2. Methods

This updated guideline was developed in accordance with the methods described in the *WHO handbook for guideline development, second edition (18)*. A summary of the process is provided here.

2.1 Groups contributing to the guideline development process

Lists of all members of the Guideline Development Group (GDG), External Review Group (ERG), systematic review teams, modelling teams and other contributors are provided in *Annex 1*, with details of their expertise and affiliations. The WHO Secretariat consisted of staff from various relevant WHO departments, and staff from the International Agency for Research on Cancer (IARC). The Steering Group of the WHO Secretariat led the coordination of the development of this guideline. Members of the Secretariat who were not part of the Steering Group were kept informed of the guideline development process and participated in the discussions, in particular during meetings of the various teams.

The GDG was established during the first half of 2019 to appraise the recommendations in the previous 2013 edition of the guideline *(12)*, prioritize the key PICO questions for which systematic reviews needed to be updated or developed, provide feedback on the evidence reviews, and make recommendations to be presented in the final guideline. There were 52 GDG members (34 women, 18 men), representing all six WHO regions as well as civil society organizations and women's groups, and women living with HIV. The members brought varied expertise on cervical screening and treatment. Two members acted as co-chairs and moderated the GDG meetings. The WHO Steering Group met regularly with the GDG chairs, the guideline methodologists, and the systematic review and modelling teams to review progress and to ensure evidence presentations and discussions were standardized.

An External Review Group (ERG) was also established. Its 18 members, none of whom was also a member of the GDG, had expertise in research, policy development, programme implementation and clinical care. Once the GDG had agreed on the recommendations, the ERG reviewed the full draft of the guideline and provided feedback.

There were multiple teams preparing evidence:

- five teams did the evidence reviews
- one team developed two mathematical models
- one team ran a survey about the feasibility of screening approaches
- one team surveyed women about their values and preferences (*Annex 2*).

The teams were based at different institutions and worked independently to prepare and present evidence during the GDG meetings. A guideline methodologist with experience of using the Grading of Recommendations Assessment, Development and Evaluation (GRADE) approach (19) coordinated the presentation of evidence and decision-making processes that facilitated the development of the recommendations, as stipulated in the *WHO handbook for guideline development* (18).

The WHO Steering Group maintained close communication with the GDG and systematic review teams using multiple platforms:

- Zoom Meetings;
- email;
- surveys and voting on the summaries of the evidence and recommendations using GRADEpro software;
- a SharePoint site for access to meeting materials, including slides and evidence summaries, and live documents for comment;
- a chat feature in SharePoint to encourage discussion among the GDG members.

2.1.1 Declarations and management of conflicts of interest

After being invited to join the GDG by the WHO Secretariat at the beginning of the guideline development process, and in accordance with the *WHO handbook for guideline development* (18), each prospective GDG member completed a written declaration of interest (DOI) form. The DOIs were reviewed by two members of the WHO Secretariat and no conflicts of interest were identified (*Annex 3*). The GDG members' names and curriculum vitae were subsequently published on the WHO website for the Department of Sexual and Reproductive Health and Research and approved by the WHO Guidelines Review Committee (GRC) in advance of participation in the process. At the beginning of every GDG meeting, members were asked to update the WHO Steering Group and other GDG members about any potential new conflicts of interest.

2.1.2 Confidentiality

Each GDG member also signed a confidentiality agreement at the beginning of the GDG process, and the WHO Secretariat restated at the start of each GDG meeting that all discussions and draft recommendations were to remain confidential until publication.

2.2 Scoping review and appraisal of the existing recommendations

In October 2019, a subgroup of the GDG met in Geneva to review the previously published recommendations and decide which should be removed, validated, edited or updated based on new evidence, and whether any new recommendations should be made for new interventions. This process was informed by a scoping review of the literature and an assessment of changes in disease burden, practice and policy. These decisions were circulated to all members of the GDG for feedback, and agreement on which recommendations to keep, update and add was reached after additional virtual meetings and electronic correspondence.

The scoping document was initially split between screening and treatment recommendations for women living with HIV (approved by the GRC in September 2019) and the general population of women (approved by the GRC in January 2020). The two scoping documents were then merged and subsequently approved by the GRC in August 2020.

2.3 Priority questions for review of evidence

The GDG identified 14 overarching questions, framed using the population (P), intervention (I), comparator (C), outcomes (O) (PICO) format, as a starting point for formulating recommendations applying to the general population of women and women living with HIV (*Table 2.1*).

Table 2.1 PICO questions for the recommendations in women (the general population of women and women living with HIV)

PICO 1	Should one screen-and-treat strategy versus another screen-and-treat strategy be used in women?
PICO 2	Should one screen, triage and treat strategy versus another screen, triage and treat strategy be used in women?
PICO 3, 4, 5, 6, 7	Should women be followed up 12 and/or 24 months, or 3, 5 or 10 years after a negative or positive test result (and treatment) with the same or a different screening test?
PICO 8i	Should women first be screened for cervical pre-cancer lesions at a specific age (or following an HIV-positive test in women living with HIV)?
PICO 8ii	Should women be screened at least twice in a lifetime or once?
PICO 9	Should treatment be within 6 or 12 months after a positive screening test, or after positive screen and triage tests (both positive), or after histologically confirmed cervical intraepithelial neoplasia (CIN)?
PICO 10	Should there be different treatments for women with histologically confirmed CIN 2/3 (including carcinoma in situ) or adenocarcinoma in situ (AIS)?
PICO 11[a]	What are the effects of health-system interventions to enable the adoption, implementation and scale-up of effective screening approaches?
PICO 12[a]	What are the effects of patient-targeted strategies to support uptake of screening approaches and follow-up care?
PICO 13[a]	What are the effects of provider-targeted strategies to support the adoption of screening approaches and follow-up care?

[a] Next phase of guideline development (Phase 3).

2.4 Priority algorithms

Since screening and treatment can be done using different primary screening and triage tests, there are numerous possible combinations or algorithms. In December 2019, GDG members were surveyed to prioritize the screening and/or triage tests and the treatments that should be evaluated. Following this prioritization exercise, a subgroup of GDG members met to review the results from the survey and to agree on the algorithms to be prioritized. They reached a consensus to address seven priority algorithms in this first phase of the guideline update (*Table 2.2*; for detailed algorithms please refer to *Annex 4*).

Table 2.2. The seven algorithms considered

Screen-and-treat approaches:	
1	VIA as the primary screening test, followed by treatment
2	HPV DNA detection (self- or clinician-collected) as the primary screening test, followed by treatment
Screen, triage and treat approaches:	
3	Cytology as the primary screening test, followed by colposcopy triage, followed by treatment
4	HPV DNA detection as the primary screening test, followed by HPV16/18 triage (when already part of the HPV test), followed by treatment, and using VIA triage for those who screen negative for HPV16/18
5	HPV DNA detection as the primary screening test, followed by VIA triage, followed by treatment
6	HPV DNA detection as the primary screening test, followed by colposcopy triage, followed by treatment
7	HPV DNA detection as the primary screening test, followed by cytology triage, followed by colposcopy and treatment

HPV: human papillomavirus; VIA: visual inspection with acetic acid.

2.5 Outcomes

The GDG agreed that the outcomes previously identified in the 2013 screening and treatment guideline *(12)* would also be the critical outcomes for the new PICO questions; the critical outcomes are listed in *Table 2.3*. To ensure coherence in the systematic reviews and modelling, a working group (subgroup of the GDG) developed standardized definitions for these outcomes (see *Annex 5*). After reviewing the evidence and modelling a limited number of outcomes, the GDG agreed to consider all outcomes together. Adverse events were defined as outcomes that were a direct consequence of pre-cancer treatment and were grouped as one category, with the exception of preterm birth, which was considered a critical outcome (see *Table 2.3*).

Table 2.3 Critical outcomes for the screening and treatment recommendations

Critical outcomes
Cervical cancer
Mortality
High-grade cervical intraepithelial neoplasia or worse (CIN2+)
HPV infection
Preterm birth
Acceptability
Pre-cancer treatments
Adverse events (direct consequence of pre-cancer treatment): – major infections or bleeding – procedure-associated pain – cervical stenosis – infertility – spontaneous abortion – perinatal deaths – premature rupture of membrane – unnecessary interventions – increased viral shedding in women living with HIV
Costs
Feasibility
Equity

See *Annex 5* for additional details.

2.6 Syntheses of evidence

Evidence was synthesized for each PICO question according to the methods in the *WHO handbook for guideline development* and the *Cochrane handbook for systematic reviews of interventions* (18,20). The literature review performed for the development of the *IARC handbooks of cancer prevention: cervical cancer screening, Vol. 18* (to be published in 2021; referred to in brief throughout this guideline as *"IARC handbook"*) (21) was also part of the evidence synthesized for the development of this guideline. We used a hierarchical approach to avoid the duplication of reviews that had been previously published. First, we searched for pre-existing systematic reviews to update (including the

systematic reviews published at the time of the previous guideline), and then searched for primary studies (including randomized and non-randomized studies) when no systematic reviews were available.

New systematic reviews or updates of systematic reviews were conducted to determine the effects of interventions (including screening tests) on outcomes, and the accuracy of screening tests in the general population of women and in women living with HIV. These systematic reviews and associated details are listed in *Table 2.4*.

Table 2.4 PICO questions with corresponding systematic reviews and reports, evidence-to-decision (EtD) tables and recommendations or good-practice statements

PICO questions	List of systematic reviews	Syntheses of the evidence (Web annex A)[a]	Evidence-to-decision tables (Web annex B)[b]	Recommendation or good practice statement (as numbered in this guideline)[c]
1, 2, 8ii	Evidence reviews for interventions for the general population of women Evidence reviews of accuracy of triage tests for the general population of women Evidence reviews for interventions and accuracy of tests for women living with HIV	IARC handbook materials Supplementary material (SM) 1 SM 2	EtD PICO 1 and 2 for the general population of women EtD PICO 1 and 2 for women living with HIV	1, 2, 3, 4, 8, 9, 10 21, 22, 23, 24, 28, 29, 30
3, 4, 5, 6, 7	Review of reviews of follow-up strategies after screening or treatment Evidence review of follow-up strategies after screening or treatment	SM 3 IARC handbook materials	EtD PICO 3, 4, 5, 6, 7	11, 12, 13, 14, 31, 32, 33, 34

AIS: adenocarcinoma in situ; CIN: cervical intraepithelial neoplasia; EtD: evidence-to-decision; IARC: International Agency for Research on Cancer; PICO: population (P), intervention (I), comparator (C), outcome (O).
[a] Web annex A is available at: https://apps.who.int/iris/bitstream/handle/10665/342366/9789240030886-eng.pdf
[b] Web annex B is available at: https://apps.who.int/iris/bitstream/handle/10665/342367/9789240030893-eng.pdf
[c] Number as listed in Tables 1 and 2 in the Executive Summary and as used throughout this guideline.

Table 2.4 (continued)

PICO questions	List of systematic reviews	Syntheses of the evidence (Web annex A)[a]	Evidence-to-decision tables (Web annex B)[b]	Recommendation or good practice statement (as numbered in this guideline)[c]
8i	Review of reviews of age to start and end screening in the general population of women A systematic literature review and narrative synthesis: age at initiation and frequency of cervical cancer screening in women living with HIV Individual patient data meta-analysis (IPD-MA): age at initiation and end of screening in women living with HIV	SM 4 SM 5 SM 6	EtD PICO 8 age at initiation for the general population of women EtD PICO 8 age at initiation for women living with HIV EtD PICO 8 age to end screening	5 25 6, 7, 26, 27
9	Systematic review for treatment within 6 to 12 months	SM 7	-	41
10	Reviews for treatment of histologically confirmed CIN2/3 and AIS	SM 8	EtD PICO 10	42
11, 12, 13	Next phase of guideline development	-	-	-
All	Understanding acceptability and client preferences for screening and treating cervical pre-cancer lesions: preliminary results of a WHO online survey Report on values, preferences, acceptability and feasibility: results of a systematic review of qualitative literature	SM 9 SM 10	-	-

AIS: adenocarcinoma in situ; CIN: cervical intraepithelial neoplasia; EtD: evidence-to-decision; IARC: International Agency for Research on Cancer; PICO: population (P), intervention (I), comparator (C), outcome (O).
[a] Web annex A is available at: https://apps.who.int/iris/bitstream/handle/10665/342366/9789240030886-eng.pdf
[b] Web annex B is available at: https://apps.who.int/iris/bitstream/handle/10665/342367/9789240030893-eng.pdf
[c] Number as listed in Tables 1 and 2 in the Executive Summary and as used throughout this guideline.

Table 2.4 (continued)

PICO questions	List of systematic reviews	Syntheses of the evidence (Web annex A)[a]	Evidence-to-decision tables (Web annex B)[b]	Recommendation or good practice statement (as numbered in this guideline)[c]
All	Survey report: feasibility concerns with priority algorithms Review of reviews of acceptability, feasibility, resources and equity	SM 11 SM 12	-	-
1, 2, 3, 4, 5, 6, 7, 8	Report of modelling	SM 13	-	-

AIS: adenocarcinoma in situ; CIN: cervical intraepithelial neoplasia; EtD: evidence-to-decision; IARC: International Agency for Research on Cancer; PICO: population (P), intervention (I), comparator (C), outcome (O).
[a] Web annex A is available at: https://apps.who.int/iris/bitstream/handle/10665/342366/9789240030886-eng.pdf
[b] Web annex B is available at: https://apps.who.int/iris/bitstream/handle/10665/342367/9789240030893-eng.pdf
[c] Number as listed in Tables 1 and 2 in the Executive Summary and as used throughout this guideline.

2.6.1 Methods used for systematic literature reviews

The detailed methods for each review are reported in the Annex A, Supplementary Materials (*Table 2.4*). In brief, the systematic review teams applied the following key methods across all systematic reviews for this guideline:

- develop a systematic review protocol with inclusion and exclusion criteria for studies based on the finalized PICO questions;
- search multiple databases (including MEDLINE, Embase and the Cochrane Library Epistemonikos) and clinical trial registries, contact investigators in the field for potentially relevant systematic reviews, and look at randomized and non-randomized trials to identify studies for new reviews or to update existing reviews;
- select literature reviews or studies based on inclusion and exclusion criteria (in duplicate or by one reviewer and verified by a second);
- extract data (in duplicate or by one reviewer and verified by another) on the benefits and harms (effects) of screening and treatment, the accuracy of screening tests used, end-user values and preferences, equity, feasibility, resources and acceptability;

- contact study authors for missing data or individual patient data when appropriate;
- assess the risk of bias in individual studies when available (in duplicate or by one reviewer and verified by another) using an appropriate risk-of-bias tool (e.g. Cochrane Risk of Bias for randomized controlled trials, ROBINS-I tools for non-randomized studies and QUADAS for diagnostic studies);
- synthesize the results (narratively or quantitatively) or analyse individual patient data when available;
- assess the certainty of the evidence using GRADE methodology *(19)*; the levels of certainty used are summarized in *Table 2.5*.

Table 2.5 Interpretation of the GRADE levels of certainty of evidence

High	We are very confident that the true effect lies close to that of the estimate of the effect
Moderate	We are moderately confident in the effect estimate: The true effect is likely to be close to the estimate of the effect, but there is a possibility that it is substantially different
Low	We have limited confidence in the effect estimate: The true effect may be substantially different from the estimate of the effect
Very low	We have very little confidence in the effect estimate: The true effect is likely to be substantially different from the estimate of the effect

Source: GRADE handbook, GRADEpro, 2021 *(19)*.

2.6.2 Individual patient data meta-analysis

We conducted an individual patient data meta-analysis (IPD-MA) to analyse age-specific data for cervical cancer and CIN in women living with HIV. We contacted the authors of the studies identified in the systematic review of screening-initiation age (*Web annex A, Supplementary material 5*) that included at least 40 women living with HIV with CIN2+. All the data sets they provided were first reviewed individually, then discrepancies were resolved with investigators, and the aligned data sets were then combined. The IPD-MA used one-stage *(22)* random study intercept models to take into account heterogeneity among studies. Generalized linear mixed models were fitted for binomial or multinomial cervical screening test responses using SAS version 9.4 (SAS Institute, Inc., United States). Random effects models were used to calculate predicted probabilities for cervical screening results by age categories, HIV status and other factors of interest (see *Web annex A, Supplementary material 6*).

2.6.3 Mathematical modelling

We used the Policy1-Cervix platform, an extensively validated dynamic model of HPV transmission, vaccination, type-specific natural history, cancer survival, screening, diagnosis and treatment *(23–31)*, to predict outcomes in women across all 78 low- and middle-income countries. The Policy-Cervix HIV-HPV model for cervical cancer among women living with HIV was used to evaluate outcomes for women living with HIV in the United Republic of Tanzania *(32)*, as there was sufficient local data available on cervical cancer control activities and HIV disease burden and control activities (including historical data). The United Republic of Tanzania has endemic HIV and is a suitable example country for evaluating optimal screening strategies for women living with HIV. The Policy1-Cervix model was one of three models used by the Cervical Cancer Elimination Modelling Consortium (CCEMC) to evaluate the impact of cervical cancer prevention interventions in 78 low- and middle-income countries *(23,24)*. We evaluated the impact of the seven algorithms considering different ages and screening intervals, as informed by the GDG (see *Table 2.2*). For the baseline analysis, we assumed that 70% of women attended screening at each routine screening event and 90% of women complied with follow-up. Outcomes were assessed over the lifetime of birth cohorts eligible for screening in 2030 onwards and included cervical cancer incidence and mortality, pre-cancer treatments, additional preterm deliveries as a result of pre-cancer treatment and cost-effectiveness. A range of sensitivity analyses were considered, including probabilistic sensitivity analysis for cost-effectiveness. The detailed methods and results of the modelling work are available in *Web annex A, Supplementary material 13*.

2.6.4 Values and preferences

A search for studies and systematic reviews was conducted that addressed, among other considerations, the values and preferences of end-users, health-care providers and other stakeholders. The literature was organized by study design and methodology, location and population, and presented to the GDG.

For primary data, all women and girls aged 15 years and older, regardless of their prior cervical cancer screening or treatment status, were eligible to participate in an anonymous, voluntary survey distributed via SurveyMonkey. The survey received approval from the WHO Ethics Review Committee and was run in English and French from 22 June to 18 September 2020. Awareness of the survey had been raised among a wide range of civil society groups through a webinar. The survey was also promoted through the Union for International Cancer Control and the WHO advisory group of women living with HIV, and shared through WHO regional focal points for the Cervical Cancer Elimination Initiative. The survey responses from the 561 respondents, including their qualitative responses to open-ended questions, were analysed. The detailed methods and results are available in *Web annex A, Supplementary materials 9 and 10*.

2.6.5 Feasibility, acceptability, resources and equity considerations

A survey of the GDG members was administered via SurveyMonkey to assess the implementation considerations for each priority algorithm. The survey was developed using the context and implementation of complex interventions (CICI) framework *(33)*. Each GDG member was asked about their level of concern about each algorithm being able to sustainably meet the large-scale goal of cervical cancer elimination. The following components of cervical cancer screening and management service delivery were queried separately according to the priority algorithm: demand generation, access to screening and the follow-up management of positives, workforce training, infrastructure development and maintenance, development and maintenance of the screening registry, and cost and integration with other priority health services. The considerations of the GDG members were assessed for the following eight stakeholder groups: health authorities at the national level, health authorities at the regional level, professional societies, providers at both the hospital and primary care levels, community health workers, clients (screened women) and the community. The detailed methods and results of the survey from the 29 respondents are provided in *Web annex A, Supplementary materials 11 and 12*.

2.7 Development of the recommendations

All the GDG meetings that focused on formulation of recommendations were held virtually. Tables to facilitate decision-making for recommendations – evidence-to-decision (EtD) tables – were produced by the guideline methodologist for each recommendation and circulated to the GDG members before each meeting. These tables included a summary of the evidence (benefits and harms), relevant values and preferences information, and other issues, including use of resources and cost, feasibility, equity and acceptability.

During the meeting, the EtD tables and evidence were discussed with the GDG. Following the meeting, all GDG members received an email through GRADEpro that solicited direct individual input. Each GDG member saw the EtD tables several times and had opportunities to ask questions and to comment both during and after the meeting. The methodologist, systematic reviewers, modellers and the WHO Steering Group assessed the GDG input and used it to write the recommendations.

Agreement on the recommendations was made by consensus during the GDG meetings, and the final written recommendations were then approved electronically. The responses solicited via email were either to approve, approve "with the following remarks" or not approve. The GDG had agreed that, if consensus could not be reached, a majority vote of 51% would have been accepted to make recommendations – yet the group did reach a consensus on all the recommendations.

Strong recommendations (worded as "WHO recommends") were made when all the desirable consequences of the intervention **clearly** outweighed the undesirable consequences in most settings.

Conditional recommendations (worded as "WHO suggests") were made when the desirable consequences of the intervention **probably** outweighed the undesirable consequences in most settings.

Table 2.6 describes how strong and conditional recommendations should be interpreted.

Additionally, the GDG provided good practice statements when it agreed that this guidance was needed, but a review of the literature was not warranted because the balance of desirable and undesirable consequences of an intervention was unequivocal, and no other criteria needed to be considered.

Table 2.6 Interpretation of strong and conditional recommendations

Implications	Strong recommendation (WHO recommends...)	Conditional recommendation (WHO suggests...)
For individuals	Most individuals in this situation would want the recommended course of action, and only a small proportion would not. Formal decision aids are not likely to be needed to help individuals make decisions consistent with their values and preferences.	The majority of individuals in this situation would want the suggested course of action, but some may not.
For health-care providers	Most individuals should receive the recommended course of action. Adherence to this recommendation (when it aligns with national guidelines) could be used as a quality criterion or performance indicator.	Clinicians should recognize that different choices may be appropriate for different individuals and that clinicians must help each individual arrive at a management decision consistent with the individual's values and preferences. Decision aids may be useful to help individuals make decisions consistent with their values and preferences.
For policy-makers	The recommendation can be adopted as policy in most situations.	Policy-making will require discussion and involvement of various stakeholders.

Source: GRADE Handbook, GRADEpro, 2021 *(19)*.

2.8 Management of the external peer review

The draft guideline document was circulated to the External Review Group (ERG) for comment. The WHO Secretariat prepared a summary table with all ERG responses and sorted the comments by topic or section. The WHO Secretariat then identified comments for discussion and presented these to the GDG, and when these issues had been resolved via email correspondence, the guideline document was finalized.