

## ORIGINAL RESEARCH ARTICLE

# Accuracy of colposcopy in the Swedish screening program

Emilia Alfonzo<sup>1</sup>  | Chenyang Zhang<sup>2</sup> | Forogh Daneshpip<sup>1</sup> | Björn Strander<sup>1,2</sup>

<sup>1</sup>Department of Obstetrics and Gynecology, Institute of Clinical Sciences, Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden

<sup>2</sup>Regional Cancer Center West, Western Healthcare Region, Gothenburg, Sweden

**Correspondence**

Emilia Alfonzo, Department of Obstetrics and Gynecology, Sahlgrenska University Hospital, 41345 Gothenburg, Sweden.

Email: [emilia.alfonzo@vgregion.se](mailto:emilia.alfonzo@vgregion.se)

**Funding information**

Hjalmar Svensson Foundation, Grant/Award Number: HJSV2020068; LUA/ALF, Grant/Award Number: 11315

**Abstract**

**Introduction:** Sensitivity and specificity of colposcopy vary greatly between studies and efficacy in clinical studies seldom corresponds with effectiveness in a real-life setting. It is unclear whether colposcopists' experience affects assessment; studies show divergent results. The study's objective was to investigate the accuracy of colposcopies in the Swedish screening program, the variability in colposcopists' assessments and whether degree of experience affects accuracy in a routine setting.

**Material and methods:** Cross-sectional register study. All colposcopic assessments with a concomitant histopathological sample from women aged at least 18 years, performed between 1999 and September 2020 in Sweden. The main outcome measure was accuracy. The accuracy of colposcopic assessments was calculated as overall agreement with linked biopsies, with three outcomes: Normal vs Atypical, Normal vs Low-Grade Atypical vs High-Grade Atypical, and Non-High-Grade Atypical vs High-Grade Atypical. A time-trend analysis was performed. The accuracy of identifiable colposcopists related to experience was analyzed.

**Results:** In total, 82 289 colposcopic assessments with linked biopsies were included for analysis of the outcome Normal vs Atypical; average accuracy was 63%. Overrating colposcopic findings was four times more common than underrating. No time trend in accuracy was noted during the study period. Accuracy in distinguishing High-Grade from Non-High-Grade lesions was better: 76%. Among identifiable colposcopists, overall accuracy was 67%. Some had significantly better accuracy than others, but no correlation with experience was found.

**Conclusions:** Colposcopy, including in a referral setting, has low accuracy in distinguishing Normal from Atypical. Increased experience alone does not lead to improvement. This is supported by the substantial differences in performance between colposcopists.

**KEYWORDS**

accuracy, cervical cancer screening program, colposcopic assessment, colposcopic experience

**Abbreviations:** CI, confidence interval; CIN, cervical intraepithelial neoplasia; HPV, human papillomavirus; HSIL, High-Grade squamous intraepithelial lesion; RCT, randomized controlled trial.

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial-NoDerivs](https://creativecommons.org/licenses/by-nc-nd/4.0/) License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2023 The Authors. *Acta Obstetrica et Gynecologica Scandinavica* published by John Wiley & Sons Ltd on behalf of Nordic Federation of Societies of Obstetrics and Gynecology (NFOG).

## 1 | INTRODUCTION

Colposcopy is an essential part of the screening program aimed at detecting precancerous lesions and cancer in the cervix. The precision of this assessment is crucial, as it determines whether and when biopsies should be taken. Colposcopy performance has been intensively studied, and sensitivity, specificity and predictive values vary greatly, often due to different ways of measuring results.<sup>1</sup> In a meta-analysis with data from 1960–1996, addressing the sensitivity and specificity of colposcopy and aimed at identifying High-Grade intraepithelial lesions (HSIL) and cancer, the average sensitivity was 85%, the average specificity was 69% and the corresponding area under the curve was 0.82.<sup>2</sup> In Swedescreen, a randomized controlled trial (RCT) of human papillomavirus (HPV) testing in a screening program, the sensitivity of colposcopy categorized as atypical for distinguishing cervical intraepithelial neoplasia (CIN)2<sup>+</sup> was 83%.<sup>3</sup> In contrast, an American RCT only found 54% sensitivity in diagnosing CIN3.<sup>4</sup> Several other studies have also reported low sensitivity, and authors have suggested that biopsies also be taken from colposcopically normal areas.<sup>5–13</sup> Many studies suffer from verification bias, as the outcome depends on the biopsies from the very lesions identified by colposcopy. Moreover, histopathological analysis, the gold standard, has been found to have only moderate interobserver reproducibility.<sup>14</sup>

The issue of whether experience or professional background affects precision is debated. Several studies have shown similar sensitivity regardless of medical training background, and specially trained nurses, general gynecologists, junior gynecologists and colposcopists have been shown to perform equally well.<sup>5,15,16</sup> Staff with a shorter education or less professional experience, such as nurses and general gynecologists, are prone to overestimate colposcopic findings and tend to compensate with multiple biopsies.<sup>5,15</sup> In contrast, an RCT showed that senior colposcopists performed better in detecting HSIL compared with junior colleagues.<sup>17</sup> A retrospective German study yielded similar findings: the respective accuracy of colposcopists with <2, 2–10 and >10 years of experience in detecting HSIL was 87%, 85.3% and 90.2%, respectively.<sup>18</sup> Studies on interobserver agreement have reported low reproducibility.<sup>19,20</sup>

The aim of this study was to investigate the accuracy of colposcopies in the Swedish screening program, the development of accuracy over time, the variability in assessment between colposcopists and how professional experience affects colposcopic accuracy in a routine setting.

## 2 | MATERIAL AND METHODS

### 2.1 | Data sources

We performed a cross-sectional study using data from the Process Registry in the Swedish National Cervical Screening Registry. This registry contains data on cervical cytology, HPV and histopathology. It also contains colposcopic assessment reports for women in

### Key message

Colposcopy in a routine setting had low accuracy in distinguishing Normal from Atypical epithelium but improved for distinguishing High-Grade. Increased experience alone did not improve accuracy. Overrating colposcopic findings and nonidentification of High-Grade lesions underline the need for biopsies.

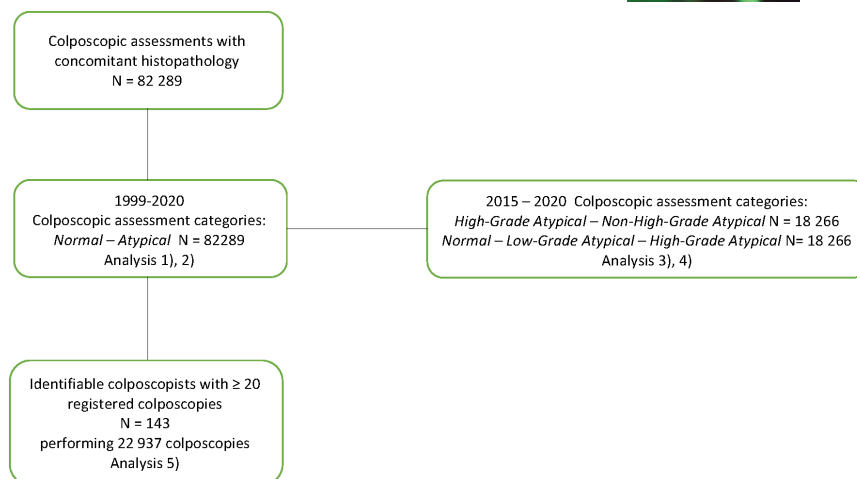
the screening program.<sup>21</sup> The Process Registry currently covers 85% of the Swedish population but colposcopy data coverage is lower and unevenly distributed, with most of the data coming from the Western Healthcare Region of Sweden. All laboratories in the affiliated regions deliver their complete histopathology, cytology and HPV data to the registry.<sup>21</sup> Data on histopathology specimen type are lacking, but most samples are derived from punch biopsies, and a minority from excisions. During the study period there has not been any recommendation in Swedish guidelines on whether biopsies should be taken when colposcopies were assessed as normal or on number of biopsies at abnormal colposcopy.

Colposcopies have been registered by the cytology form since 1999, initially with the categories Normal – Atypical – unsatisfactory. In 2008, a personal ID associated with the respective clinics was introduced by some laboratories. In 2012, this workplace-related ID was converted into a user-unique ID in parts of Western Sweden. This enables better follow-up if colposcopists change workplace. Since 2015, the Normal – Atypical categorization has gradually been replaced by a three-grade scale: Normal, Low-Grade Atypical and High-Grade Atypical. From this time forwards, Swedescore (0–10) and transformations zone type (1–3) have also been recorded in the registry. Manual registration of colposcopies, not necessitating a cytology form, was enabled in 2017. Since the start 1999, registration has successively expanded from one university laboratory in Western Sweden to be included in national guidelines since 2017. In 2020, regions outside the Western Region supplied 39% of the colposcopy data.

### 2.2 | Study population

We retrieved data on all colposcopic assessments, registered from 1999 to September 2020 in women aged at least 18 years, which could be linked to a histopathology report within ( $\pm$ ) 21 days. These colposcopies were performed by doctors as part of clinical routines in hospitals and gynecologic outpatient clinics, public and private, mainly in the Western Healthcare Region of Sweden. Reasons for colposcopy were abnormal cytology, in recent years in conjunction with positive HPV tests or follow-up of previous abnormalities, which was the less common indication. In cases of multiple discrepant histopathology diagnoses, the most severe was applied, according to the hierarchical classification in the registry. Colposcopies

**FIGURE 1** Overview of the study. Numerals (1, 2, etc.) show what population is used for analysis described in the Material and methods section.



assessed as unsatisfactory and/or with transformation zone type 3 were excluded.

### 2.3 | Statistical analyses

We performed five analyses on different subpopulations from the registry:

- Using all the data, Normal vs Atypical colposcopies with histopathological outcome were compared. In this analysis, the subset of colposcopic assessments registered as Low-Grade and High-Grade were merged into Atypical.
- These data were used for assessing accuracy over time with a time-trend analysis.
- Accuracy of the three-grade scale (see above) using an arbitrarily chosen scale of 0–1. One point was given for a perfect match and 0.5 point was given for a one-step mismatch, eg High-Grade colposcopic assessment and Low-Grade histopathology. No match was given 0 points, eg Normal colposcopy and High-Grade histopathology. In this analysis, Low-Grade histopathology was equivalent to CIN1 and koilocytosis, and High-Grade histopathology was defined as CIN2<sup>+</sup> (CIN2, CIN3, adenocarcinoma in situ [AIS] or cervical cancer).
- Dichotomized accuracy of colposcopic assessment of High-Grade vs Non-High-Grade (Low-Grade or Normal) was compared with the histopathology with the corresponding classification.
- Data with identifiable colposcopists enabled comparison of accuracy between colposcopists; accuracy related to degree of experience, calculated as number of registered colposcopies; and individual colposcopists' change in accuracy with increasing experience. Only colposcopists with  $\geq 20$  registered examinations in total were included in these analyses.

The basis for the analysis of colposcopies, as well as for analysis of the individual colposcopist's performance, comprised a linked histopathology sample. As mentioned above, each colposcopy was registered with the colposcopist's ID number. Accuracy was calculated

**TABLE 1** Accuracy of 82 289 colposcopies assessing Normal vs Atypical colposcopy.

Histology→ ↓Colposcopy	Benign		Atypical	
	n	%	n	%
Normal	11 076	13%	6 143	7%
Atypical	24 280	30%	40 790	50%

Note: Agreement in green cells. Non-agreement in red cells. Overall agreement 63% (95% CI 0.63–0.63).

as overall agreement. Differences in accuracy between colposcopists were calculated with the chi-square test. Pearson's correlation coefficient was used for calculations of accuracy related to experience, calculated as the total number of registered colposcopies per colposcopist. Linear regression was applied for analysis of change of accuracy with increasing experience in individual colposcopists. Change of accuracy with increasing experience for each colposcopist was also calculated for the three-grade scale and for High-Grade vs Non-High-Grade. Statistical analysis was performed with R.4.0.4.

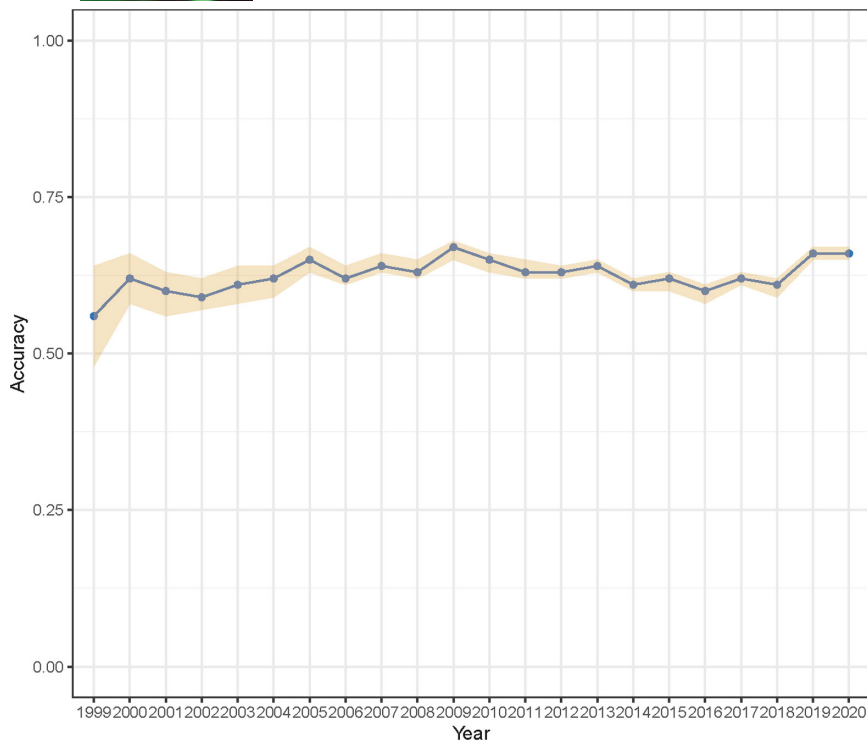
### 2.4 | Ethics statement

The Swedish Ethical Review Authority (reference number 2020-02271) approved the study on July 2, 2020.

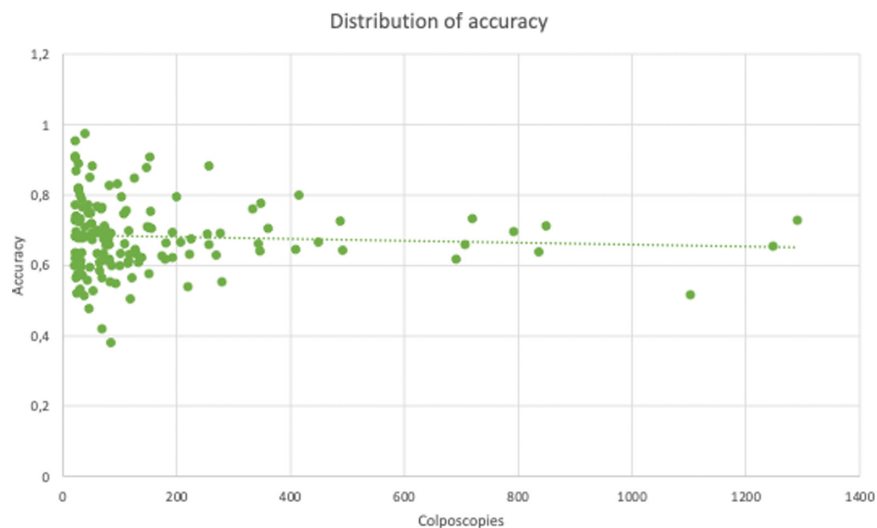
## 3 | RESULTS

The registry contained 82 289 colposcopic assessments linked to histopathological samples, performed during 1999–2020 and with outcome Normal or Atypical (Figure 1), which constituted the basis for comparison. Average accuracy was 63% (Table 1). Colposcopic findings were overrated four times more often than they were underrated. No time trend in accuracy was noted during the study period (Figure 2).

During the study period, 504 identified colposcopists performed a total of 24 362 registered colposcopies. Of these, 22 937 were



**FIGURE 2** Accuracy for assessing Normal vs Atypical in 82 289 colposcopies over time. Brown area is 95% CI.



**FIGURE 3** Accuracy for assessing Normal vs Atypical of 24 561 colposcopies performed by the 143 identifiable colposcopists with more than 20 previously registered colposcopies. Overall agreement = 0.68 (max = 0.97, min = 0.38, 2 SD = 0.21). Difference in accuracy between colposcopists: chi-square test = 734.48 ( $P < 0.0001$ ). Comparison of accuracy between colposcopists related to experience, calculated as number of registered colposcopies: Pearson's correlation coefficient = 0.0024 ( $P = 0.9786$ ; dotted regression line in the diagram). Change of accuracy with increasing experience per colposcopist: regression coefficient =  $-0.00016$ .

performed by 143 colposcopists with  $\geq 20$  previously registered examinations. Overall accuracy among the latter group for outcome Normal vs Atypical was 67% (data not shown). Some colposcopists had significantly better accuracy than others ( $P < 0.001$ ) and there was wide variability in performance (Figure 3). However, when accuracy was related to experience, calculated as the total number of colposcopies per colposcopist, no correlation was found (Figure 3) (correlation coefficient = 0.0024,  $P = 0.98$ ). Moreover, there was no

effect of increasing experience per colposcopist when the accuracy of every added colposcopy was compared with the colposcopist's previous results (regression coefficient =  $-0.00016$ ) (Figure 3). There was no effect of increased experience when it came to the three-grade scale or for the High-Grade vs Non-High-Grade comparison (data not shown).

There were records of 18 266 colposcopies with three-grade scale assessments; accuracy was 71% (Table 2). In the case of the

TABLE 2 Accuracy of 18 266 colposcopies categorized as Normal, Low-Grade Atypical or High-Grade Atypical.

Histology→ ↓Colposcopy	Benign			CIN1			CIN2 <sup>+</sup>		
	n	%	Weight	n	%	Weight	n	%	Weight
Normal	2705	15%	1	1002	5%	0.5	389	2%	0
Low-Grade	4244	23%	0.5	3855	21%	1	2253	12%	0.5
High-Grade	752	4%	0	950	5%	0.5	2116	12%	1

Note: Full agreement in green cells. Intermediate agreement in yellow cells. Non-agreement in red cells. Overall weighted agreement 71% (95% CI 0.70–0.71).

TABLE 3 Accuracy of 18 266 colposcopies categorized as High-Grade Atypical or Non-High-Grade Atypical.

Histology→ ↓Colposcopy	≤CIN1		CIN2 <sup>+</sup>	
	n	%	n	%
Non-High-Grade	11 806	65%	2 642	14%
High-Grade	1 702	9%	2 116	12%

Note: Agreement in green cells. Non-agreement in red cells. Overall agreement 76% (95% CI 0.75–0.76).

three-grade scale, overrating was twice as common as underrating. In all, 45% (8676/18 266) of the observations were totally accurate, whereas half (8449/18 266) deviated one step. Only 6% (1141/18 266) were clearly inaccurate and deviated two steps. When High-Grade or Non-High-Grade colposcopic assessment data were linked to histopathology ≤CIN1 or CIN2<sup>+</sup> data, 76% accuracy was found (Table 3). In contrast to the previous comparisons, underrating of colposcopic findings was slightly more common.

## 4 | DISCUSSION

In this study comprising 82 000 colposcopies in Sweden performed in 1999–2020, accuracy in distinguishing Normal from Atypical epithelium was found to be 63%. Overrating was four times more common than underrating. The accuracy did not change over the study period. Just over 24 000 colposcopies were attributed to an identifiable colposcopist. In this group, two-thirds of colposcopic diagnoses were correct, but colposcopists with many registered examinations performed no better than those with few registered examinations. Furthermore, no improvement in individual colposcopists' accuracy could be found with increasing experience. Colposcopic accuracy improved to 71% with the three-grade scale (Normal, Low-grade atypical and High-Grade Atypical) and reached 76% in distinguishing histopathological High-Grade from Non-High-Grade lesions.

In distinguishing Normal from Atypical, the main weakness was colposcopic overrating, which was four time more common than underrating. However, when High-Grade lesions were colposcopically distinguished from Non-High-Grade lesions, overrating was less common and underrating was more of a problem. This has clinical implications, due to the potential risk related to abstaining from biopsy. However, the finding that only 8% of High-Grade lesions

were considered colposcopically normal (Table 2) is somewhat reassuring.

To our knowledge, there is only one previous study with comparable population-based data,<sup>22</sup> comprising 84 244 patients examined by 37 colposcopists. That study found 79% accuracy in distinguishing benign from atypical lesions. The corresponding accuracy in our data was only 63%. The screening and colposcopy program in the British Columbia province of Canada is well known for its high professional standard and its results can be regarded as a benchmark. Compared with the Swedish program, the number of colposcopists involved was much lower, indicating the benefit of expertise.

Accuracy was measured for three different outcomes in our study: Normal vs Atypical, Normal vs Low-Grade Atypical vs High-Grade Atypical and Non-High-Grade Atypical vs High-Grade Atypical. The latter two exhibited superior performance. This is in accordance with the study from British Columbia, which demonstrated that the predictive accuracy of colposcopy increased with advancing severity of assessed disease.<sup>3</sup> Similar findings were described in an Australian study in which higher sensitivity in detecting CIN2<sup>+</sup> was found for High-Grade referral smears than for low-grade referral smears,<sup>16</sup> as well as in our own recent study on routine colposcopic performance with Swedescor.<sup>21</sup> The low accuracy of colposcopy for distinguishing Normal from Atypical (63%) and the high proportion of overrating concurs with the known low sensitivity for colposcopy as a screening method.<sup>23</sup> The fair ability of colposcopy, in this study, to distinguish High-Grade from other findings is clinically important, as the main purpose of colposcopy is to identify High-Grade lesions from which to take biopsies. It should be noted that the majority of CIN2<sup>+</sup> cases were assessed as Low-Grade (Tables 2 and 3), but no more than 389/4756 (8%) of the CIN2<sup>+</sup> cases were assessed as normal. Underrating was more common (14%) than overrating (9%). These signs of rather low performance of colposcopy in the Swedish program, support taking biopsies liberally, as well as adopting a realistic view of what colposcopy can detect and miss.

Our data demonstrated that some colposcopists performed significantly better than others but we failed to find any effect of experience or any effect of additional experience based on the performance of each added colposcopy for the individual colposcopist. Results were not altered when the outcomes High-Grade vs Non-High-Grade and Normal vs Low-Grade vs High-Grade were analyzed (data not shown). It is thus still unclear what characterizes more skilled colposcopists. Are some simply more gifted at pattern recognition? This hypothesis is supported by the previous finding

that superior performance may be a result of complex integration of learning.<sup>24</sup> Some researchers argue that an inverse relation may exist between duration of professional experience and quality of care.<sup>25</sup> Furthermore, a large Swedish population-based study in urology found no association between surgeon volume and surgical outcome. The authors emphasized the importance of individual feedback on results.<sup>26</sup>

Are our findings thus an effect of the lack of continuing education, so that initial mistakes and misinterpretations are simply repeated and not corrected over the years? Our results concur with previous smaller studies<sup>5,15,16</sup> in which neither professional background nor learning curve had any effect. However, the contrary was shown in an Italian RCT,<sup>17</sup> comparing juniors and seniors, in which a maximum of two biopsies per colposcopy were taken. Inexperienced colposcopists often take more biopsies, increasing sensitivity, which may explain the dissimilarity between these studies.<sup>16</sup> In a retrospective German study, with similar findings to the Italian RCT, less experienced colposcopists had a higher detection rate than more experienced colleagues.<sup>18</sup> Our study lacks data on actual colposcopic education and monitored training, and our findings do not contradict the need for formal training, certification and re-certification of colposcopists. There is a lack of prospective studies investigating whether training improves accuracy. However, in the impressively managed screening program in British Columbia<sup>22</sup> and among experts,<sup>21</sup> accuracy is substantially higher, indicating substantial potential for improvement of the Swedish program, a finding that probably is relevant for a large part of colposcopy services, globally.

Our findings could suggest a need for formalized training programs for colposcopists and conferences with discussions around colposcopy cases and images, as well as registry-based quality assurance of individual colposcopists' performance. We found no trend in accuracy during this century, which is a disappointment, since educational endeavors have increased since 2015, albeit to a relatively small extent.

There is clearly a need for quality assurance of colposcopy, which is ongoing in different countries as well as within the European Federation for Colposcopy (EFC). Accuracy, the number of correct assessments divided by all assessments, is easy to calculate and understand. However, there is no agreement concerning the criteria with which to define accuracy. We suggest, with some caution, that distinguishing High-Grade lesions, which should normally be treated, from those that should not, is the most important parameter. Indeed, these results were the best in this study, but the tendency to overrate rather than underrate and the large number of unidentified High-Grade lesions show the imperfection of colposcopy in the Swedish program and the need to take biopsies, including when lesions are assessed as low-grade or, in many instances, as benign.

There is a need for research on background factors that can explain differences in accuracy across colposcopists. Furthermore, there is a need to establish whether conventional education and training programs improve colposcopic performance, as our results indicate that this is questionable.

This is, to our knowledge, the first study based on routine colposcopies in a screening program outside British Columbia.<sup>22</sup> The strengths of our study are the population-based data, the large number of colposcopies and colposcopists and the time period of more than two decades.

This study also has several limitations. There may have been information bias. Despite national data retrieval, a majority of the identified colposcopies and colposcopists were from the Western Healthcare Region of Sweden, highlighting deficiencies in nationwide colposcopy registration. Moreover, not all colposcopies were linked to individual colposcopists by a unique ID, which may have led to underestimation of experience. There is certainly a verification bias, as the outcome was a histopathological diagnosis of samples from lesions identified by the method studied in most cases. Another selection bias derives from the inclusion criterion of a histopathological sample that could be unequivocally linked to the colposcopy. Thus, the most normal-appearing examinations, where colposcopists might have abstained from biopsy, are not represented in this study.

## 5 | CONCLUSION

Colposcopy, even in a referral setting, has low accuracy in distinguishing Normal from Atypical epithelium. Accuracy improves in identifying High-Grade lesions. Increasing experience alone does not correlate with increased accuracy. However, this does not rule out the importance of education and training, supported by the considerable differences in performance between colposcopists. Our study supports taking biopsies at routine colposcopy.

### AUTHOR CONTRIBUTIONS

EA and BS conceived and designed the study and formulated the research questions. FD reviewed medical records and matched colposcopists' IDs. CZ was responsible for data management and statistical analysis. EA, BS, CZ and FD analyzed the data. EA, BS and FD drafted the paper, which was revised and approved by all authors.

### FUNDING INFORMATION

This study was partially financed with a grant from the Hjalmar Svensson Foundation (grant number HJSV2020068), and with a LUA/ALF national grant (grant number 11315).

### CONFLICT OF INTEREST STATEMENT

None.

### ORCID

Emilia Alfonso  <https://orcid.org/0000-0002-2645-0988>

### REFERENCES

1. Brown BH, Tidy JA. The diagnostic accuracy of colposcopy – a review of research methodology and impact on the outcomes of quality assurance. *Eur J Obstet Gynecol Reprod Biol.* 2019;240:182-186.

2. 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;91:626-631.
3. Naucler P, Ryd W, Tornberg S, et al. Human papillomavirus and Papanicolaou tests to screen for cervical cancer. *N Engl J Med.* 2007;357:1589-1597.
4. 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;188:1383-1392.
5. Gage JC, Hanson VW, Abbey K, et al. Number of cervical biopsies and sensitivity of colposcopy. *Obstet Gynecol.* 2006;108:264-272.
6. Huh WK, Sideri M, Stoler M, Zhang G, Feldman R, Behrens CM. Relevance of random biopsy at the transformation zone when colposcopy is negative. *Obstet Gynecol.* 2014;124:670-678.
7. Nakamura Y, Matsumoto K, Satoh T, et al. Optimizing biopsy procedures during colposcopy for women with abnormal cervical cancer screening results: a multicenter prospective study. *Int J Clin Oncol.* 2015;20:579-585.
8. Wentzensen N, Walker JL, Gold MA, et al. Multiple biopsies and detection of cervical cancer precursors at colposcopy. *J Clin Oncol.* 2015;33:83-89.
9. Pretorius RG, Belinson JL, Burchette RJ, Hu S, Zhang X, Qiao YL. Regardless of skill, performing more biopsies increases the sensitivity of colposcopy. *J Low Genit Tract Dis.* 2011;15:180-188.
10. Pretorius RG, Bao YP, Belinson JL, Burchette RJ, Smith JS, Qiao YL. Inappropriate gold standard bias in cervical cancer screening studies. *Int J Cancer.* 2007;121:2218-2224.
11. Baasland I, Hagen B, Vogt C, Valla M, Romundstad PR. Colposcopy and additive diagnostic value of biopsies from colposcopy-negative areas to detect cervical dysplasia. *Acta Obstet Gynecol Scand.* 2016;95:1258-1263.
12. Vallapapan A, Chandeying N, Srijaipracharoen S, Uthagehaworn K. The role of random cervical biopsies in addition to colposcopy-directed biopsies in detection of CIN2. *J Obstet Gynaecol.* 2019;39:184-189.
13. Stoler MH, Vichnin MD, Ferenczy A, et al. The accuracy of colposcopic biopsy: analyses from the placebo arm of the Gardasil clinical trials. *Int J Cancer.* 2011;128:1354-1362.
14. Stoler MH, Schiffman M, Atypical Squamous Cells of Undetermined Significance-Low-grade Squamous Intraepithelial Lesion Triage Study (ALTS) Group. Interobserver reproducibility of cervical cytologic and histologic interpretations: realistic estimates from the ASCUS-LSIL Triage Study. *JAMA.* 2001;285:1500-1505.
15. Baum ME, Rader JS, Gibb RK, et al. Colposcopic accuracy of obstetrics and gynecology residents. *Gynecol Oncol.* 2006;103:966-970.
16. Bekkers RL, van de Nieuwenhof HP, Neesham DE, Hendriks JH, Tan J, Quinn MA. Does experience in colposcopy improve identification of high grade abnormalities? *Eur J Obstet Gynecol Reprod Biol.* 2008;141:75-78.
17. Bifulco G, De Rosa N, Lavitola G, et al. A prospective randomized study on limits of colposcopy and histology: the skill of colposcopist and colposcopy-guided biopsy in diagnosis of cervical intraepithelial lesions. *Infect Agent Cancer.* 2015;10:47.
18. Stuebs FA, Schulmeyer CE, Mehlhorn G, et al. Accuracy of colposcopy-directed biopsy in detecting early cervical neoplasia: a retrospective study. *Arch Gynecol Obstet.* 2019;299:525-532.
19. Massad LS, Jeronimo J, Schiffman M, National Institutes of Health/American Society for Colposcopy and Cervical Pathology (NIH/ASCCP) Research Group. Interobserver agreement in the assessment of components of colposcopic grading. *Obstet Gynecol.* 2008;111:1279-1284.
20. Jeronimo J, Massad LS, Castle PE, Wacholder S, Schiffman M, National Institutes of Health (NIH)-American Society for Colposcopy and Cervical Pathology (ASCCP) Research Group. Interobserver agreement in the evaluation of digitized cervical images. *Obstet Gynecol.* 2007;110:833-840.
21. Alfonzo E, Holmberg E, Milsom I, Strander B. Colposcopic assessment by Swedescore, evaluation of effectiveness in the Swedish screening programme: a cross-sectional study. *BJOG.* 2022;129:1261-1267.
22. Benedet JL, Maticic JP, Bertrand MA. An analysis of 84244 patients from the British Columbia cytology-colposcopy program. *Gynecol Oncol.* 2004;92:127-134.
23. Schneider A, Hoyer H, Lotz B, et al. Screening for high-grade cervical intra-epithelial neoplasia and cancer by testing for high-risk HPV, routine cytology or colposcopy. *Int J Cancer.* 2000;89:529-534.
24. Ericsson KA. Deliberate practice and acquisition of expert performance: a general overview. *Acad Emerg Med.* 2008;15:988-994.
25. Choudhry NK, Fletcher RH, Soumerai SB. Systematic review: the relationship between clinical experience and quality of health care. *Ann Int Med.* 2005;142:260-273.
26. Arnsrud Godtman R, Persson E, Bergengren O, et al. Surgeon volume and patient-reported urinary incontinence after radical prostatectomy. Population-based register study in Sweden. *Scand J Urol.* 2022;56:343-350.

**How to cite this article:** Alfonzo E, Zhang C, Daneshpaj F, Strander B. Accuracy of colposcopy in the Swedish screening program. *Acta Obstet Gynecol Scand.* 2023;102:549-555. doi:[10.1111/aogs.14538](https://doi.org/10.1111/aogs.14538)