

RESEARCH

Open Access



# Development and validation of a clinical prediction model for endocervical curettage decision-making in cervical lesions

Li Yuanxing<sup>1,2</sup>, Luo Haixia<sup>1,2</sup>, Zhang Xiu<sup>1,2</sup>, Chang Jingjing<sup>1,2</sup>, Zhao Yueyang<sup>1</sup>, Li Jing<sup>1</sup>, Li Dongyan<sup>1</sup> and Wang Wei<sup>1\*</sup>

## Abstract

**Background:** In the absence of practical and reliable predictors for whether the endocervical curettage (ECC) procedure should be performed, decisions regarding patient selection are usually based on the colposcopists' clinical judgment instead of evidence. We aimed to develop and validate a practical prediction model that uses available information to reliably estimate the need to perform ECC in patients suspected of having cervical lesions.

**Methods:** In this retrospective study, 2088 patients who underwent colposcopy, colposcopically directed biopsy (CDB) and ECC procedures between September 2019 and September 2020 at the Second Hospital of Shanxi Medical University were included. The data were analyzed with univariate and multivariable logistic regression. Least absolute shrinkage and selection operator (LASSO) was used to select predictors for ECC positivity. The ECC prediction model was presented as a nomogram and evaluated in terms of discrimination and calibration. Furthermore, this model was validated internally with cross-validation and bootstrapping.

**Results:** Significant trends were found for ECC positivity with increasing age ( $P=0.001$ ), menopause ( $P=0.003$ ), Human papillomavirus (HPV) status ( $P<0.001$ ), severity of ThinPrep Cytological Test (TCT) ( $P<0.001$ ), original squamous epithelium ectopia ( $P=0.037$ ) and colposcopy impression ( $P<0.001$ ) by multivariable logistic regression analysis. The ECC prediction model was developed based on the following predictors: age, menopause, symptom of contact bleeding, severity of TCT, HPV status, cervix visibility, original squamous epithelium ectopia, acetowhite changes and colposcopic impression. This model had satisfactory calibration and good discrimination, with an area under the receiver operator characteristic curve (AUC) of 0.869 (95% confidence interval 0.849 to 0.889).

**Conclusions:** A readily applicable clinical prediction model was constructed to reliably estimate the probability of ECC positivity in patients suspicious of having cervical lesions, which may help clinicians make decisions regarding the ECC procedure and possibly prevent adverse effects.

**Keywords:** Cervical cancer, Endocervical curettage, Preoperative factors, Nomogram, Decision-making

\* Correspondence: [drwangwei@sxmu.edu.cn](mailto:drwangwei@sxmu.edu.cn)

<sup>1</sup>Department of Obstetrics and Gynecology, the Second Hospital of Shanxi Medical University, Taiyuan 030001, Shanxi, China

Full list of author information is available at the end of the article



© The Author(s). 2021 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

## Background

Cervical cancer is the fourth most common cancer among women worldwide, with an estimated 604,127 new cases and 341,831 deaths in 2020 [1]. The detection of high-grade cervical intraepithelial neoplasia (CIN) irrefutably is of interest as a precursor to invasive cervical cancer [2]. At present, the international community follows a three-step screening method: cervical thin-layer liquid-based cytology or cytology combined with human papillomavirus (HPV) detection is performed as the primary screening [3]; cases that are suspicious or positive are referred to colposcopy; and histopathological specimens of the lesions, including colposcopically directed biopsy (CDB) and endocervical curettage (ECC) then might be taken under colposcopy to diagnose cervical lesions [4]. Obtaining the endocervical status of a lesion is a true difficulty for colposcopists. Serious lesions, including cervical cancer, could potentially be missed if ECC is not performed [5]. ECC has been included as a component of clinicians' armamentarium in the evaluation of women who have abnormal cervical test results by the American Society for Colposcopy and Cervical Pathology (ASCCP) [6]. However, not every woman would benefit from the ECC procedure equally, and there is a long-standing debate over the effectiveness of this procedure, but until now, its value has remained unclear. The additional benefit of ECC in detecting CIN has been found to vary considerably. Moreover, ECC is hard to perform in patients with a stenotic cervix or in menopausal women [7]. The inter-observer agreement of the interpretation of ECC specimens is poor [8, 9]. ECC is an uncomfortable procedure and has been rated as a 5.8 (0 to 10) on a visual analog scale by patients in prior studies [10, 11]. Additionally, the specific indications for ECC remain debated. Colposcopists usually perform ECC selectively based on personal experience instead of the evidence, possibly resulting in a higher incidence of adverse effects [12]. Further identification of subgroups of women most likely to benefit from ECC is urgently needed, but studies have lacked appropriate data or statistical power for such comparisons.

To improve population health outcomes and healthcare costs for patients with CIN or cervical cancer, we need to provide the right treatment to the right people at the right time. In the case of colposcopy, the questions are whether ECC needs to be performed and in which patients. Predicting the probability of positive ECC can be challenging given the considerable heterogeneity in the characteristics of affected individuals and their clinical course as well as the variability in the region of lesions. Reliance on clinical intuition alone might be insufficient for accurate decision-making. This creates the need for a prediction model that aims to assist clinicians in the ECC decision-making process. We focused

on a group of patients with ECC results and attempted to construct a visualized model to predict the probability of ECC positivity. Nomograms could provide predictive information tailored to individual patients by establishing a simple, reliable and pragmatic graphical representation of a complex statistical prediction model [13]. Thus, the objective of our study was to develop and validate a clinical prediction model to select patients who would more likely benefit from ECC based on our clinical data.

## Methods

### Study design, setting and population

In this retrospective study, we included all consecutive patients who underwent ECC procedures performed with colposcopy and CDB and were referred to the Second Hospital of Shanxi Medical University, China, based on abnormal co-test findings (HPV testing+cytology) or the presence of unexplained contact bleeding between September 2019 and September 2020. This study was approved by the Ethics Committees of the Second Hospital of Shanxi Medical University. Patients with an abnormal co-test result were defined as those with positive HPV testing results and/or cytology findings demonstrating atypical squamous cells of undetermined significance (ASC-US) and above. Patients were excluded if they had a history of cervical physical therapy (ablation or cryosurgery), surgical operations (loop electrosurgical excision procedure, cold knife cone or hysterectomy) or pelvic radiotherapy, had nondiagnostic or unsatisfactory sampling, or had incomplete charts.

All patients underwent HPV testing, cytology, colposcopy, CDB and ECC procedures. Demographic information, including age, menopausal age and contact bleeding, was obtained via electronic medical records. The cytology technique adopted in this study was the ThinPrep (ThinPrep Pap test, Hologic, MA) Cytological Test (TCT). Cytological results were reported according to the 2014 Bethesda System [14]. The cytological classifications were negative for intraepithelial lesions or malignancy (NILM), ASC-US, low-grade squamous intraepithelial lesion (LSIL), atypical squamous cells cannot exclude an HSIL (ASC-H), high-grade squamous intraepithelial lesion (HSIL), squamous cell carcinoma (SCC), atypical glandular cells not otherwise specified (AGC-NOS), atypical glandular cells-favor neoplasia (AGC-FN), adenocarcinoma in situ (AIS), and adenocarcinoma of the cervix (AC). The HPV testing assays were carried out with flow-through hybridization and a HybridMax gene chip (HybridBio Biotechnology Ltd., China) with the residual Pap test specimens. Twenty-one HPV genotypes can be identified using this assay, including 15 high-risk HPV (HR-HPV) genotypes (HPV16, 18, 31,33, 35, 39, 45, 51, 52, 53, 56, 58, 59, 66 and 68) and 6 low-

risk HPV (LR-HPV) genotypes (HPV6, 11, 42, 43, 44 and CP8304 (81)). All of the cytology and HPV tests were performed in the hospitals according to the procedures provided by the suppliers.

Patients with positive HPV testing results and/or abnormal cytology ( $\geq$ ASC-US) or the symptom of unexplained contact bleeding were referred to colposcopy according to the protocol of this study. Electronic colposcopy (SLC-2000, Goldway, China) was used to visually assess the cervix, including cervix visibility, normal colposcopic findings (columnar epithelium ectopy and transformation zone (TZ) type), acetowhite changes (none, thin or dense) and Lugol staining (stained or non-stained) [15]. The impression of colposcopy includes normal/benign, low-grade, high-grade and cancer [6]. CDB was performed at the squamous-columnar junction in cases of visible lesions. If the colposcopic examination showed no lesions, random 4-quadrant punch biopsies were taken. ECC was performed with a Kevorkian curette. All colposcopy, CDB and ECC procedures were performed by the same gynecologist. Histological diagnoses were graded as NILM, CIN1–3, SCC, AIS or AC by two gynecological pathologists blinded to the cytology findings, HPV testing results and colposcopic impression, based on the Lower Anogenital Squamous Terminology Standardization Project for HPV-Associated Lesions (LAST) [16] and the WHO Classification of Tumors of Female Reproductive Organs (4th edition) [17]. The worst pathological result of CDB and ECC was taken as the final diagnosis.

### Variables selection

All known risk factors for ECC positive from keynote papers and consultation with clinical experts were considered in developing the prediction model. Based on the results of the literature review, we identified the following predictor variables that are assessable before ECC procedure and are consistently associated with our outcome of the prediction model: age [8, 18–23], abnormal cytology [8, 18, 20–22, 24], HPV infection (positive/negative) [20–22], cervix visibility (adequate/inadequate) [18, 22, 24], TZ type (I/II/III) [22] and colposcopic impression [8, 18, 21, 23, 24]. Given the models will be used in general practice, we also add menopause (yes/no), symptom of contact bleeding (yes/no), cervical atrophy (yes/no), original columnar epithelium ectopy (yes/no), acetowhite changes (none, thin or dense), Lugol staining (stained/nonstained) into the prediction model after initial analysis. Meanwhile, we have classified HPV infection in more details, including HPV negative, HPV16 +, HPV18 +, HPV16 and 18 +, HR-HPV+ (non 16/18 types), LR-HPV+. The coding of these variables is outlined in Additional file 1: Table S1.

### Outcome definition

ECC was classified as negative (NILM) or positive (ASC, CIN1–3, SCC, AIS or AC). The outcome of our prediction model was based on all patients with positive ECC findings according to pathology. The diagnosis based on ECC was determined by two gynecological pathologists blinded to the cytology findings, HPV testing results and colposcopy impression based on the criteria of LAST [16]. If the diagnosis or grade based on ECC was not clear during the review, the case was presented to a third pathologist expert blinded to the patient's potential predictors.

### Sample size

Currently, there is no standard way to calculate sample sizes for prediction models. With 14 predictors for our model, the number of events per variable is over 37. This is sufficient for the development of stable models.

### Model derivation

Baseline characteristics were summarized as frequency tables to study the distributions of categorical variables and as medians and interquartile ranges for continuous variables. Restricted cubic splines were used to explore nonlinearity in the effect of continuous variables. We performed univariable analysis using the chi-square test for each potential predictor to detect any important differences in proportions, with  $\chi^2$  values and *P* values for categorical variables. Logistic regression was used to estimate multivariable regression coefficients, and the prediction strength was quantified as odds ratios (ORs) with 95% confidence intervals (CIs).

Given the large number of predictors, we used least absolute shrinkage and selection operator (LASSO) regression [25] to reduce the number of candidate predictors and to select the most significant ECC-positive predictors to build the final prediction model. To control the shrinkage procedure, 10-fold cross-validation was used to determine the penalty parameter  $\lambda$  and to develop the optimal model.

### Model performance

The performance of the prediction model was studied in terms of the model calibration (the ability to produce unbiased estimates of the outcome probability) and the model discrimination (the ability to determine whether the patients should undergo ECC). Model calibration was evaluated graphically with calibration plots (plots of predicted versus observed outcomes) [26]. Model discrimination was evaluated with the sensitivity and specificity of the ECC patient selection prediction model analyzed by the receiver operator characteristic (ROC) curve and quantified based on the area under the ROC

curve (AUC; equivalent to the *c*-statistic). The AUC is reasonable above 0.7 and strong above 0.8.

**Model validation**

Internal validation evaluates the stability of a prediction model when random changes in sample composition occur. First, 10-fold cross-validation was performed. The model was fitted on a randomly selected subset of 90% of the study patients and tested on the remaining 10% of patients. This process was repeated 10 times to estimate the extent to which the predictive accuracy of the model was overoptimistic. Second, we performed 100 bootstrap resampling, drawn with replacement from the development sample. The model was refitted in each bootstrap replicate and tested to quantify the optimism (the decrease between the performance in the bootstrap sample and the performance in the original sample) in model performance.

**Model presentation**

The final model with the shrunken regression coefficients was presented as a regression equation and converted into a nomogram to facilitate clinical application. Each predictor value in the nomogram corresponds to a regression weight such that the total score is equivalent

to the linear predictor. For this model, logistic transformation was applied to the linear predictor to produce estimates of the probability of ECC positivity.

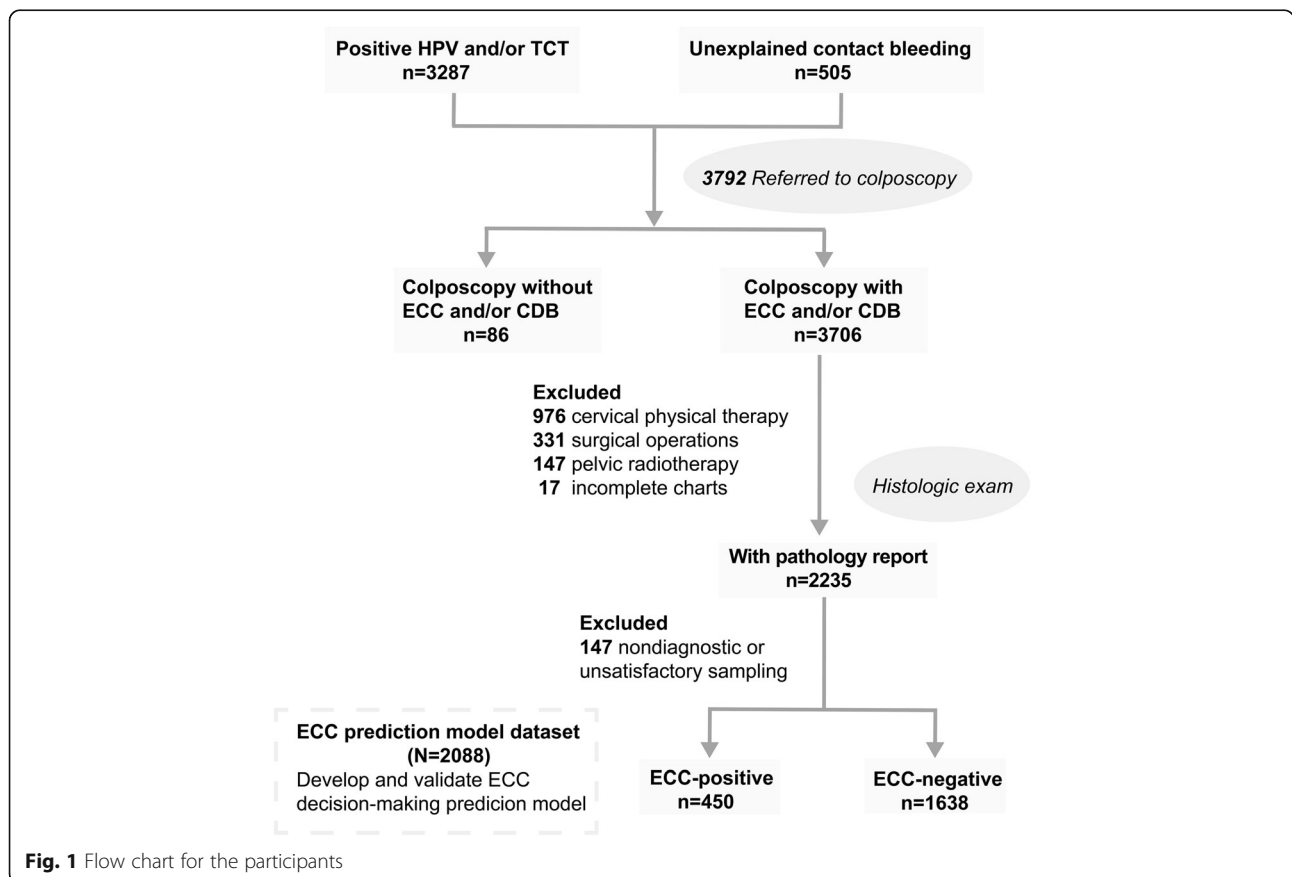
**Statistics**

All statistical tests were two-sided, and a *P*-value < 0.05 was considered statistically significant. All statistical analyses were performed in the R Studio statistics program (version 1.3.1093), and the glmnet and rms packages in the R Project (version R 4.0.3 GUI 1.73 Catalina build) were used.

**Results**

**Patient characteristics**

A total of 3792 patients were referred to colposcopy based on abnormal co-test findings (HPV testing/TCT) or the presence of unexplained contact bleeding. Then 3706 patients underwent ECC and CDB were enrolled. Of these, 1471 patients were excluded based on exclusion criteria, and 147 patients were excluded because of nondiagnostic or unsatisfactory sampling. Two thousand eighty-eight samples were finally included, with 1638 ECC-negative samples and 450 ECC-positive samples (Fig. 1). There were 994 (47.6%) samples with consistent ECC and CDB pathological findings, and 51 (2.4%)



**Fig. 1** Flow chart for the participants



samples with a worse diagnosis based on ECC than on CDB. Table 1 compares the distribution of variables between the ECC-positive group and the ECC-negative group. The median age of the patients was 52 (interquartile range 46–60) years in the ECC-positive group and 46 (interquartile range 38–54) years in the ECC-negative group. In univariate analysis, we found that the effect of age could be adequately modeled as a linear function ( $P$ -nonlinear = 0.6587) (Additional file 5: Figure S1A-B).

We found that there were statistically significant differences between the ECC-positive group and the ECC-negative group in age groups, menopause, symptom of contact bleeding, HPV Status, severity of TCT, cervix visibility, original squamous epithelium ectopia, cervical atrophy, TZ type, acetowhite changes, Lugol staining and colposcopic impression ( $P \leq 0.001$ ).

Significant trends were found for ECC positivity with increasing age ( $P = 0.001$ ), menopause ( $P = 0.003$ ), HPV status ( $P < 0.001$ ), severity of TCT ( $P < 0.001$ ), original squamous epithelium ectopia ( $P = 0.037$ ) and colposcopy impression ( $P < 0.001$ ) by multivariable logistic regression analysis (Fig. 2 and Additional file 2: Table S2). The risk of ECC positivity in women older than 60 years was 8.505 (95% CI: 2.030, 35.630) times higher than that in women younger than 30 years. Compared with patients with negative cytological findings, those with ASC-US, LSIL, ASC-H, HSIL respectively had 1.532 (95% CI: 1.081, 2.170), 1.614 (95% CI: 1.041, 2.501), 2.980 (95% CI: 1.572, 5.650) and 4.238 (95% CI: 2.493, 7.202) times higher risks of ECC positivity. Compared with patients with a colposcopic impression of normal/benign, those with low-grade, high-grade and cancer had 1.770 (95% CI: 1.020, 3.071) 3.713 (95% CI: 1.854, 7.436) and 32.837 (95% CI: 13.644, 78.932) times higher risks of ECC positivity respectively.

#### Predictor selection and model derivation

All other candidate predictor variables were included in the full model. We selected predictors using LASSO. Additional file 6: Figure S2 and Additional file 3: Table S3 show the results of the candidate variables included in the LASSO regression and their corresponding coefficients for the different values of the penalty parameter  $\lambda$ . We observed that when lambda.min was 0.003411379, 11 predictor variables remained. By increasing the  $\lambda$  value for further shrinkage, when lambda.1se was 0.01669495, 9 predictors remained in the final model. The final model included the variables age, menopause, symptom of contact bleeding, HPV status, TCT, cervix visibility, original squamous epithelium ectopia, acetowhite changes and colposcopic impression (Additional file 4: Table S4).

Figure 3 shows the nomogram for predicting the probability of ECC positivity in patients referred for colposcopy. Each predictor is assigned a specific grading value. When the grading value of the 9 predictors is determined, the total score can be obtained by adding them together and the probability of ECC positivity can be calculated.

#### Model performance and validation

The ROC curve of the nomogram in internal validation is shown in Fig. 4. The sensitivity and specificity were 82.2 and 75.4% respectively. The AUC was 0.869 (95% CI: 0.849–0.889). Overfitting was assessed by cross-validation and bootstrapping, and found to be negligible. The final model had an original  $c$ -statistic of 0.869 and a modified  $c$ -statistic of 0.863 according to 10-fold cross-validation. The apparent  $c$ -statistic was 0.859 with 0.0203 optimism according to the bootstrap procedure. Model optimism was negligible, both in the 10-fold cross-validation procedure and with bootstrapping. Calibration plots of the predicted and actual probabilities by cross-validation and bootstrapping were drawn (Fig. 5a-b). The predicted value was basically consistent with the actual ECC positivity rate.

#### Discussion

In this study, we screened out and evaluated influencing factors related to ECC positivity using mathematical methods. We proposed a novel ECC patient selection model to assess the risk of ECC positivity in patients referred for colposcopy. Using this model, patients with a higher risk of ECC positivity can be distinguished from those with a lower risk. Our ECC prediction model performed robustly in validation. The model was shown to be well calibrated and have good discrimination. The internally validated AUC was 0.869 (95% CI: 0.849–0.889) for the model to predict the probability of ECC positivity. The predictor items are assessable at hospital admission. The established nomogram could facilitate the adoption of the model into clinical practice.

#### Comparison with previous studies

Generally, the yield on ECC seems to be at an increased rate in the setting of inadequate or unsatisfactory colposcopy; in this situation, there is likely less controversy regarding the performance of ECC. In addition to the clear contraindication in certain populations, such as adolescents, immunocompromised patients, and pregnant women, there is also wide agreement that older women, in whom the cervical squamous-columnar junction is more difficult to visualize entirely [27], would benefit more from ECC; however, debate remains over the age cutoff to perform the ECC procedure [8, 18–23].

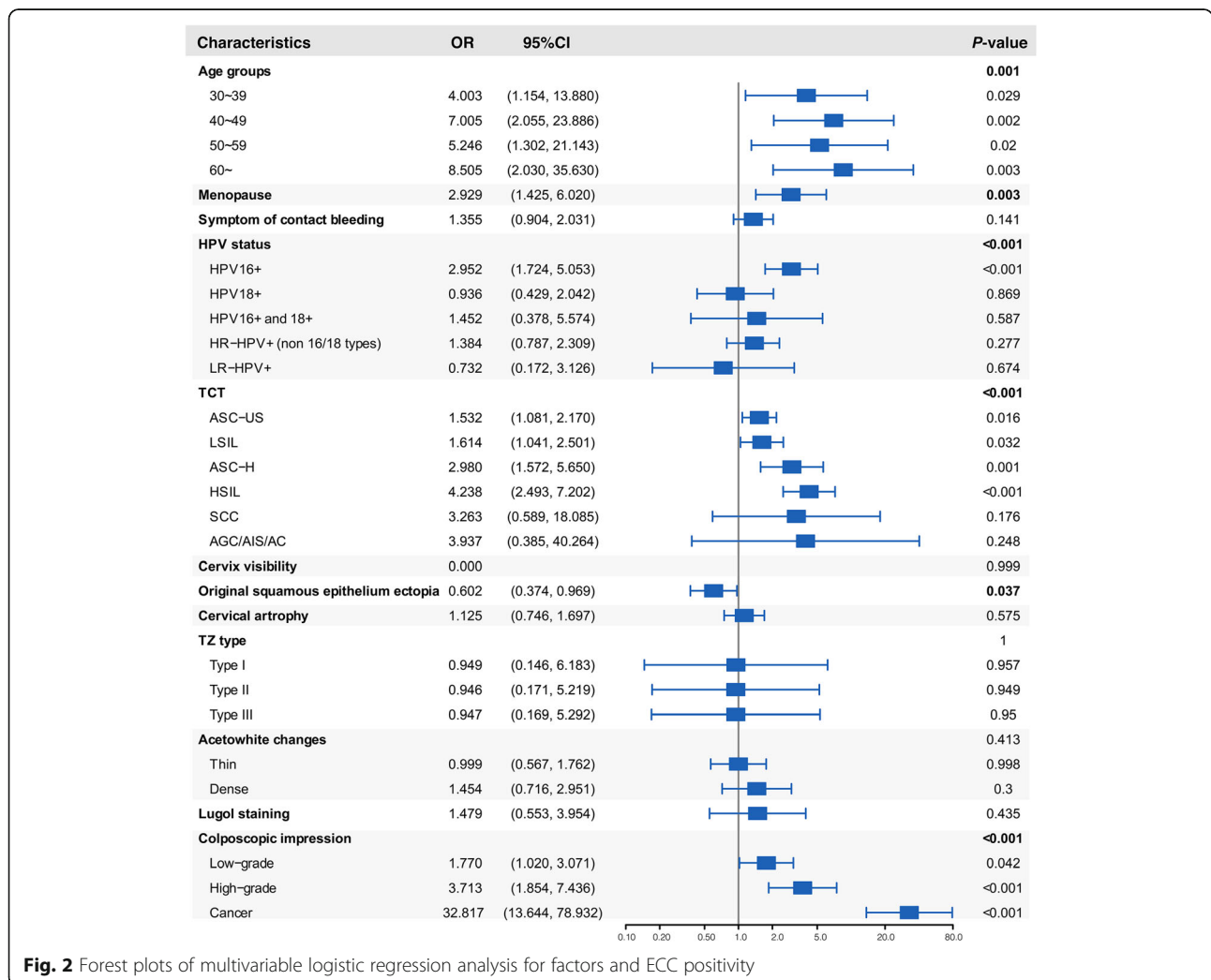
**Table 1** Baseline of clinicopathologic characteristics (N = 2088)

Characteristics	ECC positive (N = 450)	ECC negative (N = 1638)	$\chi^2$	P-value
Age <sup>a</sup>	52 (45, 60)	46 (38, 54)		
Age groups			132.498	< 0.001
~ 30	3 (0.7%)	104 (6.3%)		
30 ~ 39	54 (12%)	389 (23.7%)		
40 ~ 49	117 (26%)	549 (34%)		
50 ~ 59	152 (33.7%)	425 (26%)		
60~	124 (27.6%)	171 (10%)		
Menopause			97.411	< 0.001
No	172 (38%)	1050 (64%)		
Yes	278 (62%)	588 (36%)		
Symptom of contact bleeding			180.108	< 0.001
No	282 (63%)	1461 (89%)		
Yes	168 (37%)	177 (11%)		
HPV status			426.073	< 0.001
HPV negative	32 (7.1%)	197 (12%)		
HPV16+	264 (59%)	462 (28.2%)		
HPV18+	23 (5.1%)	145 (8.9%)		
HPV16 and 18+	7 (1.5%)	14 (0.9%)		
HR-HPV+ (non 16/18 types)	120 (27%)	790 (48.2%)		
LR-HPV+	4 (0.8%)	30 (1.8%)		
TCT			983.804	< 0.001
NILM	119 (26%)	1000 (61%)		
ASC-US	120 (27%)	379 (23%)		
LSIL	52 (12%)	175 (11%)		
ASC-H	40 (8.9%)	34 (2.1%)		
HSIL	96 (21%)	45 (2.7%)		
SCC	20 (4.4%)	2 (0.1%)		
AGC-NOS	1 (0.2%)	3 (0.2%)		
AGC-FN	1 (0.2%)	0 (0%)		
AIS	0 (0%)	0 (0%)		
AC	1 (0.2%)	0 (0%)		
Cervix visibility			28.407	< 0.001
Inadequate	9 (2.0%)	0 (0%)		
Adequate	441 (98%)	1638 (100%)		
Original squamous epithelium ectopia			23.972	< 0.001
No	414 (92%)	1353 (83%)		
Yes	36 (8.0%)	285 (17%)		
Cervical atrophy			18.697	< 0.001
No	350 (78%)	1411 (86%)		
Yes	100 (22%)	227 (14%)		
TZ type			112.127	< 0.001
Unrecognizable	16 (3.6%)	2 (0.1%)		
Type I	11 (2.4%)	122 (7.4%)		
Type II	309 (69%)	1320 (81%)		

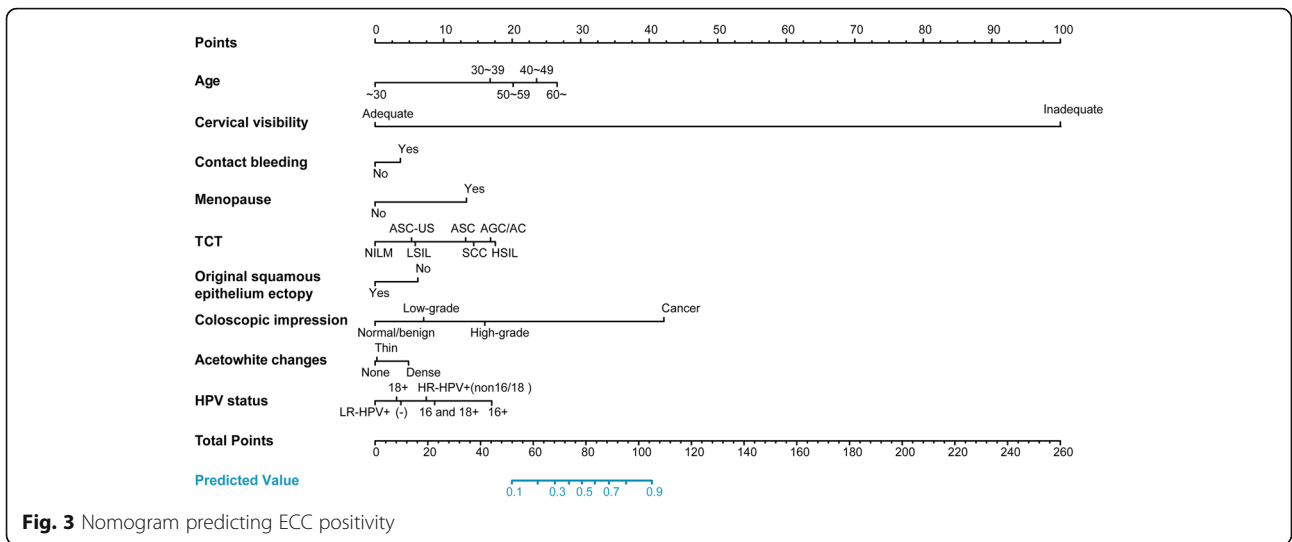
**Table 1** Baseline of clinicopathologic characteristics (N = 2088) (Continued)

Characteristics	ECC positive (N = 450)	ECC negative (N = 1638)	$\chi^2$	P-value
Type III	114 (25%)	194 (12%)		
Acetowhite changes			407.653	< 0.001
None	49 (11%)	514 (31%)		
Thin	112 (25%)	851 (52%)		
Dense	289 (64%)	273 (17%)		
Lugol staining			20.578	< 0.001
Nonstained	443 (98%)	1518 (92.7%)		
Stained	7 (1.6%)	120 (7.3%)		
Colposcopic impression			676.089	< 0.001
Normal/benign	50 (11%)	628 (38.3%)		
Low-grade	99 (22%)	739 (45.1%)		
High-grade	137 (30.4%)	253 (15.4%)		
Cancer	164 (36.4%)	18 (1.1%)		

<sup>a</sup>Statistics presented: Median (IQR); n (%)



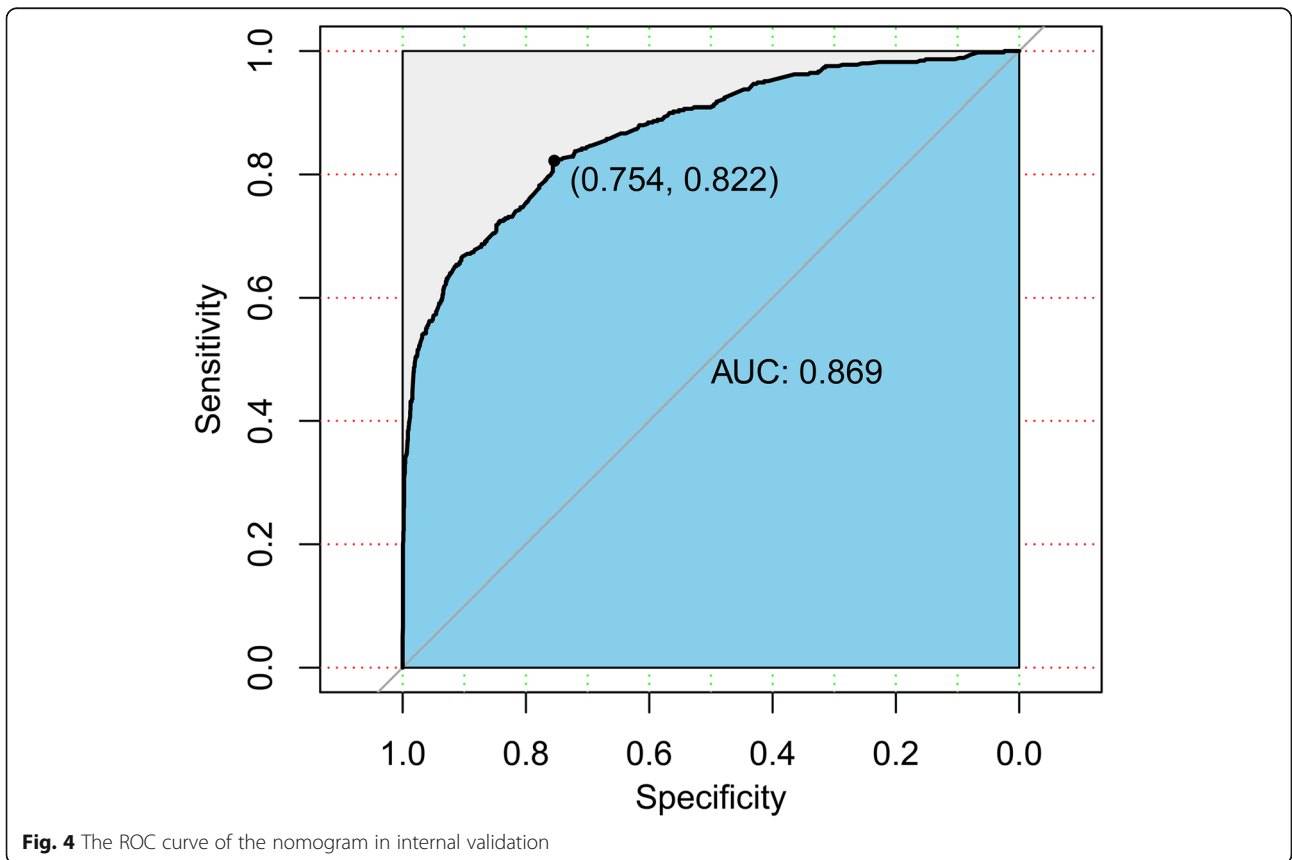
**Fig. 2** Forest plots of multivariable logistic regression analysis for factors and ECC positivity



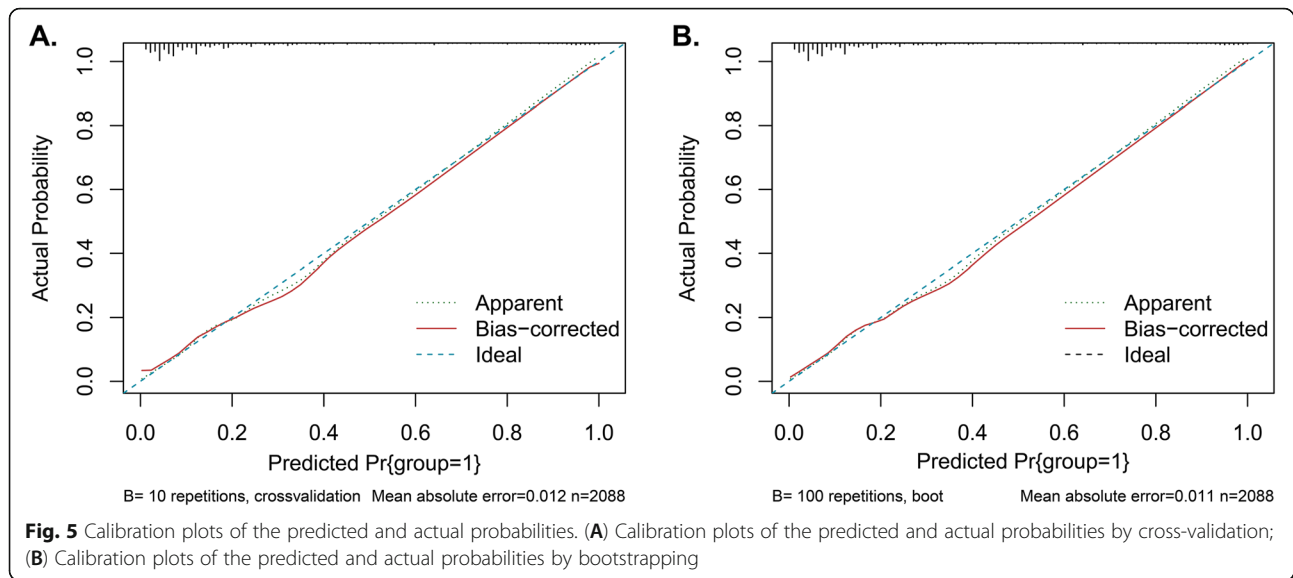
Another consideration is the grade of dysplasia. ECC was positive in more patients with severe dysplasia than in patients with mild or moderate dysplasia. It has also been reported that a positive ECC finding is directly related to the severity of the cytology results [8, 18, 20–22, 24] and colposcopy impression [8, 18, 21, 23, 24]. Some clinicians base the indication for ECC on the cytologic interpretation or colposcopic impression, although this

is not described in guidelines. However, the usefulness of ECC in the setting of high-grade dysplasia as a separate diagnostic tool is debatable. Our results confirm and extend the results from those studies, and our ECC prediction model included cytologic interpretation and colposcopic impression as important predictors.

The evidence for identifying women most likely to benefit from endocervical sampling based on TZ type







has also been inconclusive. Research has shown that even when the TZ is completely visible under colposcopy, 5% of patients still have positive ECC findings [22]. Even if the sensitivity of ECC is not strong, it may prompt colposcopists to perform the ECC procedure. Contrary to the findings of most previous studies, our study showed that the TZ type was not an independent predictor of ECC positivity. Incomplete visualization of the TZ is likely more linked to the colposcopist's experience and may not entirely raise concern for missed endocervical lesions.

Notably, we also discovered 5 powerful new predictors: the symptom of contact bleeding, HPV status, cervix visibility, original squamous epithelium ectopia and acetowhite changes. Although the most common reason for the referral of women for colposcopy is abnormal cervical cytology, there are the other indications highly recommendable for colposcopy, such as HPV16/18 positivity, persistent HR-HPV infection and the symptom of contact bleeding. In addition, a study claimed that women  $\geq 30$  years with contact bleeding should be referred for colposcopy to rule out the possibility of cervical cancer [28]. Therefore, the results of HPV testing and the presence of contact bleeding are valuable factors to determine whether ECC should be performed. HPV infection, especially HPV 16/18 subtype infection, should particularly raise concern for persistent oncogenic infection and trigger ECC. Interestingly, we subgrouped HPV status into several groups, including HPV negative, HPV16 +, HPV18 +, HPV16 and 18 +, HR-HPV+ (non 16/18 types), LR-HPV+, and HPV16 + was confirmed to be a significant predictor of ECC positive. Changes in the cervix can be highlighted with the application of exogenous contrast agents such as acetic acid and Lugol's

iodine. Moreover, prospective evaluation of the features of the cervix (original squamous epithelium mature/atrophic and columnar epithelium ectopy) was also considered in this analysis. Therefore, to our knowledge, this study is the first to assess the above factors and include them in a prediction model.

With the established model, which incorporates these and other variables, the probability of ECC positivity in women referred to colposcopy can be predicted with a high degree of accuracy.

#### Strengths and weaknesses of the study

A strength of the present study is that it was the first to develop and internally validate a model to determine whether the ECC procedure should be performed in cervical lesion screening. Second, to the best of our knowledge, no studies have yet examined the association of HR-HPV genotypes and colposcopic features, and our findings from this study attempted to fill this current knowledge gap by using multivariable logistic regression analysis. Third, LASSO analysis is a method that improves the accuracy of predictor selection and the validation of the model. Nevertheless, there are several potential limitations to our study. First, the model was derived from a retrospective analysis. Another limitation is that the prediction model has not been externally validated. Finally, the treating clinician's perception of the ECC procedure may have influenced the decision to provide aggressive support for these patients. As a result, the model may partly reflect the clinician's behavior in treating these patients. This model should be validated in a larger unselected population and re-evaluated in a multicenter study.

## Implications for patient selection, clinical practice, and research

Our findings have implications for the management of colposcopy procedures. It is not known whether patients referred to colposcopy should undergo only CDB or should be aggressively managed with ECC. The selected predictors allowed the development of a highly accurate model for predicting whether ECC should be performed, and the use of an ECC selection nomogram may aid in clinicians' empirical decision-making for such patients, thus reducing in the overuse of ECC.

## Conclusions

A readily applicable clinical prediction model was constructed to reliably estimate the probability of ECC positivity in patients suspicious of having cervical lesions, which may help clinicians make decisions regarding the ECC procedure and possibly prevent adverse effects.

## Abbreviations

ECC: The endocervical curettage; LASSO: Least absolute shrinkage and selection operator; HPV: Human papillomavirus; TCT: ThinPrep Cytological Test; AUC: Area under the receiver operator characteristic curve; CI: Confidence interval; CIN: Cervical intraepithelial neoplasia; CDB: Colposcopically directed biopsy; ASCCP: American Society for Colposcopy and Cervical Pathology; ASC-US: Atypical squamous cells of undetermined significance; NILM: Negative for intraepithelial lesions or malignancy; LSIL: Low-grade squamous intraepithelial lesion; ASC-H: Atypical squamous cells cannot exclude an HSIL; HSIL: High-grade squamous intraepithelial lesion; SCC: Squamous cell carcinoma; AGC-NOS: Atypical glandular cells not otherwise specified; AGC-FN: Atypical glandular cells-favor neoplasia; AIS: Adenocarcinoma in situ; AC: Adenocarcinoma of the cervix; HR-HPV: High-risk HPV; LR-HPV: Low-risk HPV; TZ: Transformation zone; LAST: Lower Anogenital Squamous Terminology Standardization Project for HPV-Associated Lesions; OR: Odds ratio; ROC: Receiver operator characteristic

## Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12885-021-08523-y>.

**Additional file 1: Table S1.** The coding of variables.

**Additional file 2: Table S2.** Multivariable logistic regression analysis of factors associated with ECC positivity.

**Additional file 3: Table S3.** The results of the candidate variables included in the LASSO regression and their corresponding coefficients for the different values of the penalty parameter  $\lambda$ .

**Additional file 4: Table S4.** The including predictors of final ECC prediction model.

**Additional file 5: Figure S1.** Age distribution of model development data and the association of age and ECC positivity (A-B).

**Additional file 6: Figure S2.** The progress of LASSO regression selecting the candidate variables and the penalty parameter  $\lambda$ .

## Acknowledgments

Not applicable.

## Authors' contributions

YL analyzed data and drafted the manuscript. HL, XZ and JC participated in data analysis. YZ and JL participated in data collection. DL and WW critically reviewed the manuscript. All authors read and approved the final manuscript.

## Funding

This work was supported by the grant from the Key Research and Development Program of Shanxi (201903D321152).

## Availability of data and materials

The data used in the current study are available from the corresponding author upon reasonable request.

## Declarations

### Ethics approval and consent to participate

All methods in this study were performed in accordance with the Helsinki Declaration's relevant guidelines and regulations. This study was approved by the Ethics Committees of the Second Hospital of Shanxi Medical University. Informed consent was obtained from all patients.

### Consent for publication

Not applicable.

### Competing interests

The authors declare that they have no competing interests.

### Author details

<sup>1</sup>Department of Obstetrics and Gynecology, the Second Hospital of Shanxi Medical University, Taiyuan 030001, Shanxi, China. <sup>2</sup>Shanxi Medical University, Taiyuan 030000, Shanxi, China.

Received: 15 April 2021 Accepted: 17 June 2021

Published online: 13 July 2021

## References

- Siegel RL, Miller KD, Fuchs HE, Jemal A. Cancer statistics, 2021. *CA Cancer J Clin.* 2021;71(1):7–33. <https://doi.org/10.3322/caac.21654>.
- Smith HJ, Leath CA 3rd, Huh WK, Erickson BK. See-and-treat for high-grade cytology: do young women have different rates of high-grade histology? *J Lower Genital Tract Dis.* 2016;20(3):243–6. <https://doi.org/10.1097/LGT.0000000000000229>.
- Castellsagué X. Natural history and epidemiology of HPV infection and cervical cancer. *Gynecol Oncol.* 2008;110(3 Suppl 2):S4–7. <https://doi.org/10.1016/j.jygyno.2008.07.045>.
- Wang W-J, Wang D, Zhao M, Sun X-J, Li Y, Lin H, et al. Serum lncRNAs (CCAT2, LINC01133, LINC00511) with squamous cell carcinoma antigen panel as novel non-invasive biomarkers for detection of cervical squamous carcinoma. *Cancer Manag Res.* 2020;12:9495–502. <https://doi.org/10.2147/CMAR.S259586>.
- Moniak CW, Kutzner S, Adam E, Harden J, Kaufman RH. Endocervical curettage in evaluating abnormal cervical cytology. *J Reprod Med.* 2000; 45(4):285–92.
- Perkins RB, Guido RS, Castle PE, Chelmow D, Einstein MH, Garcia F, et al. 2019 ASCCP risk-based management consensus guidelines for abnormal cervical Cancer screening tests and Cancer precursors. *J Lower Genital Tract Dis.* 2020;24(2):102–31. <https://doi.org/10.1097/LGT.0000000000000525>.
- Kishi Y, Inui S, Sakamoto Y, Mori T. Colposcopy for postmenopausal women. *Gynecol Oncol.* 1985;20(1):62–70. [https://doi.org/10.1016/0090-8258\(85\)90125-8](https://doi.org/10.1016/0090-8258(85)90125-8).
- Song Y, Zhao YQ, Li L, Pan QJ, Li N, Zhao FH, et al. A retrospective analysis of the utility of endocervical curettage in screening population. *Oncotarget.* 2017;8(30):50141–7. <https://doi.org/10.18632/oncotarget.15658>.
- Shah AA, Jeffus SK, Zhao Z, Stoler MH, Stelow EB. Adjunct p16(INK4a) immunohistochemistry aids the detection of high-grade squamous intraepithelial lesions in endocervical curettage specimens. *Am J Clin Pathol.* 2014;141(3):342–7. <https://doi.org/10.1309/AJCPDXD41YLVAZZN>.
- Driggers RW, Zahn CM. To ECC or not to ECC: the question remains. *Obstet Gynecol Clin N Am.* 2008;35(4):583–97; viii. <https://doi.org/10.1016/j.jogc.2008.09.007>.
- Korfage IJ, Essink-Bot ML, Westenberg SM, Helmerhorst T, Habbema JD, van Ballegooijen M. How distressing is referral to colposcopy in cervical cancer screening?: a prospective quality of life study. *Gynecol Oncol.* 2014;132(1): 142–8. <https://doi.org/10.1016/j.jygyno.2013.11.001>.

12. Landers EE, Erickson BK, Bae S, Huh WK. Trends in colposcopy volume: where do we go from Here? *J Lower Genital Tract Dis.* 2016;20(4):292–5. <https://doi.org/10.1097/LGT.0000000000000258>.
13. Kattan MW, Scardino PT. Evidence for the usefulness of nomograms. *Nat Clin Pract Urol.* 2007;4(12):638–9. <https://doi.org/10.1038/ncpuro0968>.
14. Nayar R, Wilbur DC. The pap test and Bethesda 2014. *Acta Cytol.* 2015;59(2):121–32. <https://doi.org/10.1159/000381842>.
15. Khan MJ, Werner CL, Darragh TM, Guido RS, Mathews C, Moscicki A-B, et al. ASCCP colposcopy standards: role of colposcopy, benefits, potential harms, and terminology for Colposcopic practice. *J Lower Genital Tract Dis.* 2017; 21(4):223–9. <https://doi.org/10.1097/LGT.0000000000000338>.
16. Darragh TM, Colgan TJ, Cox JT, Heller DS, Henry MR, Luff RD, et al. The lower Anogenital squamous terminology standardization project for HPV-associated lesions: background and consensus recommendations from the College of American Pathologists and the American Society for Colposcopy and Cervical Pathology. *J Low Genit Tract Dis.* 2012;16(3):205–42. <https://doi.org/10.1097/LGT.0b013e31825c31dd>.
17. Kurman RJ, Carcangiu ML, Herrington CS, Young RH. WHO Classification of Tumours of Female Reproductive Organs; 2014.
18. Gage JC, Duggan MA, Nation JG, Gao S, Castle PE. Detection of cervical cancer and its precursors by endocervical curettage in 13,115 colposcopically guided biopsy examinations. *Am J Obstet Gynecol.* 2010; 203(5):481.e481–9.
19. Pretorius RG, Belinson JL, Azizi F, Peterson PC, Belinson S. Utility of random cervical biopsy and endocervical curettage in a low-risk population. *J Lower Genital Tract Dis.* 2012;16(4):333–8. <https://doi.org/10.1097/LGT.0b013e3182480c18>.
20. Hu SY, Zhang WH, Li SM, Li N, Huang MN, Pan QJ, et al. Pooled analysis on the necessity of random 4-quadrant cervical biopsies and endocervical curettage in women with positive screening but negative colposcopy. *Medicine.* 2017;96(17):e6689. <https://doi.org/10.1097/MD.0000000000006689>.
21. Liu AH, Walker J, Gage JC, Gold MA, Zuna R, Dunn ST, et al. Diagnosis of cervical Precancers by Endocervical curettage at colposcopy of women with abnormal cervical cytology. *Obstet Gynecol.* 2017;130(6):1218–25. <https://doi.org/10.1097/AOG.0000000000002330>.
22. Mi L, Zhuan J, Zhang D, Hui B. Role of ECC in detection of HSIL+ under comprehensive colposcopy and multi-site biopsy. *Chin J Clin Obstet Gynecol.* 2019;20(02):128–31.
23. Zhang L, Du H, Zhang W, Yang B, Wang C, Belinson JL, et al. Diagnostic value of multiply biopsies and endocervical curettage on cervical lesions. *Zhonghua fu chan ke za zhi.* 2015;50(4):263–7.
24. van der Marel J, Rodriguez A, Del Pino M, van Baars R, Jenkins D, van de Sandt MM, et al. The value of Endocervical curettage in addition to biopsies in women referred to colposcopy. *J Lower Genital Tract Dis.* 2015;19(4):282–7. <https://doi.org/10.1097/LGT.0000000000000124>.
25. Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent. *J Stat Softw.* 2010;33(1):1–22.
26. Jaja BNR, Saposnik G, Lingsma HF, Macdonald E, Thorpe KE, Mamdani M, et al. Development and validation of outcome prediction models for aneurysmal subarachnoid haemorrhage: the SAHIT multinational cohort study. *Bmj.* 2018;360:j5745.
27. Gaffikin L, McGrath JA, Arbyn M, Blumenthal PD. Visual inspection with acetic acid as a cervical cancer test: accuracy validated using latent class analysis. *BMC Med Res Methodol.* 2007;7(1):36. <https://doi.org/10.1186/1471-2288-7-36>.
28. Xavier-Júnior JC, Vale DB, Zeferino LC, Dufloth RM. Association between concurrent genital bleeding and cervical cancer: a cross-sectional study. *Acta Obstet Gynecol Scand.* 2015;94(9):949–53. <https://doi.org/10.1111/aogs.12682>.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Ready to submit your research? Choose BMC and benefit from:**

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

**At BMC, research is always in progress.**

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

