European Journal of Obstetrics & Gynecology and Reproductive Biology 240 (2019) 182–186

Contents lists available at ScienceDirect

European Journal of Obstetrics & Gynecology and Reproductive Biology

journal homepage: www.elsevier.com/locate/ejogrb

Review article

# The diagnostic accuracy of colposcopy – A review of research methodology and impact on the outcomes of quality assurance

Brian Hilton Brown[a,*], John A. Tidy[b]

[a] Department of Infection, Immunity and Cardiovascular Disease, The University of Sheffield, Medical School, Beech Hill Road, Sheffield S10 2RX, U.K.
[b] Department of Gynaecological Oncology, Royal Hallamshire Hospital, Glossop Road, Sheffield S10 2JF

ARTICLE INFO

*Article history:*
Received 11 May 2019
Received in revised form 2 July 2019
Accepted 3 July 2019

*Keywords:*
Colposcopy
Sensitivity
Specificity
Bias
Accuracy
Methodology
Outcomes

ABSTRACT

*Objective:* To review the published diagnostic accuracy figures for the performance of colposcopy and to assess how the various forms of bias might explain the very wide range of reported values and the impact they have on quality assurance of cervical screening.
*Methods:* Publications were only selected where they contained sufficient raw data to enable diagnostic accuracy statistics to be calculated for the detection of cervical intraepithelial neoplasia grade 2+ (CIN2+), as determined by punch biopsy. In addition, both the colposcopic impression at the time of examination and the disease threshold used to determine the need for biopsy must have been reported.
*Results:* Large differences in diagnostic accuracy figures were found when the output of colposcopy was defined either, on the basis that the colposcopist thought there was CIN2+ present or, that the colposcopist considered there to be some disease present and so took a biopsy to confirm this. Weighted mean sensitivity was 68.5% (95% CI 59.9–77.1) for the first method but 95.7% (95% CI 93.4–98.0) for the second method. Weighted mean specificity was 75.9% (95% CI 69.3–82.5) for the first method but 34.2% (95% CI 27.0–41.4) for the second method. Weighted mean PPV was 68.9% (95% CI 64.2–73.6) for the first method but 54.3% (95% CI 46.5–62.1) for the second method.
*Conclusion:* The main reason for the wide range of published diagnostic accuracy figures, arises from the use of two different methods of assessing the output of colposcopy. Colposcopic Impression is appropriate when assessing the performance of a colposcopist at the time of examination, but the taking of a biopsy to confirm that Disease is Present should be used when assessing patient management. Accurate assessment of both outcomes is fundamental to any quality assurance programme.


**Contents**

Introduction . . . . . 182
Methods . . . . . 183
Results . . . . . 183
References . . . . . 186

## Introduction

Colposcopy is widely used as part of screening programmes to detect cervical cancer and pre-cancer. The main objective of colposcopy is usually the detection of high-grade disease, defined as cervical intraepithelial neoplasia grade 2+ (CIN2+). Reliable figures for the diagnostic accuracy of colposcopy are required to inform a quality control process, to assess the efficacy of changes in colposcopic practice, or to assess an adjunct to colposcopy. This review of the methodology used in published clinical trials stems from the very wide range of published diagnostic accuracy figures for clinical colposcopy. For example, Mitchell [1] in 1998 reported on 8 studies, as part of a meta-analysis, a weighted mean sensitivity of 85% and specificity 69%, but the reported values ranged from 30% to 99% for sensitivity and 39% to 93% for specificity. Underwood [2] in 2012 reported on 25 studies and

* Corresponding author.
*E-mail address:* b.h.brown@sheffield.ac.uk (B.H. Brown).

https://doi.org/10.1016/j.ejogrb.2019.07.003
0301-2115/

give pooled values of 91.3% for sensitivity and 24.6% for specificity, but the published values ranged from 65.9% to 100% for sensitivity and from 5% to 80% for specificity. The objective of this review has been to look for an explanation for this wide range of reported statistics.

Methods used for the calculation of diagnostic accuracy figures from trials of clinical colposcopy are subject to several types of bias. The FDA has published statistical guidance on reporting results from studies evaluating diagnostic tests [3]. They consider how bias can result from the absence of a reference standard or where the reference cannot be applied to the whole patient group. In colposcopy, the reference or gold standard usually applied is the histology report following one or more biopsies. The use of punch biopsies as the reference has been questioned by comparison with the results from excisional biopsy following punch biopsy [2,4]. The reduction of verification bias by the use of more than one biopsy has also been addressed [5].

Verification bias is a particular problem in studies of colposcopy. In many clinics biopsy is only performed when there is a suspicion of disease. As a result, verification by biopsy is only performed when the outcome of colposcopy is positive and not when the outcome is negative. This form of bias results in sensitivity being overestimated. Specificity will also be affected. In principle, positive predictive value (PPV) should not be affected, because the calculation of PPV only uses the statistics for patients where a biopsy has been taken. The FDA refers the reader to some methods that use maximum likelihood latent class methods to address the issue of verification bias, but these make assumptions that are themselves very difficult to verify in the context of most colposcopy studies [6].

A further cause of bias in colposcopy studies is the various methods that have been used to define the outcome of colposcopy. Some studies define the outcome based upon the Colposcopic Impression (CI) that high-grade neoplasia (CIN2+) is present. Others base the outcome on taking a biopsy because there is thought to be some Disease Present (DP). The threshold used to determine the presence of some disease is usually CIN1 + . Use of these two different outcome measures has a very significant effect on diagnostic accuracy figures.

Quality assurance of a cervical screening programme is fundamental to maintaining professional and public confidence. Outcome after referral for an abnormal screening test, detection of disease, and the performance of any clinical intervention such as the performance of colposcopy must be measured.

This paper reviews the reported diagnostic accuracy figures for colposcopy and then assesses how the various forms of bias might explain the very wide range of reported values and how this may affect quality assurance.

## Methods

Publications were only selected where they contained sufficient raw data to enable diagnostic accuracy statistics to be calculated for the detection of cervical intraepithelial neoplasia grade 2+ (CIN2+), as determined by punch biopsy. In addition, both the colposcopic impression at the time of examination and the disease threshold used to determine the need for biopsy must have been reported. The raw data had to include actual patient numbers in the various groups. Sensitivity, specificity and PPV were calculated directly from the raw data given.

Eighteen publications that met the above criteria were selected for review. Six of these were taken from the 8 reported by Mitchell [1] covering the years1973-95. Two were taken from the 25 reported by Underwood in 2012. A further 10 more recent publications were selected from a Medline/PubMed search for the years 2004 to 2019, using the keywords Colposcopy or Colposcopic associated with Diagnostic accuracy, Optical, Dynamic spectra imaging or Electrical impedance.

Diagnostic accuracy figures were calculated from the data given in the selected papers using two different methods. The outcome of colposcopy was based upon two different thresholds referred to as CI and DP. Colposcopic Impression (CI) indicates that the outcome of colposcopy was a colposcopic impression that CIN2+ was present. This records the opinion of the colposcopist and not what action was taken as a consequence. An output of DP indicates that there was disease present, usually described as CIN1+ and therefore, a biopsy was taken to confirm or exclude the presence of CIN2 + . In some of the older papers colposcopic impression and the threshold used to determine the need for a biopsy were described as mild dysplasia, moderate dysplasia, severe dysplasia, carcinoma in situ and invasive carcinoma. There is no exact way to map these thresholds to CIN1-3. For this review mild, moderate and severe dysplasia were mapped as CIN1, CIN2 and CIN3. Carcinoma in situ was included with CIN3.

## Results

Table 1 presents the diagnostic accuracy figures from the 18 selected publications. For each publication the numbers extracted and then used to calculate sensitivity, specificity and PPV are given. It was possible to calculate the statistics for both CI and DP methods from 13 of the publications. In 5 cases it was only possible to make calculations using the CI method. Mean and weighted mean figures are given for the three statistics. The weighting was on the basis of the number of patients in each study.

Results from the 18 reviewed publications were presented as mean values and 95% confidence intervals on the mean values. Mean values that were weighted according to the number of patients in each study were also calculated. Comparisons between the CI and DP methods were made using a two-tailed, non-parametric Mann-Whitney U test.

Statistical comparison of the CI and DP results shows that the mean sensitivity and specificity values are different using a Mann-Whitney two tailed U test ($p < 0.0001$). The mean values for PPV are also different ($p = 0.042$). In every one of the 13 publications where both CI and DP methods were applied the DP method produced higher sensitivity, lower specificity and lower PPV values than the CI method. For this group of 13 publications the mean sensitivity, specificity and PPV values are significantly different (p values <0.0001, <0.0001 and 0.014 respectively).

Fig. 1 presents the sensitivity and specificity data as a receiver operating characteristic (ROC) plot. This plot can be compared with similar plots given by Underwood et al [2] in their Figure 3 A and by Mitchell et al [1] in their figures 1 and 2.

It can be seen in Fig. 1 that the CI and DP points are well separated. There are a few points that overlap. These are mainly from older publications where the terminology used of mild, moderate and severe dysplasia does not map exactly onto CIN1, CIN2 and CIN3.

Discussion and conclusions

The wide range of values for both sensitivity and specificity found in the 18 reviewed publications are very similar to the ranges found in the reviews by Mitchell [1] and Underwood [2]. The main explanation for the wide range of values appears to be the use of two different ways of defining the output of colposcopy. This is clearly illustrated in Fig. 1, where the use of Colposcopic Impression (CI) as the output results in relatively low sensitivity (weighted mean 68.5%) but relatively high specificity (weighted mean 75.9%), in contrast to the Disease Present (DP) method that results in a higher sensitivity (weighted mean 95.7%) but lower specificity (weighted mean 34.2%). The CI method, based on the diagnostic opinion of the colposcopist that CIN2+ is present, is the

**Table 1**
The diagnostic accuracy figures derived from the 18 reviewed publications.

| Publication | Number of patients in each study. | Sensitivity, Specificity and PPV of colposcopy based on CI method | Sensitivity, Specificity and PPV of colposcopy based on DP method |
|---|---|---|---|
| Roensbo et al. [7] | 239 | Sen. 46/68 = 67.6%<br>Spec. 115/171 = 67.3%<br>PPV 46/102 = 45.1% | |
| Soutter et al. [8] | 308 | Sen. 35/72 = 48.6%<br>Spec. 211/236 = 89.4%<br>PPV 35/60 = 58.3% | |
| Loewers et al. [9] | 239 | Sen. 56/108 = 51.9%<br>Spec.107/131 = 81.7%<br>PPV 52/76 = 68.4% | |
| Coronado & Fasero [10] | 443 | Sen. 30/41 = 73.2%<br>Spec.371/402 = 92.3%<br>PPV 30/61 = 49.2% | |
| Huh et al. [11] | 1471 | Sen. 77/240 = 36.7%<br>Spec.1218/1261 = 96.7%<br>PPV 77/120 = 64.1% | |
| Wentzensen et al. [5] | 653 | Sen. 139/271 = 51.3%<br>Spec. 326/407 = 80.0%<br>PPV 139/220 = 63.2% | Sen. 246/252 = 97.6%<br>Spec. 221/412 = 53.6%<br>PPV 222/623 = 35.6% |
| Tidy et al. [12] | 196 | Sen. 64/87 = 73.6%<br>Spec. 91/109 = 83.5%<br>PPV 64/82 = 78.1% | Sen. 77/87 = 88.5%<br>Spec. 42/109 = 38.5%<br>PPV 77/144 = 53.5% |
| Edebiri AA, [13] | 222 | Sen. 94/114 = 82.5%<br>Spec. 74/108 = 68.5%<br>PPV 94/128 = 73.4% | Sen. 101/114 = 88.6%<br>Spec = 66/108 = 61.1%<br>PPV 101/143 = 70.6% |
| Marel et al. [14] | 610 | Sen. 163/265 = 61.5%<br>Spec. 282/343 = 82.2%<br>PPV 163/224 = 72.8% | Sen. 233/265 = 87.9%<br>Spec. 152/343 = 44.3%<br>PPV 233/424 = 55.0% |
| Zuchna et al. [4] | 244 | Sen. 135/204 = 66.2%<br>Spec. 38/40 = 95.0%<br>PPV 135/137 = 98.5% | Sen. 169/204 = 82.8%<br>Spec. 27/40 = 67.5%<br>PPV 182/240 = 75.8% |
| Huh et al. [15] | 604 | Sen. 47/70 = 67.1%<br>Spec. 34/63 = 54.0%<br>PPV 47/76 = 61.8% | Sen. 165/165 = 100.0%<br>Spec. 135/439 = 30.8%<br>PPV 165/469 = 35.2% |
| Cantor et al. [16] | 797 | Sen. 165/231 = 71.4%<br>Spec 460/566 = 81.3%<br>PPV 165/271 = 60.9% | Sen. 227/231 = 98.3%<br>Spec 255/566 = 45.1%<br>PPV 227/538 = 42.2% |
| Ferris & Miller [17] | 205 | Sen. 13/43 = 30.2%<br>Spec 150/162 = 92..6%<br>PPV = 13/25 = 52.0% | Sen. 41/43 = 95.3%<br>Spec 21/162 = 13.0%<br>PPV 41/182 = 22.5% |
| Stafl and Mattingly [18] | 659 | Sen. 323/351 = 92.0%<br>Spec 183/308 = 48.2%<br>PPV 323/448 = 72.1% | Sen. 350/351 = 99.7%<br>Spec 47/308 = 15.3%<br>PPV 350/611 = 57.3% |
| Cristoforoni et al. [19] | 188 | Sen. 21/34 = 61.8%<br>Spec 142/154 = 92.2%<br>PPV 21/33 = 63.6% | Sen. 34/34 = 100%<br>Spec 23/154 = 14.9%<br>PPV 34/165 = 20.6% |
| Benedet et al. [20] | 3252 | Sen. 1411/1730 = 81.6%<br>Spec 1046/1522 = 68.7%<br>PPV 1411/1887 = 74.8% | Sen 1670/1730 = 95.5%<br>Spec 441/1522 = 29.0%<br>PPV 1670/2751 = 60.7% |
| Benedet et al. [21] | 549 | Sen. 354/371 = 95.4%<br>Spec. 114/178 = 64.0%<br>PPV 354/418 = 84.7% | Sen. 370/371 = 99.7%<br>Spec. 61/178 = 34.3%<br>PPV 370/487 = 76.0% |
| Fung et al. [22] | 94 | Sen. 63/75 = 84%<br>Spec. 12/19 = 63.2%<br>PPV 63/70 = 90.0% | Sen. 71/75 = 94.7%<br>Spec 1/19 = 5.35%<br>PPV 71/89 = 79.8% |
| Mean figures | | Sen. 66.5% (95% CI 57.6-75.4)<br>Spec. 77.8% (95% CI 70.6-85.1)<br>PPV 68.4% (95% CI 61.5-75.3) | Sen. 94.5% (95% CI 91.1-98.0)<br>Spec. 34.8% (95% CI 23.1-46.5)<br>PPV 52.7% (95% CI 40.5-64.9) |
| Weighted mean figures | | Sen. 68.5% (95% CI 59.5-77.1)<br>Spec. 75.9% (95% CI 69.3-82.5)<br>PPV 68.9% (95% CI 64.2-73.6) | Sen 95.7% (95% CI 93.4-98.0)<br>Spec 34.2% (95% CI 27.0-41.4)<br>PPV 54.3% (95% CI 46.5-62.1) |
| Mean figures (for the 13 publications where both CI & DP were figures available). | | Sen. 70.7% (95% CI 60.0-81.3)<br>Spec. 74.9% (95% CI 65.8-83.9)<br>PPV 72.8% (95% CI 65.0-80.6) | |
| Weighted mean figures (for the 13 publications where both CI & DP figures were available). | | Sen. 75.1% (95% CI 67.5-82.7)<br>Spec. 71.0% (95% CI 64.5-77.5)<br>PPV 72.0% (95% CI 67.2-76.8) | |

appropriate method to use where the skill of the colposcopist is to be assessed. It is also the appropriate measure when immediate treatment (See and Treat – S&T) is to be carried out at the time of colposcopy. However, the use of the DP method, based upon the presence of disease (CIN1+), is appropriate when patient management is to be decided or where economic modelling of colposcopy is carried out. The DP method is based upon what the colposcopist actually does, usually the taking of a diagnostic biopsy.

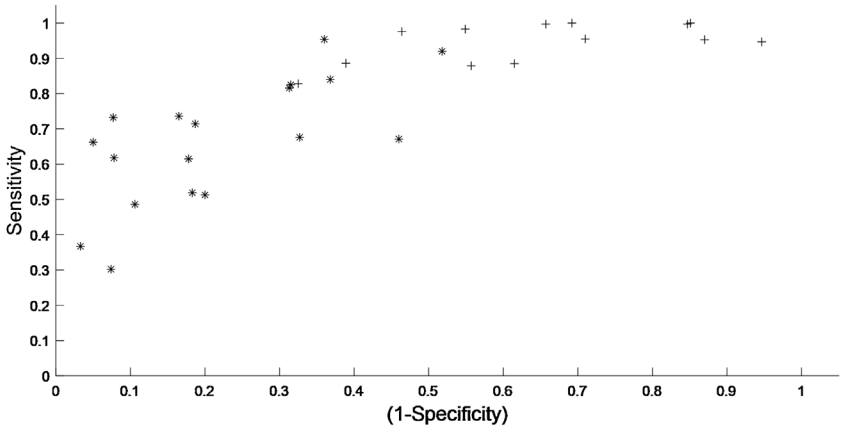


**Fig. 1.** ROC plot of the sensitivity and specificity values given in columns 3 and 4 of Table 1. The values using the CI method are shown as '*' and those using the DP method as '+'.

Even though the two methods of recording the output of colposcopy can explain a large part of the ranges in both sensitivity and specificity values, there remain considerable variations. Within the DP group sensitivities range from 82.8% to 100% and specificities from 5.4% to 67.5%. Differences in the skill of the colposcopists will be one reason for the range of diagnostic accuracy figures. Verification bias will also be a reason for some of the remaining variations in sensitivity values. If no biopsies are taken in addition to those directed by the colposcopist then, by definition, the sensitivity of the DP method will be 100%, because all the biopsy proven cases of CIN2+ will have resulted from a colposcopist directed biopsy. Several of the references [7–12] reported on the use of an adjunct to colposcopy that detected CIN2 + cases in addition to the cases detected by colposcopy alone. It is also the case that biopsies may not have been taken from the diseased tissue and hence true-positive cases will have been missed. This is further evidence that the true sensitivity of colposcopy must be lower than the reported values. The conclusion is that verification bias does lead to an overestimation of sensitivity when using the DP method. The 'true' sensitivity is probably lower than the calculated weighted mean value of 95.7%.

The effect of verification bias on the specificity values is more difficult to assess as both the numerator and denominator in the calculation may be affected. Specificity is determined by the fraction of false positives and this will be very dependent on the colposcopist's decision that a biopsy is needed. PPV should be unaffected by verification bias as the calculation only uses results that were test-positive and hence a biopsy was taken. The weighted mean values found in Table 1 were 68.9% for the CI method and 54.3% for the DP method.

In addition to verification bias, variations in the diagnostic accuracy figures will arise from differences in the study parameters of the reviewed papers. These parameters will include the size of the study, differences in the number of biopsies taken, disease incidence, the use of several colposcopists and their training/ experience. Imprecision, as opposed to bias, will be largely dictated by sample size. This is the reason for reporting weighted mean values for diagnostic accuracy. Other possible sources of bias include diagnostic review and incorporation bias. These occur when the gold standard might be influenced by the test result. This could occur if the pathologist knew which biopsies were considered by the colposcopist to be CIN2+.

Some of the variations in sensitivity and specificity within the CI and DP groups will be a consequence of differences in the incidence of disease in the study populations. This form of spectrum bias, that arises from differences in patient mix, is difficult to avoid. In principle, sensitivity and specificity do not depend upon the incidence of disease in the test population. This is because the patient numbers that contribute to sensitivity are all from the disease group and those that contribute to specificity are all disease negative. However, in addition to the skill of the colposcopist, verification bias will also have a significant effect, on the disease negative grouping, and this may well be affected by the incidence of disease.

Underwood et al [2] used a meta-analysis of 32 papers to consider the accuracy of colposcopy directed biopsies and the outcomes of excisional biopsy. They identified four studies where the punch biopsy was taken immediately before the excisional biopsy. Pooling of these four studies yielded lower sensitivity but higher specificity than the pooled results from the 32 studies. They considered the possibility that immune mediated disease regression might have resulted in false positive results from the punch biopsy and hence increased sensitivity and reduced specificity. It is also possible that the initial disease was of small volume and hence the diagnostic biopsy removed the lesion prior to the loop excision. One of the four studies identified by Underwood et al is included in the current study. None of our additional studies include further information to address this possible source of bias. In this study we have used the diagnostic biopsy as the gold standard as it predicates clinical management, but we accept that it is far from perfect. This should not detract from the fact that in every one of the studies assessed we have observed increased sensitivity, reduced specificity and reduced PPV when using the DP method rather than the CI method of analysis.

The FDA [3] in their guidance on the evaluation of diagnostic tests suggest that diagnostic accuracy statistics should be used with caution because of the effects of verification bias, where there is no confirmation of negative test results. They suggest the use of latent (hidden or missing) class methods to infer sensitivity and specificity. Walter [6] details this methodology. However, such methods can only be used where the results from two diagnostic tests are available and where the 'gold standard' has been applied to all positive results. In addition, the two diagnostic tests have to be independent and applied to the same patient cohort. These conditions are very difficult to meet as most adjuncts to colposcopy are unlikely to be completely independent from colposcopy alone.

NICE, in Table 3 of their assessment of adjuncts to colposcopy [23], use figures for the diagnostic accuracy of colposcopy alone of sensitivity 57.91% and specificity 87.4%. These figures are then used as the input to a model that is used to quantify screening and treatment pathways. These figures appear to be consistent with the CI method of describing the output of colposcopy. This method,

that is based upon the opinion of the colposcopist that CIN2+ is present, rather than the DP method based upon what action is taken, is not appropriate for a model of clinical pathways. This well illustrates the confusion that can arise from the use in the literature of different methods of assessing the output of colposcopy.

This review concludes that the main explanation for the very wide range of published diagnostic accuracy figures for colposcopy is the two different methods of defining the output or end-point of colposcopy. The CI method gives much higher specificity but lower sensitivity than the DP method. The CI method is based upon the colposcopic impression at the time of examination and is the appropriate measure of performance where immediate treatment (S & T) is considered. It may also the appropriate method where the objective is to assess the performance of a colposcopist. The DP method is based upon what action the colposcopist takes, for example taking a biopsy or returning the patient to future screening. This method is the appropriate measure as the output of colposcopy that determines subsequent patient management. Any assessment of the overall performance of colposcopy in modelling service provision should use the DP method. Both methods are needed in order to quantify all aspects of colposcopic performance.

The English Cervical screening programme publishes, on an annual basis, the outcomes of referral to colposcopy in terms of the level of disease found (CIN2+). These figures for PPV from 50 centres range between 7.3% and 26.8% for low grade cytology referrals and between 76.2% and 92.3% for high grade cytology referrals [24]. The performance of individual colposcopists is part of the national colposcopy quality assurance programme. All colposcopists are expected to achieve a positive predictive value of at least 65% for colposcopic diagnosis of high-grade lesions (CIN2 or worse) [25]. Additionally, if treatment is offered at first visit (S&T) colposcopists must achieve a PPV of at least 90% for CIN2+ in the excised tissue. These latter values are consistent with the CI method being used to assess the performance of the colposcopists, whereas the outcome of referral to colposcopy is consistent with the DP method. Outside of organised screening programmes standards for quality assurance in colposcopy do not directly require measurement of the output of colposcopic practice. Instead of quantifying the diagnostic accuracy of colposcopic diagnosis of disease or the presence of disease in any biopsy they focus instead on the process of the colposcopic examination [26,27]. This may reflect the difficulty in both obtaining and interpreting data on the output of colposcopy.

This review also concludes that verification bias is a partial explanation for the wide range of published diagnostic accuracy figures. It is not usually possible to remove verification bias, as in many situations it is not considered to be justified to take biopsies where no disease is seen on visual examination. A consequence of verification bias is that reported values for sensitivity will be higher than the true values. It is concluded that sensitivity and specificity values should be used with caution. Changes in these statistics, when modifications to colposcopic practice are being assessed in the same patient group, may be valid indicators and hence important when trying to establish new quality assurance standards.

## References


[1] Mitchell MF, Schottenfeld D, Tortolero-Luna G, Cantor SB, Richards-Kortum R. Colposcopy for the diagnosis of squamous intraepithelial lesions: a meta-analysis. Obstet Gynecol 1998;91:626–31.

[2] Underwood M, Arbyn M, Parry-Smith W, De Bellis-Ayres S, Todd R, Redman CWE, et al. Accuracy of colposcopically-directed punch biopsies: a systematic review and meta-analysis. BJOG 2012;119:1293–301.

[3] Guidance for Industry and FDA staff: Statistical guidance on reporting results from studies evaluating diagnostic tests. https://www.fda.gov/downloads/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/ucm071287.pdf.

[4] Zuchna C, Hager M, Tringler B, Georgoulopoulos A, Ciresa-Koenig A, Volgger B, et al. Diagnostic accuracy of guided cervical biopsies: a prospective multi-centre study comparing the histopathology of simultaneous biopsy and cone specimen. Am J Obstet Gynecol 2010;203(321):e1–6.

[5] Wentzensen N, Walker JL, Gold MA, Smith KM, Zuna RE, Mathews C, et al. Multiple biopsies and detection of cervical cancer precursors at colposcopy. J Clin Oncol 2015;33(1):83–94.

[6] Walter SD. Estimation of test sensitivity and specificity when disease confirmation is limited to positive results. Epidemiology 1999;10(1):67–72.

[7] Roensbo MT, Hammer A. Blaakær. Can dynamic spectral imaging system colposcopy replace conventional colposcopy in the detection of high-grade cervical lesions? Acta Obstet Gynecol Scand 2015;94:781–5.

[8] Soutter WP, Diakomanolis E, Lyons D, Ghaem-Maghami S, Ajala T, Haidopoulos D, et al. Dynamic spectral imaging: improving colposcopy. Clin Cancer Res 2009;15(5):1814–20.

[9] Louwers JA, Zaal A, Kochen M, ter Harmsel WA, Graziosi GMC, Spruijt JWM, et al. Dynamic spectral imaging colposcopy: higher sensitivity for the detection of premalignant cervical lesions. BJOG 2011;118:309–18.

[10] Coronado PJ, Fasero M. Colposcopy combined with dynamic spectral imaging. A prospective clinical study. Eur J Obstet Gynecol Reprod Biol 2016;196:11–6.

[11] Huh WK, Papagiannakis E, Gold MA. Observed colposcopy practice in US community-based clinica: the retrospective control arm of the IMPROVE-COLPO study. J Low Genit Tract Dis 2019;23(2):110–5.

[12] Tidy JA, Brown BH, Healey TJ, Daayana S, Martin M, Prendiville W, et al. Accuracy of detection of high-grade cervical intraepithelial neoplasia using electrical impedance spectroscopy with colposcopy. BJOG 2013;120:400–11.

[13] Edebiri AA. The relative significance of colposcopic descriptive appearances in the diagnosis of cervical intraepithelial neoplasia. Int J Gynecol Obstet 1990;33:23–9.

[14] van der Marel J, van Baars R, Quint WVG, Berkhof J, del Pino M, Tormé A, et al. The impact of human papillomavirus genotype on colposcopic appearance: a cross-sectional analysis. BJOG 2014;121:1117–26.

[15] Huh WK, Cestero RM, Garcia FA, Gold MA, Guido RS, McIntyre-Seltman K, et al. Optical detection of high-grade cervical intraepithelial neoplasia in vivo. Am J Obstet Gynecol 2004;190:1249–57.

[16] Cantor SB, Cárdenas-Turanzas M, Cox DD, Atkinson EN, Nogueras-Gonzalez GM, Beck JR, et al. Accuracy of colposcopy in the diagnostic setting compared with the screening setting. Obstet Gynecol 2008;111(1):7–14.

[17] Ferris DG, Miller MD. Colposcopic accuracy in residency training program: defining competency and proficiency. J Fam Pract 1993;36(5):515–20.

[18] Stafl A, Mattingly RF. Colposcopic diagnosis of cervical neoplasia. Obstet Gynecol 1973;41:169–76.

[19] Cristoforoni PM, Gerbaldo D, Perino A, Piccoli R, Montz FJ, Capitanio GL. Computerized colposcopy: results of a pilot study and analysis of its clinical relevance. Obstet Gynecol 1995;85(6):1011–6.

[20] Benedet JL, Anderson GH, Matisic JP, Miller DM. A quality-control program for colposcopic practice. Obstet Gynecol 1991;78:872–5 5(1).

[21] Benedet JL, Boyes DA, Nichols TM, Millner A. Colposcopic evaluation of patients with abnormal cervical cytology. Br J Obstet Gynaecol 1976;83(3):177–82.

[22] Fung HY, Cheung LP, Rogers MS, To KF. The treatment of cervical intra-epithelial neoplasia: when could we 'see and loop'. Eur J Obstet Gynecol Reprod Biol 1997;72(2):199–204.

[23] Adjunctive colposcopy technologies for assessing suspected cervical abnormalities: the DYSIS colposcope with DYSISmap and the ZedScan I. Section 4.36 Table 3. https://www.nice.org.uk/guidance/dg32/chapter/4-Evidence#diagnostic-accuracy.

[24] Cervical Screening Programme, England 2017-18, National Statistics, https://digital.nhs.uk/data-and-information/publications/statistical/cervical-screening-programme/england—2017-18.

[25] NHS Cervical Screening Programme, Colposcopy and Programme Management, NHSCSP Publication number 20. Third Edition March 2016 Public Health England. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/515817/NHSCSP_colposcopy_management.pdf.

[26] Ulrich Petrya KU, Nieminenb PJ, Leesonc SC, Bergerond COMA, Redmane CWE. Review article: 2017 update of the European Federation for Colposcopy (EFC) performance standards for the practice of colposcopy. Eur J Obstetr Gynecol Reprod Biol 2018;224:137–41.

[27] Wentzensen N, et al. Evidence-based consensus recommendations for colposcopy practice for cervical Cancer prevention in the United States. J Low Genit Tract Dis 2017;21(4):216–22.