes associated with carcinoma only, thus avoiding inclusion of atypia of repair or inflammatory changes in the positives.

Results

Concordance between ThinPrep and conventionally prepared Papanicolaou smears

Table 1 summarizes the overall agreement between the two methods derived by patient level, from the 17 studies containing paired (ThinPrep versus conventionally prepared Papanicolaou smears) data.

As can be seen, the ThinPrep and conventionally prepared Papanicolaou smears tend to agree overall about 89% of the time based on the five-level diagnosis classification. Adjusting for chance occurrences of concordant results produces a \( \kappa \) measure of concordance of 67%. Under a dichotomous (positive or negative) classification, the two methods agree about 92% of the time, with an adjusted \( \kappa \) of 73%. When the conventionally prepared Papanicolaou smears were reported as negative the ThinPrep prepared slide was concordant in about 93.5% of cases. The remaining 6.5% of ThinPrep slides were classified as follows: atypical, 4.55%; LGSIL, 1.56%; HGSIL, 0.36%; invasive cancer, 0.007%.

Figure 1 depicts for each of the 17 paired studies, the percentage of cases for which the two methods produced different results. It can be seen that in general the two methods tend to agree in about 90% of cases and disagree in 10% of cases, although across individual studies the disagreement was as low as 1% and as high as 20%. If the

Table 1
Overall percentage agreement obtained by summing patient counts across all studies, then calculating percentages ("patient-level" means)

<table>
<thead>
<tr>
<th>Diagnosis by ThinPrep</th>
<th>Diagnosis by conventional Papanicolaou smear</th>
</tr>
</thead>
<tbody>
<tr>
<td>Negative</td>
<td>Atypical</td>
</tr>
<tr>
<td>Negative</td>
<td>Atypical</td>
</tr>
<tr>
<td>LGSIL</td>
<td>HGSIL</td>
</tr>
<tr>
<td>Carcinoma</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Negative</th>
<th>Atypical</th>
<th>LGSIL</th>
<th>HGSIL</th>
<th>Carcinoma</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Negative</td>
<td>76.2%</td>
<td>2.30%</td>
<td>0.722%</td>
<td>0.145%</td>
<td>0.003%</td>
<td>79.3%</td>
</tr>
<tr>
<td>Atypical</td>
<td>3.71%</td>
<td>3.92%</td>
<td>0.520%</td>
<td>0.148%</td>
<td>0.014%</td>
<td>8.31%</td>
</tr>
<tr>
<td>LGSIL</td>
<td>1.27%</td>
<td>0.873%</td>
<td>4.62%</td>
<td>0.384%</td>
<td>0.006%</td>
<td>7.15%</td>
</tr>
<tr>
<td>HGSIL</td>
<td>0.29%</td>
<td>0.210%</td>
<td>0.370%</td>
<td>3.51%</td>
<td>0.071%</td>
<td>4.45%</td>
</tr>
<tr>
<td>Carcinoma</td>
<td>0.006%</td>
<td>0.011%</td>
<td>0.006%</td>
<td>0.048%</td>
<td>0.671%</td>
<td>0.742%</td>
</tr>
<tr>
<td>Total</td>
<td>81.6%</td>
<td>7.35%</td>
<td>6.24%</td>
<td>4.24%</td>
<td>0.765%</td>
<td>100%</td>
</tr>
</tbody>
</table>

Fig. 1. Discordant classification between ThinPrep and conventionally prepared Papanicolaou smears based on five-way classification (negative, atypia, LGSIL, HGSIL, carcinoma).
results are aggregated into dichotomous (positive or negative) findings, it is seen that the percentage of discordant cases is somewhat reduced, yet still centered around 10% disagreement (Fig. 2). To determine whether one method was superior at detecting abnormalities that the other missed, we analyzed discordant results (in which one method reported positive and the other reported negative).

Fig. 2. Discordant classification between ThinPrep and conventionally prepared Papanicolau smears based on two-way classification (negative/abnormal).

Fig. 3 and 4 demonstrate, for each paired study, the relative ability of the two methods to produce a positive result missed by the other. Fig. 3 demonstrates these data as ratios of ThinPrep-only to conventional-only positives. Fig. 4 demonstrates these data as a percentage (of all two-way discrepant) discordant results. It is seen that in most of the paired studies “ThinPrep positive, conventional negative” occurs

Fig. 3. Ratio of ThinPrep only to conventional only, positives.
more frequently than "ThinPrep negative, conventional positive," although the difference is significant only in four (generally the larger) of the studies.

**Concordance with histological (or other "final") findings**

**Sensitivity**

Ten studies contained paired information regarding the extent to which ThinPrep and/or conventional cytology agreed with histology or other authoritative diagnostic findings. This permitted the estimation of sensitivity and specificity values for each method, respectively. Because of the wide variability in the prevalence of abnormalities across studies (some were sampled from normal screening populations, others from high-risk subgroups), no meaningful conclusions could be drawn regarding positive and negative predictive values. Fig. 5 illustrates, for each appropriately paired study, the comparative sensitivity of ThinPrep and conventional Papanicolaou smears in detecting abnormalities. It can be seen that the different studies demonstrate a relatively wide range of sensitivity (from about 50 to 90%) but in most studies ThinPrep is more sensitive than conventionally prepared cervical cytology slides. In two of the larger studies this difference was statistically significant. An overall sensitivity of 76% for ThinPrep-prepared cervical slides versus 68% for conventionally prepared Papanicolaou smears was demonstrated.

**Specificity**

One of the 10 articles containing cytology/histology data was unsuitable for estimating specificity as it contained information only on histologically abnormal patients. Our analysis, based on the other nine articles, reveals an overall specificity of 86% for ThinPrep-prepared cervical slides versus 79% for conventionally prepared Papanicolaou smears.

Fig. 6 illustrates the overall specificity rates, which generally are in the range 80 to 90%. Again, it is seen that ThinPrep is usually more specific than conventionally prepared Papanicolaou smears (although the difference is usually not statistically significant).

**Discussion**

Our analysis of 17 studies indicates that, in general, the results of the two modalities tend to be in agreement about 90% of the time. ThinPrep classifies 6.5% of the "conventionally normal" slides as being abnormal, with 0.36% being judged HGSIL and 0.007% being judged invasive carcinoma. Accordingly, the ThinPrep modality "salvages" 6.5% of the conventionally prepared Papanicolaou smears (considered "false-negative"). To the extent that the averages obtained from our heterogeneous group of studies are meaningful, they would imply that from the 45 million conventionally prepared Papanicolaou smears that are interpreted as normal each year in the United States, uniform utilization of ThinPrep technique could lead to the additional cytological diagnosis of 162,000 cases of HGSIL and more than 3,000 cases of invasive cervical carcinoma annually.

While the fraction of HGSIL and invasive carcinoma smears revealed by the ThinPrep modality from cervical slides for which the conventionally prepared Papanicolaou...
smear diagnosis was atypical or LGSIL was higher, the latter would usually have triggered colposcopy assessment and biopsies in any event. As a result, the ThinPrep modality would not have contributed directly to the increase in the rate of detection of cases of HGSIL and invasive carcinoma, which constitute the target population to be detected to significantly reduce the incidence and related mortality of cervical carcinoma.

Fig. 5. Sensitivity of ThinPrep and conventionally prepared Papanicolaou smears, each versus a “gold standard” such as histology, subsequent surgery, biopsy, or expert consensus.

Fig. 6. Specificity of ThinPrep and conventionally prepared Papanicolaou smears, each versus a “gold standard” such as histology, subsequent surgery, biopsy, or expert consensus.
Our data indicate that in the 10 studies that compared cytology with biopsy-proven diagnosis ThinPrep tended to have about 8% higher sensitivity and about 7% higher specificity.

A recent paper by Bernstein et al. [53] describes a meta-analysis of 25 prospective studies of Thin Prep and conventionally prepared Papanicolaou smears that used the Bethesda nomenclature. They restricted their analyses to four outcome measures: ACUS, LGSIL, HGSIL, and sample adequacy. For each of the four outcomes, they compared the performance of “Direct-to-Vial,” “Split-Sample,” and “Split-Sample with ThinPrep 2000 processor” to conventionally prepared Papanicolaou smears by calculating, for each method, the percentage of all specimens falling into that specific outcome. These authors calculated a ThinPrep/ conventionally prepared Papanicolaou smear odds ratio. They did not calculate overall measures of discrepancy between the two methods, nor did they compare each method with the gold standard of histology to obtain accuracy, sensitivity, or specificity comparisons. They recommended “further studies analyzing the histologic correlation of these findings . . . to evaluate more definitively the sensitivity and specificity of ThinPrep as a more accurate screening method.”

Performing a rigorous meta-analysis of these studies is problematic because of the wide variety of conditions under which the individual investigations were performed. Nanda et al. [54] conducted an extremely rigorous, yet very restrictive, meta-analysis of the accuracy of conventionally prepared Papanicolaou smears versus “gold standard” and of ThinPrep versus “gold standard.” These authors accepted only studies that complied perfectly with their criteria. For example, they would not consider a study that compared conventionally prepared smears with ThinPrep, but only conventionally prepared with “gold standard” or ThinPrep with “gold standard.” As a result they deemed only three studies worthy of inclusion in their analysis [26,27,40]. We felt that it would be more useful to practicing physicians to collect information on the broadest possible set of studies, recast the results into consistent coding conventions, and present them in a readily interpretable graphic and tabular format, knowing that the heterogeneities inherent in such a collection will preclude the rigorous estimation of average effect sizes or the calculation of confidence intervals around these estimates.

In summary, the ThinPrep liquid-based modality appears more sensitive and specific than conventional Papanicolaou preparation. The added sensitivity results in increased detection of patients with LSIL/atypical cervical smears and HSIL. According to our analysis, wide application of the ThinPrep technology potentially may result in cytological diagnosis of an additional 162,000 patients with HSIL and 3000 patients with invasive cervical carcinoma whom otherwise would have been missed by utilization of conventionally prepared Papanicolaou smears (false-negative smears). We conclude that ThinPrep liquid-based cervical cytology should be considered the standard of care, as this technology has the potential to further reduce the incidence and increase the survival of patients with invasive cervical carcinoma as a result of early detection.

References


