

P16/Ki-67 Dual Stain zur Triage von PAP III/IIID Befunden im Screening auf Gebärmutterhalskrebs

Systematischer Review

Final



Ludwig Boltzmann Institut
Health Technology Assessment

LBI-HTA Projektbericht Nr.: 72

ISSN: 1992-0488

ISSN-online: 1992-0496

P16/Ki-67 Dual Stain zur Triage von PAP III/IIID Befunden im Screening auf Gebärmutterhalskrebs

Systematischer Review

Final



Ludwig Boltzmann Institut
Health Technology Assessment

Wien, August 2013

Projektteam

Projektleitung: Dr Ingrid Zechmeister-Koss (IZ)

Projektbearbeitung: Dr Agnes Kisser (AK)

Projektbeteiligung

Systematische Literatursuche: Tarquin Mittermayr

Externe Begutachtung: Dr. Barbara Zimmer MPH, MA, Kompetenz Centrum Onkologie, MDK Nordrhein
Ass.Prof. Dr Christoph Grimm, Abteilung für Allgemeine Gynäkologie und gynäkologische
Onkologie, Medizinische Universität Wien

Interne Begutachtung: P.D. Claudia Wild
Dr Anna Nachtnebel

Korrespondenz: Dr Agnes Kisser: agnes.kisser@hta.lbg.ac.at

Dieser Bericht soll folgendermaßen zitiert werden/This report should be referenced as follows:

Kisser A und Zechmeister-Koss I.p16/Ki-67 Dual Stain zur Triage von PAP III/IIID Befunden im Screening auf Gebärmutterhalskrebs. LBI-HTA Projektbericht Nr.: 72; 2013. Wien: Ludwig Boltzmann Institut für Health Technology Assessment.

Interessenskonflikt

Alle beteiligten AutorInnen erklären, dass keine Interessenskonflikte im Sinne der Uniform Requirements of Manuscripts Statement of Medical Journal Editors (www.icmje.org) bestehen.

IMPRESSUM

Medieninhaber und Herausgeber:

Ludwig Boltzmann Gesellschaft GmbH
Nußdorferstr. 64, 6 Stock, A-1090 Wien
<http://www.lbg.ac.at/de/lbg/impressum>

Für den Inhalt verantwortlich:

Ludwig Boltzmann Institut für Health Technology Assessment (LBI-HTA)
Garnisonngasse 7/20, A-1090 Wien
<http://hta.lbg.ac.at/>

Die LBI-HTA Projektberichte erscheinen unregelmäßig und dienen der Veröffentlichung der Forschungsergebnisse des Ludwig Boltzmann Instituts für Health Technology Assessment.

Die HTA-Projektberichte erscheinen in geringer Auflage im Druck und werden über den Dokumentenserver „<http://eprints.hta.lbg.ac.at>“, der Öffentlichkeit zur Verfügung gestellt:

LBI-HTA Projektbericht Nr.: 72

ISSN: 1992-0488

ISSN-online: 1992-0496

© 2013 LBI-HTA – Alle Rechte vorbehalten

Inhalt

Zusammenfassung	7
Summary.....	9
1 Hintergrund	11
1.1 Gebärmutterhalskrebs: Epidemiologie und Screening.....	11
1.2 Natürlicher Verlauf der Entstehung von Gebärmutterhalskrebs	12
1.3 Etablierte und neuere Methoden zur Früherkennung von Gebärmutterhalskrebs	14
1.3.1 Papanicolaou Abstrich	14
1.3.2 Dünnschichtzytologie	15
1.3.3 Automatisierte Zytologie	16
1.3.4 Kolposkopie.....	16
1.3.5 Biopsie.....	16
1.3.6 HPV High Risk Diagnostik.....	17
1.4 Screening-Strategie: Management unklarer Befunde in Österreich.....	17
1.5 P16/Ki-67 Dual Stain Immunzytologie (CINtec®Plus).....	19
2 Fragestellung	21
3 Methoden.....	25
3.1 Methodik der Nutzenbewertung	25
3.1.1 Definition der Endpunkte: Diagnostische Genauigkeit.....	25
3.1.2 Definition der Endpunkte: Klinischer Nutzen.....	26
3.1.3 Beurteilung der Qualität der Evidenz.....	27
3.1.4 Stärke der Evidenz nach GRADE	29
3.2 Systematische Literatursuche und -auswahl.....	31
4 Datenextraktion.....	33
4.1 Datenextraktionstabelle.....	34
4.2 Studienmerkmale	36
4.3 Studienergebnisse.....	37
4.4 Qualität der Studien.....	40
4.4.1 Interne Validität	40
4.4.2 Externe Validität.....	41
5 Stärke der Evidenz.....	43
6 Diskussion	45
7 Empfehlung.....	49
8 Literatur.....	51
Anhänge	55
Suchstrategie.....	55
QUADAS Auswertung	60

Abbildungsverzeichnis

Abbildung 1.2-1: Stufen in der Entstehung des Zervixkarzinoms.....	13
Abbildung 1.4-1: Management von Frauen mit PAP III ohne Atrophie oder Entzündung und Frauen mit PAP IIID	18
Abbildung 2-1: p16/Ki-67 Immunzytologie als Triagetest im Management von Frauen mit einem Primärbefund von PAP III – erwartete klinische Auswirkungen in den verschiedenen Diagnosegruppen.....	22
Abbildung 2-2: p16/Ki-67 Immunzytologie als Triagetest im Management von Frauen mit einem Primärbefund von PAP IIID – erwartete klinische Auswirkungen in den verschiedenen Diagnosegruppen.....	23
Abbildung 3.2-1: Darstellung des Auswahlprozesses (PRISMA Flow Diagram).....	32
Abbildung 4.3-1: Forest Plots der Studienergebnisse zur Sensitivität und Spezifität des p16/Ki-67 Tests	39

Tabellenverzeichnis

Tabelle 1.2-1: Klassifizierung der Zervixkarzinome.....	13
Tabelle 1.3-1: Befundwiedergabe der Zervixzytologie – Gegenüberstellung der beiden Klassifizierungssysteme München II und Bethesda.....	15
Tabelle 1.4-1: Kosten der einzelnen Leistungspakete im Zusammenhang mit Zervixkarzinomscreening und -behandlung	18
Tabelle 2-1: PICO Fragestellung.....	21
Tabelle 3.1-1: Vierfeldertafel.....	25
Tabelle 3.1-2: Zur Evaluierung der Studiendesigns verwendete Evidenzhierarchien.....	30
Tabelle 3.2-1: Ein- und Ausschlusskriterien	31
Tabelle 4.1-1: Datenextraktion: Studienmerkmale und -Ergebnisse.....	34
Tabelle 4.3-1: Übersicht der Studienergebnisse	37
Tabelle 4.4-1: Ergebnisse der Qualitätsprüfung der Einzelstudien	42
Tabelle 5-1: Evidenzprofil zur p16/Ki-67 Immunzytochemie in der Triage von Frauen mit einem PAP III oder PAP IIID Primärbefund im Screening auf Gebärmutterhalskrebs	44

Abkürzungsverzeichnis

AIS.....	Adenomakarzinom in situ
AG	Aktiengesellschaft
AGS	Atypical glandular cells/Atypische glanduläre Zellen
ASC-H.....	Atypical Squamous cells, cannot exclude HSIL
ASC-US.....	Atypical squamous cells of unknown origin
bzw.	Beziehungsweise
CI.....	Confidence interval/Konfidenzintervall
CIN.....	Cervical intraepithelial neoplasia/Zervikale intraepitheliale Neoplasien
DAS	Diagnostic Accuracy Studies/Studien zur diagnostischen Genauigkeit
DNA.....	Deoxyribonucleic Acid/Deoxyribonukleinsäure
ECC.....	Endocervical Curettage/Endozervikale Kürettage
EU	Europäische Union
FN	Falsch Negative
FP	Falsch Positive
HPV.....	Humane Papillomaviren
HR-HPV	Hochrisiko HPV
HSIL	High Grade Squamous Intraepithelial Lesions/Hochgradige Plattenepithel Läsionen
HTA	Health Technology Assessment
LSIL.....	Low Grade Squamous Intraepithelial Lesions/Leichtgradige Plattenepithel Läsionen
LEEP.....	Loop Electrosurgical Excision Procedure
LLETZ.....	Large Loop Excision of the Transformation Zone
mRNA.....	messenger RNA
NA	Not Applicable/Nicht anwendbar
NPV.....	Negativer Prädiktiver Wert
NR.....	Not Reported/Nicht Berichtet
PAP	Papanicolaou
PCR.....	Polymerase Chain Reaction/Polymerase Kettenreaktion
PPV	Positiver Prädiktiver Wert
RCT.....	Randomisiert Kontrollierte Studie
RN	Richtig Negative
RNA	Ribonucleic Acid/Ribonukleinsäure
ROC	Receiver Operator Curve/Grenzwertoptimierungskurve
RP.....	Richtig Positive
u. a.....	unter anderem
vgl.....	Vergleiche
z. B.....	zum Beispiel

Zusammenfassung

Hintergrund

Trotz der nachweislichen Erfolge des Screeningprogramms mittels „Papani-colaou-Abstrich“ (PAP-Abstrich) besteht ein wesentlicher Nachteil in der geringen Sensitivität des PAP Abstrichs. Diese führt zu einer hohen Anzahl uneindeutiger Befunde (in Österreich Befunde der Klassen PAP III und IIID), die über verschiedene Verfahren weiter abzuklären sind. Der p16/Ki-67 Dual-Stain Test verspricht, über eine immunocytochemische Doppelfärbung des Tumorsuppressorproteins p16^{INK4a} und des Proliferationsmarkers Ki-67 direkt im Abstrich, den Nachweis viral induzierter onkogener Veränderungen und positioniert sich daher als Triagetest bei unklaren Befunden und leichten und mäßigen Dysplasien im Primärscreening auf Gebärmutterhalskrebs.

unklare Befunde im Zervixkarzinom-screening; p16/Ki-67 Dual Stain Test zur Triage geeignet?

Methode

Zur Bewertung des Nutzens dieses Tests wurde eine systematische Literatursuche nach Primärstudien und relevanten Evidenzsynthesen durchgeführt. Die Qualitätsprüfung erfolgte mit dem QUADAS-2 Tool. Die Studien wurden nach der GRADE-Methode qualitativ zusammengefasst.

systematische Literatursuche und Bewertung nach GRADE

Ergebnisse

Es standen 5 diagnostische Kohortenstudien und 1 diagnostische Fall-Kontroll-Studie zur diagnostischen Genauigkeit zur Verfügung. Die Ergebnisse der Studien zeigen eine Sensitivität von 64-92,2 % in der Befundgruppe Atypical Squamous Cells of Unknown Origin (ASC-US), sowie 81,6-100 % in der Befundgruppe Low Grade Squamous intraepithelial Lesions (LSIL). Die Spezifität liegt bei 53-80,6 % in der Befundgruppe ASC-US und 43-75 % in der Befundgruppe LSIL.

Review von 6 Studien zur diagnostischen Genauigkeit

Diskussion

Fünf der sechs Studien wiesen methodische Mängel und ungenügende Berichterstattung auf. Die für eine hohe Evidenzgrundlage (Level I) notwendigen Daten aus RCTs zur Bewertung des klinischen Nutzen oder ein systematischer Review von Level II Studien zur Bewertung der diagnostischen Genauigkeit standen nicht zur Verfügung. Da das in Österreich verwendete Verfahren zur Zellgewinnung, sowie das Klassifikationssystem von den Studienmethoden abweicht, können die Ergebnisse nur bedingt auf Österreich übertragen werden. Ein optimaler Einsatz des Tests in der österreichischen Screeningroutine erfordert darüber hinaus die gegenüberstellende Evaluierung möglicher Screeningalgorithmen mit alternativen Triagemethoden.

methodische Mängel und keine Evidenz zum klinischen Nutzen

Schlussfolgerung

Auf der Basis der Evidenzgrundlage kann keine Empfehlung zur Aufnahme der p16/Ki-67 Immunzytochemie zur Triage von unklaren Primärbefunden im Zervixkarzinomscreening in den Leistungskatalog abgegeben werden. Weiterführende Studien insbesondere zum klinischen Nutzen des Tests sind angezeigt.

keine Empfehlung aufgrund mangelnder Evidenzgrundlage

Summary

Background

Cervical cancer screening has effectively reduced incidence and mortality from invasive cancer in Austria. Each year however several thousands of healthy women undergo extensive diagnostic procedures following equivocal results in the cytologic evaluation of cervical cells with the PAP test (PAP III or PAP IIID in Austria), as a consequence of the low sensitivity of this screening method. The p16/KI-67 Dual Stain test claims to detect virally induced oncogenic molecular changes in the cell through the immune cytochemical double staining of the tumour suppressor gene p16^{INK4a} and the proliferation marker Ki-67 and thereby to improve the triage of women with equivocal cytological results in cervical cancer screening.

p16/Ki-67 Test useful as triage test for equivocal results in cervical cancer screening?

Methods

A systematic literature search and evaluation of peer-reviewed literature was performed in Ovid Medline, Embase, DRB and The Cochrane Library. 411 references remained after the removal of duplicates. A further 3 articles were identified through handsearching, yielding a total of 414 references. Quality assessment of studies included in the systematic review was performed using the QUADAS-2 tool. The evidence was qualitatively summarised according to GRADE.

systematic literature search and GRADE evaluation

Results

In total 6 diagnostic accuracy studies were included in the systematic review, of which 5 were diagnostic cohort studies and 1 was a diagnostic case-control study. The studies reported sensitivities in the range 64-92.2 % for detection of CIN2+ in women with ASC-US and 81.6-100 % in women with LSIL cytology. Specificities ranged from 53 to 80.6 % for women with ASC-US cytology and from 43 % to 75 % in women with LSIL results.

review of 6 diagnostic accuracy studies

Discussion

In five of six studies we found elevated risk of bias due to flaws in the study design and insufficient reporting. There were no data from highest evidence (Level I) studies to assess clinical utility (RCT) and diagnostic accuracy (a systematic review of Level II studies). The applicability of the results is limited as the methods for cell sampling and the classification system used in the studies deviate from those in routine use in Austria. Optimal embedding of the test in the austrian screening routine would further require the comparison and evaluation of possible screening algorithms and alternative triage methods.

flaws in study design and no evidence for clinical utility

Conclusions

The evidence on diagnostic accuracy and clinical utility of the p16/Ki-67 Dual Stain immune cytochemistry is insufficient at present to support inclusion of the test in the benefits catalogue. Further clinical studies demonstrating the clinical utility of the test are indicated.

no recommendation due to missing evidence

1 Hintergrund

1.1 Gebärmutterhalskrebs: Epidemiologie und Screening

Der Papanicolaou (PAP) Abstrich oder Krebsabstrich zur Früherkennung von Zervixkarzinomen (auch: Gebärmutterhalskrebs) zählt zu den Erfolgsgeschichten der Krebsvorsorge. Diese Screeningmethode wurde in Österreich in den späten 1960er Jahren eingeführt und beruht auf der zytologischen Untersuchung eines Zellabstrichs aus Gebärmutterhals und Muttermund. Bei Verdacht auf Krebsvorstufen im Befund werden die Frauen zur weiteren histologischen Abklärung und etwaiger Behandlung überwiesen. In Österreich wird der Krebsabstrich als opportunistisches Screening bei Frauen zwischen 19 und 69 Jahren angeboten und gemäß „Vorsorge-Neu“ Dokument im 3-Jahres-Abstand empfohlen [3]. Laut „Eurobarometer“ zur Gesundheit in der Europäischen Union (2007) gaben fast zwei Drittel der Frauen in Österreich an, im vergangenen Jahr einen PAP Abstrich gemacht haben zu lassen [4].

In der Folge der Einführung des Screenings mit PAP ist die Zervixkarzinominzidenz und -mortalität in den letzten Jahrzehnten kontinuierlich gesunken. Laut Statistik Austria erkrankten im Jahr 2010 380 Frauen an Gebärmutterhalskrebs und 161 verstarben daran, im Vergleich zu 614 bzw. 211 noch im Jahr 1995. Die Inzidenz- und Mortalitätsrate lag 1995 bei 11,6 bzw. 3,3 und 2010 nur mehr bei 6,4 bzw. 2,1 pro 100.000 Frauen [5]. Damit ist Gebärmutterhalskrebs die 14. häufigste Krebserkrankung bei österreichischen Frauen. In der Altersgruppe der Frauen zwischen 15-44 Jahren ist es jedoch die viert häufigste Krebserkrankung [6].

Trotz des nachweislichen Erfolgs und protektiven Nutzens [7] dieses Früherkennungsprogramms wird die Aktualität der Methode – insbesondere im Licht des zunehmenden Wissenstandes zur Rolle der Humanen Papillomaviren (HPV) in der Krebsentstehung – kritisch diskutiert.

Tatsächlich wird (bei Cutoff LSIL/CIN2+¹) die Sensitivität eines einmaligen PAP Abstrichs mit 75,6 % (95 % CI: 66,5–83,0 %) angesetzt, die Spezifität des Test mit 81,2 % (95 % CI: 71,9–88,0 %) [8]. Das heißt, von 100 erkrankten Frauen werden 76 richtigerweise als krank erkannt und von 100 gesunden Frauen werden 81 als gesund erkannt. Um die Sensitivität des Screenings zu erhöhen, wird der Test in regelmäßigen Screeningintervallen wiederholt und ein niedriger diagnostischer Schwellenwert (PAP III+) angesetzt, ab dem PAP Befunde eine weitere Abklärung erfordern.

Diese Maßnahmen erlauben es, die niedrige Testgüte des PAP Abstrichs weitgehend zu kompensieren und eine zufriedenstellende Früherkennung bei den untersuchten Frauen zu gewährleisten. Tatsächlich ist der mit Abstand häufigste Grund für das Auftreten von Gebärmutterhalskrebs die Nicht-Teilnahme am Screening. Nur etwa 10 % der Fälle sind auf falsch-negative Ergebnisse zurückzuführen [9, 10]. Von diesen entstehen etwa zwei Drittel aufgrund von mangelhafter Probennahme und nur ein Drittel aufgrund von tat-

erfolgreiches Screening auf Gebärmutterhalskrebs mittels PAP Abstrich

Zervixkarzinom weiterhin 4. häufigste Krebserkrankung bei Frauen von 15–44 J

PAP Abstrich: niedrige Sensitivität erfordert niedrige Schwellenwerte

Großteil der Fälle von Gebärmutterhalskrebs aufgrund von Nichtteilnahme am Screening oder mangelhafter Probennahme

¹ Cutoff LSIL/CIN2+: dies bezieht sich auf die Schwellenwerte, ab wann ein Befund mittels Indextest (Zytologie) bzw. Referenzstandard (Kolposkopiegesteuerte Biopsie + Histologie) als positiv gewertet wird. CIN2+: ein Histologiebefund von CIN2, 3 oder invasives Karzinom.

sächlichen Detektionsproblemen [11]. Entsprechend fokussieren Ansätze zur Reduktion der Restmortalität² auf die Erhöhung der Screeningteilnahme und eine verbesserte Qualitätssicherung bei der Abstrichnahme [12-14].

Durch den niedrigen Schwellenwert, ab wann PAP Befunde eine weitere Abklärung erfordern, wird der Anteil unentdeckter Fälle höhergradiger Krebsvorstufen und invasiver Karzinome minimiert. Dieses Vorgehen führt jedoch zu einer großen Zahl an gesunden Frauen, die aufgrund eines positiven Befundes diagnostischen Verfahren unterzogen werden [15].

**Verfahren zur Abklärung
uneindeutiger Befunde
führen zu psychischer
Belastung und zu
Übertherapie**

Zur endgültigen Abklärung eines solchen letztlich falsch-positiven Befundes sind nach gängigen Leitlinien verschiedene Tests im Abstand mehrerer Monate vorgesehen. Dieses mehrstufige Verfahren führt aufgrund der langen Dauer der Unsicherheit zu einer hohen psychischen Belastung der betroffenen Frauen [16]. Die hohe Zahl an falsch-positiven Befunden und der häufige Wunsch der Betroffenen nach einer Abkürzung des Verfahrens führen u. U. zu Übertherapie (Konisation) und weiteren unerwünschten Folgen (z. B. erhöhtes Frühgeburtsrisiko) sowie vermeidbaren Kosten für das Gesundheitssystem.

**Bedarf an
aussagekräftigen
Triagetests**

Hier besteht Bedarf an aussagekräftigen Triagetests, die unter den auffälligen Befunden zuverlässig jene Dysplasien mit hoher Progressionswahrscheinlichkeit identifizieren können, sowie eine raschere Rückführung der Frauen ins Routinescreening ermöglichen, ohne das Risiko der Entwicklung höhergradiger Dysplasien (CIN 2/3) bzw. invasiver Zervixkarzinome zu erhöhen.

1.2 Natürlicher Verlauf der Entstehung von Gebärmutterhalskrebs

**an der
Transformationszone
(Grenze zwischen dem
Drüsenepithel des
Gebärmutterkanals und
dem Plattenepithel der
Vagina) lokalisiert**

Gebärmutterhalskrebs ist zumeist an der sogenannten Transformationszone lokalisiert, der Grenze zwischen dem Drüsenepithel des Gebärmutterkanals und dem Plattenepithel der Vagina. In der Transformationszone finden in Abhängigkeit des Hormonstatus der Frau metaplastische Umbauvorgänge statt; dadurch ist sie besonders anfällig für epitheliale Atypien. In Österreich betreffen etwa drei Viertel der bösartigen Veränderungen das Plattenepithel (Plattenepithelkarzinome, squamous lesions) und in etwa 20 % der Fälle das Drüsenepithel (Adenokarzinome, glandular lesions) (Tabelle 1.2-1). Letztere werden vom Screening weniger gut erfasst. Als Gründe dafür werden u. a. die schwierigere Abnahme der Zellen aus dem Gebärmutterhalskanal oder auch die schwierigere Erkennung der glandulären Vorstufen vermutet [17].

² Restmortalität: Mortalität trotz Screeningprogramm

Tabelle 1.2-1: Klassifizierung der Zervixkarzinome (Quelle: [18-20])

Tumortyp		Plattenepithelkarzinom			Adenokarzinom	Andere epitheliale Tumore
Häufigkeit		~70-75%			~20%	~5-10%
Ursprungsgewebe		Plattenepithel			Drüsenepithel	
Vorstufen	Histologisch	CIN1	CIN2	CIN3	Adenodysplasie AIS	Keine
	Zytologisch (Bethesda Klassifizierung)	LSIL	HSIL	HSIL	AGS	
	Zytologisch (München II Klassifizierung)	PAP IIID	PAP IIID	PAP IV	PAP IIIG	
Assoziation mit HPV Infektion		Ja			Ja	unterschiedlich

CIN – Zervikale Intraepitheliale Läsionen; ASC-US – Atypical Squamous Cells of Unknown Origin; ASC-H – Atypical Squamous Cells cannot exclude HSIL; LSIL – Low Grade Squamous Intraepithelial Lesions; HSIL – High Grade Squamous Intraepithelial Lesions; AIS – Adenocarcinoma in situ; AGS – Atypical glandular cells

Plattenepithel- wie Adenokarzinome am Gebärmutterhals sind langsam wachsende bösartige Tumore, die sich über mehrere Jahre aus Vorläuferläsionen entwickeln [2]. Unumstritten ist heute der Zusammenhang dieser Krebsarten mit der Infektion durch HPV [21]. 94,8 % aller Fälle von Gebärmutterhalskrebs sind mit einer persistierenden Infektion mit einem der kanzerogenen Hochrisikotypen der HPV assoziiert [2]: HPV 16, 18, 33, 45, 31, 58, 52, 35, 59, 56, 51, 39, 73, 68, 82. Jedoch nur 10 % der HPV Infektionen persistieren länger als zwei Jahre, die Mehrheit der Infektionen wird vom Immunsystem eliminiert oder die Viruslast unter die Detektionsgrenze gesenkt [22, 23]. Die HPV Prävalenz ist altersabhängig, am höchsten vor dem 25.-30. Lebensjahr und danach stetig absteigend [24]. Regional kann die HPV Prävalenz stark variieren [24].

94,8 % aller Fälle von Gebärmutterhalskrebs mit einer persistierenden Infektion mit einem der kanzerogenen Hochrisikotypen der HPV assoziiert

Auch aus einer persistierenden HPV Infektion entsteht nur dann Krebs, wenn diese Infektion über eine de-regulierte virale Genexpression zu exzessiver Zellproliferation, gestörter DNA Reparatur und damit verbundener Anhäufung von DNA Schäden führt [25]. Zytologisch führt das zu einer Progression der unverdächtigen sog. LSIL (Low – grade squamous intraepithelial lesions/ leichte Plattenepitheldysplasie) zu den HSIL (high-grade squamous intraepithelial lesions/schwere Plattenepitheldysplasie) und schließlich zu invasivem Krebs [23] (Abbildung 1.2-1).

aus persistierender HPV Infektion entsteht nur Krebs, wenn sie onkogene molekulare Schäden in der Zelle hervorruft

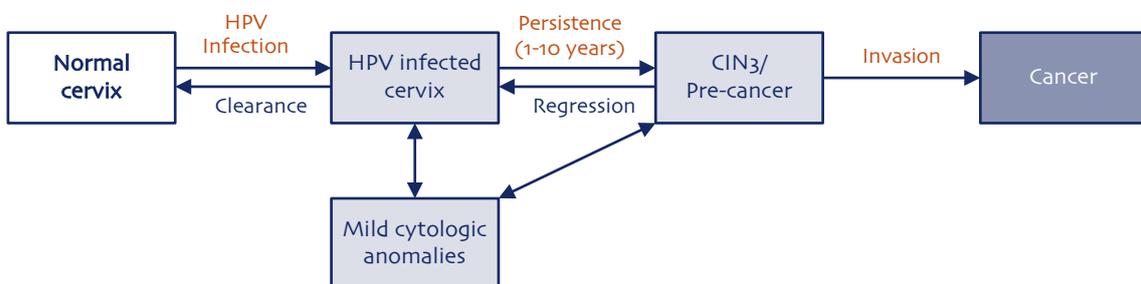


Abbildung 1.2-1: Stufen in der Entstehung des Zervixkarzinoms (Quelle: [2])

Das Krebsrisiko hängt wesentlich von der Progressionswahrscheinlichkeit der Dysplasien ab. Milde zytologische Anomalien (CIN1) weisen nach unterschiedlichen Quellen eine Progressionswahrscheinlichkeit zu CIN3 oder Krebs von nur 1,9 % [26] bis ~10 % [27] auf.

E6 und E7 sind die viralen Onkogene, die die zelleigene Zellzykluskontrolle außer Kraft setzen

Die verantwortlichen viralen Onkogene sind E6 und E7. E6 fördert die Degradation des Tumorsuppressorgens p53, während E7 an das Tumorsuppressorprotein pRB bindet und es damit funktionell deaktiviert. Beides setzt die zelleigene Kontrolle des Zellzyklus außer Kraft und initiiert eine Überproliferation der Zelle. Über eine Feedbackschleife führt dies zur Überexpression des p16^{INK4A} [23] Tumorsuppressors, der jedoch aufgrund der Bindung von pRB an E7 keine regulative Funktion mehr ausüben kann.

1.3 Etablierte und neuere Methoden zur Früherkennung von Gebärmutterhalskrebs

Erhöhung von Sensitivität und Spezifität des Primärscreenings um langwierige Abklärungsverfahren zu vermeiden

Zur Früherkennung von Gebärmutterhalskrebs kommen zur Abklärung eines auffälligen Screeningbefunds bis zur definitiven Diagnose verschiedene, zunächst nur wenig, später zunehmend invasive Methoden zum Einsatz: vom PAP Abstrich bis zur diagnostischen Konisation. Neuere Entwicklungen zielen darauf ab, die Sensitivität und Spezifität des Primärscreenings zu erhöhen, um die kostspieligeren und risikoreicheren invasiven Verfahren bei unklaren Befunden zu vermeiden. Neben der Verbesserung bestehender Verfahren gehören auch molekulare Biomarker, wie Ki-67, p16^{INK4A}, BD ProExC oder Cytoaktiv HPV L1 [28], zu den Neuentwicklungen.

1.3.1 Papanicolaou Abstrich

PAP Abstrich: Abnahme von Zellen aus dem Gebärmutterhals mittels Bürste oder Spatel; Abstrich, Fixierung und Färbung der Zellen auf Objektträger

Die konventionelle Methode des Krebsabstrichs besteht in der Abnahme von Zellen aus dem Gebärmutterhals über eine Bürste oder einen Spatel. Die Zellen werden anschließend auf einem Objektträger ausgestrichen und fixiert und in einem Zytologielabor gefärbt und befundet.

Die Screeningbefunde des PAP Abstrichs sind nicht dichotom sondern werden nach dem Grad der morphologischen Abnormität klassifiziert. Dazu stehen mehrere Nomenklatursysteme zur Verfügung.

zwei Klassifizierungssysteme für PAP Abstriche: München II (in Österreich und Deutschland) und Bethesda (international) – mit nur bedingter Entsprechung

In Österreich und Deutschland erfolgt die Klassifizierung nach dem München II System, während in der internationale Literatur üblicherweise das Bethesda System verwendet wird (Tabelle 1.3-1). Augenmerk sollte auf die Tatsache gelegt werden, dass nach der München II Klassifizierung HPV induzierte zelluläre Veränderungen (u. a. Koilozyten, koilozytäre Atypien), die vor allem bei jüngeren Frauen sehr häufig auftreten, jedoch meistens spontan remittieren, einem PAP II und damit einem noch nicht auffälligen Befund zugeordnet werden. Dieselben Veränderungen werden nach Bethesda bereits in die Kategorie LSIL eingeordnet. Weiters können mäßige Dysplasien nach München Klassifizierung sowohl in die Gruppe IIID als auch in die Gruppe IV eingeteilt werden, während nach Bethesda nur eine Klasse HSIL dafür vorgesehen ist [29].

Tabelle 1.3-1: Befundwiedergabe der Zervixzytologie –
Gegenüberstellung der beiden Klassifizierungssysteme München II und Bethesda (nach [1, 29])

München II		Bethesda-Äquivalent
o	Nicht beurteilbar	Unsatisfactory for evaluation
I	Normales altersentsprechendes Zellbild	
II	Entzündliche, regenerative, metaplastische oder degenerative Veränderungen; normale Endometrium-zellen (Angabe postmenopausal obligatorisch). Hyper- und Parakeratose; atrophisches Zellbild mit Autolyse; HPV-assoziierte Veränderungen ohne auffällige Kernveränderungen.	Negative for intraepithelial lesions or malignancy LSIL (HPV only)
III	Stärker ausgeprägte entzündliche und/oder degenerative und/oder atrophe Veränderungen mit nicht sicher beurteilbarer Dignität (CIN oder invasives Karzinom nicht auszuschließen).	ASC-US; ASC-H
IIID	Zellen einer leichten bis mäßigen Dysplasie	LSIL HSIL
IIIG	Auffällige glanduläre Zellen der Endozervix oder des Endometriums (Verdacht auf proliferative oder neoplastische Veränderungen).	Atypical glandular (NOS) cells; Atypical endocervical or glandular cells, favor neoplastic;
IV	Zellen einer mäßigen bis schweren Dysplasie oder eines Plattenepithel- oder Adenocarcinoma in situ (CIN 2-3, AIS). Kein fassbarer Anhaltspunkt für Invasion.	HSIL (without or with features suspicious for invasion); endocervical adenocarcinoma in situ (AIS)
V	Zellen eines vermutlich invasiven Plattenepithel- oder Adenokarzinoms der Zervix oder anderer maligner Tumoren.	squamous cell carcinoma; adenocarcinoma (endocervical, endometrial, extrauterine, NOS); other

CIN – Zervikale Intraepitheliale Läsionen; ASC-US – Atypical Squamous Cells of Unknown Origin; ASC-H – Atypical Squamous Cells cannot exclude HSIL; LSIL – Low Grade Squamous Intraepithelial Lesions; HSIL – High Grade Squamous Intraepithelial Lesions; AIS – Adenocarcinoma in situ; NOS – Not otherwise specified

Der Befund PAP III bezeichnet uneindeutige, aber auffällige Epithelzellen, bei denen auch höhergradige CIN oder ein invasives Karzinom nicht auszuschließen sind. Die Bethesda Klassifizierung verwendet dafür zwei Klassen: ASC-US (Atypien unbekanntes Ursprungs) und ASC-H (Atypien bei denen HSIL nicht auszuschließen sind).

PAP III:
uneindeutige, aber
auffällige Epithelzellen

PAP IIID entspricht dem Befund gering- oder mittelgradiger Dysplasien (Zervikale intraepitheliale Neoplasien (CIN) 1 oder 2). Nur bei etwa 30 % der Frauen mit PAP IIID wird eine CIN2 oder höher bestätigt [30].

PAP IIID:
gering- und
mittelgradige Dysplasien

Die Klassen haben unterschiedliche Vorhersagekraft: ein PAP IV und PAP V Befund ist hochspezifisch und benötigt keine weitere Triage [11]. Befunde der Klassen PAP III und PAP IIID sind hingegen nur wenig spezifisch.

1.3.2 Dünnschichtzytologie

Eine Weiterentwicklung der Methode stellt die flüssigkeitsbasierte Zytologie dar, bei der die Zellen nicht direkt auf einem Objektträger ausgestrichen werden, sondern als Zellsuspension gesammelt werden, von der ausgehend Zellpräparate hergestellt werden können [31, 32]. Eine erhöhte Sensitivität gegenüber der konventionellen Zytologie ist derzeit nicht erwiesen [33]. Vorteile bestehen aber in der verkürzten Auswertungszeit, der höheren Ausbeute an Zellen und der Entfernung von Verunreinigungen durch Blut oder Schleim. Zudem erlaubt sie, molekulare Tests ausgehend vom Abstrich durchzuführen ohne die Frau erneut einbestellen zu müssen, sowie automatisierte Auswertungen des Tests durchzuführen.

Flüssigkeitsbasierte
Zytologie: ermöglicht
mehrere Tests von
einem Abstrich und
automatisierte
Auswertung

Nachteilig ist der höhere Preis der Methode. Dünnschichtzytologie wird in Österreich derzeit nur als Privatleistung angeboten und teilweise von den Kassen refundiert.

1.3.3 Automatisierte Zytologie

**Auswertung durch
Computer-Imager – Ziel:
bessere und kürzere
Auswertung**

Aufgrund des hohen Anteils an negativen Befunden im Primärscreening besteht bei der Auswertung durch ZytologInnen die Gefahr der „Betriebsblindheit“. Die automatisierte Zytologie bezeichnet die Auswertung der Abstriche auf dem Objektträger durch Computer-imager. Ziel ist es, Fehler in der Auswertung aufgrund von menschlichem Versagen zu minimieren, sowie Auswertungszeiten zu verkürzen [34].

1.3.4 Kolposkopie

**visuelle Untersuchung
der Portio mittels
Spekulum und
Vergrößerungsoptik**

Bei der Kolposkopie wird mittels eines Spekulum und einer Vergrößerungsoptik die Portio untersucht. Die Untersuchung der Transformationszone kann durch eine Essigsäureprobe, die metaplastisches Gewebe weiß färbt, verfeinert werden. Weiters kann mittels der Applikation von 3 % Jodlösung atypisches Epithel identifiziert werden. Die Identifizierung der Läsionen ermöglicht die Steuerung der Biopsieentnahme [35].

Die Kolposkopie ist Teil des Leistungskatalogs in Österreich.

1.3.5 Biopsie

**Gewebeentnahme
und histologische
Untersuchung:
Einteilung der Läsionen
nach CIN1-3**

Die etablierte Diagnosemethode zur Risikoabklärung ist die kolposkopisch gesteuerte Biopsie (Gewebeentnahme) mit anschließender histologischer Untersuchung.

Die Biopsie kann über eine Knipsbiopsie erfolgen, oder über chirurgische Verfahren wie der Konisation (Entfernung eines Gewebekegels aus dem Gebärmutterhals) und der endozervikalen Kürettage (ECC, Abschabung der Auskleidung des Gebärmutterhalskanals). Konisation und ECC sind therapeutische Verfahren, die bereits eine vollständige Entfernung der Läsion zum Ziel haben.

**CIN1: HPV Infektion,
CIN2-3: Mögliche bzw.
wahrscheinliche
Krebsvorstufe**

Läsionen werden nach histologischem Befund in CIN1-3 eingeteilt. CIN1 (geringgradige Läsionen) werden heute als Manifestationen einer HPV Infektion angesehen und nicht als Krebsvorstufe. Aufgrund der hohen Rückbildungstendenz dieser Läsionen besteht daher die Option einer Kontrolle mittels Kolposkopie, Biopsie und Zytologie über maximal 2 Jahre.

CIN2 stellen möglicherweise, CIN3 hingegen mit großer Wahrscheinlichkeit bereits Krebsvorstufen dar, die ablativ behandelt werden, um invasivem Krebs vorzubeugen.

Behandlungsoptionen bei CIN2-3 sind Konisation mittels Hochfrequenz-Schlinge (LEEP, Loop-excision, Loop cone biopsy; LLETZ) oder Skalpell (Messerkonisation) [1, 15].

1.3.6 HPV High Risk Diagnostik

Mit Etablierung der HPV Infektion als Vorbedingung zu nahezu allen Fällen von Gebärmutterhalskrebs ist die HPV Diagnostik als Alternative zum PAP Abstrich als primärer Screeningmarker in Diskussion bzw. wurde in einigen europäischen Ländern bereits empfohlen [36, 37].

Das Standardverfahren in der HPV High Risk Diagnostik ist der Hybrid Capture II Test. Über die Hybridisierung von RNA Proben wird dabei DNA von 13 Hochrisiko (HR) HPV Subtypen nachgewiesen: 16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59, 68. Der Test wird als standardisierter Kit angeboten, der ein Abstrichbürstchen, das Zelltransportmedium und den Hybridisierungssassay beinhaltet.

Andere Testsysteme zum Nachweis einer HR-HPV Infektion sind in Österreich erhältlich und werden zum Teil eingesetzt. 3 davon basieren auf dem DNA Nachweis von HR-HPV Typen mittels PCR: RT HR HPV (Abbott), Cobas 4800 (Roche), Papillocheck (Greiner Bio-One) LR; eines basiert auf der Methode der Signalamplifikation: Cervista (Hologic) und eines auf mRNA PCR: Aptima (Genprobe) [38].

In Österreich wird die HPV High Risk Diagnostik bei abklärungsbedürftigen PAP-Befunden im primären Screening nicht als Triagetest vor Kolposkopie verwendet, sondern nur im Follow-up von PAP III Befunden, bei negativer histologischer Diagnose [1].

Nachweis der Infektion mit Hochrisiko HPV Typen – Hybrid Capture II Test als Standardverfahren

HPV HR Diagnostik nur im Follow-up von PAPIII bei negativer Histologie

1.4 Screening-Strategie: Management unklarer Befunde in Österreich

Das klassische Modell der Krebsfrüherkennung beinhaltet drei Schritte: Screening, Diagnose und Behandlung [15]. Screening dient der Identifizierung jener Personen in einer Population, die ein erhöhtes Risiko haben, eine Krankheit oder ein Leiden zu entwickeln. Ein Screeningbefund alleine stellt noch keine Diagnose dar. Erst die Diagnose durch das jeweils gültige Standardverfahren, unter Umständen unter Einbeziehung des Ergebnisses des Primärbefundes oder anderer abklärender Verfahren ist behandlungsentscheidend.

Die Triage ist ein zusätzlicher Schritt, der zwischen Primärscreening und Diagnose eingefügt wird, um Frauen mit verdächtigem Befund im Primärscreening nach Krankheitsrisiko weiter zu stratifizieren [15]. Oft dient die Triage der Erhöhung der Spezifität eines wenig spezifischen aber hoch sensitiven Primärscreeningtests. Frauen mit hohem Krankheitsrisiko werden zu Diagnose und Behandlung überwiesen, jene mit niedrigem Krankheitsrisiko werden, unter Umständen über ein erweitertes Follow-up, wieder ins Routinescreening überführt.

PAP III Befunde werden nach geltender Leitlinie [1] ohne Triage zur Kolposkopie überwiesen und über eine Biopsie und evtl. endozervikale Kürettage (ECC; bei nicht einsehbarer Läsion oder nicht abgrenzbarer Transformationszone) histologisch abgeklärt (Abbildung 1.4-1). Bei negativem histologischem Befund erfolgt ein erweitertes Follow-up anhand der Wiederholung der Zytologie und HPV Test, bevor die Frau wieder ins Routinescreening überführt werden kann oder im Fall von persistierenden PAP III Befunden und/oder HPV Positivität zur Konisation überwiesen wird.

Schritte in der Krebsfrüherkennung: Screening, Diagnose, Behandlung

Triage zur Stratifizierung von Frauen mit verdächtigem Primärbefund nach Krankheitsrisiko

Leitlinien Ö: derzeit keine Triage bei PAP III; Triage über Zytologie und Kolposkopie bei PAPIII

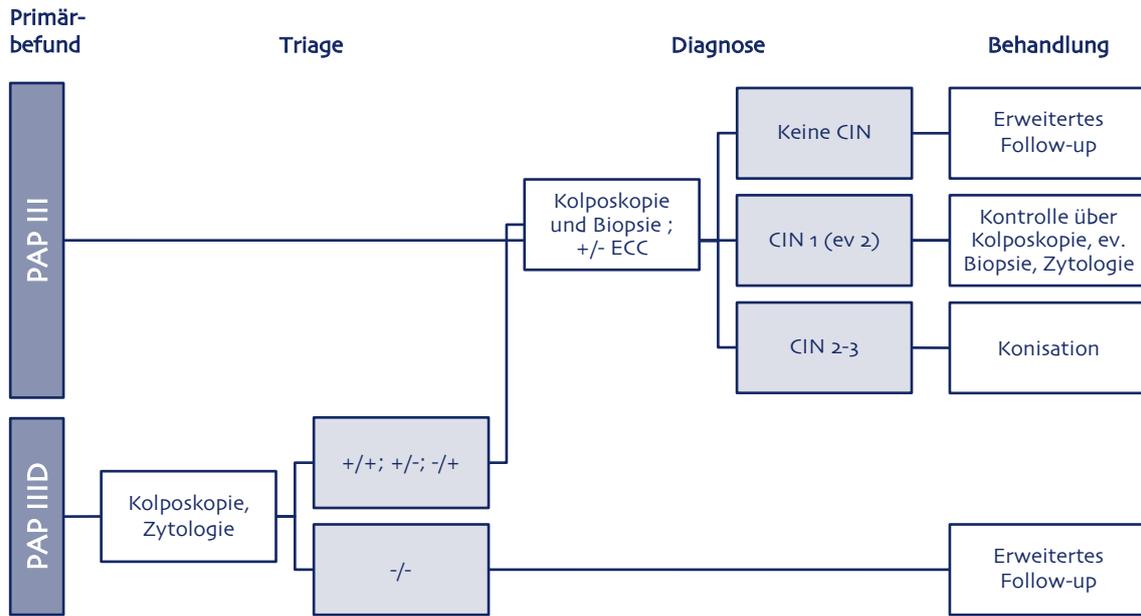


Abbildung 1.4-1: Management von Frauen mit PAP III ohne Atrophie oder Entzündung und Frauen mit PAP IIID (Quelle: [1])

Für die Triage von PAP IIID Befunden sehen die Leitlinien eine Wiederholung der Zytologie und die Kolposkopie vor (Abbildung 1.4-1). Bei negativen Befunden in der Triage (-/-) erfolgt ein erweitertes Follow-up über Zytologie. Bei positiven Befunden in einer oder beiden Triagemethoden (+/+; +/-; +/-) folgt eine histologische Abklärung nach Biopsie, endozervikaler Kürettage oder diagnostischer Konisation. In geltenden Leitlinien wird die HPV Diagnostik zur Triage von PAP IIID Befunden nicht indiziert, da in über 80 % der Frauen mit diesem Befund eine high risk HPV Infektion vorliegt und der Test daher keine zusätzlichen prognostisch relevanten Informationen liefert.

Zwar sind die genannten Triage Methoden mit geringem gesundheitlichen Risiko für die Patientin verbunden, unbefriedigend ist aber der insgesamt niedrige positive Vorhersagewert, also der niedrige Anteil jener Frauen mit positivem Befund, die tatsächlich ein Zervixkarzinom entwickeln würden [39]. Zur Einschätzung des Einsparungspotentials eines Triagetests für die Gruppe PAP III und PAP IIID sind die Kosten der verschiedenen Managementoptionen in Tabelle 1.4-1 dargestellt:

Tabelle 1.4-1: Kosten der einzelnen Leistungspakete im Zusammenhang mit Zervixkarzinomscreening und -behandlung

Leistungseinheiten	Mittelwerte der in den Honorarordnungen vereinbarten Tarife (2013) in €
Primärscreening mittels PAP	
Grundleistung (Fallpauschale, Ordination)	16,87
PAP Abstrich (Abnahme)	4,21
Zytologischer Befund (Labor)	9,28

Leistungseinheiten		Mittelwerte der in den Honorarordnungen vereinbarten Tarife (2013) in €
Methoden zur Abklärung unklarer Befunde		
PAP III	Grundleistung (Fallpauschale, Ordination)	16,87
	Kolposkopie	9,01
	Probeexcision ohne Naht	22,61
	Histologie	19,40
PAP IIID	Grundleistung (Fallpauschale, Ordination)	16,87
	Kolposkopie	9,01
	PAP Abstrich	4,21
	Zytologischer Befund	9,28
	Probeexcision ohne Naht	22,61
	Histologie	19,40
Management und Behandlung von Zervixkarzinomvorstufen		
	Konisation	2.715 (Quelle:[40], aus 2007)

1.5 P16/Ki-67 Dual Stain Immunzytologie (CINtec® Plus)

Molekulare Veränderungen in den Zellen des Gebärmutterhalses stellen eine notwendige Stufe zur Entstehung von Gebärmutterhalskrebs aus einer persistierenden HPV Infektion dar. Eine Reihe von Biomarkern werden angeboten, die den Übergang von normaler Infektion zur krankhaften Veränderung der Zelle abbilden sollen [41] und dadurch potentiell zur Triage unklarer Zytologie geeignet sind.

Gegenstand des vorliegenden Berichts ist ein Biomarker-Duo aus dieser Reihe zur Abklärung unklarer Befunde. Der p16/Ki-67 Dual-Stain Test verspricht, über die gleichzeitige immunzytochemische Färbung des Zellzyklusinhibitors p16^{INK4a} und des Proliferationsmarkers Ki-67 in einer oder mehreren Zellen im Abstrich, den Nachweis solcher viral induzierter onkogener Veränderungen.

Der Test wurde von der mtm laboratories AG in Heidelberg entwickelt, die 2011 von Roche übernommen wurde. Der Test ist in der EU als In Vitro Diagnostikum zum Nachweis von p16 und Ki-67 in PAP Abstrichen CE zertifiziert und wird als standardisierter Testkit angeboten. Aus der Zertifizierung ergeben sich darüber hinaus keine Vorgaben zur Verwendung des Tests im Zervixkarzinomscreening. Nach Angaben des Herstellers, sowie gemäß der vom Hersteller übermittelten Referenzen, bietet das CINtec® PLUS Kit eine hohe Sensitivität und Spezifität für die Identifizierung von Frauen mit hochgradigen zervikalen intraepithelialen Neoplasien (CIN2+) innerhalb der folgenden Patientinnengruppen:

- ✧ Innerhalb der Screening-Population [42-44]
- ✧ innerhalb der Gruppe von Patientinnen mit einem PAP-Befund von ASC-US oder LSIL [45-48], oder
- ✧ innerhalb der Gruppe von Frauen mit negativem PAP Befund, aber einem positiven Befund für Hochrisiko-HPV-Typen [49].

p16/Ki-67 Dual-Stain Test:
immunzytochemische
Doppelfärbung des
Tumorsuppressorprotein
s p16^{INK4a} und des
Proliferationsmarkers
Ki-67 direkt im Abstrich

Insbesondere positioniert sich der Test als Alternative zur HR-HPV Diagnostik in jenen Patientinnengruppen mit hoher HR HPV Prävalenz, in denen ein HR-HPV Test nur geringe Sensitivität aufweist (<30 Jahre, LSIL+). Da die HPV Diagnostik jedoch nach gängigen Leitlinien in Österreich in keiner der genannten Patientinnengruppen die primäre Folgeuntersuchung ist [1], wurde im vorliegenden Bericht auf eine vergleichende Beurteilung des CINtec®PLUS mit der HR-HPV Diagnostik verzichtet.

Ravarino *et al.* berichten auch eine hohe Sensitivität und Spezifität des CINtec® PLUS Tests in der Erkennung glandulärer Läsionen [50].

Listenpreis:
1.200 €/50 Tests

Nach Angaben des Herstellers liegt der Listenpreis bei 24 € pro Test (1.200 € für 50 Tests). Dazu kommen noch Laborkosten und Kosten für die Probenahme, Färbung, Auswertung, Verschickung und Speicherung der Daten. Die Färbung und Auswertung erfolgt in einem zytologischen Labor, unter Umständen ist eine entsprechende Schulung des Personals notwendig.

Nach Angaben des Herstellers kann der Test sowohl an Dünnschichtpräparaten (ThinPrep®, BD SurePath®) oder herkömmlichen PAP Abstrichen durchgeführt werden.

**Geschätzter
jährlicher Bedarf:**
30.000 Tests

Gegenstand dieses Berichts ist die Evaluation des Nutzens der p16/Ki-67 Dual-Stain Immunzytologie bei der Triage von Frauen mit einem PAP III oder PAP IIID Befund im Primärscreening. 1 % der jährlich rund 1,5 Millionen PAP Abstriche in Österreich [29] ergeben den Befund PAP III und weitere 1 % den Befund PAP IIID [40]. Bei einem Einsatz des Tests als Triagetest bei PAP III und PAP IIID kann demnach von einem Bedarf von 30.000 Tests jährlich ausgegangen werden.

2 Fragestellung

Welche Evidenz besteht zum Nutzen der Triage mit (zusätzlicher) p16/Ki-67 Dual Stain Immunzytologie im Management von Frauen mit einem PAP III oder PAP IIID Befund im Vergleich zum herkömmlichen Management?

Nutzen der p16/Ki-67 Dual Stain Immunzytologie zur Triage von PAPIII/IIID Befunden

Die Einbettung des p16/Ki-67 Dual Stain Tests in das Management von Frauen mit PAP III oder PAP IIID Befund im Primärscreening im Vergleich zum herkömmlichen Management [1] ist in Abbildung 2-1 bzw. Abbildung 2-2 (auf den nächsten beiden Seiten) dargestellt. Der Vergleich stellt die erwarteten Auswirkungen für die Patientinnen in den verschiedenen Diagnosegruppen dar. Zusammen fließen diese in die klinischen Endpunkte in der PICO Fragestellung (Tabelle 2-1) ein.

Tabelle 2-1: PICO Fragestellung

Population	Frauen mit erstmaligem PAP III Befund oder Frauen mit erstmaligem PAP IIID Befund im opportunistischen Primärscreening in Österreich
Intervention	Diagnostische Strategie mit p16/Ki67 als Triage Test
Kontroll-intervention	Herkömmliche Triage/Diagnose Strategie
Outcomes	<p>a. Patientinnenrelevanter Nutzen/Schaden:</p> <ul style="list-style-type: none"> ✿ Auswirkung auf die Inzidenz höhergradiger Vorstufen (CIN2-3) und auf das Risiko für invasive Karzinome ✿ Auswirkung auf die Lebensqualität (psychische Belastung, Angst, Akzeptanz bei den Frauen) ✿ Schäden: Nebenwirkungen durch weitere Behandlung (Biopsie, Konisation) <p>b. Auswirkungen auf das Patientinnenmanagement:</p> <ul style="list-style-type: none"> ✿ Rate der Überweisungen zur Kolposkopie gesteuerten Biopsie bzw. anderen histologischen Verfahren. ✿ Rate der Follow-up Verfahren ✿ Rate der durchgeführten Behandlungen ✿ Auswirkung auf die Screeningteilnahme <p>c. Diagnostische Genauigkeit: Sensitivität und Spezifität für die Identifizierung höhergradiger CIN (CIN2+), Positiver/Negativer Vorhersagewert</p>

Nicht Gegenstand dieses Berichts ist die Evaluierung anderer möglicher Einsatzgebiete des Tests wie a) Einsatz in der gesamten Screeningpopulation als Zusatztest zum PAP Screening, b) Einsatz in der Triage von Frauen mit negativer Zytologie aber positivem HPV Test oder c) Einsatz im Follow-up nach erfolgter Behandlung. Auch Referenzen des Herstellers, die sich auf diese Einsatzgebiete bezogen, wurden daher nicht eingeschlossen.

andere Einsatzmöglichkeiten des Tests nicht evaluiert

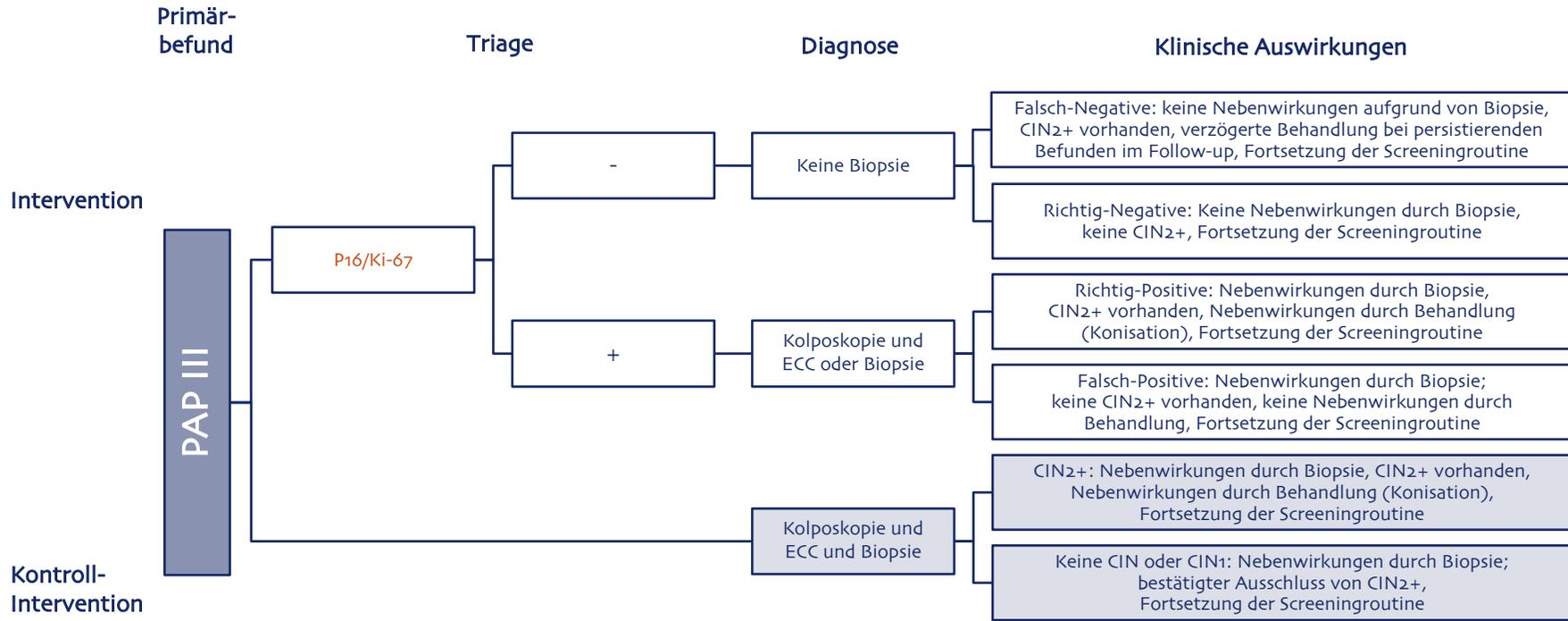


Abbildung 2-1: p16/Ki-67 Immunzytologie als Triagetest im Management von Frauen mit einem Primärbefund von PAP III – erwartete klinische Auswirkungen in den verschiedenen Diagnosegruppen

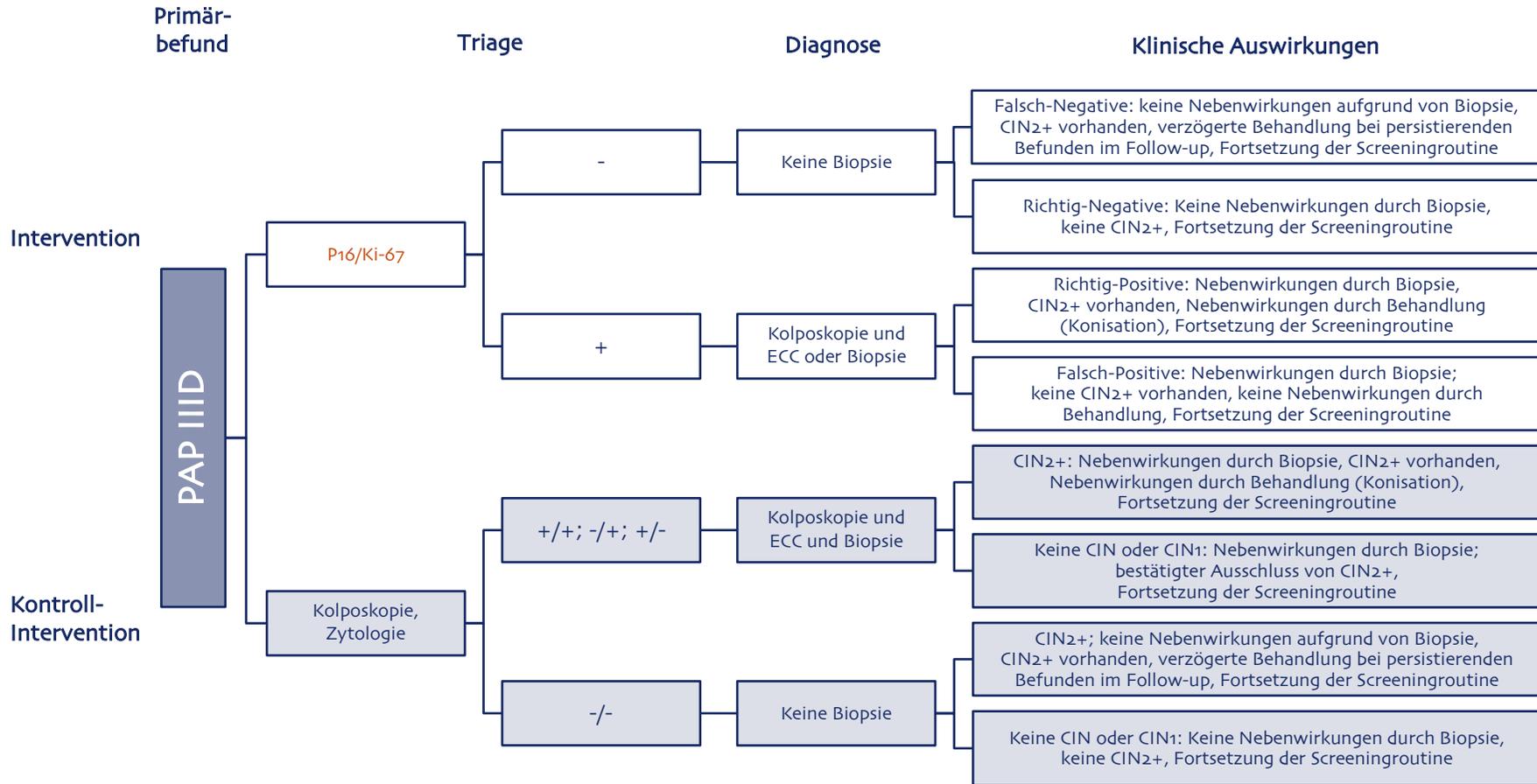


Abbildung 2-2: p16/Ki-67 Immunzytologie als Triage-test im Management von Frauen mit einem Primärbefund von PAP IIIID – erwartete klinische Auswirkungen in den verschiedenen Diagnosegruppen

3 Methoden

3.1 Methodik der Nutzenbewertung

Im Gegensatz zur Evaluierung therapeutischer Interventionen, existiert zur Zeit noch kein allgemein akzeptiertes Rahmenwerk zur Nutzenbewertung von Biomarkern. Zur Erstellung des vorliegenden Berichts wurden daher die Methodenmanuale internationaler HTA Institute herangezogen [51-58] und folgende Vorgehensweise daraus synthetisiert.

Die Methode der Wahl zur Nutzenbewertung von Biomarkern ist die systematische Übersichtsarbeit.

Die Fragestellung erfolgt nach dem PICO Schema und umfasst also die Definition der Population, der Intervention, der Kontrollintervention und der Endpunkte. In manchen Fällen ist es auch sinnvoll, Zeitpunkte und Setting der Intervention in der Fragestellung zu definieren.

Synthese der Methodik der Nutzenbewertung aus Methodenmanualen internationaler HTA Institute

3.1.1 Definition der Endpunkte: Diagnostische Genauigkeit

Der Nutzen eines Biomarkertests ergibt sich aus drei Ebenen: Technische Qualität, diagnostische Genauigkeit und klinischer Nutzen.

Parameter zur technischen Qualität geben an, wie genau und reproduzierbar der Test das molekulare Ereignis misst. Der Nutzen des Tests ist von der Robustheit des Tests im Routinesetting abhängig. Die technische Qualität wird im Rahmen des Zulassungsverfahrens geprüft und ihre Evaluierung ist nicht Teil dieses Berichts.

technische Qualität

Parameter zur diagnostischen Genauigkeit geben an, wie stark die Assoziation des Testergebnisses mit dem tatsächlich zugrundeliegenden Sachverhalt (Krankheit oder nicht vorliegende Krankheit) ist. Als Referenzstandard wird das beste verfügbare diagnostische Verfahren herangezogen. Sämtliche möglichen Ergebnisse zu diesem Zusammenhang werden in einer Vierfeldertafel zusammengefasst (Tabelle 3.1-1).

Parameter zur Diagnostischen Genauigkeit

Tabelle 3.1-1: Vierfeldertafel

		Referenzstandard	
		Krank	Gesund
Indextest	Positiv	Richtig Positiv (RP)	Falsch Positiv (FP)
	Negativ	Falsch Negativ (FN)	Richtig Negativ (RN)

Typische Parameter sind Sensitivität und Spezifität des Tests, aber auch kombinierte statistische Größen wie die Fläche unter der Grenzwertoptimierungskurve/Receiver-Operator-Curve (ROC).

$$\text{Sensitivität} = \text{RP}/(\text{RP}+\text{FN})$$

$$\text{Spezifität} = \text{RN}/(\text{RN}+\text{FP})$$

Sensitivität und Spezifität

Um eine Krankheit sicher auszuschließen (RuleOut) ist eine hohe Sensitivität des Tests notwendig. Eine hohe Sensitivität ist insbesondere wichtig, wenn ein Falsch-Negatives Resultat mit hohem Risiko (unentdeckte Krankheit) verbunden ist.

Um eine Krankheit sicher vorauszusagen (RuleIn) ist eine hohe Spezifität notwendig. Eine hohe Spezifität ist anzustreben, wenn ein Falsch-Positives Ergebnis mit hohem Risiko (unnötige Behandlung) verbunden ist.

Positiver und Negativer Vorhersagewert

Weitere Parameter sind der Positive Vorhersagewert (Positive predictive value, PPV) und der Negative Vorhersagewert (Negative predictive value, NPV). Sie geben die Wahrscheinlichkeit an, dass ein Patient krank bei positivem bzw. gesund bei negativem Testergebnis ist. Die Angabe von PPV und NPV für eine bestimmte Population ist nur zulässig, wenn die Prävalenz der Krankheit in dieser Population mit der Prävalenz in der Studienpopulation übereinstimmt.

$$PPV = \frac{RP}{(RP + FP)}$$

$$NPV = \frac{RN}{(RN + FN)}$$

Technische Qualität und diagnostische Genauigkeit sind Voraussetzungen um zu einem Nutzen des Tests zu kommen; sie stellen noch keine patientenrelevante Endpunkte dar, sondern Surrogatparameter für Endpunkte zum klinischen Nutzen.

3.1.2 Definition der Endpunkte: Klinischer Nutzen

Der klinische Nutzen des Tests (vgl. [59]) ergibt sich aus der Wahrscheinlichkeit, zu verbesserten klinischen Ergebnissen bei Einsatz des Tests zu kommen. Da der Test zur Triage eingesetzt werden soll, ergibt sich der klinische Nutzen aus dem prognostischen Potential des Tests, jene Frauen mit auffälligem Primärbefund zu identifizieren, die mit hoher Wahrscheinlichkeit höhergradige CIN (CIN2+) bzw. invasive Karzinome entwickeln werden.

Endpunkte zur Bewertung des Klinischen Nutzens: Mortalität, Lebensqualität, Nebenwirkungen, Vermeidung weiterer invasiver Tests. Soziale, psychologische Effekte; Auswirkungen auf das PatientInnenverhalten.

Klinische Auswirkungen können sich zunächst direkt aus dem Einsatz des Tests ergeben, beispielweise bei invasiven Tests oder bei Tests, die eine Bestrahlung erfordern. Vor allem aber ergeben sie sich aus den therapeutischen und weiterführenden diagnostischen Interventionen, die basierend auf den Testergebnissen vorgenommen werden oder unterlassen werden können. PatientInnenrelevante Endpunkte sind demnach beispielsweise: Mortalität, Lebensqualität, Nebenwirkungen von Test wie Behandlung, positive Konsequenzen durch die Vermeidung weiterer invasiver Tests. Weitere Effekte können soziale, psychologische Effekte oder Auswirkungen auf das PatientInnenverhalten sein (z. B. eine Veränderung des Lebensstils).

Zum Vergleich der klinischen Auswirkungen mit neuem und altem Test wurde ein Entscheidungsbaum erstellt und die klinischen Auswirkungen in den verschiedenen Diagnosegruppen identifiziert (Abbildung 2-1 und Abbildung 2-2).

Evaluierung über klinische Studien mit patientInnenrelevanten Endpunkten

Zur Evaluierung des klinischen Nutzens sind vorrangig randomisierte kontrollierte Studien (Randomised Controlled Trial/RCT) mit patientInnenrelevanten Endpunkten heranzuziehen (vgl. auch [60] für mögliche Studiendesign), die den direkte Vergleich der diagnostischen Strategie mit und ohne neuem Test erlauben. Unter bestimmten Voraussetzungen ist es möglich, aus Ergebnissen zur diagnostischen Genauigkeit Aussagen zu klinischen Endpunkten zu treffen, indem Evidenz zur diagnostischen Genauigkeit mit Evi-

denz zu klinischen Endpunkten verknüpft wird (ein sogenannter „linked-evidence“ Ansatz) [61-63]. Dieser Ansatz kann Verwendung finden, wenn:

- ✦ Der neue Test ebenso sensitiv und spezifisch wie der alte Test ist, aber weniger invasiv, kostengünstiger oder schneller/leichter auszuwerten. Die Annahme ist, dass PatientInnenmanagement und nachfolgende Behandlung zwischen altem und neuem Test vergleichbar sind.
- ✦ Der neue Test ebenso sensitiv, aber spezifischer als der alte Test ist und daher die Vermeidung unnötiger Behandlungen erlaubt. Die Annahme hier ist, dass neuer und alter Test dieselben Fälle erkennen und die Wirksamkeit der Behandlung nicht davon beeinflusst wird, welcher Test verwendet wird.
- ✦ Um Aussagen zu klinischen Endpunkten treffen zu können, muss in klinischen Studien erwiesen sein, dass die PatientInnen, die durch den neuen Test erkannt werden, dasselbe Krankheitsspektrum aufweisen wie jene, die durch den alten Test detektiert werden und ebenso wie jene auf die Behandlung ansprechen. Alle Testergebnisse sollten weiters durch denselben Referenzstandard bestätigt werden.

**Voraussetzungen für
„linked evidence“**

Üblicherweise ist die diagnostische Genauigkeit jedoch nur Surrogat für patientInnenrelevante Endpunkte und die Studien liefern daher in der Regel nach GRADE nur geringe Evidenz für den klinischen Nutzen [64]. Auch das PatientInnenmanagement dient nur als Surrogatendpunkt für den tatsächlichen klinischen Nutzen [65].

3.1.3 Beurteilung der Qualität der Evidenz

Zur Beantwortung der Fragestellung lagen ausschließlich Studien zur diagnostischen Genauigkeit vor.

Zur Qualitätsbewertung dieser Studien wurde das QUADAS-2 (Quality Assessment of Diagnostic Accuracy Studies) Tool verwendet [66]. Es handelt sich dabei um einen Fragenkatalog, anhand dessen Studien zur diagnostischen Genauigkeit hinsichtlich ihrer internen und externen Validität evaluiert werden können.

**Qualitätsbewertung
über QUADAS-2 Tool**

- ✦ In Abstimmung mit der Fragestellung wurden nachstehende Kriterien zur internen und externen Validität der Studien von den Autorinnen definiert. Um eine Einstufung der Qualität als „Hoch“ zu erhalten, müssen alle Kriterien erfüllt werden, inklusive der Kriterien mit Zusatz „OPTIMAL“.
- ✦ Um eine Einstufung der Qualität als „Mittel“ zu erhalten, müssen alle Kriterien erfüllt werden, mit Ausnahme der Kriterien mit dem Zusatz „OPTIMAL“.
- ✦ Die Qualität von Studien, die weder die Kriterien für „Hoch“ noch für „Mittel“ erfüllen, wird als „Niedrig“ eingestuft.

Kriterien zur Studienteilnahme

Selektionsbias kann entstehen, wenn die Studien nur Fälle mit eindeutig diagnostizierter Krankheit bzw. Abwesenheit von Krankheit und keine Fälle mit diagnostischer Unsicherheit einschließen. Dies führt möglicherweise zu einer Überschätzung der diagnostischen Genauigkeit.

**Kriterien zur
Studienteilnahme**

Um das Risiko für Selektionsbias zu minimieren, gilt für die Auswahl der Studienteilnehmerinnen:

- ✿ Die Selektionskriterien für die Teilnahme an der Studie müssen inklusive des Alters der Frauen und der zytologischen Klassifizierung beschrieben sein. Die durchschnittliche Persistenz der HPV Infektionen und somit das Progressionsrisiko verändern sich mit dem Alter der Frauen [6] und können dadurch die diagnostische Genauigkeit beeinflussen. Da der Test für die Verwendung als Triage nach Primärbefund evaluiert wird, müssen Frauen mit vorhergehenden Behandlungen oder persistierenden auffälligen Befunden oder Beschwerden ausgeschlossen werden.
- ✿ Fall-Kontroll Studien sind zu vermeiden und die Studien dürfen keine unzulässigen Ausschlüsse machen, also z. B. von Frauen, deren Diagnose nicht bestätigt wurde.

Um die externe Validität der Ergebnisse zu gewährleisten gilt:

- ✿ Die Zusammensetzung der Studienpopulation muss der Zielpopulation entsprechen.
- ✿ OPTIMAL: Konsekutive Aufnahme der Studienteilnehmerinnen.

Kriterien zum p16/Ki-67 Test (Indextest)

Kriterien zum p16/Ki-67 Test (Indextest)

Die Auswertung des p16/Ki-67 Test erfolgt subjektiv (Suche nach doppelpositiven Zellen in einer großen Anzahl von negativen Zellen) und daher kann Kenntnis des Referenzstandard-ergebnisses zu Bias führen.

Um das Risiko für Bias zu minimieren, gilt für den Einsatz des Indextests:

- ✿ Die Auswertung des Indextests muss gegenüber dem Referenzstandard verblindet erfolgen.
- ✿ Der Schwellenwert für ein positives Testergebnis muss eindeutig beschrieben sein. Die Verwendung unterschiedlicher Schwellenwerte kann zu Bias und Über- oder Unterschätzung der Testperformance führen.

Um die externe Validität der Ergebnisse zu gewährleisten gilt:

- ✿ Die Schwellenwerte müssen dem Schwellenwert in Routinebedingungen entsprechen. Laut Angaben des Herstellers ist *eine* doppelt gefärbte Zelle im Abstrich ausreichend für ein positives p16/Ki-67 Dual Stain Testergebnis.

Kriterien zum Referenzstandard

Kriterien zum Referenzstandard

Der Referenzstandard sollte eindeutig das Vorhandensein oder die Abwesenheit von Läsionen identifizieren können.

Um das Risiko für Bias zu minimieren, gilt für den Einsatz des Referenzstandards:

- ✿ Als Goldstandard gilt die Kolposkopie-gesteuerte Biopsie. Aufgrund der bekannt schlechten Sensitivität der PAP Zytologie kann diese nicht als valider Referenzstandard angesehen werden.
- ✿ Aufgrund der subjektiven Interpretation des Referenzstandards muss diese gegenüber der Ausgangszytologie und dem Indextest verblindet erfolgen.

Um die externe Validität der Ergebnisse zu gewährleisten gilt:

- ✧ Die Anwendung des Referenzstandards muss der Anwendung in Routinesettings entsprechen.

Kriterien zum Studienprotokoll

Ein angemessenes (zeitnahes) Intervall zwischen Referenzstandard und Index-test muss sicherstellen, dass das Ergebnis des Referenzstandard nicht durch die Progression oder Regression der Kondition beeinflusst werden kann.

Um das Risiko für Bias zu minimieren, gilt für das Studienprotokoll:

- ✧ Unter Berücksichtigung des natürlichen Verlaufs der Krebsentstehung im Gebärmutterhals darf das Intervall zwischen Index-test und Referenztest nicht größer als 6 Monate sein.
- ✧ Bei allen Studienteilnehmerinnen müssen die Index-testergebnisse über einen geeigneten Referenzstandard verifiziert werden.
- ✧ OPTIMAL: Alle Studienteilnehmerinnen erhalten den gleichen Referenzstandard.
- ✧ OPTIMAL: Alle Studienteilnehmerinnen werden in der Analyse eingeschlossen.
- ✧ Uneindeutige oder fehlende Ergebnisse müssen berichtet werden und Austritte aus der Studie müssen erklärt werden.

Kriterien zum Studienprotokoll

Vollständige Berichterstattung der Daten

Die Studie muss ausreichend Daten berichten, anhand derer die Vierfeldertafel erstellt und die Kenngrößen berechnet werden können.

Erstellung der Vierfeldertafel

3.1.4 Stärke der Evidenz nach GRADE

Die Stärke der Evidenz ergibt sich aus den folgenden Aspekten (nach [67]):

- ✧ Studiendesign
- ✧ Interne Validität der einzelnen Studien
- ✧ Direktheit
- ✧ Konsistenz der Ergebnisse
- ✧ Präzision

Zunächst wird anhand des Studiendesigns das Ausgangslevel der Evidenzstärke innerhalb einer Evidenzhierarchie definiert. Anschließend kann eine Hochstufung bzw. Abstufung innerhalb der Evidenzhierarchie je nach Beurteilung in den anderen Kategorien erfolgen, wobei derselbe Faktor nicht zur Hoch- bzw. Abstufung in verschiedenen Kategorien führen darf.

Ausgangslevel der Evidenzstärke entsprechend Studiendesign

Studiendesign

Je nach Endpunkt (Klinischer Nutzen oder diagnostische Genauigkeit) bzw. Fragestellung kommen unterschiedliche Evidenzhierarchien zum Einsatz; für diesen Bericht verwendeten wir die wie vom Oxford Center for Evidence based Medicine vorgeschlagenen Evidenzlevel [68] für Screeningtests um Studien zu Endpunkten zum klinischen Nutzen zu untersuchen und Evidenzlevel zur diagnostischen Genauigkeit für Studien mit ebendiesen Endpunkten (Tabelle 3.1-2).

Zur Evaluierung der Studiendesigns verwendete Evidenzhierarchien

Tabelle 3.1-2: Zur Evaluierung der Studiendesigns verwendete Evidenzhierarchien (nach [68])

Evidenzlevel	Screening	Diagnostische Genauigkeit
I	Systematischer Review von RCT	Systematischer Review von Level II Studien
II	RCT	Studie zur diagnostischen Genauigkeit mit unabhängigem, verblindetem Vergleich mit validem Referenzstandard, konsekutive PatientInnenkohorte, definierte PatientInnenpräsentation
III	Nicht-randomisierte kontrollierte Kohortenstudie/Follow-up Studie	Nichtkonsekutive Studien oder Studien ohne konsistent angewendeten Referenzstandard
IV	Fall-Kontroll Studie oder Fallserie	Fall-Kontroll-Studien oder Studien mit ungenauem Referenzstandard

Bewertung der internen Validität gemäß Qualitätsprüfung

Interne Validität

Interne Validität bezieht sich auf die Robustheit der einzelnen Studien gegenüber dem Auftreten von Bias. Zur Bewertung der internen Validität werden die Ergebnisse aus der Qualitätsprüfung (Kapitel 3.1.3) für die einzelnen Studien herangezogen.

Zusammenhang zwischen durchgeführtem Test und untersuchten Endpunkten

Direktheit

Direktheit bezieht sich auf den Zusammenhang zwischen dem durchgeführten Test und den untersuchten Endpunkten. In Abwesenheit von klinischen Studien, die die diagnostische Genauigkeit mit klinischen Endpunkten verknüpfen (Kapitel 3.1.2), ist die Evidenz zu diagnostischen Endpunkten als indirekt in Bezug auf den klinischen Nutzen zu werten, kann jedoch als direkt bewertet werden, wenn die diagnostische Genauigkeit als Endpunkt evaluiert wird.

Indirektheit kann auch entstehen, wenn der Vergleich zweier diagnostischer Methoden nicht der gewünschten klinischen Verwendung entspricht (z. B. Betrachtung eines Tests als Ersatz für einen anderen, wenn tatsächlich der Test als Triage oder Add-on Test gedacht ist).

Konsistenz der Ergebnisse: Bandbreite, Variabilität

Konsistenz der Ergebnisse

Zur Analyse der Konsistenz der Ergebnisse werden die Ergebnisse und Konfidenzintervalle der Studien herangezogen. Die Bandbreite der Studienergebnisse muss vertretbar eng sein. Hohe Variabilität der Studienergebnisse muss sich aus Unterschieden im Studiendesign erklären lassen.

Konfidenzintervall um den gepoolten Effektschätzer

Präzision

Präzision bezieht sich auf die Sicherheit, mit der ein Effektschätzer den wahren Wert eines bestimmten Endpunktes repräsentiert. Nach erfolgter Meta-Analyse entspricht dies dem Konfidenzintervall um den gepoolten Effektschätzer.

3.2 Systematische Literatursuche und -auswahl

Die systematische Literatursuche wurde am 08.05.2013 in folgenden Datenbanken durchgeführt:

- ✿ Medline (via Ovid)
- ✿ Embase
- ✿ The Cochrane Library
- ✿ CRD (DARE, NHS-EED, HTA)

Die systematische Suche wurde in Cochrane auf Reviews, HTAs und klinische Studien eingeschränkt. Die detaillierte Suchstrategie befindet sich im Anhang.

Nach Entfernung der Duplikate lagen insgesamt 411 bibliografische Zitate vor.

Alle vom Hersteller des CINtec®Plus bereitgestellten Referenzen waren in den Literaturergebnissen enthalten.

Durch Handsuche wurden zusätzliche 3 Arbeiten identifiziert, was die Gesamtzahl der Treffer auf 414 erhöhte.

Die Einschlusskriterien für relevante Studien sind in Tabelle 3.2-1 zusammengefasst.

**Systematische
Literatursuche in
4 Datenbanken,
Stichtag: 08.05.2013**

411 Treffer

**Handsuche:
3 zusätzliche Treffer**

Tabelle 3.2-1: Ein- und Ausschlusskriterien

Einschlusskriterien	Ausschlusskriterien
Systematische Reviews mit expliziter Suchstrategie	Anderer Indextest als p16/Ki-67 Immunzytochemie
Primärstudien	Dual Stain verwendet für andere Krebsarten
Sprache Deutsch oder Englisch	Comments, letters, unsystematic reviews, usw. P16-Ki-67 als primärer Screening marker P16-Ki-67 für Immunhistologie

Insgesamt standen 414 Quellen für die Literatursuche zur Verfügung. Die Literatur wurde von zwei Personen (AK, IZ) unabhängig voneinander begutachtet. Differenzen wurden durch Diskussion und Konsens oder die Einbindung einer dritten Person gelöst. Der Auswahlprozess ist in Abbildung 3.2-1 dargestellt.

**414 Quellen für
die Literatursuche**

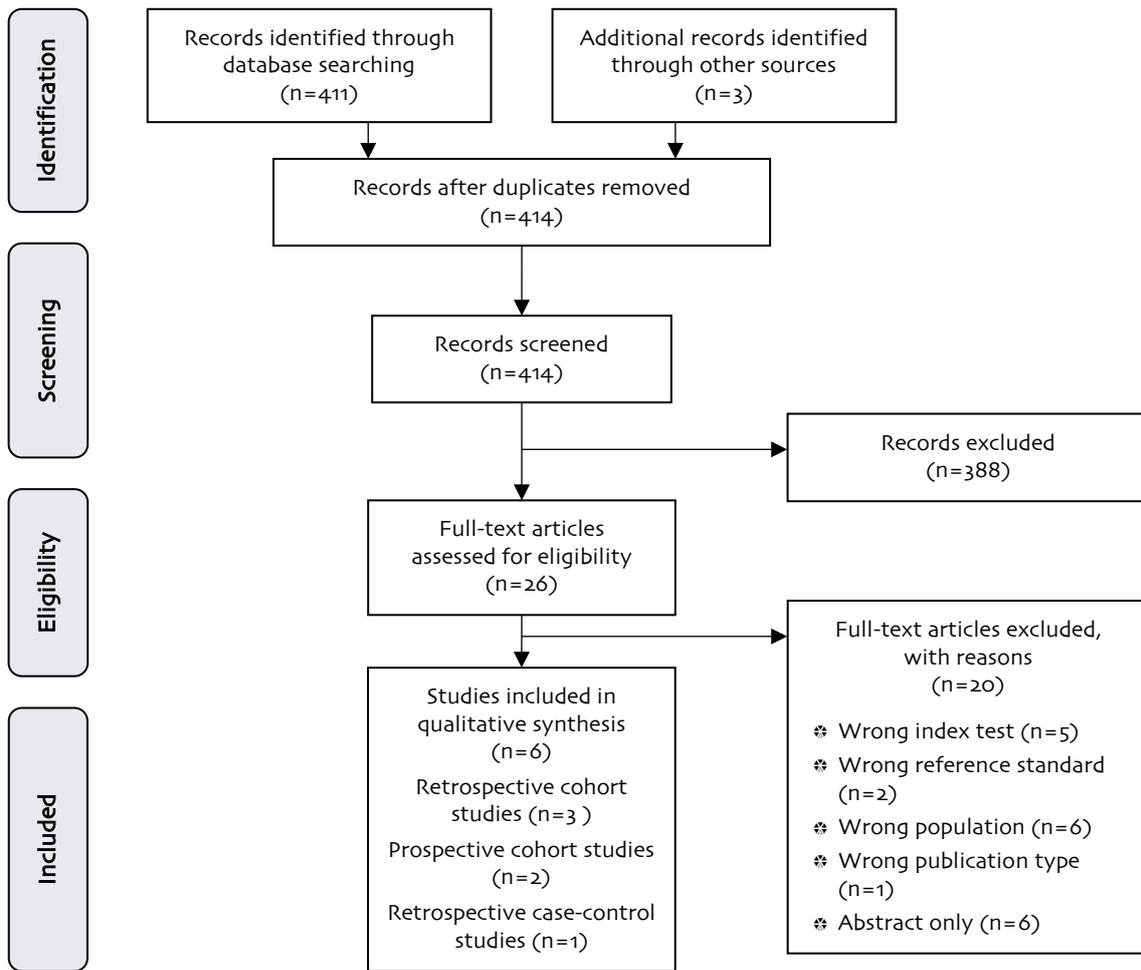


Abbildung 3.2-1: Darstellung des Auswahlprozesses (PRISMA Flow Diagram)

4 Datenextraktion

Die Datenextraktion wurde von einer Person (AK) durchgeführt. Eine zweite Person (IZ) überprüfte unabhängig die Vollständigkeit und Korrektheit der extrahierten Daten.

Die Studienmerkmale und Ergebnisse der ausgewählten Einzelstudien wurden in der Datenextraktionstabelle (Tabelle 4.1-1) zusammengefasst.

**Datenextraktion
unabhängig durch
zwei Personen**

4.1 Datenextraktionstabelle

Tabelle 4.1-1 Datenextraktion: Studienmerkmale und -Ergebnisse

Author, year, reference number	Edgerton (2011) [70]	Loghavi (2012) [46]	Schmidt (2011) [47]	Waldstrom (2013) [48]	Wentzensen (2012) [45]	Ziemke (2012) [69]
Country	United States	United States	Germany, Italy	Denmark	United States	Germany
Study sponsor	NR	NR	mtm laboratories	NR	National Cancer Institute	NR
Conflicts of interest	NR	mtm laboratories provided CINtec PLUS test kits free of charge.		No conflicts of interest disclosed.	M.Schiffmann has Other Commercial Research Support from Qiagen CareHPV and Roche free specimen testing.	No conflicts of interest.
Index test	CINtec® PLUS Dual Stain (mtm labs)	CINtec® PLUS Dual Stain (mtm labs)	CINtec® PLUS Dual Stain (mtm labs)	CINtec® PLUS Dual Stain (mtm labs)	CINtec® PLUS Dual Stain (mtm labs)	CINtec® PLUS Dual Stain (mtm labs)
Cytology preparation method	BD SurePath® liquid-based Pap slides	BD SurePath® liquid-based Pap slides	Hologic ThinPrep® liquid based cytology vials	Hologic ThinPrep® liquid based cytology vials	Hologic ThinPrep® liquid based cytology vials	conventional cytology (82) and Thinprep® liquid based cytology (140)
Description	Destained SurePath slides stained with CINtec PLUS Kit. Slides analysed by a single pathologist blinded to histology results, but not to ASC-US diagnosis.	Destained SurePath slides stained with CINtec PLUS Kit. Slides analysed by three pathologists blinded to index cytology and follow up diagnoses.	Cytology slide prepared from residual ThinPrep vial using T2000 slide processor and DualStain with CINtec PLUS Kit (mtm labs). Interpretation by trained cytotechnologist and pathologist review for pos. Slides. Blinding/ Order of testing not described.	Cytology slide prepared from residual ThinPrep vial using T2000 slide processor (Hologic) and DualStain with CINtec PLUS Kit (mtm labs). Independent interpretation by three observers (2 pathologists and 1 cytotechnician). The observers were unaware of the follow-up data and the results from the mRNA test.	Cytology slide prepared from residual ThinPrep vial using T2000 slide processor (Hologic) and DualStain with CINtec PLUS Kit (mtm labs). Interpretation by trained cytotechnologist and confirmed by a pathologist. The observers were unaware of the follow-up data and the results from the HPV test.	Destained conventional cytology slides or LBC Thin Prep slides were stained with DualStain kit. Interpretation by pathologist and blinded review by second pathologist.
Criterion for p16/Ki-67 positivity	At least one cell stained cytoplasmic brown and nuclear red	At least one cell stained cytoplasmic brown and nuclear red	At least one cell stained cytoplasmic brown and nuclear red	At least one cell stained cytoplasmic brown and nuclear red	At least one cell stained cytoplasmic brown and nuclear red	At least one cell stained cytoplasmic brown and nuclear red
Reference test Description	Follow-up colposcopic biopsy diagnosed by routine staff physicians. Most severe dysplasia was recorded (in case of multiple biopsies)	Biopsy or cytology at follow up, interpretation not described. Most severe dysplasia was recorded. Follow-up was ended when HG CIN were diagnosed.	The histologic diagnosis established on the recut tissue block specimens was used as the gold standard (ie, diagnostic accuracy criterion) for the study. A majority consensus diagnosis was established for each case by 2 or more expert pathologists. The reviewers were blinded to the original local histopathologic result and to all other test results. A combined review of H&E-stained slides and consecutive slides stained with the (p16) CINtec Histology Kit was performed by the expert pathologists.	Follow-up with histology with a punch and/or cone biopsy (48,2%) and/or cytology (52%) for at least 5 years. Most severe diagnosis from clinical follow-up with cytology or histology according to Danish guidelines. Interpretation not described. Follow-up time of at least 5 years.	Colposcopy and extended biopsy protocol with digital photographic documentation of biopsy sites. Up to 4 biopsies from acetowhite lesions or large heterogenous lesions ext over 2 quadrants. CIN2+ treated by LEEP of transformation zone. IF Fewer than 4 biopsies THEN biopsy from a quadrant w/o visible CIN was included. Interpretation was performed independently from results of index test, but cytology results often available according to standard protocols. Outcomes considered based on worst biopsy result.	Follow-up with histology with a punch and/or cone biopsy (27,7%) and/or cytology (routine patient management). Interpretation not described.

Author, year, reference number	Edgerton (2011) [70]	Loghavi (2012) [46]		Schmidt (2011) [47]		Waldstrom (2013) [48]	Wentzensen (2012) [45]		Ziemke (2012) [69]	
Study design	Retrospective cohort study	Retrospective cohort study		Retrospective, multi-center, multi-national, diagnostic case control study		Retrospective cohort study	Prospective cohort study		Prospective cohort study	
Triage Group	ASC-US	ASC-US	LSIL	ASCUS	LSIL	LSIL	LSIL	ASCUS HPV+	LSIL	PAP IIID
Number of pts	63	96	92	361	415	469	264	140	130	222
Patient (sample) collection	63 Sure Path Pap test slides diagnosed with ASCUS with correlative colposcopic biopsy collected over a period of 17 months at Emory University Hospital.	188 SurePath Pap slides diagnosed with ASCUS or LSIL retrieved from department files with histological and/or cytological follow-up.		ASC-US or LSIL LBC samples with corresponding tissue blocks from punch biopsies, cone biopsies and/or ECC within 6 months after cytology from 5 anatomic pathology labs in CH and I. Samples were included as consecutive cases within the disease (CIN2+) and no-disease (CIN1 or negative) groups until the required number of disease cases was reached for each cytologic category. Sample slides for the index test must meet minimum squamous cellularity criteria, not obscured by blood, mucus etc. and not damaged during the procedure to be included.		554 consecutive cervical cytology (Thin Prep LBC) samples diagnosed with LSIL received from Jan 2006 to Feb 2007 at the pathology dept. Of Vejle Hospital. 59 samples excluded due to lack of material, 1 sample could not be found, 25 excluded due to lack of follow-up info.	404 women with an outside referral to colposcopy Pap of ASCUS HPV+ or LSIL. Women 18+ referred to colposcopy clinic Feb 2009 to Aug 2011 for abnormal cervical cancer screening results. Exclusion criteria: previous treatment for cervical disease, prior chemotherapy or radiation therapy, pregnancy, HIV infection, inability to give informed consent.		222 patients with PAP IIID were included consecutively during routine presentation of patients at the doctor's office. 130 patients with cytomorphological categorisation in PAP IIID and low grade dysplasia were included in the study.	
Age of patients (median in years)	NR	30		NR	NR	32.3	26		34.3	33.5
Follow-up (months)	NR	1d to 83 months		up to 6m	up to 6m	up to 60m	1d		up to 36m, median 17.6m	up to 36m, median 8.9m
Outcomes										
CIN2+ Sensitivity	64% (95% CI NR)	91% (95% CI NR)	100% (95% CI NR)	92.2% (95% CI: 83.8-97.1)	94.2% (95% CI: 88.8-97.4)	88.5% (95% CI: 79.9%-94.4%)	86,8% (77.7-92.7)	81.8% (63.9-92.4)	81.58%	92.4% (95% CI: 86.2-96)
Specificity	53% (95% CI NR)	61% (95% CI NR)	43% (95% CI NR)	80.6% (95% CI: 75.6-85.1)	68.0% (95% CI: 62.2-73.4)	51.3% (95% CI: 46.2%-56.4%)	57.6% (49.8-65.0)	62.3% (52.3-71.3)	75.0%	72.8% (95% CI: 63.5-80.5)
PPV	28% (95% CI NR)	NA	NA	NA	NA	29.3% (95% CI: 23.9%-35.2%)	52.0% (43.8-60.1)	40.3% (28.7-53.0)	57.41%	79.7% (95% CI: 72.2-85.6)
NPV	84% (95% CI NR)	NA	NA	NA	NA	95.2% (95% CI: 91.3%-97.7%)	89.2% (81.5-94.0)	91.7% (82.1-96.6)	90.79%	89.3% (95% CI: 80.9-94.3)
False-Positive	23	33	41	55	89	186	NR	NR	NR	NR
True-Positive	9	10	20	71	129	77	NR	NR	NR	NR
False-Negative	5	1	0	6	8	10	NR	NR	NR	NR
True-Negative	26	52	31	229	189	196	NR	NR	NR	NR

4.2 Studienmerkmale

**6 Beobachtungsstudien,
keine RCT,
keine Meta-Analysen,
keine Systematischen
Reviews**

Zur Beantwortung der Fragestellung lagen keine kontrollierten, randomisierten Studien vor.

Insgesamt wurden 6 Beobachtungsstudien identifiziert, darunter zwei prospektive Kohortenstudien [45, 69], drei retrospektive Kohortenstudien [46, 48, 70] sowie eine retrospektive Fall-Kontroll-Studie [47].

Es wurden keine Meta-Analysen, bereits erfolgte Assessments oder Systematische Reviews zur Fragestellung identifiziert.

Länder

**Studien durchgeführt in
USA, D, DK, I**

Drei der Studien wurden in den USA durchgeführt [45, 46, 70], zwei in Deutschland [47, 69], eine in Dänemark [48] und eine in Deutschland und Italien [47].

Population

Mit Ausnahme der Studie von Ziemke *et al.* [69] verwendeten alle Studien das Bethesda System zur Klassifizierung der Studienpopulation.

**520 Frauen mit ASC-US
Befund (3 Studien);
140 Frauen ASC-US und
HPV+ (1 Studie);
1.240 Frauen mit LSIL
Befund (4 Studien);
222 Frauen mit PAP IIID
Befund (1 Studie)**

Drei der analysierten Studien untersuchten die Wirksamkeit des CINtec®-Plus Tests in der Triage von insgesamt 520 Frauen mit einem ASC-US Befund [46, 47, 70], der am ehesten der Kategorie PAP III entspricht, eine weitere Studie untersuchte 140 Frauen mit einem ASC-US Befund bei gleichzeitigem positiven HPV Befund [45].

Vier Studien untersuchten die Triage bei insgesamt 1.240 Frauen mit LSIL Befund [45-48], der, wie in Kapitel 1.3.1 beschrieben, einem PAP IIID Befund nur ungenau entspricht. Die Studie von Ziemke *et al.* [69] ist die einzige Studie, die 222 Frauen mit PAP IIID Befund untersucht. Die Frauen werden allerdings weiter nach PAP IIID mit leichten (130 Frauen) und PAP IIID mit mäßigen Dysplasien (92 Frauen) unterteilt, wobei erstere Gruppe als Annäherung an die LSIL Klassifizierung angesehen werden kann.

Endpunkte

**Sensitivität und Spezifität
des Tests zur Erkennung
höhergradiger Vorstufen
(6 Studien);**

Keine der identifizierten Studien betrachtete klinische Endpunkte wie Auswirkung auf die Inzidenz höhergradige Vorstufen (CIN2/CIN3) oder invasiver Karzinome oder die Auswirkung auf die Inzidenz von Nebenwirkungen und auf die Lebensqualität.

**PPV und NPV (3 Studien);
Keine klinischen
Endpunkte.**

Alle Studien berichteten die Sensitivität und Spezifität des Tests zur Erkennung höhergradiger Vorstufen (Tabelle 4.3-1).

In drei Studien entsprach die Studienpopulation einer Population in Routine-settings [45, 48, 69] und erlaubte daher die Angabe von Positivem und Negativem Prädiktivem Wert.

Referenztest

Alle Studien verwendeten die Kolposkopie-gesteuerte Biopsie als Referenztest: drei der Studien [45, 47, 70] ausschließlich; in den anderen drei Studien [46, 48, 69] wurde alternativ auch die Zytologie als Referenzstandard verwendet.

Referenztest:
Kolposkopie-gesteuerte Biopsie (6 Studien);
alternativ auch Zytologie (3 Studien)

Indextest

In einer Studie wurde der Indextest (p16/Ki-67 Immunzytochemie) sowohl in Dünnschichtpräparaten als auch in konventionellen PAP Abstrichen durchgeführt [69], in den 5 anderen Studien ausschließlich in Dünnschicht-Zytologiepräparaten.

Indextest in Dünnschicht-Zytologiepräparaten (5 Studien);
alternativ auch in konventionellen Abstrichen (1 Studie)

4.3 Studienergebnisse

Tabelle 4.3-1: Übersicht der Studienergebnisse

Studie	Anzahl der Studienteilnehmerinnen	Falsch Positive	Richtig Positive	Falsch Negative	Richtig Negative	Sensitivität	95% CI	Spezifität	95% CI	PPV	95% CI	NPV	95% CI
ASC-US/CIN2+													
Edgerton [70]	63	23	9	5	26	64%	NR	53%	NR	28%	NR	84%	NR
Loghavi [46]	96	33	10	1	52	91%	NR	61%	NR	NA		NA	
Schmidt [47]	361	55	71	6	229	92.2%	83.8-97.1	80.6%	75.6-85.1	NA		NA	
Wentzensen ³ [45]	140	NR	NR	NR	NR	81.8%	63.9-92.4%	62.3%	52.3-71.3%	40.3%	28.7-53.0%	91.7%	82.1-96.6%
LSIL/CIN2+													
Loghavi [46]	92	41	20	0	31	100%	NR	43%	NR	NA		NA	
Schmidt [47]	415	89	129	8	189	94.2%	88.8-97.4	68.0%	62.2-73.4	NA		NA	
Waldstrom [48]	469	186	77	10	196	88.5%	79.9%-94.4%	51.3%	46.2-56.4%	29.3%	23.9-35.2%	95.2%	91.3-97.7%
Wentzensen [45]	264	NR	NR	NR	NR	86.8%	77.7-92.7	57.6%	49.8-65.0	52.0%	43.8-60.1%	89.2%	81.5-94.0%
Ziemke ⁴ [69]	130	NR	NR	NR	NR	81.6%	NR	75.0%	NR	57.4%	NR	90.8%	NR

LSIL – Low Grade Squamous Intraepithelial Lesions; ASC-UC – Atypical Squamous Cells of Unknown Origin; CIN – Cervical Intraepithelial Neoplasia; NR – Not Reported; NA – Not Applicable; CI – Confidence Interval; PPV – Positive Predictive Value; NPV – Negative Predictive Value

³ Nur HPV positive ASC-US

⁴ PAP IIID mit leichten Dysplasien

<p>keine direkte Evidenz zum klinischen Nutzen</p>	<p>Die Studienergebnisse wurden in Tabelle 4.3-1 zusammengefasst. Es gibt keine direkte Evidenz aus RCT zum klinischen Nutzen des p16/Ki-67 Tests. Die vorliegenden Studien können nur hinsichtlich der diagnostischen Genauigkeit des Tests ausgewertet werden.</p>
<p>Sensitivität ASC-US/CIN2+: 64-92,2 % (4 Studien)</p>	<p>Die Sensitivität des p16/Ki-67 zur Erkennung höhergradiger CIN (CIN2+) in Frauen mit einem ASC-US Ausgangsbefund wurde von 4 Studien erhoben und lag innerhalb der Bandbreite 64-92,2 % [45-47, 70]. Von 100 Frauen mit CIN2+ hatten also 64 bis 92 Frauen ein positives Testergebnis; 10-30 jedoch wurden trotz vorliegender Erkrankung negativ getestet.</p>
<p>ASC-US/CIN2+: Spezifität: 53-80,6 % (4 Studien). PPV: 28-40,3 % (2 Studien). NPV: 84-91,7 % (2 Studien)</p>	<p>Die Spezifität für diesen Ausgangsbefund lag innerhalb der Bandbreite 53-80,6 %. Von 100 Frauen ohne CIN oder mit CIN1 hatten also 53-81 Frauen ein negatives p16/Ki-67 Testergebnis; 20-47 Frauen jedoch ohne vorliegende Erkrankung ein positives. Der PPV, also die Wahrscheinlichkeit einer bestehenden Erkrankung bei positivem Testergebnis in dieser Befundgruppe lag zwischen 28-40,3 % [45, 70]. Der NPV, also die Wahrscheinlichkeit, bei einem negativen Testergebnis tatsächlich gesund zu sein, lag in den Studien zwischen 84-91,7 % [45, 70].</p>
<p>LSIL/CIN2+: Sensitivität: 81,6-100 % (5 Studien). Spezifität 43-75 % (5 Studien). PPV: 29,3-57,4 % (3 Studien). NPV: 89,2-95,2 % (3 Studien)</p>	<p>Die Sensitivität des p16/Ki-67 zur Erkennung höhergradiger CIN (CIN2+) in Frauen mit einem LSIL Ausgangsbefund wurde von 5 Studien erhoben und die Effektschätzer reichten von 81,6 % bis 100 % [45-48, 69]. Die Effektschätzer zur Spezifität für diesen Ausgangsbefund lagen innerhalb der Bandbreite 43-75 %. Die Effektschätzer zum PPV in dieser Befundgruppe lagen zwischen 29,3 % und 57,4 %, jene zum NPV zwischen 89,2 % und 95,2 % [45, 48, 69].</p> <p>Um die Bandbreite und Konsistenz der Ergebnisse darzustellen, wurden Forest Plots der Sensitivität und Spezifität in den zwei Befundgruppen erstellt (Abbildung 4.3-1:).</p> <p>Drei von sechs Studien [46, 69, 70] berichteten ihre Ergebnisse ohne Angabe eines Konfidenzintervalls. Es wurde daher auf die Berechnung eines gepoolten Effektschätzers verzichtet.</p> <p>Die Ergebnisse zur Sensitivität (LSIL/CIN2+) weisen eine relative geringe Streuung auf, die einzelnen Effektschätzer haben kleine Konfidenzintervalle und die Konfidenzintervalle (wenn vorhanden) überlappen.</p> <p>Bei allen anderen Kenngrößen finden wir eine große Streuung der Ergebnisse. Bei den Ergebnissen zur Spezifität überlappen die Konfidenzintervalle der einzelnen Studien in beiden Befundgruppen nicht.</p>

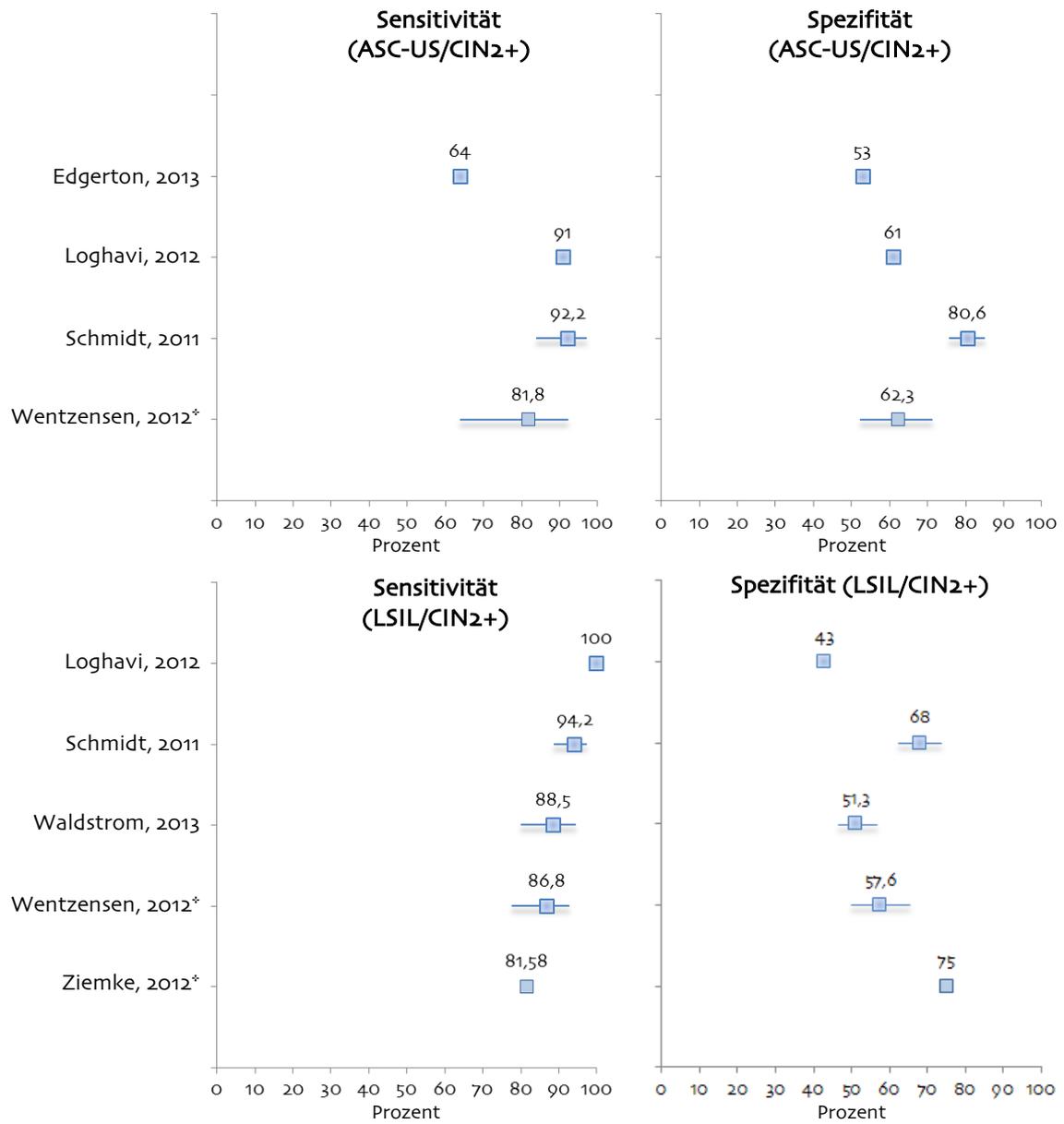


Abbildung 4.3-1: Forest Plots der Studienergebnisse zur Sensitivität und Spezifität des p16/Ki-67 Tests
(Prospektive Studien wurden mit einem * markiert)

4.4 Qualität der Studien

Beurteilung der internen und externen Validität der Studien mittels QUADAS-2 Tool

Zur Beurteilung der internen und externen Validität der Studien wurde das QUADAS-2 (Quality Assessment of Diagnostic Accuracy Studies) Tool mit den in Kapitel 3.1.3 angeführten Kriterien angewendet. Die Beurteilung erfolgte durch zwei WissenschaftlerInnen (AK, IZ), unabhängig voneinander. Differenzen wurden durch Diskussion und Konsens oder die Einbindung einer dritten Person gelöst.

Das vervollständigte QUADAS-2 Tool befindet sich im Anhang.

4.4.1 Interne Validität

Die interne Validität der Studien wurde mittels des in Kapitel 3.1.3 beschriebenen Kriterienkatalogs beurteilt. Die Ergebnisse der internen Validitätsprüfung sind in Tabelle 4.4-1 zusammengefasst.

Die Analyse ergab folgende kritischen Punkte der Studien:

Studienteilnahme

Biasrisiko durch Auswahl der Studienteilnehmer in 5 von 6 Studien

Nur in 2 der 6 Studien bestand die Studienpopulation aus einer konsekutiven Kohorte von Frauen, die im Routinescreening einen auffälligen Befund aufwiesen und prospektiv analysiert wurde [45, 69]. Die Selektionskriterien wurden nur in einer von 6 Studien ausführlich beschrieben [45]. Vier von 6 Studien beschränkten sich retrospektiv auf Fälle mit verfügbarem Biopsiematerial. Dies führt zu einem erheblichen Bias-Risiko in den Studien [46-48, 69, 70]: es ist nicht auszuschließen, dass jene Frauen, die zur Biopsie überwiesen wurden, aufgrund nicht näher beschriebener Symptome, Ko-morbidität, vorangegangener Interventionen oder Tests bereits eine stärkere Verdachtsgrundlage aufwiesen als Frauen mit einem erstmaligen PAP III oder PAP IIID im Routinescreening.

Index Test

Biasrisiko durch nicht-verblindete Auswertung des Indextests (1 Studie)

Während in 4 von 6 Studien der Indextest ohne Kenntnis der Ergebnisse des Referenzstandards durchgeführt wurde, fehlten bei einer Studie dazu Angaben [47]. Alle Studien verwendeten als Schwellenwert für ein positives Testergebnis eine einzelne doppelt gefärbte Zelle.

Referenztest

Biasrisiko durch Verwendung eines inadäquaten Referenzstandards (3 Studien)

In drei von sechs Studien wurde das Follow-up im Routinebetrieb durchgeführt und dabei eine zytologische ODER histologische Abklärung als Referenzstandard verwendet [46, 48, 69]. Die Zytologie ist nicht als Gold Standard zur Identifizierung der CIN geeignet und Bias kann daher bei diesen Studien nicht ausgeschlossen werden. Eine Studie machte keine Angaben, ob die Auswertung des Referenzstandards verblindet gegenüber den Ergebnissen der Ausgangszytologie oder des Indextests erfolgte [45].

Studienprotokoll

Die vorliegenden Studien untersuchten die diagnostische Genauigkeit des p16/Ki-67 Tests. Das bedeutet, dass der p16/Ki-67 in Hinsicht auf seine Eignung zur Identifizierung einer Grunderkrankung an höhergradigen Dysplasien getestet werden soll – ein kurzer Follow-up Zeitrahmen (bis zu 6 Monaten) ist daher zu einer akkuraten Bestimmung der Sensitivität und Spezifität notwendig, um den Einfluss von Weiter- oder Rückdifferenzierung auszuschließen. Ein Follow-up von mehr als 6 Monaten (Verlaufsbeobachtungen) – wie in 4 von 6 Studien [46, 48, 69, 70]- wäre nur in prospektiven Studien adäquat, die die Eignung des Tests zur Risikostratifizierung mit entsprechenden klinischen Endpunkten untersuchen.

In drei Studien wurden nicht alle Frauen mit demselben Referenzstandard diagnostiziert [46, 48, 69]. In zwei Studien wurden einige Frauen gar nicht diagnostiziert, ohne Erklärung für die Austritte aus der Studie [47, 48].

Zwei Studien lieferten nicht genug Daten, um die Vierfeldertafel auszufüllen [45, 69].

Biasrisiko durch inadäquaten Follow-up Zeitrahmen (4 von 6 Studien)

Biasrisiko aufgrund von unvollständiger oder uneinheitlicher Referenzdiagnose

4.4.2 Externe Validität

Die externe Validität der Studien wurde mittels des in 3.1.3 beschriebenen Kriterienkatalogs beurteilt. Die Ergebnisse der Validitätsprüfung sind in Tabelle 4.1-1 tabellarisch zusammengefasst.

Studienteilnahme

Zwei Studien [45, 69] inkludieren eine konsekutive Kohorte von Frauen im Routinesetting. Davon führen nur Wentzensen *et al.* [45] detaillierte Einschluss/Ausschlusskriterien an, die z. B. Frauen mit vorangegangener Behandlung wegen Krankheiten des Gebärmutterhalses von der Studie ausschließen. Keine der Studien lieferte Angaben zur Anamnese der Studienteilnehmerinnen. Dadurch lassen sich in 4 der 6 Studien [46-48, 70] keine und in den zwei anderen nur bedingt Rückschlüsse auf die Anwendbarkeit der Studienergebnisse im Routinescreening zur Triage auffälliger Befunde treffen.

Eine weitere Hürde bei der Interpretation der Ergebnisse stellt die Verwendung unterschiedlicher Klassifizierungssysteme dar. 5 von 6 Studien verwenden Patientinnenpopulationen, die nach Bethesda System in ASC-US oder LSIL eingeteilt wurden. ASC-US entspricht in etwa PAP III, während LSIL PAP IIID Befunden entspricht, die zusätzlich zytomorphologisch auf niedriggradige Läsionen eingestuft wurden, aber auch PAP II Befunde mit HPV Positivität einschließt.

Eine Generalisierbarkeit auf österreichische Settings ist daher in 5 von 6 Studien nicht direkt gegeben.

ungenügende Beschreibung der Einschlusskriterien und Anamnese der Studienteilnehmer (5 von 6 Studien)

Verwendung eines in Ö nicht verwendeten Klassifizierungssystems (5 von 6 Studien)

Indextest

In allen Studien wurde der p16/Ki-67 Test an flüssigkeitsbasierten Dünnschichtpräparaten vorgenommen. Nur in einer Studie [69] wurde der Test auch an konventionellen Zytologiepräparaten durchgeführt. Da die Dünnschichtzytologie nicht routinemäßig im Primärscreening in Österreich eingesetzt wird, ergeben sich daraus bei allen Studien Schwierigkeiten bei der Umsetzbarkeit der Studienergebnisse auf Routinebedingungen.

Verwendung einer in Ö nicht routinemäßig verwendeten Zellabnahmemethode

Referenztest

Alle Studien verwendeten einen Referenztest, der auch im Routinebetrieb eingesetzt wird.

Tabelle 4.4-1: Ergebnisse der Qualitätsprüfung der Einzelstudien

	Biasrisiko				Bedenken bzgl. Externe Validität			Qualität der Studie
	Studien- teilnahme	Index Test	Referenz- standard	Studien- protokoll	Studien- teilnahme	Index Test	Referenz- standard	
Edgerton [70]	Hoch	Niedrig	Niedrig	Unklar	Hoch	Hoch	Niedrig	Niedrig
Loghavi [46]	Hoch	Niedrig	Hoch	Hoch	Hoch	Hoch	Niedrig	Niedrig
Schmidt [47]	Hoch	Unklar	Niedrig	Hoch	Hoch	Hoch	Niedrig	Niedrig
Waldstrom [48]	Hoch	Niedrig	Hoch	Hoch	Hoch	Hoch	Niedrig	Niedrig
Wentzensen [45]	Niedrig	Niedrig	Unklar	Niedrig	Hoch	Hoch	Niedrig	Niedrig
Ziemke [69]	Niedrig	Niedrig	Hoch	Hoch	Niedrig	Hoch	Niedrig	Niedrig

5 Stärke der Evidenz

Zur Beurteilung der Stärke der Evidenz wird ein GRADE Evidenzprofil adaptiert auf Diagnostische Tests verwendet (Table 7.3 in [63]) und für jedes Outcome die Stärke der Evidenz nach GRADE beurteilt:

GRADE Evidenzprofil adaptiert auf Diagnostische Tests

Stärke der Evidenz:

- ✧ *hoch*: Es ist unwahrscheinlich, dass neue Studien einen wichtigen Einfluss auf die Einschätzung des Effektes haben werden
- ✧ *mittel*: Neue Studien werden möglicherweise einen wichtigen Einfluss auf die Einschätzung des Effektes haben
- ✧ *niedrig*: Neue Studien werden sehr wahrscheinlich einen wichtigen Einfluss auf die Einschätzung des Effektes haben
- ✧ *sehr niedrig*: Jegliche Einschätzung des Effektes ist sehr unsicher

Die Anwendung des GRADE-Schemas für die vorliegende Fragestellung ist in der Tabelle 5-1 dargestellt.

Evidenzlage zum klinischen Nutzen nicht vorhanden; zur diagnostischen Genauigkeit sehr niedrig

Insgesamt ist die Evidenzlage a) zum klinischen Nutzen des p16/Ki-67 Tests in der Triage unklarer Befunde **nicht vorhanden**; b) zur diagnostischen Genauigkeit des p16/Ki-67 Tests in der Triage unklarer PAP Befunde ist die Evidenzlage **sehr niedrig**.

Tabelle 5-1: Evidenzprofil zur p16/Ki-67 Immunzytochemie in der Triage von Frauen mit einem PAP III oder PAP IIID Primärbefund im Screening auf Gebärmutterhalskrebs

Outcome	No and type of studies/ No of patients	Findings (Lowest; Highest reported effect size) ⁵	Starting GRADE ⁶	GRADE modifying factors					GRADE of evidence for outcome
				Internal Validity	Consistency of results	Directness	Precision	Other	
Lebensqualität	Keine Evidenz								
Inzidenz höhergradiger CIN	Keine Evidenz								
Nebenwirkungen weiterer Behandlungen	Keine Evidenz								
Screeningteilnahme	Keine Evidenz								
Einfluss auf PatientInnenmanagement	Keine Evidenz								
Sensitivität (ASC-US/CIN2+)	4 DAS ⁷ /660	64%; 92.2%	II-III	-1 ⁸	-1 ⁹	-1 ¹⁰	-1 ¹¹		Very low
Spezifität ASC-US/CIN2+	4 DAS ⁷ /660	53%; 80.6%	II-III	-1 ⁸	-1 ⁹	-1 ¹⁰	-1 ¹¹		Very low
Positiver Prädiktiver Wert (ASC-US/CIN2+)	2 DAS ⁷ /203	28%; 40.3%	II-III	-1 ⁸	-1 ⁹	-1 ¹⁰	-1 ¹¹		Very low
Negativer Prädiktiver Wert (ASC-US/CIN2+)	2 DAS ⁷ /203	84%; 91.7%	II-III	-1 ⁸	0	-1 ¹⁰	-1 ¹¹		Very low
Sensitivität LSIL/CIN2+	5 DAS ⁷ /1,370	81.6%; 100%	II-III	-1 ⁸	0	-1 ¹⁰	-1 ¹¹		Very low
Spezifität LSIL/CIN2+	5 DAS ⁷ /1,370	43%; 75%	II-III	-1 ⁸	-1 ⁹	-1 ¹⁰	-1 ¹¹		Very low
Positiver Prädiktiver Wert (LSIL/CIN2+)	3 DAS ⁷ /863	29.3%; 57.4%	II-III	-1 ⁸	-1 ⁹	-1 ¹⁰	-1 ¹¹		Very low
Negativer Prädiktiver Wert (LSIL/CIN2+)	3 DAS ⁷ /863	89.2%; 95.2%	II-III	-1 ⁸	0	-1 ¹⁰	-1 ¹¹		Very low

⁵ Berechnung des gepoolten Effektschätzers aufgrund fehlender Konfidenzintervalle nicht möglich

⁶ Basierend auf Tabelle 3.1-2:

⁷ DAS – Diagnostic Accuracy Studies /Studien zur diagnostischen Genauigkeit

⁸ Interne Validität: Biasrisiko entsprechend der Qualitätsanalyse der einzelnen Studien (Kapitel 4.4.1)

⁹ Hohe Bandbreite der Effektschätzer aus den einzelnen Studien, die sich nicht aufgrund von Variabilität im Studiendesign erklären lassen

¹⁰ Kein direkter Vergleich mit Zytologie alleine oder Zytologie/Kolposkopie.

¹¹ Fehlende Konfidenzintervalle in zwei Studien

6 Diskussion

Die p16/Ki-67 Immunzytochemie (CINtec®Plus) wird in Österreich zur Differentialdiagnostik von Befunden der Gruppen PAP III und PAP IIID bereits unsystematisch eingesetzt. Die vorliegende systematische Übersichtsarbeit hat die Evidenzgrundlage zum klinischen Nutzen des Tests zur Triage dieser Befundklassen untersucht. Es wurden keine systematischen Reviews oder Primärstudien identifiziert, die den klinischen Nutzen für patientinnenrelevante Endpunkte, wie die Reduktion der Inzidenz höhergradiger CIN, Mortalität oder Lebensqualität, untersuchen. Es konnten auch keine Studien identifiziert werden, die Auswirkungen auf das Patientinnenmanagement oder die Screeningteilnahme untersuchen. Es sind hierzu auch keine laufenden Studien registriert.

Bei den im vorliegenden Bericht eingeschlossenen Studien handelt es sich um 6 Studien zur diagnostischen Genauigkeit, darunter 2 prospektive Kohortenstudien [45, 69], drei retrospektive Kohortenstudien [46, 48, 70], sowie eine retrospektive Fall-Kontroll-Studie [47] mit insgesamt 2.252 Studienteilnehmerinnen. Die Fall-Kontroll-Studie wurde vom Hersteller des Tests gesponsert, zwei weitere gaben potentielle Interessenskonflikte über die kostenfreie Bereitstellung der Tests durch den Hersteller an [45, 46].

Die Ergebnisse der Studien zeigen eine Sensitivität von 64-92,2 % in der Befundgruppe ASC-US, sowie 81,6-100 % in der Befundgruppe LSIL. Die Spezifität ist wesentlich niedriger und liegt bei 53-80,6 % in der Befundgruppe ASC-US und 43-75 % in der Befundgruppe LSIL.

Bis auf die Studie von Wentzensen *et al.* [45], wiesen alle Studien erhebliche methodische Mängel auf. Diese beinhalteten insbesondere fehlende Beschreibung der Studienteilnehmerinnen, mangelnde Verblindung von Index- bzw. Referenzstandardergebnissen, die Verwendung inadäquater Referenzstandards, sowie inadäquate Zeitabstände zwischen der Durchführung des p16/Ki-67 Tests und dem Referenzstandard. Insgesamt waren die Ergebnisse der einzelnen Studien (mit Ausnahme der Sensitivität in der LSIL Befundgruppe) sehr heterogen. Die Evidenzgrundlage für die Endpunkte zur diagnostischen Genauigkeit Sensitivität und Spezifität ist entsprechend niedrig.

Für eine hohe Evidenzgrundlage (Level I in Tabelle 3.1-2) werden für den klinischen Nutzen Daten aus RCTs oder für die diagnostische Genauigkeit ein systematischer Review von Level II Studien zur diagnostischen Genauigkeit gefordert.

Methodisch hochwertige prospektive diagnostische Studien sind daher zunächst notwendig, um die Sensitivität und Spezifität des Dual Stains gesichert zu belegen. Der Hersteller wies auf eine bereits abgeschlossene (Datum des Studienabschlusses: 13.01.2010 [71]), jedoch noch nicht publizierte Studie hin: PALMS (Primary, ASC-US, LSIL Marker Study) ist eine multinationale, multizentrische, prospektive, diagnostische Studie zur Evaluierung des p16/Ki-67 Dual Stains an >27.000 Frauen im Screening. Die Ergebnisse werden im Lauf des Jahres 2013 erwartet.

keine Evidenzgrundlage für den klinischen Nutzen anhand von Primärstudien oder Systematischen Reviews mit patientinnenrelevanten Endpunkten

6 Studien zur Diagnostischen Genauigkeit mit insgesamt 2.252 Studienteilnehmerinnen

5 von 6 Studien mit erheblichen methodischen Mängeln: niedrige Evidenzgrundlage

Bedarf an methodisch hochwertigen Studien zur Diagnostischen Genauigkeit

auch bei niedriger Spezifität ein klinischer Nutzen durch Reduktion der Überweisungsrate zur Biopsie erwartbar

Der potentielle klinische Nutzen der p16/Ki-67 Dual Stain Immunzytochemie als Triagetest ergibt sich aus dem Vergleich der klinischen Auswirkungen im klinischen Pfad (Abbildung 2-1 und Abbildung 2-2). Nach den derzeit geltenden Managementrichtlinien werden in Österreich uniform alle Frauen mit einem PAP III oder PAP IIID Befund zur Kolposkopie überwiesen, in der Befundgruppe PAP III ist zudem in jedem Fall eine histologische Abklärung nach ECC oder Biopsie durchzuführen. Trotz der relativ niedrigen Spezifität des p16/Ki-67 Tests wäre also ein klinischer Nutzen durch Reduktion der Falsch-Positiven Befunde zu erwarten: weniger Frauen ohne Krebsvorstufen müssten sich den (invasiveren) diagnostischen Verfahren unterziehen.

Kein größerer Nutzen oder Schaden ist in der Gruppe der Richtig-Positiven zu erwarten. Es ist zwar nicht völlig auszuschließen, dass p16/Ki-67 positive CIN ein spezifisches Spektrum der CIN darstellen, beispielsweise aufgrund unterschiedlicher Expressionslevel in verschiedenen HPV Typen. Ein Einfluss des p16/Ki-67 Testergebnisses auf das Ansprechen der CIN auf Behandlung ist dennoch nicht zu erwarten – eine Behandlung erfolgt in jedem Fall nur auf der Basis einer nachgeschalteten diagnostischen Abklärung durch den Goldstandard Kolposkopie-gesteuerte Biopsie.

potentieller Schaden in Falsch-Negativen Gruppe über klinische Studien zu untersuchen

Der Schaden in der Gruppe der Falsch-Negativen (Auswirkungen auf das Risiko höhergradiger CIN), aber auch der Richtig-Negativen (kein Ausschluß höhergradiger CIN mittels Referenzstandard) ist nicht direkt aus der diagnostischen Genauigkeit ableitbar und müsste über weiterführende klinische Studien mit entsprechenden Endpunkten nachgewiesen werden.

Nur prospektive klinische Studien können die Überlegenheit bzw. Gleichwertigkeit des Patientinnenmanagements mit p16/Ki-67 Triage zur bestehenden Managementstrategie hinsichtlich der Inzidenz behandlungsbedürftiger Diagnosen belegen. Die Akzeptanz des Tests bei den betroffenen Frauen und eine tatsächliche Verminderung psychischer Belastung durch den auffälligen PAP Befund müsste ebenfalls nachgewiesen werden.

Schwierigkeiten bei der Umsetzung der Ergebnisse auf österreichische Verhältnisse

Schwierigkeiten ergeben sich bei der Umsetzung der Ergebnisse internationaler Studien auf österreichische Verhältnisse. Zum einen werden die meisten Ergebnisse in Dünnschichtzytologie-Präparaten gewonnen, während in Österreich im Routinebetrieb konventionelle PAP Abstriche gewonnen werden. Inwieweit die beiden Zytologiemethoden die Performance des Tests beeinflussen, bleibt abzuklären.

Zum zweiten evaluieren die meisten Studien die Performance des Tests in Bethesda PAP Klassen, die nur bedingt mit den Klassen nach München II System übereinstimmen (siehe Kapitel 1.3.1). Die Performance des Tests müsste daher über Modellierungsverfahren auf München II Patientinnengruppen umgelegt werden und/oder aber die Indikation über zusätzliche zytologische Kriterien differenziert werden.

Triagetests in Kontext mit anderen aufkommenden Technologien zu setzen

Aus gesundheitspolitischer Sicht ist für den Einsatz eines Triagetests in bevölkerungsbezogenen relevanter Krebsfrüherkennungsuntersuchungen eine hohe Evidenzgrundlage notwendig: die evidenzbasierte Überprüfung der Testanwendung sollte im Vergleich mit anderen Screeningstrategien unter Berücksichtigung anderer aufkommender Technologien zum HPV Screening oder zur HPV Infektionsprävention durchgeführt werden. Insbesondere sind hier die HPV Impfung und der HPV Test anzuführen. Die beiden HPV Impfstoffe (gegen die HPV Typen 16 und 18, bzw. zusätzlich HPV 6 und HPV 11) werden in Österreich ab Februar 2014 als Teil des Gratskinderimpfprogramms für Buben und Mädchen in der 4. Klasse Volksschule angeboten

werden [72]. Mit zunehmender Durchimpfung der Bevölkerung und sinkender Prävalenz von HPV Infektionen und Persistenz ist es denkbar, dass die gegenwärtigen (und neuen) Screeningmethoden an Effizienz verlieren.

Aufgrund der deutlich höheren Spezifität des HPV Tests im Vergleich zum PAP Abstrich haben bereits mehrere HTA Institute Empfehlungen für den Einsatz des HPV Tests zusätzlich zur Zytologie als primären oder sekundären Screeningmarker ausgesprochen [36, 37, 73]. In Österreich ist der Test aufgrund der bestehenden Kassenleistung Kolposkopie in den Managementrichtlinien nur zur Abklärung negativer Biopsien bei auffälligem Befund vorgesehen. Sowohl HPV Tests und HPV Impfung sollte in einer vergleichenden Analyse zur Identifikation optimaler Screeningalgorithmen Berücksichtigung finden (siehe zum Beispiel [74]).

Die Screeningteilnahme beträgt in Österreich nur etwa 60 % [4, 29]. Qualitätssichernde Maßnahmen des Screenings wurden initiiert [13, 14], die Zielvorgaben jedoch noch nicht erreicht. Neben der angemessenen Einführung neuer effektiverer Screening-Algorithmen ist klar, dass die wesentlichere Erhöhung der Bevölkerungsgesundheit durch die Verbesserung der Teilnahmeraten, sowie durch die Qualitätssicherung der einzelnen Screeningmethoden innerhalb einer Gesamtstrategie zu erzielen ist.

**vergleichende Analyse
neuer Technologien zur
Identifikation optimaler
Screeningalgorithmen**

**andere Maßnahmen:
Erhöhung der
Teilnahmeraten und
Qualitätssicherung
des Screenings**

7 Empfehlung

Aufgrund der Evidenzlage ist der Einsatz der p16/Ki-67 Dual Stain Immunzytologie in der Triage unklarer PAP Befunde (PAP III und PAP IIID) in Österreich **nicht zu empfehlen**.

aufgrund niedriger Evidenzlage keine Empfehlung

8 Literatur

- [1] Leitlinie für die Diagnose und Therapie von Cervikalen Intraepithelialen Neoplasien (CIN) und Mikrokarzinomen der Cervix uteri. *Speculum – Zeitschrift für Gynäkologie und Geburtshilfe*. 2005;23(3):20-5.
- [2] Schiffman M, Castle PE, Jeronimo J, Rodriguez AC, Wacholder S. Human papillomavirus and cervical cancer. *Lancet*. 2007;370(9590):890-907. Epub 2007/09/11.
- [3] U P. VORSORGE NEU – Internationale wissenschaftliche Grundlagen zum Programm der Österreichischen Vorsorgeuntersuchungen. 2005.
- [4] Health in the European Union European Commission Special Eurobarometer 272e/Wave 662 2007.
- [5] Statistik Austria, Österreichisches Krebsregister. Stand 24.09.2012. Erstellt am: 05.10.2012. Jeweils auf 100.000 Personen/Männer/Frauen, Standardbevölkerung = WHO-Weltbevölkerung, 2001.
- [6] PreHdict. Human Papillomavirus and Related Diseases. Summary Report June 3, 2013 Austria.
- [7] Peirson L, Fitzpatrick-Lewis D, Ciliska D, Warren R. Screening for cervical cancer: a systematic review and meta-analysis. *Systematic reviews*. 2013;2:35. Epub 2013/05/28.
- [8] Arbyn M, Bergeron C, Klinkhamer P, Martin-Hirsch P, Siebers AG, Bulten J. Liquid compared with conventional cervical cytology: a systematic review and meta-analysis. *Obstetrics and gynecology*. 2008;111(1):167-77. Epub 2008/01/01.
- [9] Schneider V. Gynäkologische Krebsvorsorge in Deutschland. *Der Pathologe*. 2012;33(4):286-92.
- [10] Marquardt K, Buttner HH, Broschewitz U, Barten M, Schneider V. Persistent carcinoma in cervical cancer screening: non-participation is the most significant cause. *Acta Cytol*. 2011;55(5):433-7. Epub 2011/10/12.
- [11] Nanda K, McCrory DC, Myers ER, Bastian LA, Hasselblad V, Hickey JD, et al. Accuracy of the Papanicolaou test in screening for and follow-up of cervical cytologic abnormalities: a systematic review. *Ann Intern Med*. 2000;132(10):810-9. Epub 2000/05/20.
- [12] Schröder J, Wilfing C, Regitnig P. Qualitätsoffensive PAP-Abstrich: Zwei Projekte – ein gemeinsamer Erfolg Soziale Sicherheit. 2008.
- [13] European Commission, Directorate-General for Health and Consumers. European Guidelines for Quality Assurance in Cervical Cancer Screening. 2008.
- [14] ARGE Qualitätssicherung der Österreichischen Gesellschaft für Zytologie und der Österreichischen Gesellschaft für Pathologie. Leitlinie zur Qualität in der gynäkologischen Zytologie. 2010.
- [15] Solomon D. Chapter 14: Role of triage testing in cervical cancer screening. *Journal of the National Cancer Institute Monographs*. 2003(31):97-101. Epub 2003/06/17.
- [16] Herzog TJ, Wright JD. The impact of cervical cancer on quality of life--the components and means for management. *Gynecol Oncol*. 2007;107(3):572-7. Epub 2007/10/30.
- [17] Berrington A, Green J. Comparison of risk factors for invasive squamous cell carcinoma and adenocarcinoma of the cervix: collaborative reanalysis of individual data on 8,097 women with squamous cell carcinoma and 1,374 women with adenocarcinoma from 12 epidemiological studies. *Int J Cancer*. 2007;120(4):885-91. Epub 2006/11/30.
- [18] Lax S. Histopathology of cervical precursor lesions and cancer. *Acta dermatovenerolog*. 2011;20(3):125-33.
- [19] WHO histological classification of tumours of the uterine cervix. WHO International Agency for Research on Cancer – IARC Screening Group; 2003 [cited 2013]; Available from: <http://screening.iarc.fr/atlasclassifwho.php?lang=1>.
- [20] Zielonke N. Krebsinzidenz und Krebsmortalität in Österreich. 2012.

- [21] Walboomers JM, Jacobs MV, Manos MM, Bosch FX, Kummer JA, Shah KV, et al. Human papillomavirus is a necessary cause of invasive cervical cancer worldwide. *J Pathol.* 1999;189(1):12-9. Epub 1999/08/19.
- [22] Stanley MA. Immune responses to human papilloma viruses. *The Indian journal of medical research.* 2009;130(3):266-76. Epub 2009/11/11.
- [23] Doorbar J. Molecular biology of human papillomavirus infection and cervical cancer. *Clin Sci (Lond).* 2006;110(5):525-41. Epub 2006/04/07.
- [24] De Vuyst H, Clifford G, Li N, Franceschi S. HPV infection in Europe. *Eur J Cancer.* 2009;45(15):2632-9. Epub 2009/08/28.
- [25] Doorbar J, Quint W, Banks L, Bravo IG, Stoler M, Broker TR, et al. The biology and life-cycle of human papillomaviruses. *Vaccine.* 2012;30 Suppl 5:F55-70. Epub 2012/12/05.
- [26] Pretorius RG, Peterson P, Azizi F, Burchette RJ. Subsequent risk and presentation of cervical intraepithelial neoplasia (CIN) 3 or cancer after a colposcopic diagnosis of CIN 1 or less. *Am J Obstet Gynecol.* 2006;195(5):1260-5. Epub 2006/11/01.
- [27] Ostor AG. Natural history of cervical intraepithelial neoplasia: a critical review. *Int J Gynecol Pathol.* 1993;12(2):186-92. Epub 1993/04/01.
- [28] Brown CA, Bogers J, Sahebali S, Depuydt CE, De Prins F, Malinowski DP. Role of protein biomarkers in the detection of high-grade disease in cervical cancer screening programs. *J.* 2012;2012:289315.
- [29] Reiner-Concin A. Der Pap-Abstrich. *Österreichische Ärztezeitung.* 2011;7.
- [30] Ziemke P. Der prädiktive Wert der Gruppe IIID und eines zusätzlichen HPV-Tests. *Geburtshilfe und Frauenheilkunde.* 2012.
- [31] Whitlock EP, Vesco KK, Eder M, Lin JS, Senger CA, Burda BU. Liquid-based cytology and human papillomavirus testing to screen for cervical cancer: a systematic review for the U.S. Preventive Services Task Force. *Ann Intern Med.* 2011;155(10):687-97, W214-5. Epub 2011/10/19.
- [32] Hoda RS, Loukeris K, Abdul-Karim FW. Gynecologic cytology on conventional and liquid-based preparations: a comprehensive review of similarities and differences. *Diagn Cytopathol.* 2013;41(3):257-78. Epub 2012/04/18.
- [33] Ronco G, Confortini M, Maccallini V, Naldoni C, Segnan N, Sideri M, et al. [Health technology assessment report. Use of liquid-based cytology for cervical cancer precursors screening]. *Epidemiol Prev.* 2012;36(5 Suppl 2):e1-e33. Epub 2012/11/21. Health Technology assessment Report. Uso della citologia in fase liquida nello screening dei precursori del cancro del collo uterino.
- [34] Dalla Palma P, Moresco L, Giorgi Rossi P. Health technology assessment of computer-assisted pap test screening in Italy. *Acta Cytol.* 2013;57(4):349-58. Epub 2013/07/19.
- [35] Empfehlungen der AG-CPC zur Kolposkopie bei Erkrankungen des unteren Genitaltraktes der Frau. Deutsche Gesellschaft für Gynäkologie und Geburtshilfe Juli 2008.
- [36] Mittendorf; T, Nocon; M, Roll; S, Mühlberger; N, Sroczynski; G, Siebert; U, et al. HPV-DNA-Diagnostik zur Zervixkarzinomfrüherkennung. *Schriftenreihe Health Technology Assessment (HTA) in der Bundesrepublik Deutschland* 2007.
- [37] Vesco K, Whitlochh E, Eder M. Screening for cervical cancer: a systematic evidence review for the U.S. preventive service task force. Evidence synthesis No 86 AHRQ Publication No11-015156-EF-1. 2011.
- [38] Grimm C. Die Bedeutung der typenspezifischen HPV-Typisierung. *Speculum – Zeitschrift für Gynäkologie und Geburtshilfe (Ausgabe für Österreich).* 2012;30(1):6-9.
- [39] Dalla Palma P, Giorgi Rossi P, Collina G, Buccoliero AM, Ghiringhello B, Lestani M, et al. The risk of false-positive histology according to the reason for colposcopy referral in cervical cancer screening: a blind revision of all histologic lesions found in the NTCC trial. *Am J Clin Pathol.* 2008;129(1):75-80. Epub 2007/12/20.

- [40] Zechmeister; I, Blasio; BFD, Radlberger; P, Wild; C, Kvas; E, Garnett; G, et al. Ökonomische Evaluation der Impfung gegen humane Papillomaviren (HPV-Impfung) in Österreich. HTA Projektbericht 2007;2007(9).
- [41] Sahasrabudde VV, Luhn P, Wentzensen N. Human papillomavirus and cervical cancer: biomarkers for improved prevention efforts. *Future Microbiol.* 2011;6(9):1083-98. Epub 2011/10/01.
- [42] Dona MG, Vocaturo A, Giuliani M, Ronchetti L, Rollo F, Pescarmona E, et al. p16/Ki-67 dual staining in cervico-vaginal cytology: correlation with histology, Human Papillomavirus detection and genotyping in women undergoing colposcopy. *Gynecol Oncol.* 2012;126(2):198-202.
- [43] Singh M, Mockler D, Akalin A, Burke S, Shroyer AL, Shroyer KR. Immunocytochemical colocalization of P16(INK4a) and Ki-67 predicts CIN2/3 and AIS/adenocarcinoma. *Cancer cytopathol.* 2012;120(1):26-34.
- [44] Yoshida T, Sano T, Kanuma T, Inoue H, Itoh T, Yazaki C, et al. Usefulness of CINtec PLUS p16/Ki-67 double-staining in cytological screening of cervical cancer. *Acta Cytol.* 2011;55(5):413-20.
- [45] Wentzensen N, Schwartz L, Zuna RE, Smith K, Mathews C, Gold MA, et al. Performance of p16/Ki-67 immunostaining to detect cervical cancer precursors in a colposcopy referral population. *Clin Cancer Res.* 2012;18(15):4154-62.
- [46] Loghavi S, Walts AE, Bose S. CINtec(R) plus dual immunostain: A triage tool for cervical pap smears with atypical squamous cells of undetermined significance and low grade squamous intraepithelial lesion. *Diagn Cytopathol.* 2012. Epub 2012/07/27.
- [47] Schmidt D, Bergeron C, Denton KJ, Ridder R, European CCSG. p16/ki-67 dual-stain cytology in the triage of ASCUS and LSIL papanicolaou cytology: results from the European equivocal or mildly abnormal Papanicolaou cytology study. *Cancer cytopathol.* 2011;119(3):158-66.
- [48] Waldstrom M, Christensen RK, Ornskov D. Evaluation of p16(INK4a)/Ki-67 dual stain in comparison with an mRNA human papillomavirus test on liquid-based cytology samples with low-grade squamous intraepithelial lesion. *Cancer cytopathol.* 2013;121(3):136-45.
- [49] Petry KU, Schmidt D, Scherbring S, Luyten A, Reinecke-Luthge A, Bergeron C, et al. Triage Pap cytology negative, HPV positive cervical cancer screening results with p16/Ki-67 Dual-stained cytology. *Gynecol Oncol.* 2011;121(3):505-9.
- [50] Ravarino A, Nemolato S, Macciocu E, Frascini M, Senes G, Faa G, et al. CINtec PLUS immunocytochemistry as a tool for the cytologic diagnosis of glandular lesions of the cervix uteri. *Am J Clin Pathol.* 2012;138(5):652-6. Epub 2012/10/23.
- [51] Schunemann HJ, Oxman AD, Brozek J, Glasziou P, Jaeschke R, Vist GE, et al. Grading quality of evidence and strength of recommendations for diagnostic tests and strategies. *Bmj.* 2008;336(7653):1106-10. Epub 2008/05/17.
- [52] Atkins D, Best D, Briss PA, Eccles M, Falck-Ytter Y, Flottorp S, et al. Grading quality of evidence and strength of recommendations. *Bmj.* 2004;328(7454):1490. Epub 2004/06/19.
- [53] Nachtnebel A. Evaluation von Diagnostika – Hintergrund, Probleme, Methoden. HTA Projektbericht Nr 36 2010.
- [54] Derksen J. Medical tests (assessment of established medical science and medical practice). 2011;CVZ Report 293.
- [55] Medical Services Advisory Committee. Guidelines for the assessment of diagnostic technologies. 2005.
- [56] National Institute for Health and Clinical Excellence. Diagnostics Assessment Programme Manual. 2011.
- [57] Canadian Agency for Drugs and Technologies in Health. Evaluation Framework for Genetic Tests. 2012(37).
- [58] Febbo PG, Ladanyi M, Aldape KD, De Marzo AM, Hammond ME, Hayes DF, et al. NCCN Task Force report: Evaluating the clinical utility of tumor markers in oncology. *Journal of the National Comprehensive Cancer Network: JNCCN.* 2011;9 Suppl 5:S1-32; quiz S3. Epub 2011/12/22.
- [59] Bossuyt PM, Reitsma JB, Linnet K, Moons KG. Beyond diagnostic accuracy: the clinical utility of diagnostic tests. *Clinical chemistry.* 2012;58(12):1636-43. Epub 2012/06/26.

- [60] Janatzek S. [The benefit of diagnostic tests--from surrogate endpoints to patient-relevant endpoints]. *Zeitschrift für Evidenz, Fortbildung und Qualität im Gesundheitswesen*. 2011;105(7):504-9. Epub 2011/10/01. Nutzen diagnostischer Tests – vom Surrogat zur Patientenrelevanz.
- [61] Lord SJ, Irwig L, Simes RJ. When is measuring sensitivity and specificity sufficient to evaluate a diagnostic test, and when do we need randomized trials? *Ann Intern Med*. 2006;144(11):850-5. Epub 2006/06/07.
- [62] Staub LP, Dyer S, Lord SJ, Simes RJ. Linking the evidence: intermediate outcomes in medical test assessments. *International journal of technology assessment in health care*. 2012;28(1):52-8. Epub 2012/01/11.
- [63] AHRQ. *Methods Guide for Medical Test Reviews*. AHRQ Publication No 12-EHC017. 2012.
- [64] Schunemann HJ, Oxman AD, Brozek J, Glasziou P, Bossuyt P, Chang S, et al. GRADE: assessing the quality of evidence for diagnostic recommendations. *ACP journal club*. 2008;149(6):2. Epub 2008/12/17.
- [65] Staub LP, Lord SJ, Simes RJ, Dyer S, Houssami N, Chen RY, et al. Using patient management as a surrogate for patient health outcomes in diagnostic test evaluation. *BMC medical research methodology*. 2012;12:12. Epub 2012/02/16.
- [66] Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med*. 2011;155(8):529-36. Epub 2011/10/19.
- [67] Schunemann HJ, Oxman AD, Brozek J, Glasziou P, Bossuyt P, Chang S, et al. GRADE: assessing the quality of evidence for diagnostic recommendations. *Evidence-based medicine*. 2008;13(6):162-3. Epub 2008/12/02.
- [68] OCEBM Levels of Evidence Working Group*. „The Oxford 2011 Levels of Evidence“. Oxford Centre for Evidence-Based Medicine. <http://www.cebm.net/index.aspx?o=5653>.
* OCEBM Table of Evidence Working Group = Jeremy Howick, Iain Chalmers (James Lind Library), Paul Glasziou, Trish Greenhalgh, Carl Heneghan, Alessandro Liberati, Ivan Moschetti, Bob Phillips, Hazel Thornton, Olive Goddard and Mary Hodgkinson.
- [69] Ziemke P, Marquardt K. [Immunocytochemistry of p16(INK4a) and Ki-67 as adjunctive method for routine gynecological cytology of mild and moderate dysplasia.]. *Pathologe*. 2012. Epub 2012/07/04. p16(INK4a)- und Ki-67-Immunzytochemie als Zusatzmethode für die Gruppe IIID der gynäkologischen Zytologie.
- [70] Edgerton N, Cohen C, Siddiqui MT. Evaluation of CINtec PLUS testing as an adjunctive test in ASC-US diagnosed SurePath preparations. *Diagn Cytopathol*. 2013;41(1):35-40.
- [71] Denton KJ, Bergeron C, Klement P, Trunk MJ, Ridder R. P16 immuno-cytochemistry for the triage of ASC-US and LSIL pap cytology cases – Results from a large retrospective study. *Laboratory Investigation*. 2009;89:211A-2A.
- [72] Bundesministerium für Gesundheit. *Impfplan*. 2013.
- [73] KCE Belgium Health Care Knowledge Center. *Cervical Cancer Screening and Human Papillomavirus (HPV) Testing KCE reports vol38C*. 2006.
- [74] Goldhaber-Fiebert JD, Stout NK, Salomon JA, Kuntz KM, Goldie SJ. Cost-effectiveness of cervical cancer screening with human papillomavirus DNA testing and HPV-16,18 vaccination. *Journal of the National Cancer Institute*. 2008;100(5):308-20. Epub 2008/03/04.

Anhänge

Suchstrategie

Medline (via Ovid)

Database: Ovid MEDLINE(R) <1946 to April Week 4 2013>, Ovid MEDLINE(R) In-Process & Other Non-Indexed Citations <May 07, 2013>, Ovid MEDLINE(R) Daily Update <May 07, 2013>, Ovid OLDMEDLINE(R) <1946 to 1965>		
Search Strategy		
1	PAP.mp.	14,424
2	Papanicolaou.mp.	5,086
3	cytological screening.mp.	418
4	(p16* adj15 cytology).mp.	109
5	*Vaginal Smears/	9,227
6	1 or 2 or 3 or 4 or 5	24,354
7	*Papillomaviridae/	13,137
8	Human papillomavirus*.mp.	22,829
9	HPV*.mp.	24,590
10	*Carcinoma, Squamous Cell/	71,412
11	atypical squamous cells of undetermined significance.mp.	940
12	ASCUS*.mp.	1,156
13	Low-grade squamous intraepithelial lesion*.mp.	1,188
14	LSIL*.mp.	1,094
15	(Cervi* adj1 (Cancer* or Carcinoma*)).mp.	35,469
16	*Uterine Cervical Neoplasms/	46,190
17	*Cervical Intraepithelial Neoplasia/	5,362
18	7 or 8 or 9 or 10 or 11 or 12 or 13 or 14 or 15 or 16 or 17	139,489
19	6 and 18	8,899
20	Dual stain*.mp.	558
21	*Triage/	3,802
22	triage test*.mp.	87
23	*Cyclin-Dependent Kinase Inhibitor p16/	2,625
24	*Genes, p16/	1,045
25	INK4*.mp.	3568
26	(p16 adj1 ^{INK4a}).mp.	2,286
27	*Ki-67 Antigen/	2,305
28	Ki67.mp.	5,361
29	(p16 adj5 Ki-67).mp.	261
30	CinTec.mp.	25
31	20 or 21 or 22 or 23 or 24 or 25 or 26 or 27 or 28 or 29 or 30	17,303
32	19 and 31	189

08.05.2013

Embase

Embase Session Results		
No. Query Results	Results	Date
#37. pap OR 'papanicolaou test'/mj OR 'cytological screening' OR 'vagina smear'/mj AND ('wart virus'/mj OR 'human papillomavirus' OR hpv* OR 'squamous cell carcinoma'/mj OR 'atypical squamous cells of undetermined significance' OR ascus* OR 'low-grade squamous intraepithelial lesion' OR 'low-grade squamous intraepithelial lesions' OR lsil* OR cervi* NEAR/1 (cancer* OR carcinoma*) OR 'uterine cervix tumor'/mj OR 'uterine cervix carcinoma in situ'/mj OR 'cervical intraepithelial neoplasia') AND ('dual stain' OR 'dual staining' OR triage OR 'cyclin dependent kinase inhibitor 2a'/mj OR ('tumor suppressor gene'/exp AND p16*) OR ^{INK4a*} OR p16 NEAR/1 ink4* OR 'ki 67 antigen'/mj OR p16 NEAR/3 'ki-67' OR p16 ^{INK4a} OR cintec:dn OR cintec) AND ('clinical article'/de OR 'clinical study'/de OR 'clinical trial'/de OR 'cohort analysis'/de OR 'comparative study'/de OR 'controlled clinical trial'/de OR 'controlled study'/de OR 'cross-sectional study'/de OR 'diagnostic test accuracy study'/de OR 'major clinical study'/de OR 'multicenter study'/de OR 'pilot study'/de OR 'prospective study'/de OR 'randomized controlled trial'/de OR 'retrospective study'/de)	227	8 May 2013
#36. pap OR 'papanicolaou test'/mj OR 'cytological screening' OR 'vagina smear'/mj AND ('wart virus'/mj OR 'human papillomavirus' OR hpv* OR 'squamous cell carcinoma'/mj OR 'atypical squamous cells of undetermined significance' OR ascus* OR 'low-grade squamous intraepithelial lesion' OR 'low-grade squamous intraepithelial lesions' OR lsil* OR cervi* NEAR/1 (cancer* OR carcinoma*) OR 'uterine cervix tumor'/mj OR 'uterine cervix carcinoma in situ'/mj OR 'cervical intraepithelial neoplasia') AND ('dual stain' OR 'dual staining' OR triage OR 'cyclin dependent kinase inhibitor 2a'/mj OR ('tumor suppressor gene'/exp AND p16*) OR ^{INK4a*} OR p16 NEAR/1 ink4* OR 'ki 67 antigen'/mj OR p16 NEAR/3 'ki-67' OR p16 ^{INK4a} OR cintec:dn OR cintec)	420	8 May 2013
#35. 'dual stain' OR 'dual staining' OR triage OR 'cyclin dependent kinase inhibitor 2a'/mj OR ('tumor suppressor gene'/exp AND p16*) OR ^{INK4a*} OR p16 NEAR/1 ink4* OR 'ki 67 antigen'/mj OR p16 NEAR/3 'ki-67' OR p16 ^{INK4a} OR cintec:dn OR cintec	24,042	8 May 2013
#34. cintec	76	8 May 2013
#33. cintec:dn	12	8 May 2013
#32. p16 ^{INK4a}	1,383	8 May 2013
#31. p16 NEAR/3 'ki-67'	373	8 May 2013
#30. 'ki 67 antigen'/mj	3,200	8 May 2013
#29. p16 NEAR/1 ink4*	3,872	8 May 2013
#28. ^{INK4a*}	4,269	8 May 2013
#27. 'tumor suppressor gene'/exp AND p16*	2,993	8 May 2013
#26. p16*	21,229	8 May 2013
#25. 'tumor suppressor gene'/exp	43,370	8 May 2013
#24. 'cyclin dependent kinase inhibitor 2a'/mj	2,047	8 May 2013
#23. triage	12,005	8 May 2013
#22. 'dual staining'	612	8 May 2013
#21. 'dual stain'	57	8 May 2013

#20. pap OR 'papanicolaou test'/mj OR 'cytological screening' OR 'vagina smear'/mj AND ('wart virus'/mj OR 'human papillomavirus' OR hpv* OR 'squamous cell carcinoma'/mj OR 'atypical squamous cells of undetermined significance' OR ascus* OR 'low-grade squamous intraepithelial lesion' OR 'low-grade squamous intraepithelial lesions' OR lsil* OR cervi* NEAR/1 (cancer* OR carcinoma*) OR 'uterine cervix tumor'/mj OR 'uterine cervix carcinoma in situ'/mj OR 'cervical intraepithelial neoplasia')	8,450	8 May 2013
#19. 'wart virus'/mj OR 'human papillomavirus' OR hpv* OR 'squamous cell carcinoma'/mj OR 'atypical squamous cells of undetermined significance' OR ascus* OR 'low-grade squamous intraepithelial lesion' OR 'low-grade squamous intraepithelial lesions' OR lsil* OR cervi* NEAR/1 (cancer* OR carcinoma*) OR 'uterine cervix tumor'/mj OR 'uterine cervix carcinoma in situ'/mj OR 'cervical intraepithelial neoplasia'	152,284	8 May 2013
#18. 'cervical intraepithelial neoplasia'	6,615	8 May 2013
#17. 'uterine cervix carcinoma in situ'/mj	6,198	8 May 2013
#16. 'uterine cervix tumor'/mj	19,884	8 May 2013
#15. cervi* NEAR/1 (cancer* OR carcinoma*)	71,840	8 May 2013
#14. lsil*	1,650	8 May 2013
#13. 'low-grade squamous intraepithelial lesions'	690	8 May 2013
#12. 'low-grade squamous intraepithelial lesion'	816	8 May 2013
#11. ascus*	1,516	8 May 2013
#10. 'atypical squamous cells of undetermined significance'	1,106	8 May 2013
#9. 'squamous cell carcinoma'/mj	54,246	8 May 2013
#8. hpv*	30,739	8 May 2013
#7. 'human papillomavirus'	26,038	8 May 2013
#6. 'wart virus'/mj	10,564	8 May 2013
#5. pap OR 'papanicolaou test'/mj OR 'cytological screening' OR 'vagina smear'/mj	28,899	8 May 2013
#4. 'vagina smear'/mj	4,775	8 May 2013
#3. 'cytological screening'	531	8 May 2013
#2. 'papanicolaou test'/mj	2,608	8 May 2013
#1. pap	23,716	8 May 2013

The Cochrane Library

Search Name: p16/Ki-67 PAP Test final (AK o80513)	
Last Saved: 08/05/2013 15:55:21.238	
Description: AK 05/2013	
ID	Search
#1	PAP (Word variations have been searched)
#2	Papanicolaou (Word variations have been searched)
#3	cytological screening (Word variations have been searched)
#4	p16* near cytology
#5	MeSH descriptor: [Vaginal Smears] explode all trees
#6	#1 or #2 or #3 or #4 or #5
#7	MeSH descriptor: [Papillomaviridae] explode all trees
#8	Human papillomavirus* (Word variations have been searched)
#9	MeSH descriptor: [Carcinoma, Squamous Cell] explode all trees
#10	atypical squamous cells of undetermined significance (Word variations have been searched)
#11	ASCUS* (Word variations have been searched)
#12	Low-grade squamous intraepithelial lesion* (Word variations have been searched)
#13	LSIL* (Word variations have been searched)
#14	Cervi* near (Cancer* or Carcinoma*)
#15	MeSH descriptor: [Uterine Cervical Neoplasms] explode all trees
#16	MeSH descriptor: [Cervical Intraepithelial Neoplasia] explode all trees
#17	#7 or #8 or #9 or #10 or #11 or #12 or #13 or #14 or #15 or #16
#18	#6 and #17
#19	Dual stain* (Word variations have been searched)
#20	MeSH descriptor: [Triage] explode all trees
#21	Triage (Word variations have been searched)
#22	MeSH descriptor: [Cyclin-Dependent Kinase Inhibitor p16] explode all trees
#23	MeSH descriptor: [Genes, p16] explode all trees
#24	p16* (Word variations have been searched)
#25	INK4* (Word variations have been searched)
#26	MeSH descriptor: [Ki-67 Antigen] explode all trees
#27	p16* near Ki-67*
#28	Ki67* (Word variations have been searched)
#29	CinTec (Word variations have been searched)
#30	#19 or #20 or #21 or #22 or #23 or #24 or #25 or #26 or #27 or #28
#31	#18 and #30 in Cochrane Reviews (Reviews and Protocols), Other Reviews, Trials and Technology Assessments
68 Hits	

CRD

#### p16/Ki67 PAP-Test (AK 05/3013)	
1	Papanicolaou
2	PAP
3	cytological screening
4	MeSH DESCRIPTOR Papillomaviridae EXPLODE ALL TREES
5	Human papillomavirus*
6	HPV*
7	MeSH DESCRIPTOR Carcinoma, Squamous Cell EXPLODE ALL TREES
8	atypical squamous cells of undetermined significance
9	ASCUS*
10	Low-grade squamous intraepithelial lesion*
11	LSIL*
12	MeSH DESCRIPTOR Vaginal Smears EXPLODE ALL TREES
13	Cervi* NEAR (Cancer* OR Carcinoma*)
14	MeSH DESCRIPTOR Uterine Cervical Neoplasms EXPLODE ALL TREES
15	MeSH DESCRIPTOR Cervical Intraepithelial Neoplasia EXPLODE ALL TREES
16	#1 OR #2 OR #3 OR #4 OR #5 OR #6 OR #7 OR #8 OR #9 OR #10 OR #11 OR #12 OR #13 OR #14 OR #15
17	Dual stain*
18	MeSH DESCRIPTOR Triage EXPLODE ALL TREES
19	trriage
20	MeSH DESCRIPTOR Cyclin-Dependent Kinase Inhibitor p16 EXPLODE ALL TREES
21	MeSH DESCRIPTOR Genes, p16 EXPLODE ALL TREES
22	p16
23	^{INK4a}
24	p16 NEAR ^{INK4a}
25	MeSH DESCRIPTOR Ki-67 Antigen EXPLODE ALL TREES
26	p16 NEAR Ki-67
27	CinTec
28	#18 OR #19 OR #20 OR #22 OR #23 OR #24 OR #25
29	#16 AND #28
48 Hits	

08.05.2013

QUADAS Auswertung

Domain 1: Patient selection									
id	reviewer	author_year_pub	patients_rob_desc	patient_sample	patient_CC	patient_ex	patient_rob	patients_app_desc	patient_applicability
			<i>Describe methods of patient selection.</i>	<i>Was a consecutive or random sample of patients enrolled?</i>	<i>Was a case-control design avoided?</i>	<i>Did the study avoid inappropriate exclusions?</i>	<i>Could the selection of patients have introduced bias?</i>	<i>Describe included patients (prior testing, presentation, intended use of index test and setting)</i>	<i>Is there concern that the included patients do not match the review question?</i>
	AK	Edgerton 2011	63 Sure Path Pap test slides diagnosed with ASCUS with correlative colposcopic biopsy collected over a period of 17 months at Emory University Hospital.	No	Yes	Yes	HIGH	Setting, Prior testing and presentation not described; Use as triage test in ASC-US population for detection of underlying high-grade disease.	HIGH
	AK	Loghavi 2012	188 SurePath Pap slides diagnosed with ASCUS or LSIL retrieved from department files with histological and/or cytological follow-up.	No	Yes	Yes	HIGH	Setting, Prior testing and presentation not described; Use as triage test in ASC-US /LSIL population for detection of underlying high-grade disease.	HIGH
	AK	Schmidt 2011	ASC-US or LSIL LBC samples with corresponding tissue blocks from punch biopsies, cone biopsies and/or ECC within 6 months after cytology from 5 anatomic pathology labs in CH and I. Samples were included as consecutive cases within the disease (CIN2+) and no-disease (CIN1 or negative) groups until the required number of disease cases was reached for each cytologic category. Sample slides for the index test must meet minimum squamous cellularity criteria, not obscured by blood, mucus etc. and not damaged during the procedure to be included.	Yes	No	Yes	HIGH	Setting, Prior testing and presentation not described; Use as triage test in ASC-US /LSIL population for detection of underlying high-grade disease.	HIGH
	AK	Waldstrom 2012	554 consecutive cervical cytology (Thin Prep LBC) samples diagnosed with LSIL received from Jan 2006 to Feb 2007 at the pathology dept. Of Vejle Hospital. 59 samples excluded due to lack of material, 1 sample could not be found, 25 excluded due to lack of follow-up info.	Yes	Yes	No	HIGH	Setting, Prior testing and presentation not described; Use as triage test in LSIL population for detection of underlying high-grade disease.	HIGH
	AK	Wentzensen 2012	404 women with an outside referral to colposcopy Pap of ASCUS HPV+ or LSIL.	Yes	Yes	Yes	LOW	Women 18+ referred to colposcopy clinic Feb 2009 to Aug 2011 for abnormal cervical cancer screening results. Exclusion criteria: previous treatment for cervical disease, prior chemotherapy or radiation therapy, pregnancy, HIV infection, inability to give informed consent. Setting, Prior testing and presentation not described; Use as triage test in ASC-US HPV+ /LSIL referral population for detection of underlying high-grade disease.	HIGH
	AK	Ziemke 2012	222 patients with PAP IIID were included consecutively during routine presentation of patients at the doctor's office. 130 patients with cytomorphological categorisation in PAP IIID and low grade dysplasia were included in the study.	Yes	Yes	Yes	LOW	Routine presentation of patients at gynecologists, patient management according to guidelines. Use as prognostic factor for high grade disease	LOW

Domain 2: Index test(s)							
id	reviewer	author year_pub	index_desc	index_blind	index_threshold	index_rob	index_applicability
			<i>Describe the index test and how it was conducted and interpreted.</i>	<i>Were the index test results interpreted without knowledge of the results of the reference standard?</i>	<i>If a threshold was used, was it pre-specified?</i>	<i>Could the conduct or interpretation of the index test have introduced bias?</i>	<i>Is there concern that the index test, its conduct, or interpretation differ from the review question?</i>
	AK	Edgerton 2011	Destained SurePath slides stained with CINtec PLUS Kit. Slides analysed by a single pathologist blinded to histology results, but not to ASC-US diagnosis.	Yes	Yes	LOW	HIGH
	AK	Loghavi 2012	Destained SurePath slides stained with CINtec PLUS Kit. Slides analysed by three pathologists blinded to index cytology and follow up diagnoses.	Yes	Yes	LOW	HIGH
	AK	Schmidt 2011	Cytology slide prepared from residual ThinPrep vial using T2000 slide processor (Hologic) and DualStain with CINtec PLUS Kit (mtm labs). Interpretation by trained cytotechnologist and pathologist review for pos. Slides. Blinding/Order of testing not described.	Unclear	Yes	UNCLEAR	HIGH
	AK	Waldstrom 2012	Cytology slide prepared from residual ThinPrep vial using T2000 slide processor (Hologic) and DualStain with CINtec PLUS Kit (mtm labs). Interpretation by trained cytotechnologist and pathologist review for pos. Slides. The observers were unaware of the follow-up data and the results from the mRNA test.	Yes	Yes	LOW	HIGH
	AK	Wentzensen 2012	Cytology slide prepared from residual ThinPrep vial using T2000 slide processor (Hologic) and DualStain with CINtec PLUS Kit (mtm labs). Interpretation by trained cytotechnologist and confirmed by a pathologist. The observers were unaware of the follow-up data and the results from the HPV test.	Yes	Yes	LOW	HIGH
	AK	Ziemke 2012	Destained conventional cytology slides or LBC Thin Prep slides were stained with DualStain kit. Interpretation by pathologist and blinded review by second pathologist.	Yes	Yes	LOW	HIGH

Domain 3: Reference standard							
id	reviewer	author year_pub	ref_desc	ref_acceptable	ref_blind	ref_rob	ref_applicability
			<i>Describe the reference standard and how it was conducted and interpreted.</i>	<i>Is the reference standard likely to correctly classify the target condition?</i>	<i>Were the reference standard results interpreted without knowledge of the results of the index test?</i>	<i>Could the reference standard, its conduct, or its interpretation have introduced bias?</i>	<i>Is there concern that the target condition as defined by the reference standard does not match the review question?</i>
	AK	Edgerton 2011	Follow-up biopsy diagnosed by routine staff physicians. Most severe dysplasia was recorded (in case of multiple biopsies)	Yes	Yes	LOW	LOW
	AK	Loghavi 2012	Biopsy or cytology at follow up, interpretation not described. Most severe dysplasia was recorded. Follow-up was ended when HG CIN were diagnosed.	No	Yes	HIGH	LOW
	AK	Schmidt 2011	The histologic diagnosis established on the recut tissue block specimens was used as the gold standard (ie, diagnostic accuracy criterion) for the study. A majority consensus diagnosis was established for each case by 2 or more expert pathologists. The reviewers were blinded to the original local histopathologic result and to all other test results. A combined review of H&E-stained slides and consecutive slides stained with the (p16) CINtec Histology Kit was performed by the expert pathologists.	Yes	Yes	LOW	LOW
	AK	Waldstrom 2012	Follow-up with histology with a punch and/or cone biopsy (48,2%) and/or cytology (52%) for at least 5 years. Most severe diagnosis from clinical follow-up with cytology or histology according to Danish guidelines. Interpretation not described. Follow-up time of at least 5 years.	No	Yes	HIGH	LOW
	AK	Wentzensen 2012	Colposcopy and extended biopsy protocol wit digital photographic documentation of biopsy sites. Up to 4 biopsies from acetowhite lesions or large heterogenous lesions ext over 2 quadrants. CIN2+ treated by LEEP of transformation zone. IF Fewer than 4 biopsies THEN biopsy from a quadrant w/o visible CIN was included. Interpretation not described. Outcomes considered based on worst biopsy result.	Yes	Unclear	UNCLEAR	LOW
	AK	Ziemke 2012	Follow-up with histology with a punch and/or cone biopsy (27,7%) and/or cytology (routine patient management) . Interpretation not described.	No	Yes	HIGH	LOW

Domain 4: Flow and timing									
id	reviewer	author_year_pub	flow_rob_desc	flow_rob_desc2	flow_time	flow_partver_bias	flow_difver_bias	flow_analysis	flow_rob
			<i>Describe any patients who did not receive the index test(s) and/or reference standard or who were excluded from the 2*2 table (refer to flow diagram):</i>	<i>Describe the time interval and any interventions between index test(s) and reference standard:</i>	<i>Was there an appropriate interval between index test(s) and reference standard?</i>	<i>Did all patients receive a reference standard?</i>	<i>Did patients receive the same reference standard?</i>	<i>Were all patients included in the analysis?</i>	<i>Could the patient flow have introduced bias?</i>
	AK	Edgerton 2011	None.	Not described-	Unclear	Yes	Yes	Yes	UNCLEAR
	AK	Loghavi 2012	13 excluded due to high background staining.	Analysis of archived specimens with time interval of up to 86 months following collection of samples.	No	Yes	No	No	HIGH
	AK	Schmidt 2011	135 samples not included to enrich for disease cases. 34 cases were excluded due to insufficient amounts of residual material to perform the analysis.	up to 6 months.	Yes	Yes	Yes	No	HIGH
	AK	Waldstrom 2012	59 samples excluded due to lack of material, 1 sample could not be found, 25 excluded due to lack of follow-up info.	Analysis of archived specimens with time interval of up to 5 years following collection of samples.	No	No	No	No	HIGH
	AK	Wentzensen 2012	17 excluded due to insufficient amounts of residual material, 48 excluded due to low cellularity, 221 had referral cytology ≠ ASCUS/LSIL.	No interventions between index test and reference standard – time interval 1d.	Yes	Yes	Yes	Yes	LOW
	AK	Ziemke 2012	None.	Follow up according to routine german patient management guidelines. (includes repeat cytology) up to 36months.	No	Yes	No	Yes	HIGH