UCSF
UC San Francisco Previously Published Works

Title

Reporting and Assessing the Quality of Diagnostic Accuracy Studies for Cervical Cancer Screening and Management

Permalink

https://escholarship.org/uc/item/5qg6q31j

Journal

Journal of Lower Genital Tract Disease, 24(2)

ISSN

1089-2591

Authors

Clarke, Megan A
Darragh, Teresa M
Nelson, Erin
et al.

Publication Date

2020-04-01

DOI

10.1097/lgt.0000000000000527

Copyright Information

This work is made available under the terms of a Creative Commons Attribution-NonCommercial-NoDerivatives License, available at https://creativecommons.org/licenses/by-nc-nd/4.0/

Peer reviewed

eScholarship.org Powered by the California Digital Library
University of California

OPEN

Reporting and Assessing the Quality of Diagnostic Accuracy Studies for Cervical Cancer Screening and Management

Megan A. Clarke, PhD, MHS,[1] Teresa M. Darragh, MD,[2] Erin Nelson, MD,[3] Elizabeth R. Unger, MD, PhD,[4] Rosemary Zuna, MD,[5] Miriam Cremer, MD, MPH,[6] Colleen K. Stockdale, MD, MS,[7] Mark H. Einstein, MD, MS,[8] and Nicolas Wentzensen, MD, PhD, MS[1]

Objective: We adapted the Quality Assessment of Diagnostic Accuracy Studies 2 (QUADAS-2) tool for studies of cervical cancer screening and management and used the adapted tool to evaluate the quality of studies included in a systematic review supporting the 2019 Risk-Based Management Consensus Guidelines.

Methods: We evaluated the quality of all studies included in our systematic review for postcolposcopy (n = 5) and posttreatment (n = 23) surveillance using QUADAS-2 criteria. Subsequently, we adapted signaling questions to indications of cervical cancer screening and management. An iterative process was carried out to evaluate interrater agreement between 2 study authors (M.A.C. and N.W.). Discrepant ratings were discussed, and criteria were adapted accordingly. We also evaluated the influence of study quality on risk estimates and between study variation using stratified subgroup meta-analyses.

Results: Twelve signaling questions for bias assessment that were adapted to or newly developed for cervical cancer screening and management are described here. Interrater agreement on bias assessment increased from 70% to 83% during the adaptation process. Detailed assessment of bias and applicability showed that all studies on postcolposcopy management and 90% of studies on posttreatment management had high risk of bias in at least 1 domain. Most commonly, high risk of bias was observed for the patient selection domain, indicating the heterogeneity of study designs and clinical practice in reported studies.

Conclusions: The adapted QUADAS-2 will have broad application for researchers, evidence evaluators, and journals who are interested in designing, conducting, evaluating, and publishing studies for cervical cancer screening and management.

Key Words: cervical cancer screening and management, guidelines, systematic review, quality assessment, QUADAS

(*J Low Genit Tract Dis* 2020;24: 157–166)

The 2019 American Society of Colposcopy and Cervical Pathology (ASCCP) Risk-Based Management Consensus Guidelines address management of patients with abnormal cervical cancer screening results and management of patients who are under surveillance after colposcopy or treatment of cervical precancers. Two components of the clinical guidelines development process were separated: a clinical consensus process that defined risk thresholds for surveillance intervals, colposcopy referral, and treatment. In parallel, risk estimates from screening and triage tests and prior screening information were calculated to assess a patient's risk of cervical precancer based on robust clinical databases and published data.

Systematic reviews of diagnostic accuracy studies are an important component of evidence assessment for developing clinical practice guidelines and can also demonstrate where current knowledge gaps exist to guide future research. The strength of evidence is largely dependent on the availability of high-quality studies that can be synthesized by meta-analyses to ensure generalizability and support recommendations.[1–3] Cervical cancer prevention is based on a multistep process involving population screening and management of screen-positive individuals to decide who needs treatment.[4] Consequently, the performance evaluation of tests for cervical cancer screening is context dependent and underlying populations and study designs vary depending on which step of the screening and management process is being evaluated. This context dependence of cervical cancer screening and management can complicate the quality assessment of diagnostic studies, because study designs and procedures for various clinical management scenarios of screen-positive individuals may differ from those designed to evaluate screening tests and vice versa. For instance, one cannot extrapolate the use of a test approved for screening to be equally as useful in posttreatment surveillance. Therefore, using standardized, context-adapted approaches to perform quality assessment is important to assure high-quality evidence reviews that are clinically meaningful.

Several tools have been developed to assess and improve the quality of diagnostic accuracy studies. The Standards for Reporting Diagnostic Accuracy statement, designed to improve the quality of reporting of diagnostic accuracy studies, is primarily targeted at researchers who publish their findings.[5] The Quality Assessment of Diagnostic Accuracy Studies 2 (QUADAS-2) tool is used to evaluate the risk of bias and applicability of diagnostic accuracy studies in systematic reviews.[6] These tools provide an important foundation for researchers to design and assess high-quality studies, but they are very general, and many criteria unique to cervical cancer screening and management studies are not addressed by these tools.

The creators of the QUADAS-2 tool intended for their generic guidance to be adapted for specific questions or systematic reviews.[6] The field of cervical cancer screening and management, with an ever-growing number of assays and a wide range of

[1]Division of Cancer Epidemiology & Genetics, National Cancer Institute/NIH, Bethesda, MD; [2]Department of Pathology, University of California San Francisco, San Francisco, CA; [3]Department of Obstetrics and Gynecology, University of Texas San Antonio, San Antonio, TX; [4]National Center for Emerging and Zoonotic Infectious Diseases, Centers for Disease Control and Prevention, Atlanta, GA; [5]Department of Pathology, College of Medicine, University of Oklahoma, Oklahoma City, OK; [6]Women's Health, Cleveland Clinic, Cleveland, OH; [7]Department of Obstetrics and Gynecology, Carver College of Medicine, University of Iowa, Iowa City, IA; and [8]Department of Obstetrics, Gynecology & Women's Health, Rutgers New Jersey Medical School, Newark, NJ

Reprint requests to: Megan Clarke, PhD, MHS, Clinical Genetics Branch, Division of Cancer Epidemiology & Genetics, National Cancer Institute, 9609 Medical Center Dr, Rm 6E552, Rockville, MD 20892. E-mail: megan.clarke@nih.gov; Nicolas Wentzensen, MD, PhD, MS, Clinical Genetics Branch, Division of Cancer Epidemiology & Genetics, National Cancer Institute, 9609 Medical Center Dr, Rm 6E448, Rockville, MD 20892. E-mail: wentzenn@mail.nih.gov

The authors have declared they have no conflicts of interest.

M.H.E. and N.W. share co-senior authorship.

Supported by American Society of Colposcopy and Cervical Pathology (ASCCP) and the Intramural Research Program of the National Cancer Institute.

The conclusions, findings, and opinions expressed by authors do not necessarily reflect the official position of the US Department of Health and Human Services, National Institutes of Health, or the Centers for Disease Control and Prevention.

Supplemental digital content is available for this article. Direct URL citations appear in the printed text and are provided in the HTML and PDF versions of this article on the journal's Web site (www.jlgtd.com).

Copyright © 2020 The Author(s). Published by Wolters Kluwer Health, Inc. on behalf of the ASCCP. This is an open-access article distributed under the terms of the Creative Commons Attribution-Non Commercial-No Derivatives License 4.0 (CCBY-NC-ND), where it is permissible to download and share the work provided it is properly cited. The work cannot be changed in any way or used commercially without permission from the journal.

DOI: 10.1097/LGT.0000000000000527

indications,[4,7] can especially benefit from a formal adaptation of quality criteria to improve the reliability, precision, and reproducibility of these criteria. As part of the 2019 ASCCP Risk-Based Management Consensus Guidelines effort, the New Technologies Working Group conducted a systematic review of the literature evaluating assays for management of abnormal cervical cancer screening results. In that context, a team of content experts reviewed the QUADAS-2 criteria and adapted them for use with studies of cervical cancer screening and management. This article describes the adaptation process and the application of these standard methods to the systematic review. Beyond the application to the current guidelines process, these criteria will support future evidence reviews and provide guidance for researchers and reviewers to improve design and reporting of diagnostic accuracy studies in the setting of cervical cancer screening and management.

METHODS

Adaptation of QUADAS-2 Criteria to Studies of Cervical Cancer Screening and Management

The QUADAS-2 tool was developed to assess the quality of diagnostic accuracy studies and is recommended for use in systematic reviews of diagnostic accuracy by the Agency for Healthcare Research and Quality, Cochrane Collaboration, the UK National Institute for Health and Clinical Excellence, and among others.[6] The tool covers 4 domains including: (1) Patient selection, (2) index test, (3) reference standard, and (4) flow and timing (i.e., the flow of patients through the study and timing of the index test [s] and reference standard). Each domain includes signaling questions that are designed to help the reviewer(s) assess the risk of bias and the first 3 domains are also assessed in terms of concerns regarding applicability. We formally adapted the QUADAS-2 criteria for diagnostic accuracy studies in cervical cancer screening and management. In support of the update of the 2019 ASCCP Risk-Based Management Consensus Guidelines, the New Technologies Working Group conducted systematic reviews of tests for postcolposcopy and posttreatment surveillance. All working group members participated in the assessment of study eligibility and full-text abstraction of relevant articles (Clarke et al., in this issue). During the abstraction process, working group members evaluated the quality of all 5 postcolposcopy studies[8–12] and 23 posttreatment studies[13–35] included in the systematic reviews using the generic QUADAS-2 criteria that uses 11 signaling questions across the 4 domains (Supplemental Table 1 http://links.lww.com/LGT/A147).[6] After this initial quality assessment, clinical criteria and signaling questions were discussed among the working group and adapted to cervical cancer screening and management questions, according to the guidance from QUADAS-2 (first adaptation, supplemental Table 1 http://links.lww.com/LGT/A147). This process relied on having a diverse working group with experts in different areas, including gynecology, pathology, epidemiology, and assay evaluation that are all relevant to different areas of the quality assessment. From the original QUADAS-2 questions, Q2, Q5, and Q8 were reworded. One question each was added to the index test and reference test domains. Q10 was dropped from the flow and timing domain, because the content was included in Q9. Subsequently, 2 authors (M.A.C. and N.W.) applied the adapted criteria for quality assessment of a random third of the articles (n = 7). Agreement between both observers was evaluated and discrepant ratings were discussed and resolved. This process identified that the domain of patient selection required further clarification. The signaling questions Q1 to Q3, and the related explanations were updated accordingly (final adaptation, supplemental Table 1 http://links.lww.com/LGT/A147). In addition, some descriptions for signaling question in the flow and timing category were updated as well. The final adaption included 3 signaling questions for each of the 4 domains. Five questions were kept unchanged from QUADAS-2 (Q4, Q7, Q9, Q11, Q12), 5 were adapted (Q1, Q2, Q3, Q5, Q10), and 2 were newly developed (Q6, Q8). Subsequently, 2 authors (M.A.C. and N.W.) evaluated 10 articles and achieved high agreement on the quality assessment. Consensus ratings were recorded for all studies that were evaluated by 2 observers. The remaining articles were divided between the 2 authors and underwent quality assessment using the final adapted criteria outlined hereinafter.

Study Quality Assessment and Statistical Analysis

Each study was assessed using 12 signaling questions (3 from each domain) and 3 questions regarding study applicability (1 each from the first 3 domains). The rating for each question is yes/no/unclear. "Yes" indicates a small risk of bias, whereas "no" indicates a high risk of bias for the specific question. "Unclear" indicates that the risk of bias could not be assessed because of missing information. We assessed agreement between both evaluators using 3 (yes/no/unclear) and 2 (yes or combined unclear/no) response levels. Agreement was calculated for each question, for each domain, and for the overall assessment.

Studies that were judged as "low" on all domains regarding bias or applicability were rated as having an overall low risk of bias or low concern regarding applicability. Studies that were judged as having high risk of bias in 1 or more domains were rated as having an overall high risk of bias or high concern regarding applicability.

To demonstrate the impact of quality ratings on test performance, we conducted subgroup meta-analyses to evaluate differences in the overall baseline risk of cervical intraepithelial neoplasia grade 2 or worse (CIN) 2+, the risks in test positives and test negatives, and between study variance quantified using the τ^2 statistic[36,37] in the subset of studies evaluating human papillomavirus (HPV) tests for posttreatment (n = 21) according to study quality. For these analyses, we evaluated the effects of having any risk of bias (1 or more domains ranked as high risk) and having 2 or more domains ranked as high risk. We also evaluated the effects of risk of bias and concerns about applicability for each domain. Where feasible, we assessed the influence of domain-specific signaling questions that had demonstrated variability in responses (Q2, Q3, Q6, and Q11). Statistical analyses were conducted in Stata SE Version 15 using the metaprop_one package.[37]

Role of the Funding Source

The guidelines effort received support from the National Cancer Institute and ASCCP. Participating organizations supported travel for their participating representatives. All participating consensus organizations, including the primary funders, had equal and balanced roles in the consensus process including data analysis and interpretation, writing of manuscript, and decision to submit for publication. No industry funds were used in the development of these guidelines. The corresponding authors had final responsibility for the submission decision.

RESULTS

Interobserver Agreement of Adapted Quality Assessment

For a first set of 7 studies, we evaluated the interrater agreement between both observers as an initial assessment of the adapted signaling questions. Across all 7 studies, the percent agreement was 70.2% when comparing individual responses of "yes" versus "unclear" versus "no" and 76.2% when comparing responses of "yes" versus "unclear/no." For individual signaling

© 2020 The Author(s). Published by Wolters Kluwer Health, Inc. on behalf of the ASCCP.

questions, the percent agreement ranged from 14.3% (Q9) to 100.0% (Q1, Q2, Q4, and Q7) when comparing individual responses. Based on these results, questions 1 and 2 were combined because these were interdependent and nondiscriminatory as standalone questions. To evaluate the adapted QUADAS-2 criteria, we compared quality assessment of 10 articles from a systematic review of posttreatment surveillance between 2 observers. Two reviewers independently assessed the criteria (see Table 2). The agreement using the revised signaling questions and clarifications increased to 83% for individual responses and to 87% for "yes" versus "unclear/no." For the first set of 7 studies, agreement for applicability ratings ranged from 57% for patient selection, 86% for index test, and 100% for reference standard. For the second set of 10 studies, agreement improved to 90% for patient selection and index test and 100% for reference standard.

The components of the quality assessment are summarized in Figure 1. The assessment domains include patient selection, index test, reference standard, and flow and timing. In the following sections, each domain and signaling questions are described within the context of cervical cancer screening and management studies.

Domain: Patient Selection

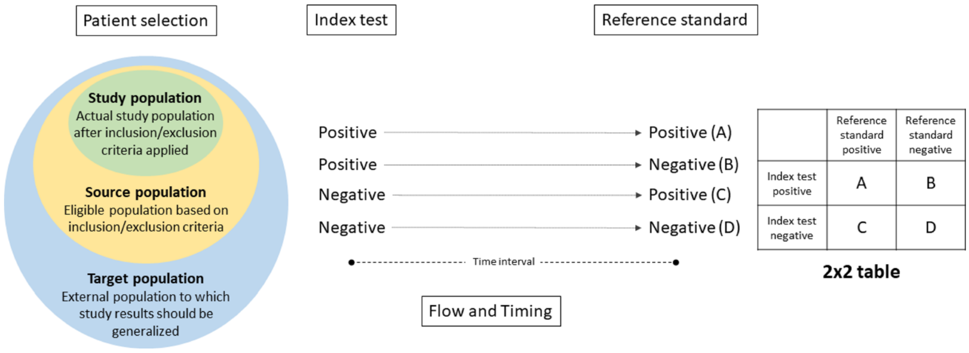
The patient selection domain addresses the question: "Could the selection of patients or study participants have introduced bias?" The constitution of the study population is centrally important to a high-quality study. We distinguished 3 populations, study, source, and target. The study population is the population that is reported on in an article, sampled from a larger source population. The target population is the group for whom study results are supposed to inform (see Figure 1). The bias assessment evaluates differences between the study population and the source population, whereas the applicability question evaluates whether findings from the study population will apply to the target population.

Signaling Question 1: Was a Consecutive or Random Sample of Patients Enrolled? If Sampling Was Performed, Was the Sampling Frame Adequately Described?. The study population should be representative of the underlying source population. Enrolling a consecutive series of a large number of patients can ensure a representative sample, unless there are specific variations over time, like changes in screening or clinical practice, that would alter the constitution of the source population. A random sample taken from the source population for an extended period can also yield a study population that is representative of the source population. Study designs that enrich for individuals at higher risk of precancer can be very efficient, particularly for studies of cervical cancer screening and triage. However, it is critical that the enriched population is weighted back to the source population for the purpose of reporting of accuracy and risk estimates. Successful weighting requires unbiased sampling of cases and controls, as well as knowledge of the sampling frame.

Signaling Question 2: Did the Study Avoid Inappropriate Exclusions?. All exclusions made when creating the study population from the source population need to be documented and described. Exclusions should be prespecified and should not be related to specific index test or reference test results. Exclusions are considered inappropriate if they alter the study population in such a way that potentially introduces bias and/or does not address the intended research question. Examples of inappropriate exclusions that are likely to introduce bias include: excluding unsatisfactory index test results or exclusions based on a specific screening result.

Signaling Question 3: Did the Study Avoid Exclusion of a Substantial Number From the Source Population Due to Missing Data?. When patients are included from routine clinical practice or electronic medical records, important data needed for study inclusion may be missing in eligible participants. Any exclusions need to be clearly described and enumerated. If the data are missing at random, risk of bias should be low. As the reasons for missing information cannot be sufficiently evaluated, large proportions (>15–20%) of missing data may introduce bias. Examples that may introduce bias if the proportion of missing data is substantial include missing index test results and patients lost to follow-up.

Applicability: Are There Concerns That the Included Patients Do Not Match the Research Question?. This question addresses whether the findings from the study population will be applicable to the target population. Applicability may be affected when the age range and age distribution of the study



	Reference standard positive	Reference standard negative
Index test positive	A	B
Index test negative	C	D

FIGURE 1. Components of quality assessment for diagnostic accuracy studies. The assessment domains include patient selection, index test, reference standard, and flow and timing. The patient selection domain addresses potential sources of bias in the selection of patients included in the study (i.e., the study population). The study population is sampled from a larger source population (the population that is eligible based on study inclusion criteria). The target population is the external population to which results are intended to inform. The index test domain addresses potential sources of bias for the assay/test under investigation in the study. A simple schematic shows possible outcomes of positive or negative index test results. The reference standard domain addresses potential sources of bias in the ascertainment and/or measurement of study outcomes. A simple schematic shows possible outcomes of positive or negative results for the reference standard. Together, the results from the index test and reference standard can be included in a 2 × 2 table to evaluate the diagnostic accuracy of an index test. Flow and timing addresses potential sources of bias in the flow and timing of procedures carried out in the study.

© 2020 The Author(s). Published by Wolters Kluwer Health, Inc. on behalf of the ASCCP.

population differ from what is expected in routine clinical practice, when the extent of previous screening and management is not comparable with the target population, and/or if the study is restricted to individuals with certain screening results. In addition, there are other population characteristics that may influence the risk of disease, for example, HIV prevalence or immunosuppression, which may introduce concerns about applicability if the intended target is the general population.

Considerations for Specific Indications

For screening and triage, the source population should reflect individuals undergoing routine cervical cancer screening in the appropriate age range (e.g., 21–65 years), and the study population should include a series of consecutively screened patients or a random sample that can be weighted back to the full screening population. If the population is enriched for disease endpoints (e.g., from a colposcopy referral population), then a sampling frame and weighting scheme should be described that allows extrapolation back to the full screening population. Importantly, when colposcopy clinics are used for enrichment, use of referral criteria that differ from the study question risks the introduction of bias (e.g., if the colposcopy population is based on cytology screening, but the study is evaluating HPV screening). A screening population should include individuals without prior screening abnormalities because previous test results are an important indicator of risk. If the screening history of patients is not known, this should be specifically stated. For postcolposcopy surveillance, the source population is individuals who have undergone colposcopy without treatment; including those with CIN 2 or less. The study population should include a series of consecutive patients with previous colposcopy within 6 to 18 months. Previous screening methods and postcolposcopy management algorithms should be clearly described (e.g., cytology-based screening vs co-testing). For posttreatment surveillance, the source population should reflect individuals who have undergone excisional treatment for histologically confirmed CIN 2, CIN 3, or adenocarcinoma in situ and should not include patients treated for cervical cancer, as well as no or few (≤10%) treated less than CIN 2. The study population should include a series of consecutive patients with previous treatment in the last 6 to 18 months, and the treatment modality should be described.

Domain: Index Test

The index test domain addresses the question: “Could the conduct or interpretation of the index test have introduced bias?” The index test results are one central component of a 2 × 2 table that is evaluated in diagnostic studies (see Figure 1). The index test is the assay under investigation in the study, and a study may evaluate 1 or more index tests in the same population or among population subsets. The study methods should provide a clear description of the index test, how it was conducted, and how it was and interpreted. Index tests that are currently used in clinical practice for cervical cancer screening and management include cervical cytology and HPV testing (including HPV DNA and mRNA testing). Current candidate diagnostic tests for cervical cancer screening and management include technologies such as extended HPV genotyping,[38] p16/Ki-67 dual stain,[39–42] cellular and viral DNA methylation,[43–47] and automated visual evaluation.[48]

Signaling Question 4: Were the Index Test Results Interpreted Without Knowledge of the Results of the Reference Standard?. In clinical practice, cervical cancer screening tests and management (e.g., cytology or HPV testing) are conducted before performing the reference standard (colposcopy and biopsy) being performed. Therefore, if the index test is being conducted as part of routine clinical practice, it is reasonable to assume that it was interpreted without knowledge of the reference test results, even if this is not explicitly stated in the Methods. In studies that use previously collected, banked samples to evaluate an index test that was not part of routine clinical care, results of the candidate index test should be interpreted without knowledge of the results from the reference standard. The Methods should clearly state that reference standard results were not known to those interpreting the index tests under evaluation.

Signaling Question 5: Was the Threshold for the Index Test Prespecified?. It is presumed that all index tests used in cervical cancer screening have a threshold; therefore, a description of the prespecified threshold (i.e., test positivity cutoffs or result categories of the index test) should be clearly stated in the methods section. If the index test is a commercially available kit, it is sufficient to state that it was carried out according to routine practice and/or the manufacturer's instructions with the appropriate references included.

Signaling Question 6: Were Patients With Inadequate, Indeterminate, or Missing Index Test Results Reported?. Exclusion of patients in the study population with inadequate, indeterminate, or missing index test results may introduce bias. Therefore, these results should be enumerated in the Methods or Results section and ideally reported with results from the reference standard to assess the risk of disease in these patients.

Applicability: Are There Concerns That the Index Test, Its Conduct, or Its Interpretation Differ From the Research Question?. Factors that may affect the diagnostic accuracy results of the index test, and therefore the applicability of study findings, include variations in test technology (e.g., manual vs automated cytology or dual stain testing, sampling methods, assays used for high-risk HPV testing, and the number of HPV types classified as “high-risk”), training and expertise of those performing and evaluating the index test if expected setting for use differs from the study (e.g., commercial laboratory vs a clinical setting), and variation in sample storage times and conditions (e.g., comparing results from freshly collected samples or extracts to those stored or frozen before or after extraction).

Domain: Reference Standard

The reference standard domain addresses the question: “Could the reference standard, its conduct, or its interpretation have introduced bias?” The reference test results are the other central component of the 2 × 2 table that is evaluated in diagnostic studies, describing the study endpoints or outcomes (see Figure 1). The study methods should provide a clear description of the reference test, how it was conducted, and how it was interpreted. The reference standard for cervical cancer screening and management studies is based on histologic evaluation of cervical biopsies and/or excisional treatment specimens. However, not all patients have results for the reference test as they may not undergo biopsy or excisional treatment. Depending on other clinical characteristics (e.g., negative HPV and negative cytology and/or normal colposcopic impression), these patients may be considered as noncases in screening and management studies.

Signaling Question 7: Is the Reference Standard Likely to Correctly Classify the Outcomes?. The reference standard in cervical cancer screening and management studies is histologic evaluation of biopsy specimens collected during colposcopy or excisional treatment. This reference standard requires a combination of high-quality colposcopy and histology.[49] To obtain complete disease ascertainment, colposcopy evaluation needs to be conducted

© 2020 The Author(s). Published by Wolters Kluwer Health, Inc. on behalf of the ASCCP.

according to high standards and biopsies need to be taken from all acetowhite lesions. Similarly, histologic evaluation requires high standards and interpretation should follow established nomenclature and diagnostic criteria. When using the lower anogenital squamous terminology (LAST) nomenclature,[50] the use of p16 or other immunohistochemical assays in diagnostic framework should follow guidelines systematically and be clearly described.

Signaling Question 8: Are the Distributions of CIN 2, CIN 3, and Cancer Endpoints Reflective of the Case Mix That Would Be Encountered in Clinical Practice?. A study that is skewed toward having too little or too few CIN 3 endpoints, for example, may signal a risk of bias in the conduct and interpretation of the reference standard. At a minimum, CIN 3 should be enumerated separately from CIN 2 outcomes in the results section so that readers can determine the distribution of these two endpoints. Cancers are typically very rare in screening and management studies. Ideally, primary endpoints for diagnostic accuracy should include CIN 3+, with CIN 2+ reported separately. If LAST criteria are used, it is important to report high-grade squamous intraepithelial lesion results as qualified by –CIN 2 or –CIN 3, according to the options given by the LAST guidelines.[50]

Signaling Question 9: Was the Interpretation of the Reference Test Carried Out Without Being Influenced by the Specific Results of the Index Test?. Typically, the reference test should be interpreted independent of results of the index test. However, in cervical cancer screening and management, clinical algorithms may reveal some index test results (e.g., when HPV-positive individuals are referred to colposcopy). Furthermore, pathology practice may include evaluating the complete clinical picture when histologic assessment occurs, i.e., all screening test results are revealed to the pathologist and the pathologist may review prior cytology or histology slides. This practice may vary in different settings (e.g., academic teaching centers versus commercial diagnostic laboratory) as well as between individual pathologists and cannot be controlled in studies that rely on clinical databases. Detailed reporting of pathology practice is important to assess the risk of bias in these situations.

Applicability: Are There Concerns That the Reference Standard, Its Conduct, or Its Interpretation Differ From the Research Question?. Current risk-based screening and management guidelines are based on CIN 3+ endpoints; therefore, studies reporting only CIN 2+ may be difficult to extrapolate to risk estimates used for clinical decision-making. Furthermore, the constitution of the group of endpoints can affect the applicability of the study findings. For example, the number of cancers included in CIN 3+ endpoints can vary substantially between studies and may affect risk estimates for various tests.

Domain: Flow and Timing

The flow and timing domain addresses the question: “Could the study flow and timing have introduced bias?” The Methods and Results sections should provide a clear description of clinical referral algorithms (i.e., patients who did/did not receive the index tests or reference standard, respectively) and of any patients excluded from the analyses. The follow-up time interval and any interventions between the index test(s) and the reference standard should also be clearly described.

Signaling Question 10: Was the Time Interval Between Sample Collection for Testing and Application of the Reference Standard Acceptable and Similar to What Would Be Done in Clinical Practice?. The time interval between sample collection for index testing and the application of the reference standard should reflect what is done in routine clinical practice (see considerations for specific indications hereinafter). At a minimum, the results should report summary measures of follow-up time for all participants (e.g., mean/median with standard deviations and/or ranges). If the interval is outside of the range of what is considered normal by routine clinical standards for a given indication, this could introduce bias. Ideally, a description of the clinical referral algorithm should be provided so that the timing of index testing and application of the reference standard can be assessed by reviewers.

Signaling Question 11: Did All Patients Receive the Reference Standard?. Generally, in diagnostic studies, it is expected that all study participants receive a reference standard. However, because HPV-negative women in the general population have a very low risk of cervical precancer, and most a general population is negative for HPV, the reference standard is typically not applied to the whole population in cervical cancer screening studies. Referring all individuals with negative screening tests to colposcopy would be an extensive and unnecessary burden. For management indications, the tradeoff between complete disease ascertainment and unnecessary referral is altered and the requirements may be different, as outlined hereinafter. When the reference standard of colposcopy and biopsy is not applied to all patients, there needs to be a clear description of what approach was used to determine disease endpoints (see considerations for specific indications hereinafter). Importantly, the referral algorithm should not introduce bias. When evaluating an index test, unbiased comparisons require colposcopy referral at some point for all patients testing positive. When evaluating multiple index tests, all individuals testing positive for any test should be referred to colposcopy equally.

In some studies, patients who were supposed to receive a reference standard may be lost to follow-up. This can introduce bias when the risk in the patients who did not undergo the reference standard differs from the risks in patients who do. The degree to which this may introduce bias is increased as the proportion of individuals who are lost to follow-up increases (captured in Q3 in the patient selection domain).

Signaling Question 12: Were All Enrolled Study Patients Included in the Analysis?. There is potential for bias if the number of patients enrolled differs from the number of patients included in the 2 × 2 table analysis and results. Some patients may not be included in the analysis because of missing results of 1 or more tests. Such differences should be clearly stated or described in a Consolidated Standards of Reporting Trials diagram.

Considerations for Specific Indications

For screening and triage, different study designs exist. Clinical or registration trials typically refer all individuals positive for any test to colposcopy immediately and conduct follow-up for at least 3 years. For studies integrated in clinical practice, referral algorithms and times may differ depending on screening test results. Therefore, the reference standard may not be equally applied to the whole population. For example, in a co-testing setting, patients who are HPV positive with atypical squamous cell of undetermined significance or worse undergo immediate colposcopy, whereas patients who are HPV+ with negative for intraepithelial lesions or malignancy may return for repeat testing in 1 year. Similarly, individuals with negative cytology and negative HPV results do not need to undergo colposcopy, and biopsies are not taken. If their screening results are negative, it is acceptable to assume no precancer or cancer in these individuals. Verification bias adjustment

© 2020 The Author(s). Published by Wolters Kluwer Health, Inc. on behalf of the ASCCP.

has been proposed to account for the lack of histologic endpoints in individuals with negative screening results. Verification bias adjustment involves conducting tests for the reference standard (colposcopy/biopsy and histology) in a subset of individuals with negative screening results, followed by weighting back to the full population. However, the risk of precancer is very low in screen-negative individuals, and there is a risk of detecting morphologic look-alike lesions that are not caused by carcinogenic HPV and are not true precancers, leading to incorrect adjustments of precancer prevalence. In postcolposcopy studies, similar arguments can be made and it is not required that all women are sent to colposcopy if the tests used for management provide sufficient reassurance against disease. For posttreatment studies, it is expected that individuals are evaluated with colposcopy at some point during follow-up. Because the size of the population is relatively small, and the baseline risk is higher compared with a screening population, unnecessary colposcopy is much less of a concern.

Quality Assessment of Studies Included in Systematic Review of Postcolposcopy and Posttreatment Management

Consensus ratings for the 5 postcolposcopy surveillance studies and the 23 posttreatment studies included in the systematic review are shown in Tables 1 and 2, respectively. Overall, studies tended to be of low quality, with all 5 of the postcolposcopy studies having at least one domain rated as high risk of bias and 21 (91%) of the posttreatment studies having at least one domain with a high risk of bias. All 5 postcolposcopy studies and 12 of the posttreatment studies (52%) were rated as having high risk of bias in at least 2 domains.

Evaluation of domain-specific results for the 23 posttreatment articles demonstrated variability with respect to overall domain ratings as well as responses to individual signaling questions. For example, only 3 studies (13%) were rated as having a low risk of bias for the patient selection domain. Studies were more likely to be considered as having a low risk of bias for the index test and reference standard domains (n = 12, 52%, respectively) and the flow and timing domain (n = 9, 39%). Within domains, ratings varied substantially by signaling question. For example, within the patient selection domain, only one study (5%) was judged as having high risk of bias for Q1, whereas for Q2 and Q3, those numbers increased to 9 (43%) and 13 (62%), respectively. This demonstrates that inappropriate exclusions and exclusions of a high percentage from the base population are common in the literature. Very few studies were judged as high risk of bias for index test domain signaling questions Q4 (n = 0) and Q5 (n = 1, 5%), reference standard domain signaling questions Q7 (n = 1, 5%) and Q9 (n = 0, 0%), and flow and timing signaling questions Q10 (n = 0, 0%) and Q12 (n = 1, 5%), whereas Q8 (reference standard) and Q11 (flow and timing) had more studies judged as high risk of bias (n = 11 [52%] and n = 9 [43%], respectively). Concerns about applicability were observed in 7 (33%) studies for the patient selection, 6 (29%) studies for the index test, and 3 (14%) studies for the reference standard domains. A total of 5 (24%) posttreatment studies had concerns about applicability in more than one domain.

Influence of Study Quality on Between-Study Variance and Risk Estimates

To demonstrate how study quality assessment may influence performance estimates reported in the systematic review, we performed analyses of the baseline and posttest risks as well as τ^2 estimates of between study variation, stratified by different measures of study quality (see Table 3). Overall for studies with 2 or more domains with high risk of bias, the risk estimates were not

TABLE 1. Consensus Quality Assessment Ratings Using the Adapted QUADAS-2 Criteria for Postcolposcopy Surveillance Studies (n = 5)

	Patient selection					Index test					Reference standard					Flow and timing			
Signaling questions	**Q1**	**Q2**	**Q3**	**Risk of bias**	**Applicability concerns**	**Q4**	**Q5**	**Q6**	**Risk of bias**	**Applicability concerns**	**Q7**	**Q8**	**Q9**	**Risk of bias**	**Applicability concerns**	**Q10**	**Q11**	**Q12**	**Risk of bias**
Cortecchia S, 2013[8]	Y	Y	N	High	High	Y	Y	Y	Low	Low	Y	N	Y	High	Low	U	Y	Y	Unclear
Gurumurthy M, 2014[9]	Y	Y	Y	Low	High	Y	Y	N	High	Low	Y	N	Y	High	Low	Y	Y	N	High
Kang WD, 2018[10]	Y	N	U	High	High	Y	Y	N	High	High	Y	Y	Y	Low	Low	Y	Y	Y	Low
Tverelv LR, 2018[11]	Y	N	N	High	Low	Y	Y	N	High	Low	Y	Y	Y	Low	Low	Y	U	Y	Unclear
Ye J, 2017[12]	Y	Y	U	Unclear	High	Y	Y	U	Unclear	Low	Y	N	Y	High	Low	N	U	U	High

N indicates no; Y, yes.

© 2020 The Author(s). Published by Wolters Kluwer Health, Inc. on behalf of the ASCCP.

TABLE 2. Consensus Quality Assessment Ratings Using the Adapted QUADAS-2 Criteria for Posttreatment Studies (n = 23)

	Patient selection					Index test					Reference standard					Flow and Timing			
Signaling questions	**Q1**	**Q2**	**Q3**	**Risk of bias**	**Applicability concerns**	**Q4**	**Q5**	**Q6**	**Risk of bias**	**Applicability concerns**	**Q7**	**Q8**	**Q9**	**Risk of bias**	**Applicability concerns**	**Q10**	**Q11**	**Q12**	**Risk of bias**
Bruhn LV, 2018[13]	Y	Y	Y	Low	Low	Y	Y	Y	Low	Low	Y	Y	Y	Low	Low	Y	U	Y	Unclear
Byun JM, 2018[14]	Y	Y	N	High	Low	Y	Y	N	High	High	Y	N	Y	High	Low	Y	U	Y	Unclear
Ceballos KM, 2017[15]	Y	Y	N	High	Low	Y	Y	Y	Low	Low	Y	Y	Y	Low	Low	Y	Y	Y	Low
Cubie HA, 2014[16]	U	Y	U	Unclear	Low	Y	Y	N	High	Low	Y	Y	Y	Low	Low	Y	N	Y	High
Du R, 2013[17]	Y	N	N	High	High	Y	Y	Y	Low	High	Y	Y	Y	Low	Low	Y	Y	Y	Low
Fan A, 2018[18]	Y	N	N	High	High	Y	N	Y	High	High	Y	N	Y	High	High	Y	Y	Y	Low
Friebe K, 2017[19]	Y	Y	N	High	Low	Y	Y	N	High	Low	N	N	Y	High	Low	Y	U	Y	Unclear
Gosvig CF, 2015[20]	Y	N	N	High	Low	Y	Y	Y	Low	Low	Y	N	Y	High	Low	Y	N	Y	High
Hansen J, 2017[21]	U	U	U	Unclear	Low	Y	Y	U	Unclear	Low	Y	Y	Y	Low	Low	U	Y	Y	Unclear
Herfs M, 2015[22]	Y	Y	N	High	Low	Y	Y	Y	Low	Low	Y	Y	Y	Low	Low	Y	Y	Y	Low
Innamaa A, 2015[23]	Y	N	N	High	High	Y	Y	Y	Low	High	Y	N	Y	High	Low	Y	Y	Y	Low
Kalampokas E, 2018[24]	Y	N	U	Unclear	High	Y	Y	U	Unclear	Low	Y	Y	Y	Low	Low	U	Y	Y	Unclear
Kang WD, 2016[25]	Y	N	U	High	High	Y	Y	N	High	High	Y	Y	Y	Low	Low	Y	N	Y	High
Khunamornpong S, 2015[26]	N	U	U	High	Low	Y	Y	Y	Low	Low	Y	N	Y	High	Low	Y	N	Y	High
Kong TW, 2014[27]	Y	N	N	High	Low	Y	Y	Y	Low	Low	Y	N	Y	High	Low	Y	Y	Y	Low
Lubrano A, 2012[28]	Y	U	N	High	Low	Y	Y	Y	Low	Low	Y	N	Y	High	High	Y	Y	Y	Low
Molloy M, 2016[29]	Y	Y	Y	Low	Low	Y	Y	Y	Low	Low	Y	Y	Y	Low	Low	Y	N	Y	High
Persson M, 2012[30]	Y	U	U	Unclear	High	Y	Y	U	Unclear	High	Y	N	Y	High	Low	U	U	N	High
Polman NJ, 2017[31]	Y	Y	Y	Low	Low	Y	Y	Y	Low	Low	Y	Y	Y	Low	Low	Y	N	Y	High
Ryu A, 2012[32]	Y	N	N	High	Low	Y	Y	Y	Low	Low	Y	Y	Y	Low	Low	Y	Y	Y	Low
Torné A, 2013[33]	Y	U	U	Unclear	High	Y	Y	U	Unclear	Low	Y	N	Y	High	Low	Y	N	Y	High
Wu J, 2016[34]	Y	Y	N	High	Low	Y	Y	N	High	Low	Y	N	Y	High	Low	Y	N	Y	High
Zhao C, 2014[35]	Y	N	N	High	Low	Y	Y	N	High	Low	Y	Y	Y	Low	Low	Y	N	Y	High

N indicates no; Y, yes.

© 2020 The Author(s). Published by Wolters Kluwer Health, Inc. on behalf of the ASCCP.

statistically significantly different than those studies with one or fewer domains having high risk of bias (p value for heterogeneity >0.1 for all comparisons). Among studies with low risk of bias overall and for each domain, τ^2 estimates tended to be lower than those with high risk of bias and compared with the overall τ^2 estimates (see Table 3). Bias within individual domains had different effects on baseline and posttest risk estimates, but these differences were not statistically significant.

For applicability measures, we observed differences in the baseline and posttest risk estimates, but these differences were not statistically significant. Among studies with low concerns about applicability for each domain, τ^2 estimates tended to be lower than those with high concerns, with the exception of the reference standard domain.

DISCUSSION

A thorough assessment of the quality of diagnostic accuracy studies included in systematic reviews and meta-analyses is essential for interpreting the strength and robustness of evidence used for clinical decision-making in cervical cancer screening and management.[6] Issues with the design, conduct, or reporting of diagnostic accuracy studies can potentially lead to bias and/or results that are not applicable to the research question of interest. To assess the quality of diagnostic accuracy studies included in the systematic review for the 2019 ASCCP Risk-Based Management Consensus Guidelines, we adapted a widely used tool for evaluating the quality of diagnostic accuracy studies, the QUADAS-2,[6] for studies of cervical cancer screening and management. Our adapted tool provides specific guidance for how to evaluate the risk of bias and applicability of studies in the context of the various cervical cancer screening and management settings (screening and triage, postcolposcopy surveillance, and posttreatment). In a meta-analysis of risk estimates, 2 factors are important: the actual risk estimate and the CI around the risk estimate (i.e., precision). Both factors can be affected in biased studies. If studies are all judged to have the same risk of bias that acts toward a specific direction (e.g., higher risk), the bias should have a predictable effect on the overall risk estimate within this subgroup (higher risk compared with studies without this bias). However, if the bias affects the risk estimate unpredictably in both directions, the subgroup-specific risk estimate may not be different from the overall risk estimate, but the precision could be lower, and the between-study variation will be higher. In context of the guidelines, it is important how close a risk estimate is in relation to a clinical management threshold. If the risk is very close to the threshold, the precision of the risk estimate from a systematic review and meta-analysis becomes more important than if the risk is very different from the threshold (Cheung et al., in this issue).

The iterative process we applied for adapting the QUADAS-2 tool resulted in high interobserver agreement for evaluating risk of bias and concerns about study applicability in studies evaluating assays for postcolposcopy and posttreatment management. Importantly, we identified a high risk of bias in at least one domain for most

TABLE 3. The Influence of Study Quality on CIN 2+ Risk Estimates and Between Study Variation, Posttreatment Studies Evaluating HPV Tests (n = 21)

	n	Baseline risk	τ^2	Risk in negatives	τ^2	Risk in positives	τ^2
Risk of bias							
<2 domains	12	0.049 (0.034–0.069)	0.2788	0.006 (0.002–0.018)	1.5406	0.161 (0.100–0.251)	0.7259
≥2 domains	9	0.048 (0.025–0.090)	0.9374	0.008 (0.003–0.023)	1.9964	0.211 (0.106–0.378)	1.3994
Patient selection							
Low	3	0.066 (0.033–0.126)	0.3055	0.009 (0.002–0.050)	0.9790	0.217 (0.124–0.353)	0.2276
High	13	0.048 (0.030–0.075)	0.6684	0.008 (0.003–0.018)	1.7193	0.163 (0.057–0.390)	1.0943
Unclear	5	0.049 (0.026–0.092)	0.4772	0.008 (0.001–0.051)	2.9129	0.204 (0.122–0.321)	1.6534
Index test							
Low	11	0.040 (0.024–0.066)	0.6689	0.007 (0.003–0.016)	1.3692	0.181 (0.105–0.295)	0.9741
High	6	0.066 (0.039–0.108)	0.3959	0.005 (0.001–0.031)	3.7472	0.217 (0.112–0.377)	0.8518
Unclear	4	0.062 (0.033–0.114)	0.3383	0.029 (0.016–0.053)	0.000	0.195 (0.055–0.503)	1.9348
Reference standard							
Low	12	0.055 (0.037–0.081)	0.3835	0.005 (0.001–0.022)	3.0756	0.170 (0.098–0.279)	1.0322
High	9	0.042 (0.023–0.074)	0.7624	0.008 (0.004–0.018)	1.0164	0.200 (0.106–0.345)	1.0872
Unclear	0						
Flow and timing							
Low	8	0.039 (0.021–0.072)	0.7424	0.005 (0.001–0.014)	1.4547	0.193 (0.096–0.350)	1.2500
High	8	0.063 (0.039–0.101)	0.4361	0.010 (0.003–0.031)	2.0888	0.233 (0.141–0.359)	0.6170
Unclear	5	0.054 (0.031–0.094)	0.3160	0.016 (0.005–0.049)	0.4756	0.143 (0.051–0.343)	1.5296
Applicability							
Patient selection							
Low concern	14	0.054 (0.040–0.073)	0.2481	0.009 (0.004–0.021)	1.6770	0.189 (0.129–0.268)	0.5613
High concern	7	0.040 (0.018–0.088)	1.1370	0.004 (0.001–0.016)	1.5184	0.177 (0.067–0.392)	2.0179
Index test							
Low concern	15	0.055 (0.041–0.074)	0.2664	0.010 (0.005–0.022)	1.3716	0.175 (0.112–0.263)	0.8379
High concern	6	0.038 (0.016–0.088)	1.1711	0.003 (0.001–0.014)	1.9617	0.206 (0.082–0.429)	1.5688
Reference standard							
Low concern	19	0.047 (0.032–0.070)	0.6221	0.007 (0.003–0.016)	2.2601	0.177 (0.112–0.268)	1.1648
High concern	2	0.057 (0.041–0.079)	0.0000	0.009 (0.003–0.023)	0.000	0.208 (0.150–0.281)	0.1161

studies included in our systematic review. The patient selection and flow and timing domains were most likely to have a high risk of bias for at least one of the signaling questions, reflecting the wide variation in clinical practice and the lack of standards for conducting postcolposcopy and posttreatment studies. Using these adapted criteria, we demonstrated that meta-analyses of studies with lower risk of bias and concern about applicability generally have less between-study variation, producing more robust and precise risk estimates. Conversely, we did not observe statistically significant differences in the absolute risk measures in studies with high versus low risk of bias, suggesting that the potential biases do not affect risk estimates in a specific direction, and that the point estimates reflect the actual risks from the source population. It is important to point out that bias in certain domains and for certain signaling questions could have differential impact on the resulting risk estimates, and it is challenging to predict in which direction the bias will act. Even where we observed nonstatistically significant differences in risk estimates, these differences generally did not cross a risk threshold, with few marginal exceptions.

Our study was limited by the relatively few primary studies that were eligible for inclusion for the quality assessment. In the future, we will validate this tool using additional studies for other indications (e.g., triage of HPV-positive individuals), which will allow for greater power to detect associations between study quality and risk estimates and may enable a more detailed assessment of specific signaling questions and further adaptation of this tool.

Although use of the adapted QUADAS-2 was tightly linked to the systematic review for the ASCCP guidelines effort, this tool will have broad application for researchers, evidence evaluators, and journals who are interested in designing, conducting, evaluating, and publishing studies for cervical cancer screening and management. Going forward, these criteria will serve a dual purpose for evaluating the quality of evidence, as well as supporting and improving future planning and reporting of studies.

REFERENCES

1. Higgins JPT, Green S (editors). Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated March 2011]. The Cochrane Collaboration, 2011. Available at: www.handbook.cochrane.org. Accessed March 3, 2020.
2. Arbyn M, Redman CWE, Verdoodt F, et al. Incomplete excision of cervical precancer as a predictor of treatment failure: a systematic review and meta-analysis. *Lancet Oncol* 2017;18:1665–79.
3. Arbyn M, Smith SB, Temin S, et al. Detecting cervical precancer and reaching underscreened women by using HPV testing on self samples: updated meta-analyses. *BMJ* 2018;363:k4823.
4. Wentzensen N, Arbyn M, Berkhof J, et al. Eurogin 2016 Roadmap: how HPV knowledge is changing screening practice. *Int J Cancer* 2017;140:2192–200.
5. Bossuyt PM, Reitsma JB, Bruns DE, et al. STARD 2015: an updated list of essential items for reporting diagnostic accuracy studies. *BMJ* 2015;351:h5527.
6. Whiting PF, Rutjes AW, Westwood ME, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med* 2011;155:529–36.
7. Cuschieri K, Ronco G, Lorincz A, et al. Eurogin roadmap 2017: triage strategies for the management of HPV-positive women in cervical screening programs. *Int J Cancer* 2018;143:735–45.
8. Cortecchia S, Galanti G, Sgadari C, et al. Follow-up study of patients with cervical intraepithelial neoplasia grade 1 overexpressing p16Ink4a. *Int J Gynecol Cancer* 2013;23:1663–9.
9. Gurumurthy M, Cotton SC, Sharp L, et al. Postcolposcopy management of women with histologically proven CIN 1: results from TOMBOLA. *J Low Genit Tract Dis* 2014;18:203–9.
10. Kang WD, Ju UC, Kim SM. Is human papillomavirus genotype important in predicting disease progression in women with biopsy-proven negative or CIN1 of atypical squamous cell of undetermined significance (ASC-US) cytology? *Gynecol Oncol* 2018;148:305–10.
11. Tverelv LR, Sorbye SW, Skjeldestad FE. Risk for cervical intraepithelial neoplasia grade 3 or higher in follow-up of women with a negative cervical biopsy. *J Low Genit Tract Dis* 2018;22:201–6.
12. Ye J, Cheng B, Cheng YF, et al. Prognostic value of human papillomavirus 16/18 genotyping in low-grade cervical lesions preceded by mildly abnormal cytology. *J Zhejiang Univ Sci B* 2017;18:249–55.
13. Bruhn LV, Andersen SJ, Hariri J. HPV-testing versus HPV-cytology co-testing to predict the outcome after conization. *Acta Obstet Gynecol Scand* 2018;97:758–65.
14. Byun JM, Jeong DH, Kim YN, et al. Persistent HPV-16 infection leads to recurrence of high-grade cervical intraepithelial neoplasia. *Medicine* 2018; 97:e13606.
15. Ceballos KM, Lee M, Cook DA, et al. Post-loop electrosurgical excision procedure high-risk human papillomavirus testing as a test of cure: the British Columbia Experience. *J Low Genit Tract Dis* 2017;21:284–8.
16. Cubie HA, Canham M, Moore C, et al. Evaluation of commercial HPV assays in the context of post-treatment follow-up: Scottish Test of Cure Study (STOCS-H). *J Clin Pathol* 2014;67:458–63.
17. Du R, Meng W, Chen ZF, et al. Post-treatment human papillomavirus status and recurrence rates in patients treated with loop electrosurgical excision procedure conization for cervical intraepithelial neoplasia. *Eur J Gynaecol Oncol* 2013;34:548–51.
18. Fan A, Wang C, Han C, et al. Factors affecting residual/recurrent cervical intraepithelial neoplasia after cervical conization with negative margins. *J Med Virol* 2018;90:1541–8.
19. Friebe K, Klapdor R, Hillemanns P, et al. The value of partial HPV genotyping after conization of cervical dysplasias. *Geburtshilfe Frauenheilkd* 2017;77:887–93.
20. Gosvig CF, Huusom LD, Andersen KK, et al. Long-term follow-up of the risk for cervical intraepithelial neoplasia grade 2 or worse in HPV-negative women after conization. *Int J Cancer* 2015;137:2927–33.
21. Hansen J, Waibel J, Timme S, et al. Validity parameters of the human papillomavirus detection test Hybrid Capture 2 with and without cytology after laser destruction and large loop excision of the transformation zone treatment of high-grade cervical intraepithelial neoplasia lesions. *J Low Genit Tract Dis* 2017;21:289–93.
22. Herfs M, Somja J, Howitt BE, et al. Unique recurrence patterns of cervical intraepithelial neoplasia after excision of the squamocolumnar junction. *Int J Cancer* 2015;136:1043–52.
23. Innamaa A, Dudding N, Ellis K, et al. High-risk HPV platforms and test of cure: should specific HPV platforms more suited to screening in a 'test of cure' scenario be recommended? *Cytopathology* 2015;26:381–7.
24. Kalampokas E, Wilson J, Gurumurthy M, et al. Effect of high-risk human papillomavirus but normal cytology at test of cure on achieving colposcopy standards. *J Low Genit Tract Dis* 2018;22:110–4.
25. Kang WD, Kim SM. Human papillomavirus genotyping as a reliable prognostic marker of recurrence after loop electrosurgical excision procedure for high-grade cervical intraepithelial neoplasia (CIN2-3) especially in postmenopausal women. *Menopause* 2016;23:81–6.
26. Khunamornpong S, Settakorn J, Sukpan K, et al. Application of HPV DNA testing in follow-up after loop electrosurgical excision procedures in Northern Thailand. *Asian Pac J Cancer Prev* 2015;16:6093–7.
27. Kong TW, Son JH, Chang SJ, et al. Value of endocervical margin and high-risk human papillomavirus status after conization for high-grade cervical intraepithelial neoplasia, adenocarcinoma in situ, and microinvasive carcinoma of the uterine cervix. *Gynecol Oncol* 2014;135:468–73.
28. Lubrano A, Medina N, Benito V, et al. Follow-up after LLETZ: a study of 682 cases of CIN 2-CIN 3 in a single institution. *Eur J Obstet Gynecol Reprod Biol* 2012;161:71–4.

© 2020 The Author(s). Published by Wolters Kluwer Health, Inc. on behalf of the ASCCP.

29. Molloy M, Comer R, Rogers P, et al. High risk HPV testing following treatment for cervical intraepithelial neoplasia. *Ir J Med Sci* 2016;185:895–900.

30. Persson M, Brismar Wendel S, Ljungblad L, et al. High-risk human papillomavirus E6/E7 mRNA and L1 DNA as markers of residual/recurrent cervical intraepithelial neoplasia. *Oncol Rep* 2012;28:346–52.

31. Polman NJ, Uijterwaal MH, Witte BI, et al. Good performance of p16/ki-67 dual-stained cytology for surveillance of women treated for high-grade CIN. *Int J Cancer* 2017;140:423–30.

32. Ryu A, Nam K, Kwak J, et al. Early human papillomavirus testing predicts residual/recurrent disease after LEEP. *J Gynecol Oncol* 2012;23:217–25.

33. Torne A, Fuste P, Rodriguez-Carunchio L, et al. Intraoperative post-conisation human papillomavirus testing for early detection of treatment failure in patients with cervical intraepithelial neoplasia: a pilot study. *BJOG* 2013;120:392–9.

34. Wu J, Jia Y, Luo M, et al. Analysis of residual/recurrent disease and its risk factors after loop electrosurgical excision procedure for high-grade cervical intraepithelial neoplasia. *Gynecol Obstet Invest* 2016;81:296–301.

35. Zhao L, Wentzensen N, Zhang RR, et al. Factors associated with reduced accuracy in Papanicolaou tests for patients with invasive cervical cancer. *Cancer Cytopathol* 2014;122:694–701.

36. Higgins JP, Thompson SG, Spiegelhalter DJ. A re-evaluation of random-effects meta-analysis. *J R Stat Soc Ser A Stat Soc* 2009;172:137–59.

37. Nyaga VN, Arbyn M, Aerts M. Metaprop: a Stata command to perform meta-analysis of binomial data. *Arch Public Health* 2014;72:39.

38. Schiffman M, Hyun N, Raine-Bennett TR, et al. A cohort study of cervical screening using partial HPV typing and cytology triage. *Int J Cancer* 2016; 139:2606–15.

39. Clarke MA, Cheung LC, Castle PE, et al. Five-year risk of cervical precancer following p16/Ki-67 dual-stain triage of HPV-positive women. *JAMA Oncol* 2018.

40. Wentzensen N, Clarke MA, Bremer R, et al. Clinical evaluation of human papillomavirus screening with p16/Ki-67 dual stain triage in a large organized cervical cancer screening program. *JAMA Intern Med* 2019;179:881–8.

41. Wentzensen N, Fetterman B, Castle PE, et al. p16/Ki-67 dual stain cytology for detection of cervical precancer in HPV-positive women. *J Natl Cancer Inst* 2015;107:djv257.

42. Wentzensen N, Schwartz L, Zuna RE, et al. Performance of p16/Ki-67 immunostaining to detect cervical cancer precursors in a colposcopy referral population. *Clin Cancer Res* 2012;18:4154–62.

43. Clarke MA, Gradissimo A, Schiffman M, et al. Human papillomavirus DNA methylation as a biomarker for cervical precancer: consistency across 12 genotypes and potential impact on management of HPV-positive women. *Clin Cancer Res* 2018;24:2194–202.

44. Clarke MA, Luhn P, Gage JC, et al. Discovery and validation of candidate host DNA methylation markers for detection of cervical precancer and cancer. *Int J Cancer* 2017;141:701–10.

45. Clarke MA, Wentzensen N, Mirabello L, et al. Human papillomavirus DNA methylation as a potential biomarker for cervical cancer. *Cancer Epidemiol Biomarkers Prev* 2012;21:2125–37.

46. Mirabello L, Sun C, Ghosh A, et al. Methylation of human papillomavirus type 16 genome and risk of cervical precancer in a Costa Rican population. *J Natl Cancer Inst* 2012;104:556–65.

47. Wentzensen N, Sun C, Ghosh A, et al. Methylation of HPV18, HPV31, and HPV45 genomes and cervical intraepithelial neoplasia grade 3. *J Natl Cancer Inst* 2012;104:1738–49.

48. Hu L, Bell D, Antani S, et al. An observational study of deep learning and automated evaluation of cervical images for cancer screening. *J Natl Cancer Inst* 2019;111:923–32.

49. Wentzensen N, Massad LS, Mayeaux EJJr., et al. Evidence-based consensus recommendations for colposcopy practice for cervical cancer prevention in the United States. *J Low Genit Tract Dis* 2017;21:216–22.

50. Darragh TM, Colgan TJ, Thomas Cox J, et al, Members of the LAST Project Work Groups. 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. *Int J Gynecol Pathol* 2013;32:76–115.

© 2020 The Author(s). Published by Wolters Kluwer Health, Inc. on behalf of the ASCCP.