



Effect of study design and quality on unsatisfactory rates, cytology classifications, and accuracy in liquid-based versus conventional cervical cytology: a systematic review

Elizabeth Davey, Alexandra Barratt, Les Irwig, Siew F Chan, Petra Macaskill, Patricia Mannes, A Marion Saville

Lancet 2006; 367: 122–32

See [Comment](#) page 88

Screening and Test Evaluation Program, School of Public Health, University of Sydney, Australia (E Davey MPH, A Barratt FAFPHM, Prof L Irwig FFPHM, S F Chan MAppStat, P Macaskill PhD, P Mannes MPH); and Victorian Cytology Service, Victoria, Australia (A M Saville MBChB)

Correspondence to:

Dr Elizabeth Davey, Screening and Test Evaluation Program, School of Public Health, Edward Ford Building, A27, University of Sydney, NSW 2006, Australia
daveye@health.usyd.edu.au

Background Liquid-based cytology is reported to increase the sensitivity of cervical cytology and the proportion of slides that are satisfactory for assessment, in comparison with conventional cytology. Although some countries have changed to liquid-based cytology for cervical screening, controversy remains. We reviewed the published work to assess the performance of liquid-based cytology relative to conventional cytology in primary studies assessed to be of low, medium, or high methodological quality.

Methods 56 primary studies were reviewed and assessed with strict methodological criteria. Liquid-based cytology and conventional cytology were compared in terms of the percentage of slides classified as unsatisfactory, the percentage of slides classified in each cytology category, and the accuracy of detection of high-grade disease. Data were examined for studies overall and in strata to examine the effect of study quality on results.

Findings The median difference in the percentage of unsatisfactory slides between liquid-based cytology and conventional cytology was 0·17%. Only one small study was a randomised controlled trial. The classification of high-grade squamous epithelial lesion varied according to study quality ($p=0\cdot04$), with conventional cytology classifying more slides in this category than did liquid-based cytology in high-quality studies ($n=3$) only. In medium-quality ($n=30$) and high-quality studies, liquid-based cytology classified more slides as atypical squamous cells of unknown significance than did conventional cytology when compared with low-quality studies ($n=17$; $p=0\cdot05$). Only four studies provided sufficient verified data to allow estimation of sensitivity and specificity and comparison of test accuracy.

Interpretation We saw no evidence that liquid-based cytology reduced the proportion of unsatisfactory slides, or detected more high-grade lesions in high-quality studies, than conventional cytology. This review does not lend support to claims of better performance by liquid-based cytology. Large randomised controlled trials are needed.

Introduction

For over 30 years, screening for cervical cancer has used the conventional Papanicolaou (Pap) smear. Despite limited accuracy of the test,^{1,2} the incidence of cervical cancer has fallen substantially.³ Liquid-based cytology has since been developed as an alternative to conventional cytology.

Liquid-based cytology involves rinsing the sampling tool into a vial of liquid to produce a suspension of cells, from which a monolayer of cells on a slide is prepared. Slides produced in this way can be read more quickly than conventional cytology slides^{4,5} and the liquid sample can be used for human papillomavirus (HPV) DNA testing. Automated reviews can be done with both conventional cytology and liquid-based cytology.

Liquid-based cytology has been compared with conventional cytology in many studies; most report increased sensitivity for detecting pathological changes and a higher proportion of slides that are adequate for assessment.^{6–9} Several systematic reviews have been done^{6,10–15} with diverse conclusions. The UK National Institute for Clinical Excellence¹⁶ reported that liquid-based cytology improved sensitivity slightly and that, in a pilot study in England of 178 000 slides, percentages of unsatisfactory slides decreased from 9·1% to an average

of 1·6% after conventional cytology was replaced with liquid-based cytology. (That report did not satisfy the criteria for inclusion in our review.) Other reviews reached similar conclusions.^{6,13,17} However, Sulik and colleagues¹² reported little difference in performance between liquid-based cytology and conventional cytology. Nanda and colleagues¹¹ concluded that there are insufficient high-quality data to compare liquid-based cytology with conventional cytology, and Moseley and Paget's extensive review¹⁴ concluded that the data do not support the implementation of liquid-based cytology and recommended further studies.

Some countries, including the USA and the UK, are incorporating liquid-based cytology into screening programmes. Many countries have been reluctant to adopt liquid-based cytology without definitive evidence of higher or at least equivalent accuracy. If equivalence can be shown, other characteristics such as greater reproducibility, lower cost, or the capacity for HPV DNA testing could make liquid-based cytology more desirable than conventional cytology in screening programmes.

Most reviews noted that the quality of primary studies varies and that many studies had methodological deficiencies, including inadequate application of reference standards and inadequate follow-up of negative

cytology. The quality of studies that assess tests can affect conclusions,¹⁸ and low-quality studies consistently overestimate the accuracy of tests.¹⁹ Standards have been established for the design and reporting of medical tests.²⁰ Two reviews^{15,17} systematically assessed the quality of primary studies, but neither examined whether the results of test comparison (cytological classifications or accuracy) varied according to study quality.

Furthermore, the accuracy of many tests is a trade-off between sensitivity and specificity. Thus, even if liquid-based cytology does improve sensitivity (true-positive rate) for high-grade abnormalities, it could simultaneously increase the number of low-grade abnormalities (false positives), which are less likely to represent serious disease but might trigger clinical investigation. These false positives are undesirable in a screening programme.

We did a systematic review of studies that assessed the use of liquid-based cytology as a replacement for conventional cytology in cervical screening. The aims of our review were to examine the relative performance of the two approaches, and to assess how results varied by study design characteristics, mainly quality, with regard to: the proportion of unsatisfactory slides; the proportions of slides classified as normal, atypical squamous cells of undetermined significance (ASCUS), low-grade squamous intraepithelial lesion (LSIL), and high-grade squamous intraepithelial lesion (HSIL); and the accuracy for detecting reference standard HSIL.

Methods

Inclusion criteria

For inclusion in the review, reports had to be of primary studies published in peer-reviewed journals and had to describe direct comparison of liquid-based cytology as a replacement for conventional cytology, with both techniques done by manual reading (not an automated screening system).

Search strategy

We searched MEDLINE from 1966 to Jan 15, 2004 with the MESH headings: Exp cervix neoplasms/or cervical cancer [Text Word]; Exp mass screening/or screening [Text Word]; Thinprep [Text Word]; Liquid [Text Word]; Fluid [Text Word]; Cytorich [Text Word]; Autocyte [Text Word]; (Thinprep [Text Word] OR Liquid [Text Word] OR Fluid [Text Word] OR Cytorich [Text Word] OR Autocyte [Text Word]); (Exp cervix neoplasms/or cervical cancer [Text Word]) AND (Exp mass screening/or screening [Text Word]) AND (Thinprep [Text Word] OR Liquid [Text Word] OR Fluid [Text Word] OR Cytorich [Text Word] OR Autocyte [Text Word]).

The most recent search was done on Jan 15, 2004. This search yielded 145 articles, from which we retrieved 29 that were relevant. 116 were excluded because the studies they described did not meet the inclusion criteria or they were duplicate reports of the same study. In the case of duplicates, we extracted data from the paper

presenting the most comprehensive results. To ensure capture of all trade names for liquid-based cytology, an additional search was run including the term SurePath; no additional eligible articles were obtained. We also reviewed reference lists of published reviews and primary studies. This strategy yielded 27 additional articles. Two reviewers assessed eligibility, and articles were included after consensus was reached. From two papers we extracted two full data sets from each, representing distinct populations. An identical search was run in EMBASE from 1973 until March 30, 2004, but did not identify any additional articles.

Assessment of study design characteristics and quality

Two reviewers, unaware of each other's results or the study results, did the appraisal independently. Each article was assessed with a detailed list of appraisal items, which included study question, year of publication, liquid-based cytology proprietary name (ThinPrep, AutocytePrep, CytoRich, Cytoscreen, or Papspin), study design, and items relating to validity and applicability. Study design was based on either paired (generally split-sample) or independent (direct-to-vial) samples. Unlike independent studies, split-sample design might disadvantage liquid-based cytology because residual cells are used to prepare the liquid-based cytology slide (webappendix).

Validity included whether a reference standard had been used, type of reference standard (consensus cytology, colposcopy, biopsy, and histology, or a combination of these), method of allocating participants to tests, masked reading of slides, and masked (to test) application of the reference standard (table 1). Applicability items covered study setting (screening, referral clinic, or mixed settings), reproducibility, and the country of study. Disagreements were resolved by consensus and, if necessary, review by a third person. The authors of two papers were contacted to verify details about masking.^{21,22} Both responded.

Four categories of study quality were defined according to methodological criteria. Table 2 defines the categories and how criteria differ for paired and independent studies. In high-quality studies, both tests and the reference standard were read without knowledge of any other results. Medium-quality studies provided evidence that a reference standard had been applied to at least some cases, but methods of masking or verification or both did not meet requirements for high quality. By definition, studies of high or medium quality could potentially provide data from which sensitivity and specificity could be calculated. Low-quality studies did not use a reference standard, so could provide no data on test accuracy (webappendix).

Extraction of results

After the quality appraisal, the full articles were made available and results were again extracted by two

See Lancet Online for webappendix

Information extracted	
Study design	Independent-sample or paired-sample design If independent: whether groups were randomised*
Study setting	Family practices/referral clinics/both/other
Study validity	
Reference standard	
Was a reference standard used?	Type used: histology/cytology/whether consensus reading was used
Application of reference standard	In paired-sample studies: were all tests, all positive tests, or all discordant tests verified, or was verification partial or unclear? In independent-sample studies: were all positive tests or all positive tests plus a random sample of negative tests verified, or was verification partial or unclear?
Was reference standard assessed without knowledge of test results?	Yes/no/unclear
Were different reference standards applied within the study?	Yes/no/unclear
Allocation	In independent-sample studies: were women randomly allocated the test? If not allocated randomly, were controls concurrent or historical?
Masking of test reading	In paired-sample studies: were tests read without knowledge of the other's results?

*See webappendix.

Table 1: Criteria for assessment of study quality

reviewers. Differences were resolved by consensus and, in some cases, review by a third person. In eight cases, authors were contacted for confirmation of details of tabulated data. Five responded and provided clarification.^{21,23–26}

Analysis

Unsatisfactory slides

For each study, the percentage of slides classified as unsatisfactory by conventional cytology was subtracted from the percentage classified as unsatisfactory by liquid-based cytology, to give a difference in percentages. These differences, sorted by sign and magnitude, were displayed in a forest plot. The χ^2 statistic²⁷ was used to test for heterogeneity between studies. A summary estimate of the differences (and 95% CI) was obtained by the random-effects method of Dersimonian and Laird.²⁷

Studies were stratified by quality, setting, design (paired or independent), and liquid-based cytology proprietary name, both individually and grouped in two categories; one was ThinPrep, the other AutocytePrep and CytoRich combined. Random-effects meta-regression was used to test for association, thereby taking account of study size. Because the normality assumption for the random effects was not met (because the data were skewed), non-parametric methods were also applied (Mann-Whitney test for two strata, Kruskal-Wallis test for three or more strata). P values less than 0.05 were taken as statistically significant. Significance tests, and the direction of the effects, were in close agreement for the two approaches, except where indicated in the results. The non-parametric estimates are reported in the results.

An improvement in cytology reading could have occurred over time as cytologists became more

experienced with liquid-based cytology and its use became more widespread. None of the papers quantified cytologists' training and experience in sufficient detail for us to include this measure in our analysis. Several studies reported having trained staff in liquid-based cytology, and in one study cytologists' experience was reported to be extensive in conventional cytology but limited in liquid-based cytology.²¹ We used year of publication as a proxy for liquid-based cytology reading experience. We acknowledge the limitations of this proxy because selection of readers could have occurred within studies, and laboratories introduced liquid-based cytology at different times, both within and between countries. Association between differences in percentages and year of publication was assessed with Spearman rank correlation analysis.

Cytology classifications

For each study the percentage of slides classified as normal, ASCUS, LSIL, and HSIL by conventional cytology was subtracted from the percentage classified in each category by liquid-based cytology, to give a difference in percentage for each cytological classification. HSIL was chosen as a category, instead of HSIL and cancer combined because: many studies did not present numbers of cancer cases separately and, of those that did, numbers generally included adenocarcinoma; and numbers of separately identified cancer cases were small (0.07%). Studies were stratified by quality, study design, setting, and liquid-based cytology proprietary name as above. Random-effects meta-regression was used as above to test for association. Again, the distribution of study estimates was skewed in some analyses, therefore non-parametric methods were also used.

To investigate the effect of paired versus independent study design on cytological classifications, we examined differences in percentages within medium-quality studies, because this stratum was the only one that provided adequate and balanced numbers of studies of both designs to allow a valid comparison.

Accuracy

Studies of medium and high quality that presented data from which we could calculate both sensitivity and specificity of liquid-based cytology and conventional cytology to detect high-grade reference standard disease were plotted in receiver operating characteristic (ROC) space. An additional high-quality study that quoted sensitivity and specificity, without presenting the raw data, was also included in the plot. Because of the small number of studies and substantial heterogeneity of these results, summary ROC analysis was not undertaken.

Role of the funding source

The sponsor had no role in study design, data collection, data analysis, data interpretation, or the writing of the

report. The corresponding author had full access to all the data in the study and had final responsibility for the decision to submit for publication.

Results

56 eligible studies (58 datasets) were identified.^{4,5,7-9,21-26,29-73} In addition, the search yielded two potentially eligible articles in languages other than English. Neither provided data that would contribute to assessment of accuracy, so they were not included in the review.

Study design was generally poor, with few studies that used adequate methods to compare tests. Table 2 presents the studies according to our quality classification. No study was classified as ideal quality, because none randomly assigned women to liquid-based cytology or conventional cytology and verified at least all positive slides with a masked reference standard. Five studies (six datasets) met criteria for high quality. All five were paired-sample studies, which reported tests read without knowledge of the other, a reference standard applied without knowledge of test results, and verification of at least all discordant slides. All but one used split-sample specimen collection. One study²² (which provided one dataset) was paired and used direct-to-vial sampling, since two specimens were taken from each woman. Randomisation was used to establish whether the liquid-based cytology or conventional cytology specimen was taken first. 32 studies of either design type were classified as medium quality. These studies reported the use of a reference standard but did not meet masking or verification, or both, requirements for classification as high quality. The remaining 19 were classified as low quality because they used no reference standard.

Table 3 shows the distribution of studies according to study design, quality, and setting. Apart from the five paired high-quality studies, the remaining 34 paired studies were evenly distributed between medium and low quality categories. Of the 32 medium-quality studies, reference standard assessment was reported to be done without knowledge of test results in only one (paired) study³⁸ but results of verification of all discordant test results were not presented. Of the 17 studies that used an independent-sample study design, 15 were classified as medium-quality and two as low-quality. Only one used randomisation to allocate women, but it did not adequately apply a reference standard.⁵¹ None of the 15 medium-quality independent studies reported the use of masked assessment of reference standard and none verified at least a random sample of test negatives and all test positives.

22 studies were set in referral clinics, six in screening settings, 14 in mixed screening and clinic settings, and setting was unspecified in 14. Of the 58 datasets, the liquid-based cytology proprietary name was ThinPrep in 39 datasets, AutocytePrep in eight, CytoRich in nine, Cytoscreen in one, and Papspin in one.

	Terminology within review*	Number of studies
Independent randomised sample study with:		
Verification, by a masked reference standard, of at least all positive slides	Ideal	0
Paired-sample study with:		
Tests read without knowledge of the other's results	High quality	5
Masked reference standard		
Verification of at least all discordant slides		
Independent non-randomised sample study with:		
Masked reference standard	High quality	0
Verification of all positive and of at least a random sample of negative slides		
Either paired-sample or independent-sample study with:		
Reference standard applied but methods of masking and/or verification that do not meet requirements for high quality	Medium quality	32
Either paired-sample or independent sample study with:		
No reference standard	Low quality	19

*High quality=references 4, 21-24; medium quality=7-9,25,26,29-31,34-38,42,44,48,50-53,55,58,61-66,68,69,71,73; low quality=5,32,33,39-41,43,45-47,49,54,56,57,59,60,67,70,72.

Table 2: Categories of quality

Data were pooled from 48 datasets (46 studies) that provided information on percentages of unsatisfactory slides. Overall, 3646 (0.75%) of a total of 483 050 liquid-based cytology slides were unsatisfactory, whereas 5389 (0.81%) of a total of 662 401 conventional cytology slides were unsatisfactory.

A forest plot of the differences in percentages of unsatisfactory slides (liquid-based cytology – conventional cytology) is shown in figure 1. Differences ranged from –10.80% to 8.91%. However, the larger studies (with the narrower confidence intervals) all had differences close to zero. There was strong evidence of heterogeneity ($p < 0.0001$). Only four studies reported more than 5% of liquid-based cytology slides as unsatisfactory, and only three studies reported more than 5% of conventional cytology slides as unsatisfactory. Only one study reported more than 5% of both liquid-based cytology and conventional cytology as unsatisfactory.

The summary estimate of the differences was –0.14% (95% CI –0.33% to 0.06%). The median of the

	Paired n=39(41)	Independent n=17	Total n=56(58)
Quality			
Ideal	0	0	0
High	5 (6)	0	5 (6)
Medium	17	15	32
Low (no reference standard)	17 (18)	2	19 (20)
Setting			
Screening	3	3	6
Referral clinic	19	3	22
Mixed	10 (11)	4	14 (15)
Unspecified	7 (8)	7	14 (15)

Data are numbers of studies. Two studies contained two datasets. Where these studies appear, the number of datasets is given in parentheses.

Table 3: Description of studies

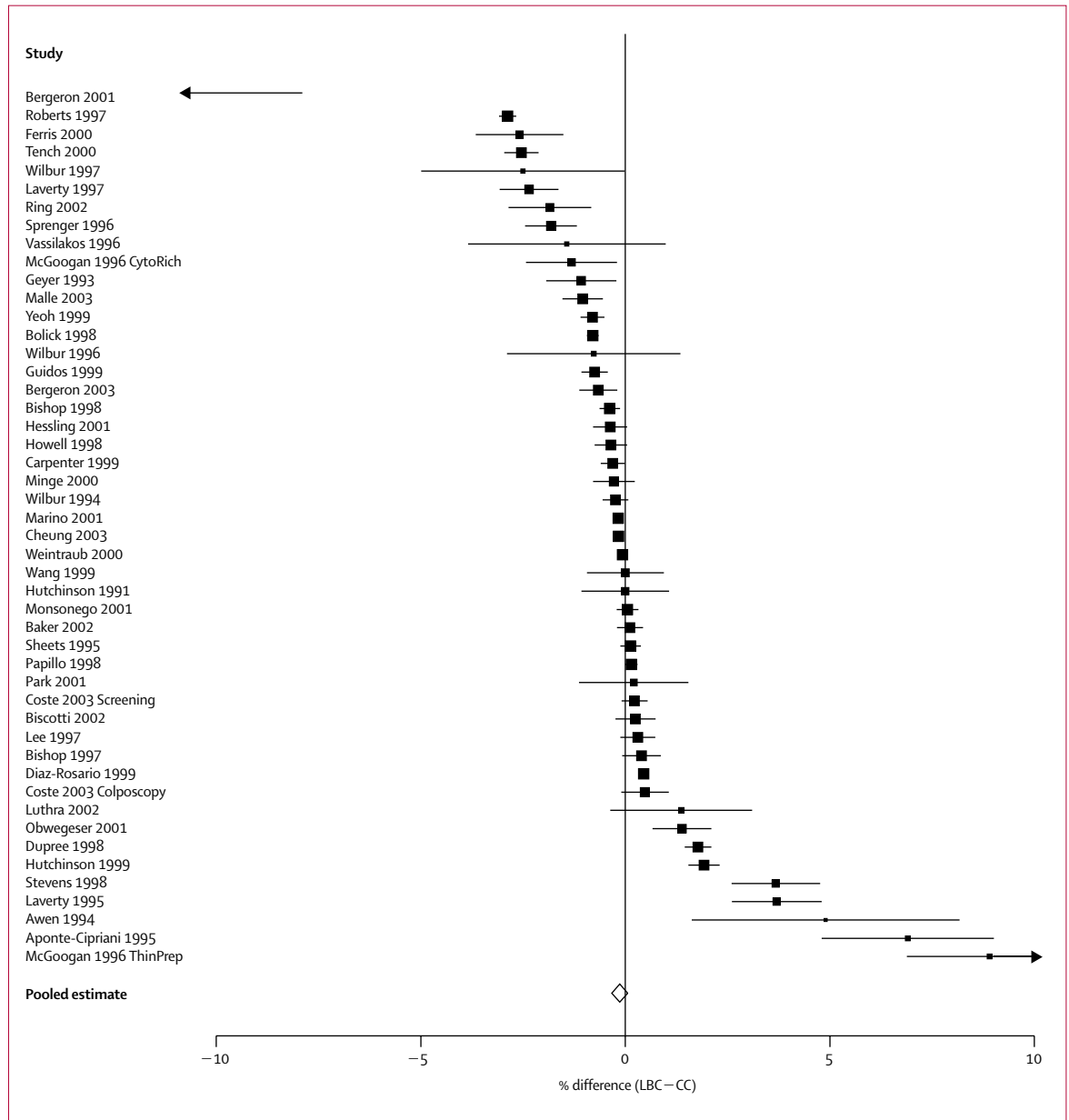


Figure 1: Differences in percentages of unsatisfactory slides (LBC-CC)
 CC=conventional cytology. LBC=liquid-based cytology. Data are derived from 48 datasets that contained data on unsatisfactory slides.

differences across all studies (liquid-based cytology–conventional cytology) was -0.17% (IQR -0.98% to 0.37%). There were no significant differences between the medians of the differences in percentages of unsatisfactory slides according to study quality, study design, or setting. Estimates of medians of the differences in percentages of unsatisfactory slides when examined by quality were: -0.07% , -0.17% , and -0.12% for high, medium, and low quality, respectively. When examined by study design, estimates were zero for paired studies and -0.17% for independent studies. Estimates according to setting

were 1.02% , 0.03% , -0.24% , and -0.38% for screening, referral clinic, mixed screening and referral, and unspecified settings, respectively. Meta-regression results were consistent with these findings, except for setting, for which evidence of an association was noted; however, the differences were not clinically significant, ranging from 0.4% for screening settings to -0.4% in mixed settings.

A significant difference was noted according to liquid-based cytology proprietary name when examined individually ($p=0.01$; ThinPrep $n=32$; AutocytePrep $n=7$; CytoRich $n=8$; Cytoscreen was omitted because $n=1$) and

	Liquid-based cytology (%) (n=562 662)	Conventional cytology (%) (n=688 035)
Normal	518878 (92.22%)	646014 (93.89%)
ASCUS	22986 (4.09%)	26561 (3.86%)
LSIL	15041 (2.67%)	10236 (1.49%)
HSIL	5348 (0.95%)	4771 (0.69%)
Cancer	409 (0.07%)	453 (0.07%)

Data are derived from the 52 datasets included in table 5.

Table 4: Cytology classifications by liquid-based cytology and conventional cytology

when grouped according to manufacturer ($p=0.003$; ThinPrep $n=32$; AutocytePrep or CytoRich $n=15$). However, these differences were small (0.12% for ThinPrep, and -1.08% for AutocytePrep or CytoRich). When differences between percentages of unsatisfactory slides in liquid-based cytology and conventional cytology were examined according to year of publication, there was no evidence of correlation ($p=0.65$).

Cytological classifications from the 52 datasets (50 studies) for which complete data were available are shown in table 4. Overall, more than 1.25 million slides were included. Since cytological classifications are a continuum, a change in the percentage of slides in one category must change the percentage in at least one other category. The distributions shown in table 4 might, therefore, result from several shifts in classification in either, or both, tests. Conventional cytology classified more slides as normal and ASCUS than did liquid-based cytology, whereas liquid-based cytology classified more slides as LSIL and HSIL than did conventional cytology (table 5). However, there was evidence of variation in the classification of HSIL by study quality ($p=0.04$), with conventional cytology classifying more slides as HSIL than did liquid-based cytology in high-quality studies only.

Liquid-based cytology classified more slides as LSIL than did conventional cytology in all quality categories. The median of the differences was greater in high-quality than in low-quality studies, but this finding was not significant ($p=0.10$). Overall, conventional cytology classified more slides as ASCUS than did liquid-based cytology. However, this effect was apparent only among low-quality studies, whereas in studies of medium and high quality, liquid-based cytology classified more slides as ASCUS than did conventional cytology ($p=0.05$).

Differences between liquid-based cytology and conventional cytology classifications of normal, ASCUS, and LSIL slides did not vary between paired and independent studies ($p=0.17$, 0.44 , and 0.08). Although liquid-based cytology classified more slides as HSIL than did conventional cytology in both study designs, the difference was greater in independent than in paired studies ($p=0.03$). There was no evidence of differences between liquid-based cytology and conventional cytology

in any cytological classification according to setting. Results obtained with meta-regression were consistent with those obtained with non-parametric methods for analyses by quality, study design, and setting. No significant difference was noted in any cytological classification when examined by liquid-based cytology proprietary names, individually, or grouped as ThinPrep, or either AutocytePrep or CytoRich ($p=0.35$, $p=0.35$, $p=0.39$, and $p=0.20$ for normal, ASCUS, LSIL, and HSIL, respectively, in group analyses).

Within medium-quality studies, differences (liquid-based cytology minus conventional cytology) in percentages of slides classified as normal, ASCUS, LSIL, and HSIL were -1.20% , 0.14% , 1.30% , and 0.30% , respectively, for independent studies ($n=15$), and -1.06% , 0.37% , 0.61% , and 0.09% , respectively, for paired studies ($n=15$). Although liquid-based cytology detected more LSIL and HSIL than did conventional cytology in both study designs, there was no difference between study designs for any cytological classification: ($p=1.00$, 0.49 , 0.12 , and 0.25 , respectively).

We examined the accuracy of two test thresholds (ASCUS and LSIL cytology) against one reference standard threshold (HSIL). Of the 56 studies, 37 used a reference standard. Five studies used cytology alone, 26 used colposcopy with or without histology, and six used a combination of cytology and histology. Of these studies, 33 did not provide clear thresholds or adequate verification data. Of the five high-quality studies (all paired), two provided data from which we could calculate sensitivity and specificity. One study provided data for sensitivity but not specificity, one presented data against a reference standard threshold of LSIL, and one quoted sensitivity and specificity but did not present data from which they could be calculated.

	Normal	ASCUS	LSIL	HSIL
All studies	-0.48 (IQR -2.45 to 0.42)	-0.19 (IQR -1.47 to 0.68)	0.80 (IQR 0.40 to 1.85)	0.12 (IQR -0.09 to 0.39)
Study quality				
High	-0.33	0.08	1.53	-0.69
Medium	-1.13	0.26	1.08	0.26
Low	0.23	-0.72	0.49	0.04
p^*	0.01	0.05	0.10	0.04
Study design				
Paired	-0.30	-0.25	0.61	0.07
Independent	-1.20	0.14	1.14	0.28
p^*	0.17	0.44	0.08	0.03
Study setting				
Screening	-1.85	0.74	0.98	0.13
Referral clinic	-0.35	-0.39	0.68	0.09
Both	-1.00	-0.18	1.51	0.23
Unspecified	0.20	-0.37	0.58	0.17
p^*	0.24	0.32	0.15	0.91

Analysis uses 52 datasets that provided data for all cytology categories. IQR=inter-quartile range. *Based on non-parametric significance tests: Mann-Whitney test for variables with two categories; Kruskal-Wallis test for variables with three or more categories.

Table 5: Medians of the differences in percentages of cytology classifications (liquid-based minus conventional cytology)

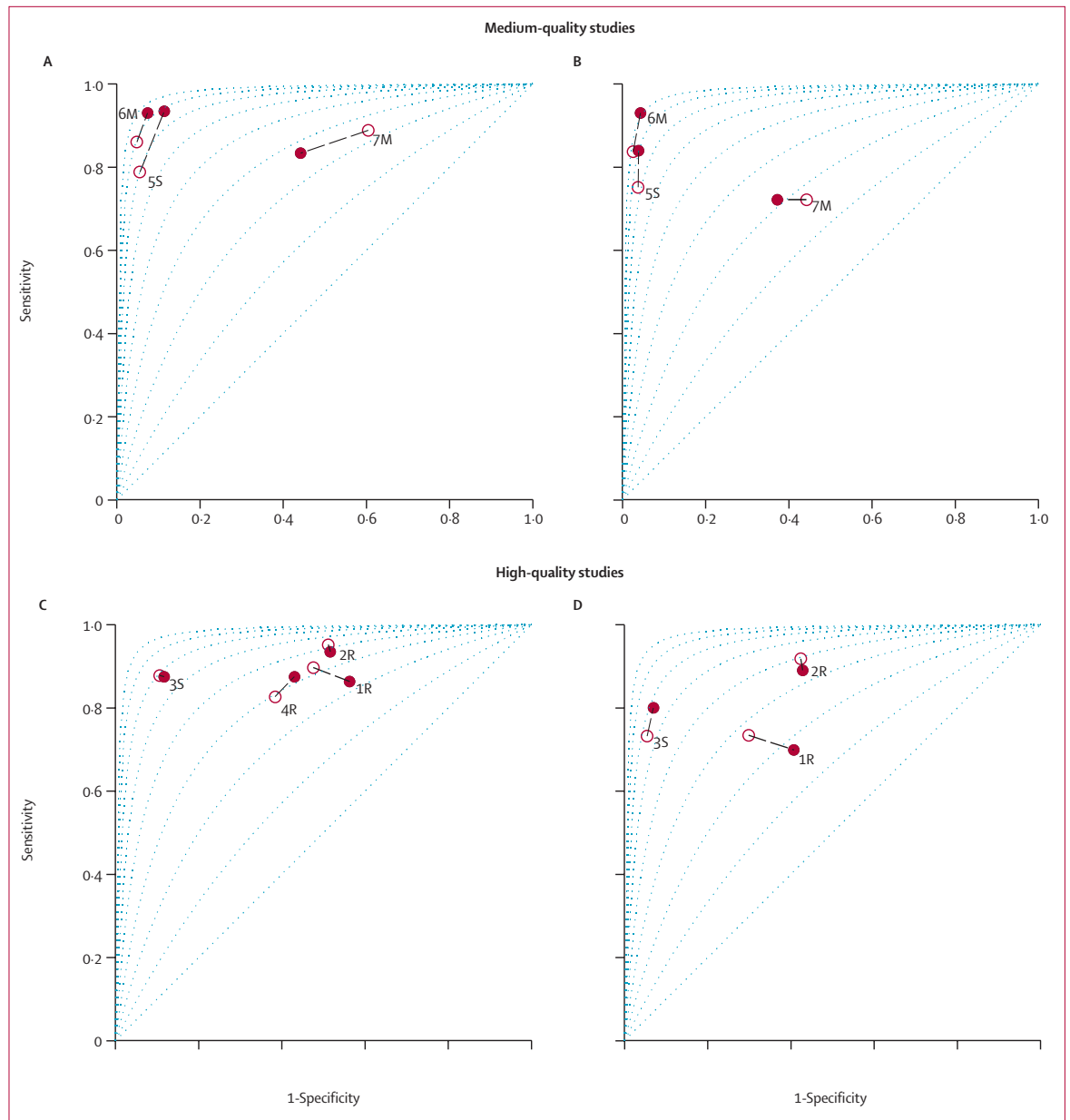


Figure 2: Accuracy to detect reference standard HSIL

A and B show medium-quality studies (A with a test threshold of ASCUS, B with a test threshold of LSIL) in ROC space. C and D show high-quality studies with the same thresholds.

High quality:

1: cone biopsy population, ref std=histology (cone biopsy);²² 2: colposcopy population, ref std=colposcopy, abnormalities biopsied;²¹ 3: screening population, ref std=colposcopy, abnormalities biopsied;²¹ 4: colposcopy population, ref std=histology (raw data not provided).⁴

Medium quality:

5: screening population, colposcopy on all \geq ASCUS, ref std=all screening tests and histology;³⁴ 6: screening and colposcopy population, ref std=consensus cytology;³⁶ 7: screening and colposcopy population, ref std=histology.³⁶

Open symbols=conventional cytology. Solid symbols=liquid-based cytology. R=referred for investigation; S=screening; M=mixed screening and referred populations.

Of the 17 medium-quality paired studies, two provided data from which sensitivity and specificity could be calculated. Relative true-positive and relative false-positive rates could be calculated in another two (webappendix). However, these were meaningless owing

to very small number of cases with discordant cytology. The remaining studies did not use an appropriate reference standard threshold (six studies), provided insufficient data (five studies), or could not be assessed (two studies).

Of the 15 independent medium-quality studies, none verified a sufficient percentage of positive cytology slides and a random sample of negative cytology slides to allow valid estimation of sensitivity and specificity.

The four studies (six datasets) that provided data from which sensitivity and specificity could be calculated all provided data at both ASCUS and LSIL cytological (test) thresholds. Five datasets used colposcopy, histology, or both, as a reference standard, and one used consensus cytology. The study that quoted sensitivity and specificity provided these only at an ASCUS cytological threshold and used histology as the reference standard.

For each study identifier shown in figure 2, there are two points (one for liquid-based cytology and one for conventional cytology) joined by a dashed line. For example, the study represented by point 1²² estimated sensitivity 90% and specificity 53% for conventional cytology and sensitivity 86% and specificity 44% for liquid-based cytology at a cytology threshold of ASCUS (figure 2, C). The same study estimated sensitivity 73% and specificity 70% for conventional cytology and sensitivity 70% and specificity 59% for liquid-based cytology at a cytology threshold of LSIL (figure 2, D).

The dotted curves represent contours defined by a constant diagnostic odds ratio, a measure of test accuracy. These dotted grid curves enable differentiation between changing thresholds (overall test positivity rate) and a true improvement in accuracy. If points identifying studies move along grid curves, the threshold is changing without a change in accuracy. Points that lie on curves that are closer to the top left-hand corner of the graph represent higher accuracy.

The seven datasets shown represent different settings (populations) and the plots thus display variability in test accuracy over populations. One study²¹ provided datasets separately for screening and colposcopy groups. One study³⁶ assessed a mixed screening and colposcopy population with two reference standards, one study³⁴ a screening population only, one⁴ a colposcopy population, and another²² a population undergoing cone biopsy.

The small number of studies that provided evidence on accuracy did not allow examination of liquid-based cytology technique by proprietary name.

None of the pairs of points in figure 2 provides evidence that liquid-based cytology improves accuracy compared with conventional cytology in the detection of high-grade disease.

Discussion

Although liquid-based cytology has been claimed to reduce the proportion of slides judged unsatisfactory for assessment, the results of this review of 1 145 451 slides do not support this claim. We noted no meaningful difference in the percentages of unsatisfactory slides between liquid-based cytology and conventional

cytology. The median of the differences (liquid-based cytology minus conventional cytology across all studies was -0.17% ; the pooled estimate of the differences based on a random-effects model was -0.14% . All large studies, and most studies overall, had differences near zero. Furthermore, we noted no important variation in this result across strata of study quality, study design, setting, liquid-based cytology proprietary name, or publication date. Our finding of no difference between methods differs from previous reviews, which used much smaller datasets.^{16,17}

Our examination of cytological classifications over 1 250 697 slides showed that in high-quality studies conventional cytology classified more slides as HSIL than did liquid-based cytology, whereas low quality studies did not show this difference. In studies of medium and high quality, liquid-based cytology classified more slides as ASCUS than did conventional cytology, whereas conventional cytology classified more slides as ASCUS than did liquid-based cytology only in low-quality studies. In summary, in higher-quality studies conventional cytology classified more slides as HSIL and fewer slides as ASCUS than did liquid-based cytology. This evidence does not, therefore, support claims that liquid-based cytology detects more high-grade cytological lesions than does conventional cytology. Although not verified abnormalities, classifications of ASCUS, LSIL, or HSIL define the rate of investigation in a screening population, and thus have clinical, personal, and financial importance.

The validity of colposcopy and biopsy as a reference standard has been questioned.²⁸ However, even an imperfect reference standard, if applied without knowledge of the two tests being compared, will provide an unbiased reference comparison of the accuracy of the two tests. That colposcopy and biopsy is the most widely used reference standard in clinical practice was reflected in the papers in this review.

Although many primary studies have concluded that liquid-based cytology is better than conventional cytology, our assessment showed that very few studies were adequately designed to compare these tests validly. Of the 56 studies in our review, only five were satisfactorily designed to compare performance (webappendix). Overall, a third of studies did not use a reference standard and, of the two-thirds that did, most did not adequately apply masking and verification by the reference standard. Of 37 studies that used a reference standard, only four provided sufficient data to allow sensitivity and specificity to be validly estimated. Among these, there is no evidence that liquid-based cytology is more accurate than conventional cytology.

Conventional cytology classified more slides as HSIL than did liquid-based cytology in high-quality studies. Four of the five high-quality studies (five of six datasets) were paired studies that used the split-sample method.

In split-sample studies, liquid-based cytology slides are prepared from cells remaining after conventional cytology slides have been made and therefore the performance of liquid-based cytology might be adversely affected. However, in medium quality studies, there was no evidence of poorer liquid-based cytology performance in the split-sample studies than in the independent studies.

Recent studies have not shown improvement in design; we noted only two independent-sample studies published after 2002;^{26,71} both used historical controls. The lack of independent studies in our high-quality category, and the fact that our review included only one underpowered randomised controlled trial, highlight the paucity of well-designed studies in this field.

Although we did not find liquid-based cytology to be more accurate than conventional cytology, equivalent performance might be sufficient if liquid-based cytology has other advantages, such as the opportunity for concurrent HPV DNA testing, reduces reading times, or is more economical than conventional cytology. This review highlights the need for large-scale randomised controlled trials. These trials should incorporate colposcopy and biopsy of women who have positive results on either test, and histology read without knowledge of cytology results as a reference standard. Ideally, colposcopy and biopsy should also be done without knowledge of cytology results.

Although they would be expensive and might pose ethical challenges and practical difficulties (such as obtaining consent from large numbers of women undergoing screening), large-scale randomised trials could be integrated into routine services, as has happened with faecal-occult-blood testing for bowel cancer in Australia.⁷⁴ Such trials might, however, be cost-effective compared with unnecessary introduction of new technologies. Furthermore, studies that concurrently assess the development of new technologies and assess their performance while being implemented, have been proposed to inform management and policy decisions at an early stage.⁷⁵ Such an approach might be useful in assessment of cervical cytology.

The evidence presented here does not lend support to a conclusion that liquid-based cytology is better than conventional cytology. Liquid-based cytology did not reduce the percentage of unsatisfactory slides compared with conventional cytology. There are very few studies with which to estimate the relative performance of the two methods validly and there is no evidence that liquid-based cytology is more accurate than conventional cytology at detecting high-grade disease in high-quality studies. There is a clear need for large-scale randomised trials to assess liquid-based cytology.

Contributors

Elizabeth Davey designed, developed, and did most of the review, assessed study quality, extracted data, did most of the analyses, and wrote and reviewed the paper. Alexandra Barratt designed the review, assessed study quality, extracted data, reviewed analyses, and assisted in

writing and reviewing the paper. Les Irwig developed methodological approaches, designed the review, and reviewed analyses and the paper. Siew Chan developed and did analyses, and reviewed the paper. Petra Macaskill designed the review, directed and reviewed analyses, interpreted results, and edited the paper. Patricia Mannes participated in designing the review, assessed study quality, extracted data, and reviewed the paper. Marion Saville participated in the design of the review and reviewed the paper.

Conflict of interest statement

We declare that we have no conflicts of interest.

Acknowledgments

This work was supported in part by program grant No. 211205 from the National Health and Medical Research Council, Australia.

References

- Walter SD, Irwig L, Glasziou PP. Meta-analysis of diagnostic tests with imperfect reference standards. *J Clin Epidemiol* 1999; **52**: 943–51.
- Fahey MT, Irwig L, Macaskill P. Meta-analysis of Pap test accuracy. *Am J Epidemiol* 1995; **141**: 680–89.
- Franco EL, Duarte-Franco E, Ferenczy A. Cervical cancer: epidemiology, prevention and the role of human papillomavirus infection. *Can Med Assoc J* 2001; **164**: 1017–25.
- Ferenczy A, Robitaille J, Franco E, Arseneau J, Richart RM, Wright TC. Conventional cervical cytologic smears vs ThinPrep smears: a paired comparison study on cervical cytology. *Acta Cytol* 1996; **40**: 1136–42.
- Laverty CR, Farnsworth A, Thurloe JK, Grieves A, Bowditch R. Evaluation of the CytoRich slide preparation process. *Anal Quant Histol* 1997; **19**: 239–45.
- Bernstein SJ, Sanchez-Ramos L, Ndubisi B. Liquid-based cervical cytologic smear study and conventional Papanicolaou smears: a meta-analysis of prospective studies comparing cytologic diagnosis and sample adequacy. *Am J Obstet Gynecol* 2001; **185**: 308–17.
- Roberts JM, Gurley AM, Thurloe JK, Bowditch R, Laverty CR. Evaluation of the ThinPrep Pap test as an adjunct to the conventional Pap smear. *Med J Aust* 1997; **167**: 466–69.
- Carpenter AB, Davey DD. ThinPrep Pap Test: performance and biopsy follow-up in a university hospital. *Cancer* 1999; **87**: 105–12.
- Guidos BJ, Selvaggi SM. Use of the Thin Prep Pap Test in clinical practice. *Diagn Cytopathol* 1999; **20**: 70–73.
- Arbyn M, Abarca M, Delvenne P, et al. Meta-analysis of the test accuracy of liquid-based versus conventional cytology. XV International Congress of Cytology, Chile, April, 2004 (abstr).
- Nanda K, McCrory DC, Myers ER, et al. Accuracy of the Papanicolaou test in screening for and follow-up of cervical cytologic abnormalities: a systematic review. *Ann Intern Med* 2000; **132**: 810–19.
- Sulik SM, Kroeger K, Schultz JK, Brown JL, Becker LA, Grant WD. Are fluid-based cytologies superior to the conventional Papanicolaou test? A systematic review. *J Fam Pract* 2001; **50**: 1040–46.
- Abulafia O, Pezzullo JC, Sherer DM. Performance of ThinPrep liquid-based cervical cytology in comparison with conventionally prepared Papanicolaou smears: a quantitative survey. *Gynecol Oncol* 2003; **90**: 137–44.
- Moseley RP, Paget S. Liquid-based cytology: is this the way forward for cervical screening? *Cytopathology* 2002; **13**: 71–82.
- Klinkhamer PJ, Meerding WJ, Rosier PF, Hanselaar AG. *Liquid-based cervical cytology* *Cancer* 2003; **99**: 263–71.
- National Institute for Clinical Excellence. Guidance on the use of liquid-based cytology for cervical screening. Technology Appraisal Number 69. http://www.nice.org.uk/pdf/TA69_LBC_review_Full_Guidance.pdf (accessed Sept 10, 2004).
- Noorani HZ, Brown A, Skidmore B, Stuart GC. Liquid-based cytology and human papillomavirus testing in cervical cancer screening. Ottawa: Canadian Coordinating Office for Health Technology Assessment, November, 2003.
- Whiting P, Rutjes AW, Reitsma JB, Glas AS, Bossuyt PM, Kleijnen J. Sources of variation and bias in studies of diagnostic accuracy: a systematic review. *Ann Intern Med* 2004; **140**: 189–202.
- Lijmer JG, Mol BW, Heisterkamp S, et al. Empirical evidence of design-related bias in studies of diagnostic tests. *JAMA* 1999; **282**: 1061–66.

- 20 Bossuyt P, Reitsma JB, Bruns DE, et al. Towards complete and accurate reporting of studies of diagnostic accuracy: the STARD initiative. *BMJ* 2003; **326**: 41–44.
- 21 Coste J, Cochand-Priollet B, de Cremoux P, et al. Cross sectional study of conventional cervical smear, monolayer cytology, and human papillomavirus DNA testing for cervical cancer screening. *BMJ* 2003; **326**: 733–37.
- 22 Bergeron C, Bishop J, Lemarie A, et al. Accuracy of thin-layer cytology in patients undergoing cervical cone biopsy. *Acta Cytol* 2001; **45**: 519–24.
- 23 Sherman ME, Mendoza M, Lee KR, et al. Performance of liquid-based, thin-layer cervical cytology: correlation with reference diagnoses and human papillomavirus testing. *Mod Pathol* 1998; **11**: 837–43.
- 24 Hessling JJ, Raso DS, Schiffer B, Callicott J, Husain M, Taylor D. Effectiveness of thin-layer preparation vs conventional Pap smears in a blinded, split-sample study: extended cytologic evaluation. *J Reprod Med* 2001; **46**: 880–86.
- 25 Weintraub J, Morabia A. Efficacy of a liquid-based thin layer method for cervical cancer screening in a population with a low incidence of cervical cancer. *Diagn Cytopathol* 2000; **22**: 52–59.
- 26 Cheung AN, Szeto EF, Leung BS, et al. Liquid-based cytology and conventional cervical smears: a comparison study in an Asian screening population. *Cancer* 2003; **99**: 331–35.
- 27 Systematic reviews in health care: meta-analysis in context. Egger M, Davey Smith G, Altman DG, eds. London: BMJ Publishing Group, 2001.
- 28 The ASCUS–LSIL Triage Study (ALTS) Group. Results of a randomized trial on the management of cytology interpretations of atypical squamous cells of undetermined significance. *Am J Obstet Gynecol* 2003; **188**: 1383–92.
- 29 Monsonego J, Autillo-Touati A, Bergeron C, et al. Liquid-based cytology for primary cervical cancer screening: a multi-centre study. *Br J Cancer* 2001; **84**: 360–66.
- 30 Diaz-Rosario LA, Kabawat SE. Performance of a fluid-based, thin-layer papanicolaou smear method in the clinical setting of an independent laboratory and an outpatient screening population in New England. *Arch Pathol Lab Med* 1999; **123**: 817–21.
- 31 Tezuka F, Oikawa H, Shuki H, Higashiiwai H. Diagnostic efficacy and validity of the ThinPrep method in cervical cytology. *Acta Cytol* 1996; **40**: 513–18.
- 32 Awen C, Hathway S, Eddy W, Voskuil R, Janes C. Efficacy of ThinPrep preparation of cervical smears: a 1,000-case, investigator-sponsored study. *Diagn Cytopathol* 1994; **11**: 33–36.
- 33 Hutchinson ML, Agarwal P, Denault T, Berger B, Cibas ES. A new look at cervical cytology: ThinPrep multicenter trial results. *Acta Cytol* 1992; **36**: 499–504.
- 34 Hutchinson ML, Zahniser DJ, Sherman ME, et al. Utility of liquid-based cytology for cervical carcinoma screening: results of a population-based study conducted in a region of Costa Rica with a high incidence of cervical carcinoma. *Cancer* 1999; **87**: 48–55.
- 35 Bishop JW, Bigner SH, Colgan TJ, et al. Multicenter masked evaluation of AutoCyte PREP thin layers with matched conventional smears including initial biopsy results. *Acta Cytol* 1998; **42**: 189–97.
- 36 Bishop JW. Comparison of the CytoRich system with conventional cervical cytology: preliminary data on 2,032 cases from a clinical trial site. *Acta Cytol* 1997; **41**: 15–23.
- 37 Bolick DR, Hellman DJ. Laboratory implementation and efficacy assessment of the ThinPrep cervical cancer screening system. *Acta Cytol* 1998; **42**: 209–13.
- 38 Corkill M, Knapp D, Hutchinson ML. Improved accuracy for cervical cytology with the ThinPrep method and the endocervical brush–spatula collection procedure. *J Lower Genital Tract Dis* 1998; **2**: 12–16.
- 39 Howell LP, Davis RL, Belk TI, Agdigos R, Lowe J. The AutoCyte preparation system for gynecologic cytology. *Acta Cytol* 1998; **42**: 171–77.
- 40 Laverty CR, Thurloe JK, Redman NL, Farnsworth A. An Australian trial of ThinPrep: a new cytopreparatory technique. *Cytopathology* 1995; **6**: 140–48.
- 41 Lee KR, Ashfaq R, Birdsong GG, Corkill ME, McIntosh KM, Inhorn SL. Comparison of conventional Papanicolaou smears and a fluid-based, thin-layer system for cervical cancer screening. *Obstet Gynecol* 1997; **90**: 278–84.
- 42 Minge L, Fleming M, VanGeem T, Bishop JW. AutoCyte Prep system vs conventional cervical cytology: comparison based on 2,156 cases. *J Reprod Med* 2000; **45**: 179–84.
- 43 McGoogan E, Reith A. Would monolayers provide more representative samples and improved preparations for cervical screening? Overview and evaluation of systems available. *Acta Cytol* 1996; **40**: 107–19.
- 44 Sprenger E, Schwarzmann P, Kirkpatrick M, et al. The false negative rate in cervical cytology: comparison of monolayers to conventional smears. *Acta Cytol* 1996; **40**: 81–99.
- 45 Stevens MW, Nespolon WW, Milne AJ, Rowland R. Evaluation of the CytoRich technique for cervical smears. *Diagn Cytopathol* 1998; **18**: 236–42.
- 46 Takahashi M, Naito M. Application of the CytoRich monolayer preparation system for cervical cytology: a prelude to automated primary screening. *Acta Cytol* 1997; **41**: 1785–89.
- 47 Vassilakos P, Cossali D, Albe X, Alonso L, Hohener R, Puget E. Efficacy of monolayer preparations for cervical cytology: emphasis on suboptimal specimens. *Acta Cytol* 1996; **40**: 496–500.
- 48 Vassilakos P, Schwartz D, de Marval F, et al. Biopsy-based comparison of liquid-based, thin-layer preparations to conventional Pap smears. *J Reprod Med* 2000; **45**: 11–16.
- 49 Wang TY, Chen HS, Yang YC, Tsou MC. Comparison of fluid-based, thin-layer processing and conventional Papanicolaou methods for uterine cervical cytology. *J Formos Med Assoc* 1999; **98**: 500–05.
- 50 Wilbur DC, Facik MS, Rutkowski MA, Mulford DK, Atkison KM. Clinical trials of the CytoRich specimen-preparation device for cervical cytology: preliminary results. *Acta Cytol* 1997; **41**: 24–29.
- 51 Obwegeser JH, Brack S. Does liquid-based technology really improve detection of cervical neoplasia? A prospective, randomized trial comparing the ThinPrep Pap Test with the conventional Pap Test, including follow-up of HSIL cases. *Acta Cytol* 2001; **45**: 709–14.
- 52 Papillo JL, Zarka MA, St John TL. Evaluation of the ThinPrep Pap test in clinical practice: a seven-month, 16,314-case experience in northern Vermont. *Acta Cytol* 1998; **42**: 203–08.
- 53 Bur M, Knowles K, Pekow P, Corral O, Donovan J. Comparison of ThinPrep preparations with conventional cervicovaginal smears: practical considerations. *Acta Cytol* 1995; **39**: 631–42.
- 54 Wilbur DC, Cibas ES, Merritt S, James LP, Berger BM, Bonfiglio TA. ThinPrep Processor: clinical trials demonstrate an increased detection rate of abnormal cervical cytologic specimens. *Am J Clin Pathol* 1994; **101**: 209–14.
- 55 Luthra UK, Chishti M, Dey P, et al. Performance of monolayered cervical smears in a gynecology outpatient setting in Kuwait. *Acta Cytol* 2002; **46**: 303–10.
- 56 Aponte-Cipriani SL, Teplitz C, Rorat E, Savino A, Jacobs AJ. Cervical smears prepared by an automated device versus the conventional method: a comparative analysis. *Acta Cytol* 1995; **39**: 623–30.
- 57 Geyer JW, Hancock F, Carrico C, Kirkpatrick M. Preliminary evaluation of Cyto-Rich: an improved automated cytology preparation. *Diagn Cytopathol* 1993; **9**: 417–22.
- 58 Hatch KD, Sheets E, Kennedy A, Ferris DG, Darragh T, Twigg L. Multicenter direct to vial evaluation of a liquid-based Pap test. *Journal of Lower Genital Tract Disease* 2004; **8**: 308–12.
- 59 Hutchinson ML, Cassin CM, Ball HG 3rd. The efficacy of an automated preparation device for cervical cytology. *Am J Clin Pathol* 1991; **96**: 300–05.
- 60 Dupree WB, Suprun HZ, Beckwith DG, Shane JJ, Lucente V. The promise and risk of a new technology: the Lehigh Valley Hospital's experience with liquid-based cervical cytology. *Cancer* 1998; **84**: 202–07.
- 61 Sheets EE, Constantine NM, Dinisco S, Dean B, Cibas ES. Colposcopically directed biopsies provide a basis for comparing the accuracy of ThinPrep and Papanicolaou smears. *J Gynecol Tech* 1995; **1**: 27–34.
- 62 Wilbur DC, Dubeshter B, Angel C, Atkison KM. Use of thin-layer preparations for gynecologic smears with emphasis on the cytomorphology of high-grade intraepithelial lesions and carcinomas. *Diagn Cytopathol* 1996; **14**: 201–11.

- 63 Park IA, Lee SN, Chae SW, Park KH, Kim JW, Lee HP. Comparing the accuracy of ThinPrep Pap tests and conventional Papanicolaou smears on the basis of the histologic diagnosis: a clinical study of women with cervical abnormalities. *Acta Cytol* 2001; **45**: 525–31.
- 64 Marino JF, Fremont-Smith M. Direct-to-vial experience with AutoCyte PREP in a small New England regional cytology practice. *J Reprod Med* 2001; **46**: 353–58.
- 65 Yeoh GPS, Chan KW, Lauder I, Lam MB. Evaluation of the ThinPrep Papanicolaou test in clinical practice: 6-month study of 16541 cases with histological correlation in 220 cases. *Hong Kong Med J* 1999; **5**: 233–39.
- 66 Biscotti CV, O'Brien DL, Gero MA, Gramlich TL, Kennedy AW, Easley KA. Thin-layer Pap test vs. conventional Pap smear: analysis of 400 split samples. *J Reprod Med* 2002; **47**: 9–13.
- 67 Ring M, Bolger N, O'Donnell M, et al. Evaluation of liquid-based cytology in cervical screening of high-risk populations: a split study of colposcopy and genito-urinary medicine populations. *Cytopathology* 2002; **13**: 152–59.
- 68 Ferris DG, Heidemann NL, Litaker MS, Crosby JH, Macfee MS. The efficacy of liquid-based cervical cytology using direct-to-vial sample collection. *J Fam Pract* 2000; **49**: 1005–11.
- 69 Baker JJ. Conventional and liquid-based cervicovaginal cytology: a comparison study with clinical and histologic follow-up. *Diagn Cytopathol* 2002; **27**: 185–88.
- 70 Malle D, Pateinakis P, Chakka E, Destouni C. Experience with a thin-layer, liquid-based cervical cytologic screening method. *Acta Cytol* 2003; **47**: 129–34.
- 71 Bergeron C, Fagnani F. Performance of a new, liquid-based cervical screening technique in the clinical setting of a large French laboratory. *Acta Cytol* 2003; **47**: 753–61.
- 72 Tench W. Preliminary assessment of the AutoCyte Prep: direct-to-vial performance. *J Reprod Med* 2000; **45**: 912–16.
- 73 Weynand B, Berliere M, Haumont E, et al. A new liquid-based cytology technique. *Acta Cytol* 2003; **47**: 149–53.
- 74 Australian Government Department of Health and Ageing. Bowel Cancer Screening Program. <http://www.cancerscreening.gov.au/bowel> (accessed Sept 30, 2004)
- 75 Hunink MG, Krestin GP. Study design for concurrent development, assessment, and implementation of new diagnostic imaging technology. *Radiology* 2002; **222**: 604–14.