publiziert bei: AWMF online
Das Portal der wissenschaftlichen Medizin

Zusammenstellung: Evidenzberichte und Zusatzmaterialien von M. Arbyn et al. für die Leitliniengruppe Prävention des Zervixkarzinoms

Unit of Cancer Epidemiology
Scientific Institute of Public Health
Juliette Wytsmanstreet 14
B1050 Brussels

Die Erstellung der Evidenzberichte und Zusatzmaterialien wurde mit Mitteln der Deutschen Krebshilfe im Rahmen des Leitlinienprogramms Onkologie finanziert.

Content

1. Question: performance of liquid-based versus conventional cervical cytology to detect cervical precancer

1.1. Introduction

Thin-layer cytology or *liquid-based cytology* (LBC) is a new technique for transferring the cellular material collected with a spatula and/or a brush from the transformation zone of the uterine cervix. The cells are not spread directly onto a slide to obtain a conventional Pap (CP) smear but transferred into a vial with a fixative liquid. This container is then sent to a specially equipped laboratory. So far, two commercially available LBC systems, ThinPrep (Cytyc, Boxborough, MA) and SurePath (formerly, AutoCyte PREP or CytoRich, TriPath Imaging Inc., Burlington, NC) are FDA approved, allowing the claim of increased detection of squamous intraepithelisal lesions and a reduction of the number of unsatisfactory smears compared to the CP [1,2].

Several systematic reviews regarding the performance of LBC to detect cervical cancer precursors were performed[3-16]. Conclusions formulated by the reviewing authors were disparate and depended largely on selection criteria to include individual studies and the considered performance parameters. Studies comparing detection rates for low grade cytological abnormalities often yielded more favorable results for LBC [4,6,9,17], whereas in studies focusing on accuracy for biopsy-confirmed high-grade CIN (cervical intraepithelial neoplasia), no significant differences between the CP and LBC were found [10,12,18]. In the framework of the preparation of the 2nd edition of the European Guidelines for Quality Assurance[19] in cervical cancer screening, a new comprehensive systematic review was performed which concluded that **"Liquid-based cervical cytology is neither more sensitive nor more specific for detection of high-grade cervical intraepithelial neoplasia compared with the conventional Pap test"[20].** This systematic review will be updated by including new randomised trials comparing both types of cytology and published since 2007. It will also address specimen adequacy and duration of microscopic inspection.

1.2. Materials and Methods

1.2.1. Clinical Questions

1.2.1.1. Accuracy of LBC

What is the absolute clinical accuracy (sensitivity and specificity) to identify or exclude high-grade cervical precancer or worse (CIN2+) using liquid-based cytology or conventional cytology and what is the relative accuracy of liquid-based versus conventional cytology?

1.2.1.2. Specimen quality

What is the percentage of cervical cell specimen that is judged as unsatisfactory for microscopic interpretation? Does the use of LBC result in less unsatisfactory specimen? Or, in other words: what is the relative inadequacy rate of LBC compared to CP (ratio of percentages of unsatisfactory samples [LBC/CP])?

1.2.1.3. Duration of microscopic interpretation

What is the average time needed to interpret an LBC specimen versus a CP? Does the interpretation of LBC requires less time?

1.2.2. Literature Retrieval

Previously, meta-analyses comparing the performance of liquid-based cytology[21-24] versus conventional cytology[21,25] have been performed and published by the Unit of Cancer Epidemiology. To update these systematic reviews, an additional systematic literature search in Medline and EMBASE, performed on May 31st, 2013 resulted in 320 unique articles. The search strategy used is shown in Box 1 which was, for each database, adapted to the relevant syntax.

“cervix neoplasm” or “cervical intraepithelial neoplasia” or “cervix dysplasia” or "cin" AND “monolayer” or “thin layer” or “liquid-based” or “Thin-Prep” or "Thinprep" or “CytoRich” or “Autocyte” or “SurePath” AND "RCT" or "randomised trial" or "randomised controlled trial"

Box 1. Search string for literature retrieval.

We refer to the published reviews for details on the strategy for literature retrieval. Two study designs were considered: 1) concomitant testing design and 2) two-cohort design. In the concomitant testing design, two cervical cell samples are prepared from the same patient. Most often a single sample is taken from the uterine cervix, a conventional Pap is prepared, and the residual cellular material remnant on the sampling device is then transferred into a vial with fixative liquid ("split-sample"). Occasionally, two separate samples are collected: one for the conventional Pap and another one for liquid-based cytology. In the two-cohort design, conventional Pap samples and liquid-based cytology samples are taken from women belonging to separate but similar populations. The current review is restricted to studies where all subjects were submitted to gold standard verification, based on colposcopy and histology of colposcopy-targeted biopsies, allowing evaluation of the absolute and relative test validity for cervical intraepithelial neoplasia grade 2 or worse (CIN 2+) without verification bias. Randomised controlled studies with at least 90% completeness in follow-up confirmation of cytologically positive women were added to the metaanalysis of the relative sensitivity. This addition is justified because, in randomised controlled trials, the ratio of the detection rate of CIN 2_ in the liquid-based cytology arm over that in the conventional Pap arm is equivalent to the ratio of the absolute sensitivities derived from studies with complete gold standard verification [26,27].

1.2.3. Outcome measures

The following outcomes were considered regarding diagnostic test accuracy:
1) absolute accuracy of LBC and CP to detect CIN2+, considering the cut-offs for test-positivity of ASC-US+, LSIL+ and HSIL+, for split-sample studies or two-cohort studies with complete verification with the reference test.
2) relative accuracy of LBC compared to CP, considering the same test-positivity cut-offs and studies but including also randomised trials.
Absolute and relative accuracy were computed using the STATA procedures *metaprop*[b] and *metan*[28], respectively. Overall pooled measures, with 95% confidence intervals were calculated using random effects models[29]. The statistical heterogeneity was assessed by the p-value for heterogeneity (following a chi2 distribution) as well as by the I^2 statistic, which measures the proportion of the variation that is due to inter-study heterogeneity.

[b] Metaprop is a statistical procedure in STATA developed at the Unit of Cancer Epidemiology (IPH Brussels) to pool proportions based on binomial distributions.

Regarding the specimen adequacy, the proportion of unsatisfactory cervical cell specimen in LBC samples and in conventional Pap smears as well as the ratio of these proportions were computed and pooled using the same . For studies where more than one LBC system was described, rates of unsatisfactory specimen were compared by LBC preparation method. The STATA procedures metaprop and metan were used for meta-analytical pooling as described in the previous paragraph.

Data on the average duration of microscopic inspection of an LBC specimen and a CP (measured in seconds) were extracted and compared.

1.3. Results

1.3.1. Study selection and characteristics

The process of literature retrieval and study selection of studies are shown in Figure 1. In total, 11 studies could be included, which fulfilled the selection criteria for assessment of the absolute accuracy (all enrolled women received both CP and LBC and were subsequently submitted to verification with the reference test)[30-40]. Two of them were randomised trials[38,40] and the nine others were diagnostic test accuracy studies.

In addition, four randomised[41,42] or quasi-randomised[41,42] trials were selected where women were allocated ad random to CP or LBC. In these four randomised trials, women received the reference test if CP or LBC were abnormal and therefore they contributed only to the assessment of the relative sensitivity. The eleven studies used for the meta-analysis of the absolute accuracy contributed also to the that of the relative sensitivity and specificity.

The main study characteristics are summarised in Table 2. The ThinPrep system was used in ten studies[30,32,33,35,37,39,41-44], and the AutoCye/SurePath procedure in two other reports[31,38]. CellSlide[34], DNA Citoliq[36], and Liqui-Prep[40] were each applied in only one study. Most studies enrolled women who were referred to colposcopy for diagnostic work-up because of prior cytological abnormalities[30-40], whereas four trials enrolled only women who attended primary cervical cancer screening[41-44]. Coste et al recruited a screening and a referral population[32]. The study size (number of enrolled women) varied between 151[34] to over 85,000[42].

Hundred and eleven studies were identified where the proportion of unsatisfactory specimen was identified in CP and LBC. Eight studies contained

information on the number of unsatisfactory smears in different LBC systems (see 3.2.2).

Twelve studies were identified where the average duration of microscopic inspection in LBC and CP was reported (see 3.2.3).

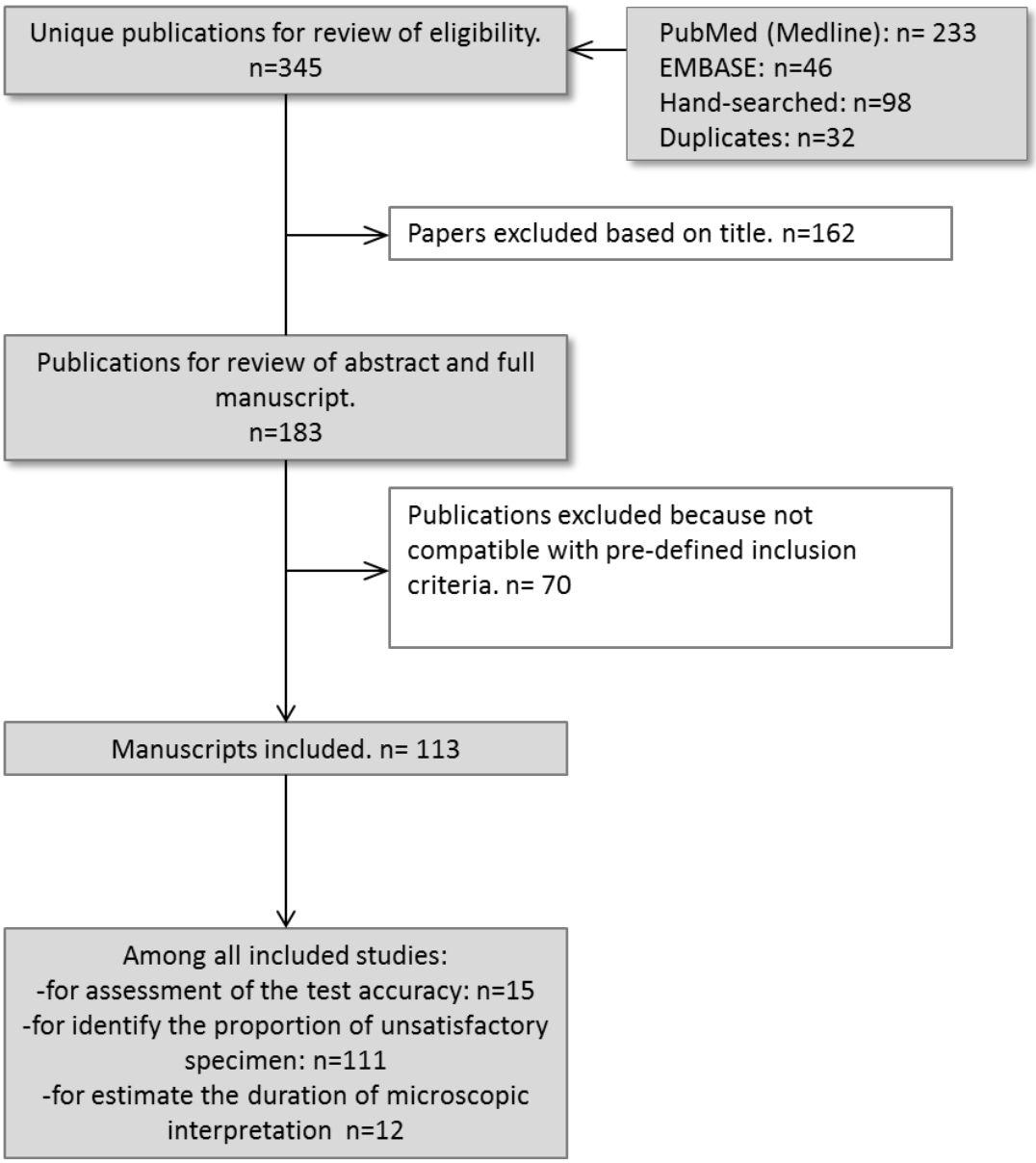


Figure 1. PRISMA flow chart for the retrieval of literature in meta-analysis

1.3.2. Accuracy to detect high-grade CIN

1.3.2.1. 3.2.1. Absolute accuracy to detect CIN2+

Table 1 shows the pooled absolute sensitivity and specificity of LBC and CP to detect underlying CIN2+ at cut-offs ASC-US+, LSIL+, and HSIL+, computed by a bivariate model. The joint variation of sensitivity and specificity is displayed by the HSROC curves with summary points, and 95% confidence ellipses around them (see Figure 2).

Table 1. Pooled absolute sensitivity of LBC and CP [in % (with 95% CI between brackets] to detect CIN2+.

Cut-off	Parameter	LBC		CP		References
ASC-US+	Sensitivity	90.7	(84.9-94.5)	88.6	(82.1-92.9)	30-32,35-40
	Specificity	59.6	(48.1-70.1)	63.2	(50.4-74.4)	
LSIL+	Sensitivity	77.7	(66.0-86.2)	74.6	(63.0-83.5)	30-37,39,40
	Specificity	78.2	(68.1-85.7)	79.2	(69.1-86.7)	
HSIL+	Sensitivity	59.4	(46.6-71.1)	56.5	(45.1-67.2)	30-37,39,40
	Specificity	95.2	(92.2-97.1)	95.8	(93.4-97.3)	

At cut-offs ASC-US+, LSIL+ and HSIL+, the absolute pooled sensitivity of LBC to detect CIN2+ was 91%, 78% and 59%, whereas the sensitivity of CP was 89%, 75% and 57%, respectively.

The pooled specificity to exclude CIN2+ was 60%, 78% and 95% for LBC and 63%, 79% and 96% for CP, for the cut-offs of ASC-US+, LSIL+, and HSIL+, respectively.

Table 2. Study characteristics of included studies.

Author	Country	Study population	Study design	Study size	LBC procedure	Collection method	Gold standard	Independence LBC/CP interpretation	Blinding gold standard towards type of cytology
Ferenczy, 1996	Canada US	Women referred for colposcopy.	Concomitant testing, split-sample.	364	TP-beta	Acellon Combi sampler	Colposcopy on all. Punch biopsy or LEEP and ECC for all cases.	Blinded	Blinded
Bergeron, 2001	France	Women referred for cone biopsy.	Concomitant testing, direct-to-vial.	500	AutoCyte	Cervex-Brush	Colposcopy and cone biopsies for all subjects.	Blinded	Blinded
Coste, 2003	France	1) Women referred for colposcopy. 2) Screening population	Concomitant testing, split-sample.	2585 (=828+1757)	TP2000	Not documented	Colposcopy on all subjects, biopsies if colposcopically positive.	Blinded	Blinded
Confortini, 2004	Italy	Women referred for colposcopy. LBC was taken just before colpo-scopy, 30-60days after abnormal CP.	Concomitant testing, direct-to-vial.	297	TP2000	Not documented	Colposcopy on all. Histology of punch or excision biopsies, negative colposcopy accepted as TN.	Blinded	Blinded
Confortini, 2005	Italy Spain	Women referred for colposcopy. LBC just before colposcopy, 30-60days after an abnormal CP.	Concomitant testing, direct-to-vial.	151	CellSlide	Spatula & EC brush	Colposcopy on all, histology documented for 22 CIN2+ cases.	Blinded	Blinded
Hussein, 2005	UK	Follow-up of screen-positive women.	Concomitant testing, split-sample.	441	TP2000	Broom	Colposcopy on all, biopsy (punch or excision) if suspicion of HSIL (all) or LSIL (partial).	Not documented.	Not documented.
Longatto Filho, 2005	Brazil	Follow-up or screen-positive women (VIA, Pap smear).	Concomitant testing, split-sample.	1,095	DNA-Citoliq	CP: Ayre spatula & EC brush. LBC:EC brush.	Colposcopy on all, biopsies taken when indicated. Conization & hysterectomy were taken into account.	Blinded†	Blinded
Taylor, 2006	South-Africa	High-risk population, included in a see & treat trial (15% treated with cryotherapy).	2-cohort, LBC & CP rotated every 6 months.	LBC:3,184 CCP:2,463	TP2000	Spatula & EC brush	Colposcopy on all. ECC & biopsy of colposcopic abnormalities	Not of application, because of 2-cohort design.	Blinded*
Ronco, 2007	Italy	Women invited for organised screening.	RCT	LBC-HC2: 22,708 CP: 22,466	TP2000	Plastic spatula & EC brush	Colposcopy according to age, HPV test result and study site. Biopsies if colposcopic suspicion.	Not of application, because RCT.	Not blinded, quality review of CIN cases was blinded.

Author	Country	Study population	Study design	Study size	LBC procedure	Collection method	Gold standard	Independence LBC/CP interpretation	Blinding gold standard towards type of cytology
Strander, 2007	Sweden	Women invited for organised screening.	Quasi-randomised trial, alternating every week LBC & CP.	LBC: 8,810 CP: 4,674	TP2000	CP: wooden Ayre spatula & EC-brush LBC: plastic spatula & EC-brush	Repeat cytology if ASC-US or LSIL. Colposcopy referral if HSIL or ASC-US+ at repeat cytology.	Not of application since quasi-randomised trial.	Colposcopists and histologists were blinded.
Sykes, 2008	New Zealand	Women referred to colposcopy	RCT	LBC: 451 CP: 453	SurePath	Not documented	Colposcopy on all, colposcopic biopsies or subsequent biopsy <12 months later.	Not documented	Not documented
Angstetra, 2009	Australia	Women referred to colposcopy	Concomitant testing, split-sample.	1,961	ThinPrep	Cervex-Brush	Colposcopy on all women with biopsies taken if suspected colposcopic high-grade abnormalities.	Not documented	Not documented
Siebers, 2009	Netherlands	Women attending organised screening	RCT	LBC: 46,066 CP: 39,010	TP2000	Cervex-Brush	Repeat cytology if ASC-US/LSIL. Colposcopy if HSIL+, or ASC-US+ at repeat cytology. Biopsies from colposcopically abnormal areas.	Blinded	Blinded
Jesdapatarakul, 2011	Thailand	Women referred to colposcopy	RCT: arm1 first LBC, then CP and vice-versa in arm2.	194	Liqui-Prep	LBC: Cervex. CP:extended tip spatula.	Colposcopy on all, colposcopic directed biopsies from any suspicious lesions and ECC if no colposcopy was negative or unsatisfactory.	Blinded	Blinded
Klug, 2013	Germany	Women participating in opportunistic cervical cancer screening	Quasi-randomised trial, alternating every week LBC & CP.	LBC: 11,331 CP: 9,296	TP2000	LBC & CP: spatula and EC-brush.	All women with LSIL+ were referred to colposcopy. Biopsies from colposcopically abnormal areas.	Not of application since quasi-randomised trial.	Blinded

ASC-US: atypical squamous cells of undetermined significance; colpo: colposcopy; CP: conventional Pap smear; EC brush: endocervical brush; ECC: endo-cervical curettage; HC2: Hybrid Capture-2 assay (Digene Corp., Gaithersburg, MD); hr: high-risk HPV types; HSIL: high-grade intraepithelial lesion; LBC: liquid-based cytology; LEEP: loop electrosurgical excision procedure; LSIL: low-grade intraepithelial lesion; RCT: randomised controlled trial; VIA: visual inspection after application of acetic acid.

† Longatto Filho, 2005: authors report that CP and LBC were interpreted blindly, but the same cytologists interpreted the two preparations from the same patient.

* Taylor, 2006: cytology and colposcopy/histology were blinded to each other but given study design with CP & LBC performed in separate periods, blinding cannot be considered as complete.

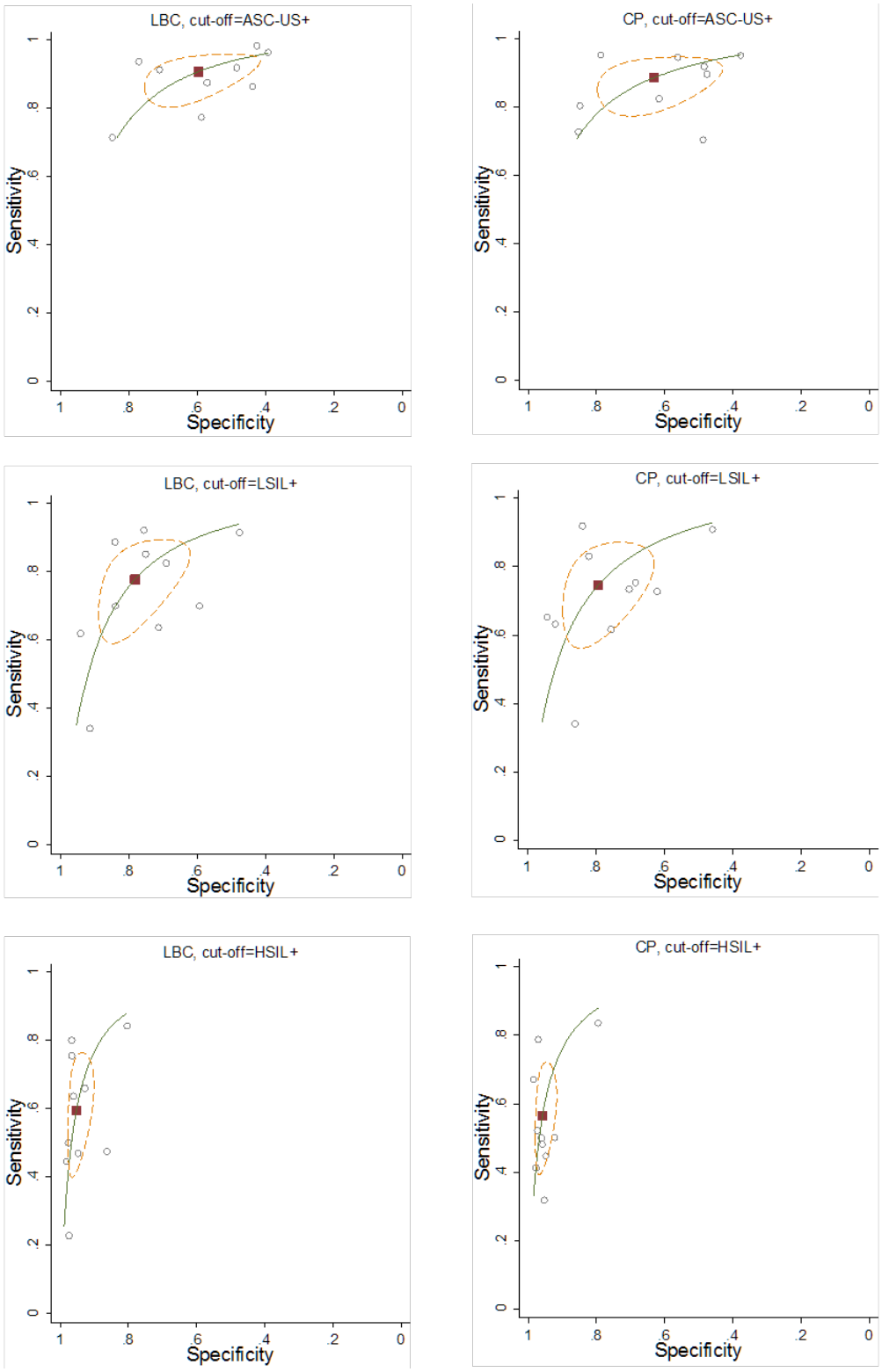


Figure 2. HSROC curves showing the joint variation of the absolute sensitivity and specificity of LBC (left) and CP (right), at cut-off ASC-US= (top), LSIL+ (middle), and HSIL+ (bottom) to detect CIN2+.

1.3.2.2. Relative accuracy of LBC compared with the to detect CIN2+

For the assessment of the relative sensitivity, all cross-sectional test accuracy studies with complete verification as well as the randomised controlled trials were included in the meta-analysis. In the meta-analysis of the relative specificity, the RCTs were not included. The forest plots displaying the relative sensitivity and specificity are shown in Figure 3, Figure 4 and Figure 5, for ASC-US+, LSIL+ and HSIL+, respectively. Not all studies reported the data for all possible cut-offs of test positivity.

Overall, LBC was not more sensitive than LBC, neither at cut-off ASC-US+ (RR=1.02 [95% CI: 0.99-1.06]), nor at LSIL+ (RR=1.05 [0.99-1.12]), nor at HSIL+ (RR=1.04 [95% CI: 0.95-1.15]) (see Table 3). Significant inter-study heterogeneity in the sensitivity values was observed with the large majority of studies demonstrating similarity of sensitivity at the exception of Longatto-Filho 2007 (at cut-off ASC-US+), Strander 2007 and Klug 2013 (at cut-off LSIL+) and Confortini 2004 and Hussein 2005 (at cut-off HSIL+). LBC never was less sensitive than CP, excepted in the study of Jesdapatarakul 2013, where LIQUI-PREP was used (at cut-off HSIL+). The variation in the sensitivity was not explained by the use of different preparation systems for LBC (p for inter-group comparisons always >0.05).

Table 3. Meta-analysis of the relative sensitivity (top) and specificity (bottom) of LBC compared to CP to detect underlying CIN2+.

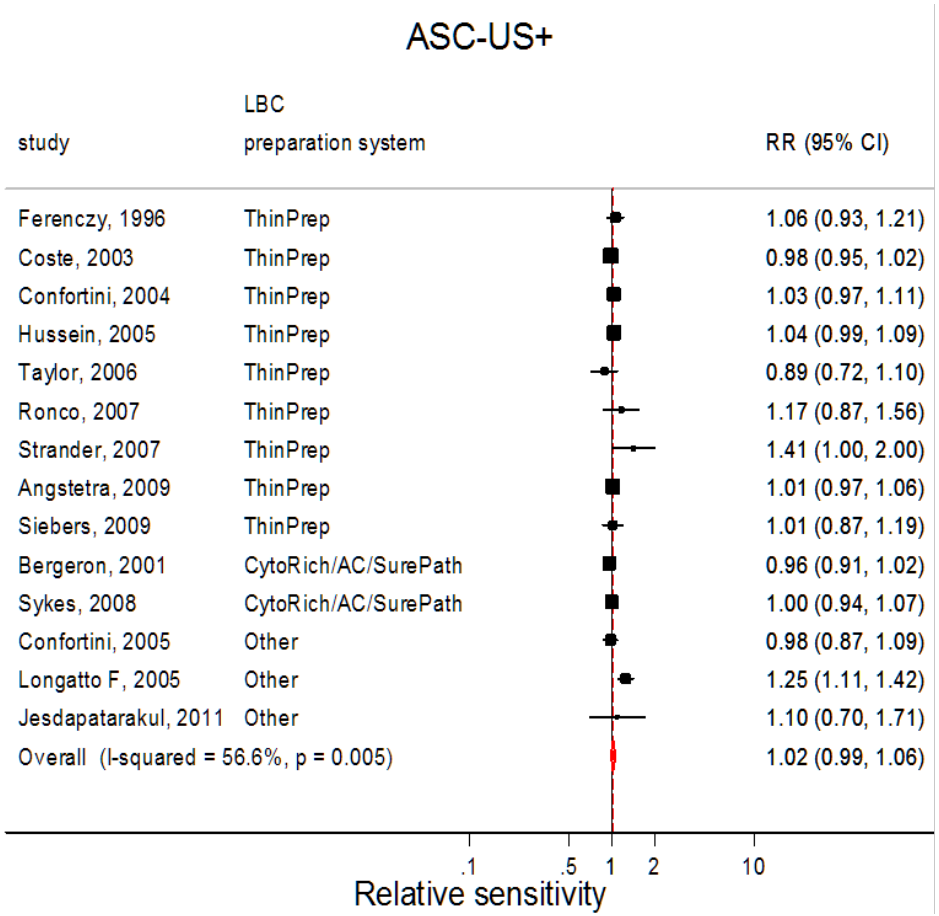
Cut-off	Nb of studies	Relative sensitivity (95% CI)	p* (heterogeneity)	p† (inter-group)
ASC-US+	14	1.02 (0.99-1.06)	0.005	0.063
LSIL+	14	1.05 (0.99-1.12)	0.005	0.414
HSIL+	12	1.04 (0.95-1.15)	0.055	0.673

Cut-off	Nb of studies	Relative specificity (95% CI)	P* (heterogeneity)	p† (inter-group)
ASC-US+	8	0.94 (0.88-1.01)	<0.001	<0.001
LSIL+	10	0.98 (0.95-1.02)	<0.001	<0.001
HSIL+	10	0.99 (0.98-1.01)	0.002	0.005

* test for overall inter-study heterogeneity.

† test for differences in accuracy between LBC preparation systems (ThinPrep, CytoRich/AutoCyre/SurePath or other systems).

Also the relative specificity (LBC/CP) never was statistically significantly different from unity with ratios of 0.94 (95% CI: 0.88-1.01) at cut-off ASC-US+, 0.98 (95% CI: 0.95-1.02) at cut-off LSIL+, and 0.99 (95% CI: 0.98-1.01) at HSIL+ (see Table 3). The inter-study heterogeneity in the relative specificity was large and varied also significantly by LBC preparation system. In the large majority of studies, the confidence interval around the specificity included unity. However, Hussein et al [35] (at cut-off ASC-US+), Longatto-Filho et al[36] (at all cut-offs) and Bergeron et al[31] (at cut-off LSIL+) observed a significantly lower specificity in LBC. Only Taylor et al[37] found a higher specificity for LBC at cut-off HSIL+.



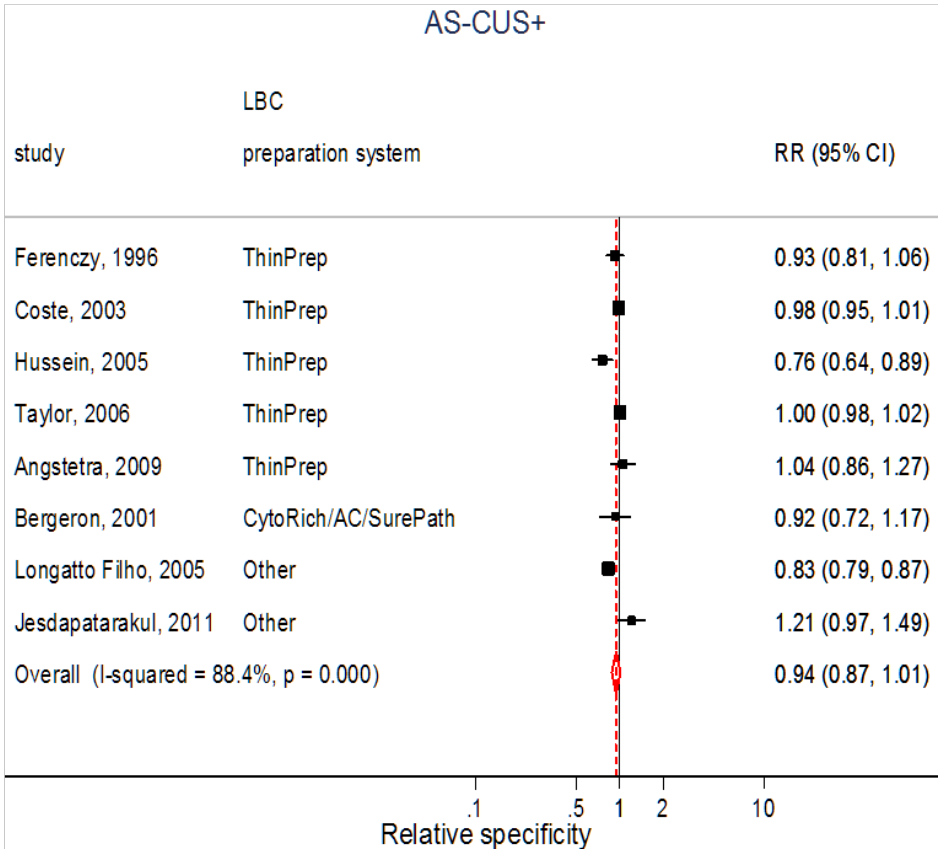
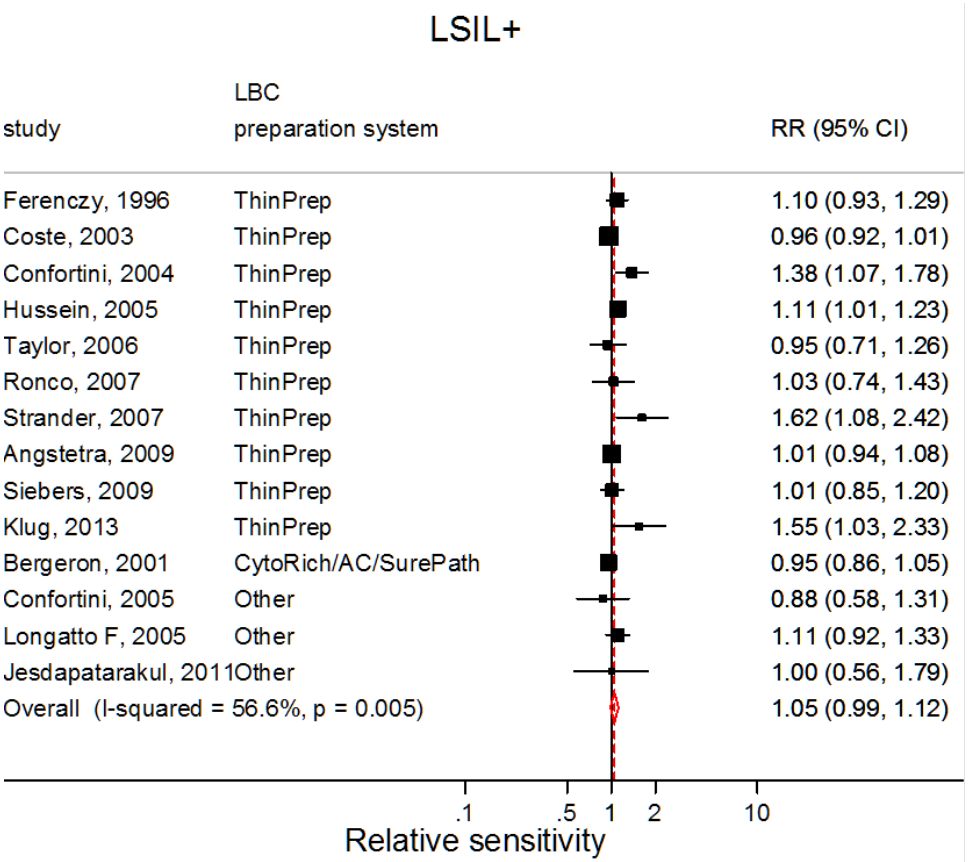


Figure 3. Forest plot of the relative sensitivity (left) and specificity (right) of LBC compared to CP, considering ASC-US+ as a positive test, to detect CIN2+.



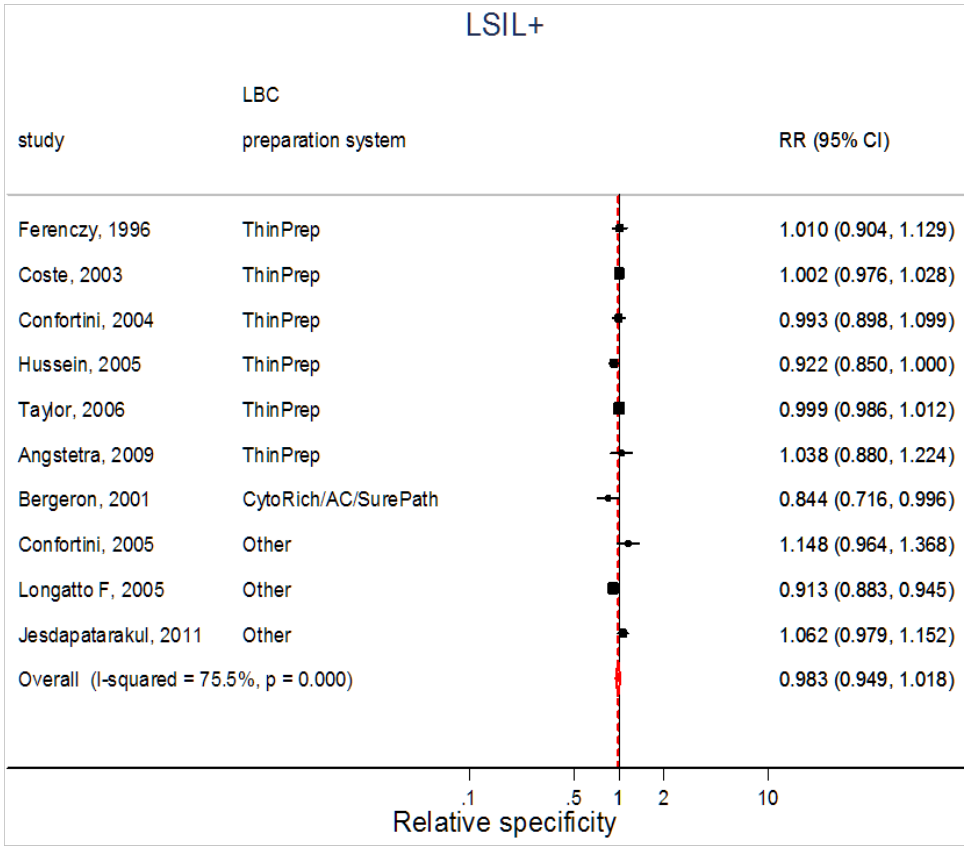


Figure 4. Forest plot of the relative sensitivity (left) and specificity (right) of LBC compared to CP, considering LSIL+ as a positive test, to detect CIN2+.

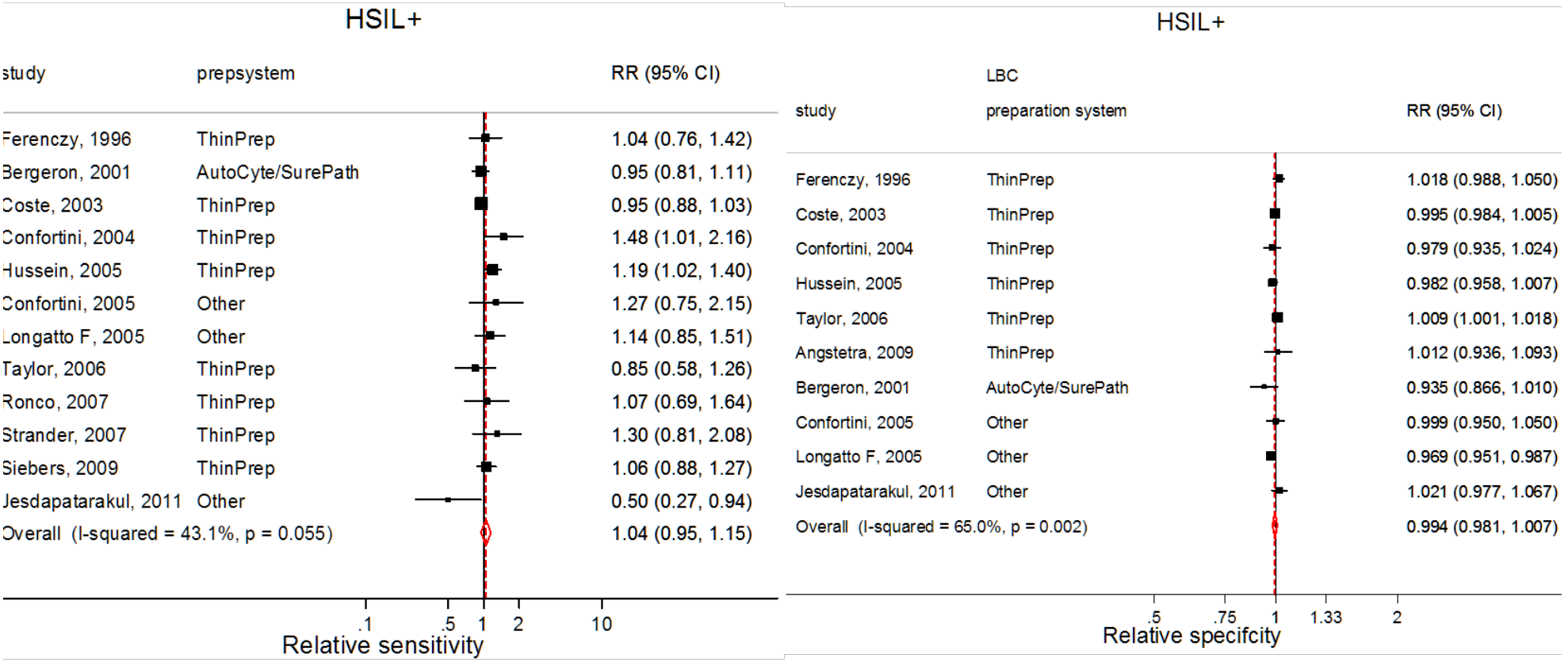


Figure 5. Forest plot of the relative sensitivity (left) and specificity (right) of LBC compared to CP, considering HSIL+ as a positive test, to detect CIN2+.

1.3.3. Judgement of specimen adequacy of liquid-based cytology specimen

The proportion of inadequate smears in LBC and CP as well as the ratio of these proportions are shown in forest plots and tables, separately by study design and by system used for the preparation of LBC. First, results are displayed for studies where a split-sample design was applied, where the LBC prepetation system was ThinPrep® (Table 4), SurePath®/AutoCyte®/ CytoRich® (Table 5) or other systems (Table 6), respectively. Subsequently the same type of results are shown for direct-to-vial studies.

1.3.3.1. Comparison of the adequacy of LBC and conventional Pap smears in split-sample studies

In split-sample studies, the average of unsatisfactory specimen was 1.6% (95% CI: 1.2%-2.1%, Table 4), 1.0% (95% CI: 0.5%-1.4%, Table 5) and 0% (95% CI: 0.0%-1.0%, Table 6), when ThinPrep, Autocyte or other systems were used for the preparation of an LBC specimen, respectively, whereas in the these proportions for CP in the comparative studies were 2.4% (95% CI: 1.5%-3.3%, Figure 6), 2.0% (95% CI: 1.2%-3.0%, Figure 8) and 1.5% (95% CI: 0.0%-6.9%, Figure 10). The use of AutoCyte resulted, on average, in a significant reduction in the proportion of inadequate specimen compared to CP (RR=0.46 [95% CI: 0.28–0.74], Figure 9). However for ThinPrep (RR=0.74 [95% CI: 0.54–1.02], Figure 7) and for the other LBC systems (RR=0.38 [95% CI: 0.08–1.84], Figure 11) this reduction was non-significant.

Table 4: Proportion of inadequate cervical cell samples using liquid-based cytology (LBC) or the conventional Pap smear (CP), and ratio of these proportions. Only studies where a split-sample design was used and where the applied LBC system was ThinPrep®.

	ThinPrep®		CP		Ratio LBC/CP	
Study	Proportion	95% CI	Proportion	95% CI	Ratio	95% CI
Hutchinson, 1991[45]	0.7%	0.2% - 2.0%	0.7%	0.2% - 2.0%	1.00	0.20 - 4.93
Wilbur, 1994[46]	0.3%	0.2% - 0.6%	0.6%	0.4% - 0.9%	0.56	0.26 - 1.20
Laverty, 1995[47]	5.2%	4.3% - 6.2%	1.5%	1.0% - 2.1%	3.50	2.34 - 5.23
McGoogan, 1996[48]	13.4%	11.8% - 15.2%	4.5%	3.6% - 5.7%	2.99	2.29 - 3.90
Wilbur, 1996[49]	1.2%	0.4% - 3.4%	1.9%	0.8% - 4.4%	0.60	0.15 - 2.49
Lee, 1997[50]	1.9%	1.6% - 2.2%	1.6%	1.3% - 1.9%	1.19	0.93 - 1.53
Roberts, 1997[51]	0.7%	0.6% - 0.8%	3.5%	3.4% - 3.7%	0.19	0.16 - 0.22
Corkill, 1998[52]	0.6%	0.3% - 1.2%	0.3%	0.1% - 0.7%	2.23	0.69 - 7.23
Boman, 1999[53]	1.1%	0.5% - 2.5%	1.7%	0.9% - 3.3%	0.63	0.21 - 1.90
Hutchinson, 1999[54]	2.5%	2.2% - 2.9%	0.6%	0.5% - 0.8%	4.25	3.15 - 5.72
McGoogan, 1999[55]	2.4%	1.4% - 4.2%	8.0%	5.9% - 10.7%	0.30	0.16 - 0.57
Shield, 1999[56]	6.3%	4.1% - 9.7%	17.3%	13.5% - 22.0%	0.37	0.22 - 0.60
Wang, 1999[57]	1.1%	0.6% - 2.0%	1.1%	0.6% - 2.0%	1.00	0.44 - 2.30
Monsonego, 2001[58]	0.5%	0.4% - 0.8%	0.5%	0.3% - 0.7%	1.12	0.66 - 1.89
Park, 2001[59]	1.0%	0.4% - 2.4%	1.2%	0.6% - 2.7%	0.83	0.26 - 2.71
Armstrong, 2002[60]	2.9%	2.4% - 3.7%	0.9%	0.6% - 1.4%	3.26	2.05 - 5.19
Biscotti, 2002[61]	0.2%	0.0% - 1.4%	0.2%	0.0% - 1.4%	1.00	0.06 - 15.93
Grace, 2002[62]	1.3%	0.8% - 2.2%	11.0%	9.2% - 13.1%	0.12	0.07 - 0.21
Luthra, 2002[63]	4.9%	3.7% - 6.4%	3.5%	2.6% - 4.8%	1.39	0.91 - 2.11
Ring, 2002[64]	0.8%	0.5% - 1.5%	2.7%	1.9% - 3.7%	0.31	0.16 - 0.62
Coste, 2003[32]	0.4%	0.2% - 0.8%	0.1%	0.0% - 0.3%	3.67	1.02 - 13.13
Malle, 2003[65]	0.3%	0.1% - 0.6%	1.3%	0.9% - 1.9%	0.21	0.09 - 0.48
Confortini, 2004[33]	0.7%	0.2% - 2.4%	0.3%	0.1% - 1.9%	2.00	0.18 - 21.94
Hussein, 2005[35]	0.7%	0.2% - 2.0%	4.3%	2.8% - 6.6%	0.16	0.05 - 0.53
Tuncer, 2005[66]	1.7%	1.3% - 2.1%	2.3%	1.9% - 2.8%	0.73	0.54 - 0.98
Davey, 2007[67]	1.8%	1.7% - 1.9%	3.1%	3.0% - 3.2%	0.58	0.53 - 0.62
Rahimi, 2008[68]	2.0%	1.0% - 3.7%	2.0%	1.0% - 3.7%	1.00	0.40 - 2.50

Treacy, 2009[69]	3.2%	3.1% - 3.4%	11.0%	10.7% - 11.3%	0.30	0.28 - 0.31
Halford, 2010[70]	0.9%	0.8% - 0.9%	3.6%	3.5% - 3.8%	0.24	0.22 - 0.26
Overall	**1.6%**	**1.2% - 2.1%**	**2.4**	**1.5% - 3.3%**	**0.74**	**0.54 – 1.02**

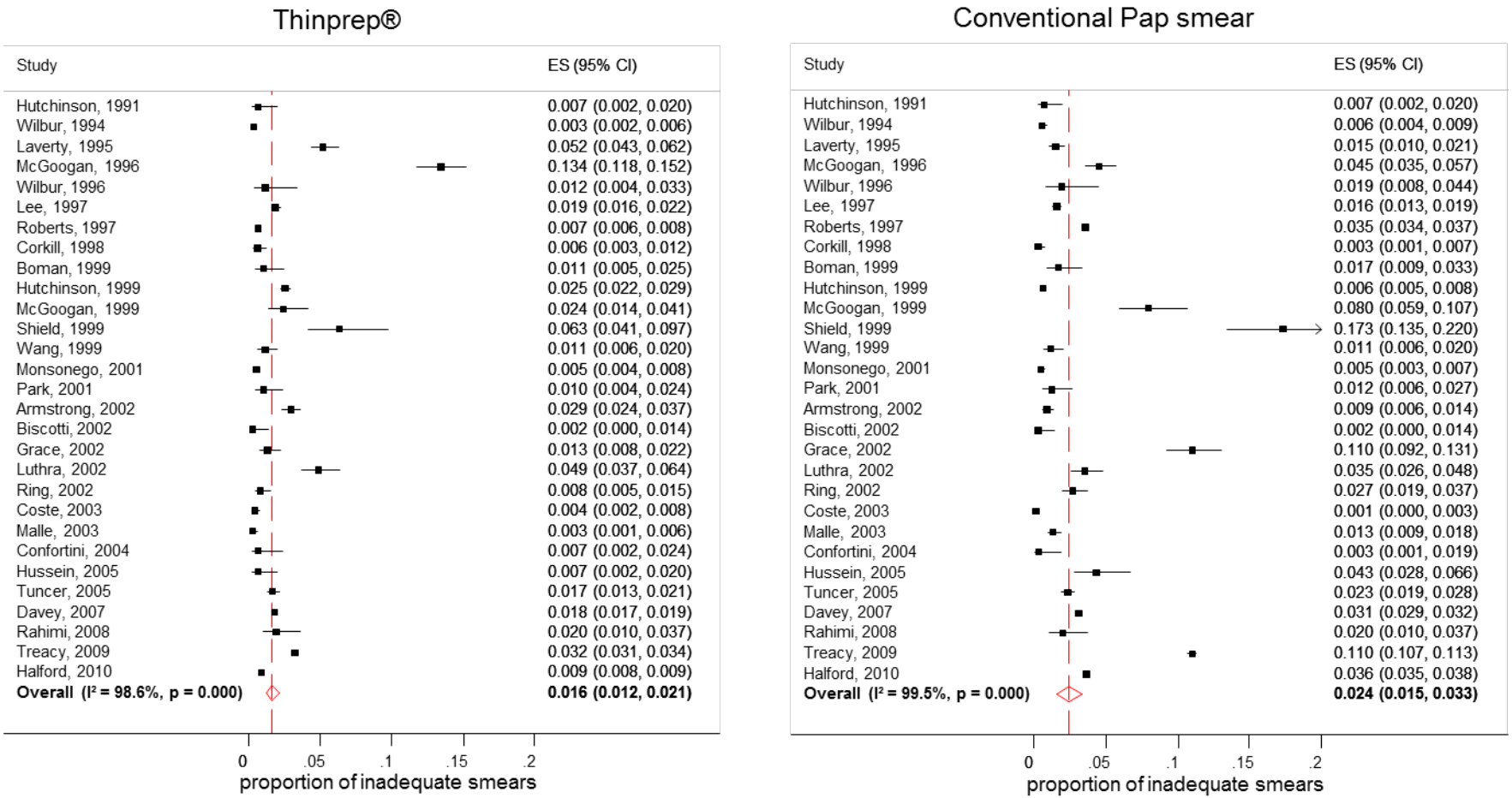


Figure 6: Proportion of inadequate cervical cell samples using ThinPrep® (left) or the conventional Pap smear (right), in studies with a split-sample design.

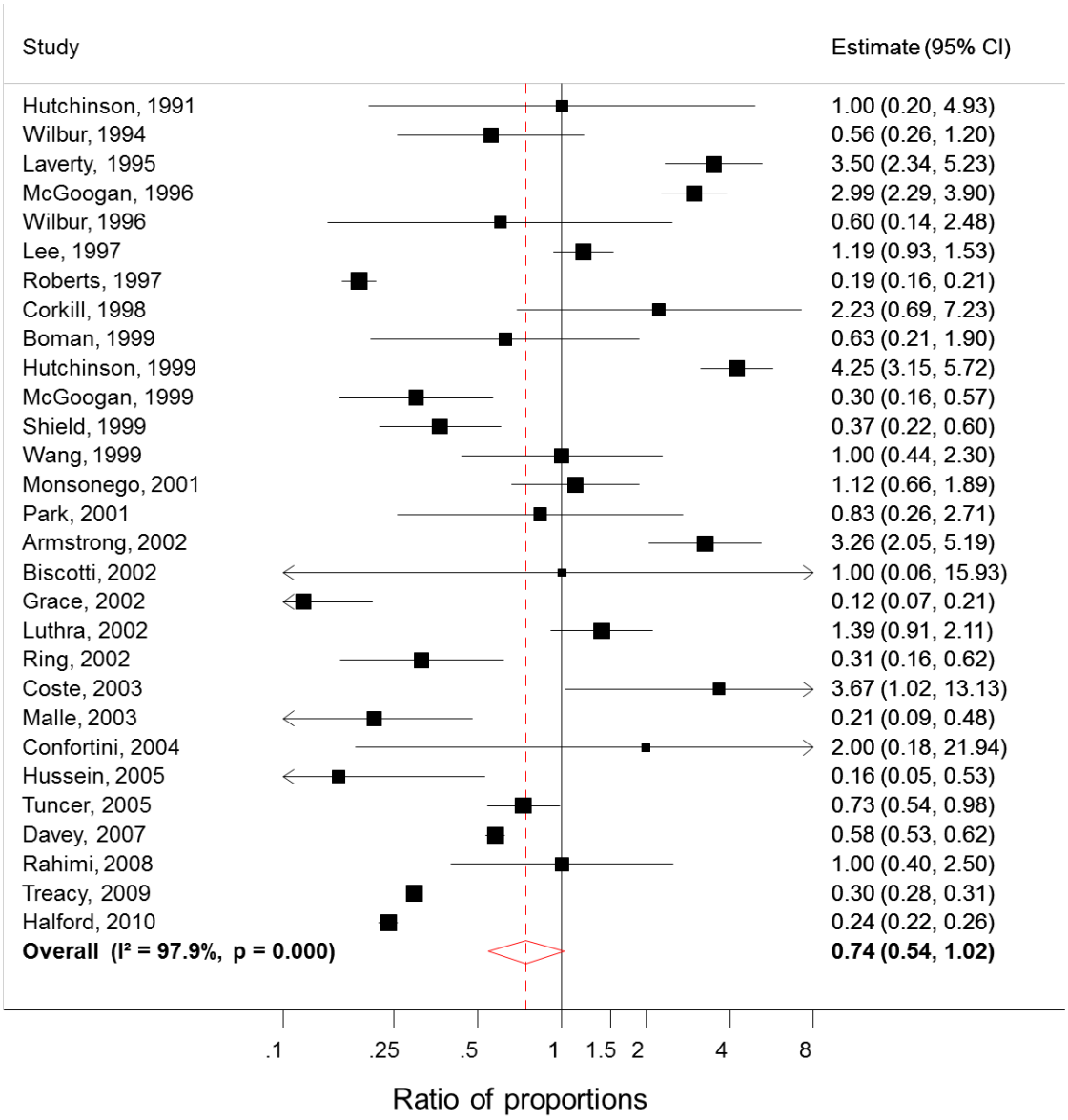


Figure 7: Figure 8. Ratio of the proportions of inadequate cervical cell samples of liquid based cytology (LBC) versus conventional Pap smear (CP). Only studies where a split-sample design was used and where the applied LBC system was ThinPrep®.

Table 5. Proportion of inadequate cervical cell samples using liquid-based cytology (LBC) or the conventional Pap smear (CP), and ratio of these proportions. Only studies where a split-sample design was used and where the applied LBC system was SurePath® (CytoRich® or AutoCyte®).

	Surepath®		**CP**		**Ratio LBC/CP**	
Study	Proportion	95% CI	Proportion	95% CI	Ratio	95% CI
Jensen, 1991[71]	0.1%	0.0% - 0.4%	1.8%	1.2% - 2.5%	0.07	0.02 - 0.28
Geyer, 1993[72]	0.2%	0.0% - 1.0%	1.1%	0.5% - 2.3%	0.17	0.02 - 1.38
McGoogan, 1996[48]	2.0%	1.4% - 2.8%	3.3%	2.5% - 4.3%	0.60	0.39 - 0.93
Sprenger, 1996[73]	0.6%	0.4% - 1.0%	2.4%	1.9% - 3.1%	0.26	0.15 - 0.43
Vassilakos, 1996[74]	3.8%	2.5% - 5.7%	5.2%	3.6% - 7.3%	0.72	0.42 - 1.25
Bishop, 1997[75]	0.8%	0.5% - 1.3%	0.4%	0.2% - 0.8%	2.00	0.86 - 4.66
Laverty, 1997[76]	0.3%	0.1% - 0.6%	2.6%	2.0% - 3.4%	0.11	0.05 - 0.25
Wilbur, 1997[77]	1.0%	0.4% - 3.0%	3.5%	1.9% - 6.3%	0.30	0.08 - 1.08
Bishop, 1998[78]	0.6%	0.5% - 0.8%	1.0%	0.8% - 1.2%	0.61	0.43 - 0.85
Howell, 1998[79]	0.1%	0.0% - 0.7%	0.4%	0.1% - 1.0%	0.33	0.04 - 3.20
Kunz, 1998[80]	0.2%	0.0% - 1.0%	0.2%	0.0% - 1.0%	1.00	0.06 - 15.95
Stevens, 1998[81]	4.0%	3.1% - 5.2%	0.4%	0.2% - 0.8%	11.20	4.50 - 27.87
Minge, 2000[82]	0.6%	0.4% - 1.0%	0.9%	0.6% - 1.4%	0.68	0.34 - 1.38
Bergeron, 2001[31]	0.8%	0.3% - 2.0%	11.6%	9.1% - 14.7%	0.07	0.03 - 0.19
Hessling, 2001[83]	0.4%	0.2% - 0.7%	0.7%	0.5% - 1.2%	0.50	0.23 - 1.11
Robyr, 2002[84]	3.7%	2.6% - 5.3%	1.3%	0.7% - 2.4%	2.80	1.37 - 5.72
Harkness, 2003[85]	2.7%	2.2% - 3.3%	5.9%	5.1% - 6.8%	0.45	0.35 - 0.59
Bowditch, 2012[86]	0.2%	0.1% - 0.5%	4.1%	3.3% - 5.0%	0.04	0.02 - 0.12
Overall	**0.9%**	**0.5% - 1.4%**	**2.0%**	**1.2% - 3.0%**	**0.46**	**0.28 – 0.74**

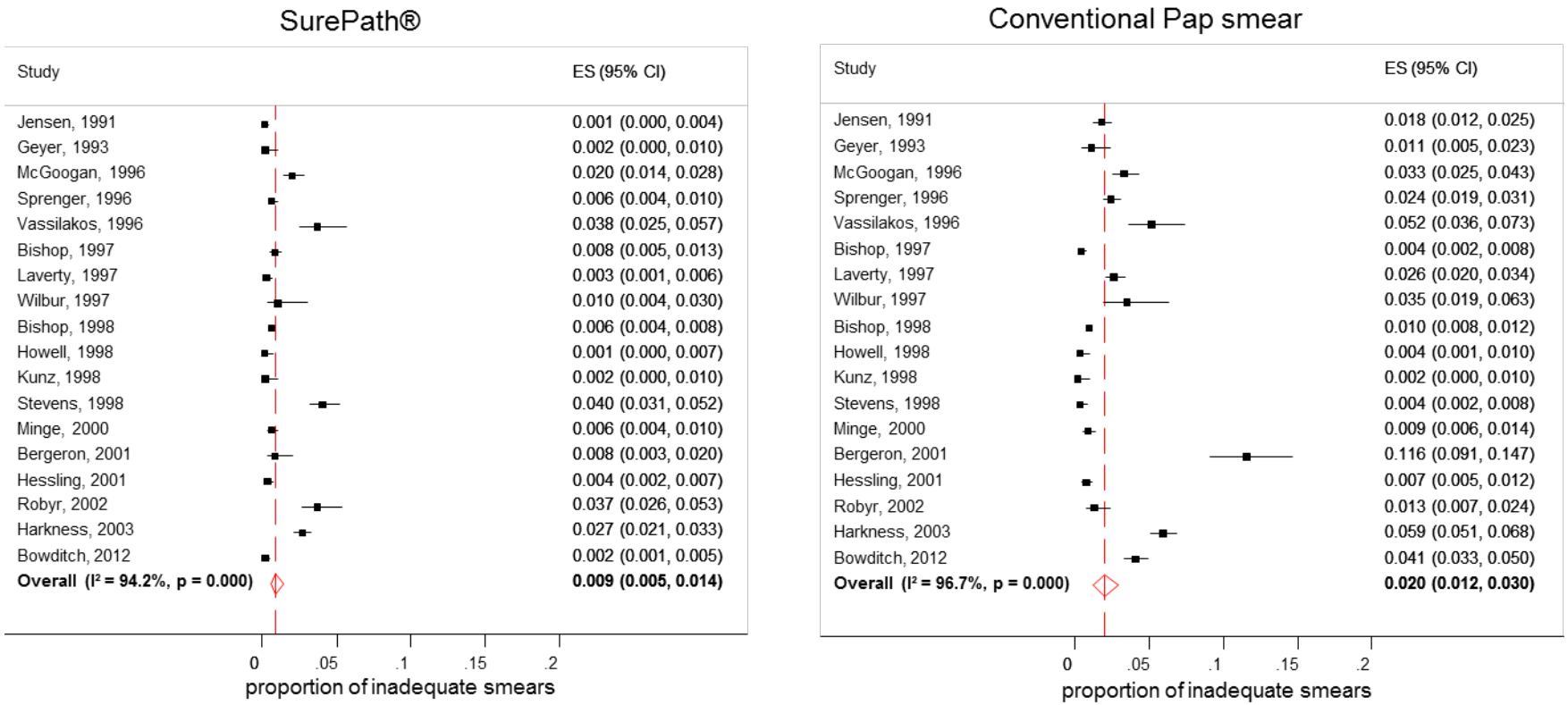


Figure 9: Proportion of inadequate cervical cell samples using SurePath® (CytoRich® or AutoCyte®, left) or the conventional Pap smear (right), in studies with a split-sample design.

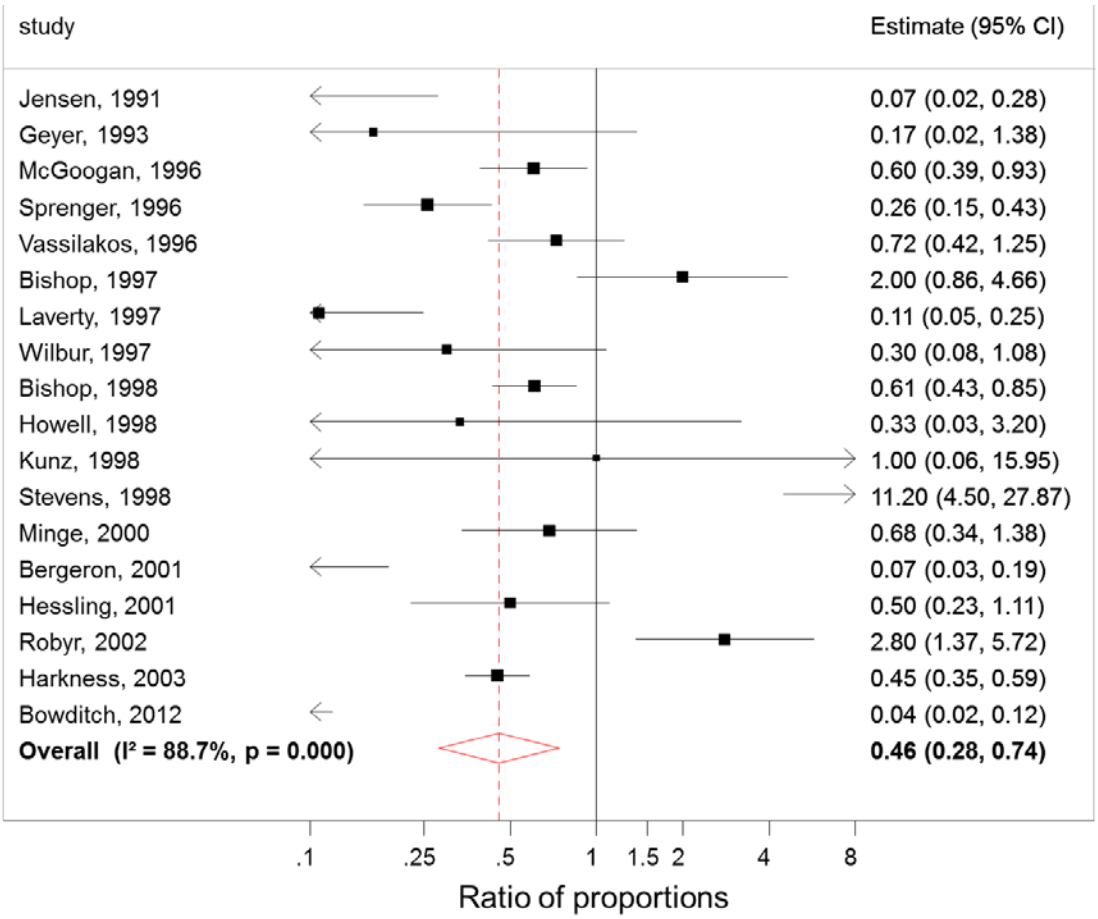


Figure 10. Ratio of the proportions of inadequate cervical cell samples of liquid based cytology (LBC) versus conventional Pap smear (CP). Only studies where a split-sample design was used and where the applied LBC system was SurePath® (CytoRich® or AutoCyte®).

Table 6. Proportion of inadequate cervical cell samples using liquid-based cytology (LBC) or the conventional Pap smear (CP), and ratio of these proportions. Only studies where a split-sample design was used and where the applied LBC system was neither ThinPrep® nor SurePath® (CytoRich® or AutoCyte®).

		Other LBC		**CP**		**Ratio LBC/CP**	
Study	LBC-system	Proportion	95% CI	Proportion	95% CI	Ratio	95% CI
Khalbuss, 2000[87]	ThinSpin®	0.0%	0.0% - 0.5%	0.1%	0.0% - 0.7%	0.33	0.01 - 8.16
Mattosinho 2004[88]	In house	0.6%	0.3% - 1.5%	0.0%	0.0% - 0.5%	11.1	0.61 - 199.9
Longatto-F, 2005[36]	DNACitoliq®	1.4%	0.8% - 2.3%	11.6%	9.8% - 13.8%	0.12	0.07 - 0.20
Rosenthal, 2006[89]	PapSpin®	0.2%	0.0% - 1.2%	0.4%	0.1% - 1.5%	0.50	0.05 - 5.50
Jesdapatarakul, 2011[40]	Liqui-Prep®	0.0%	0.0% - 1.9%	2.1%	0.8% - 5.2%	0.11	0.01 - 2.05
Overall		**0.3%**	**0.0%- 1.0%**	**1.5%**	**0.0% - 6.9%**	**0.38**	**0.08 – 1.84**

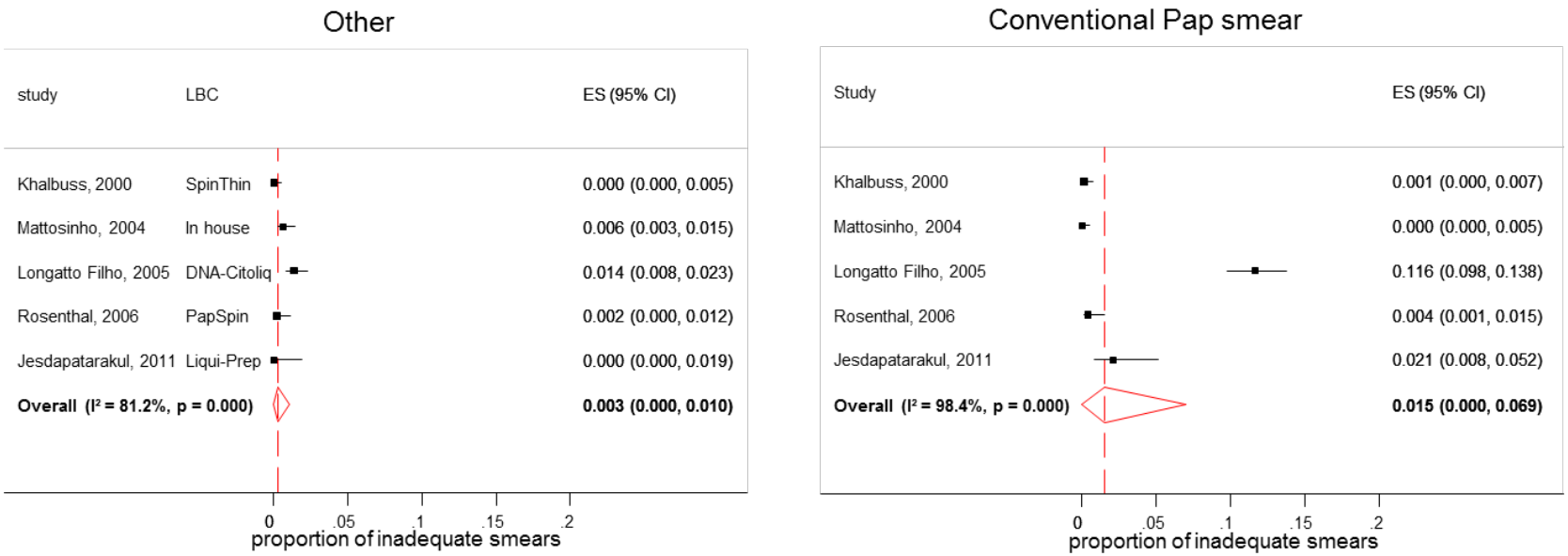


Figure 11. Proportion of inadequate cervical cell samples using different liquid based cytology (LBC) systems (left) or the conventional Pap smear (right), in studies with a split-sample design.

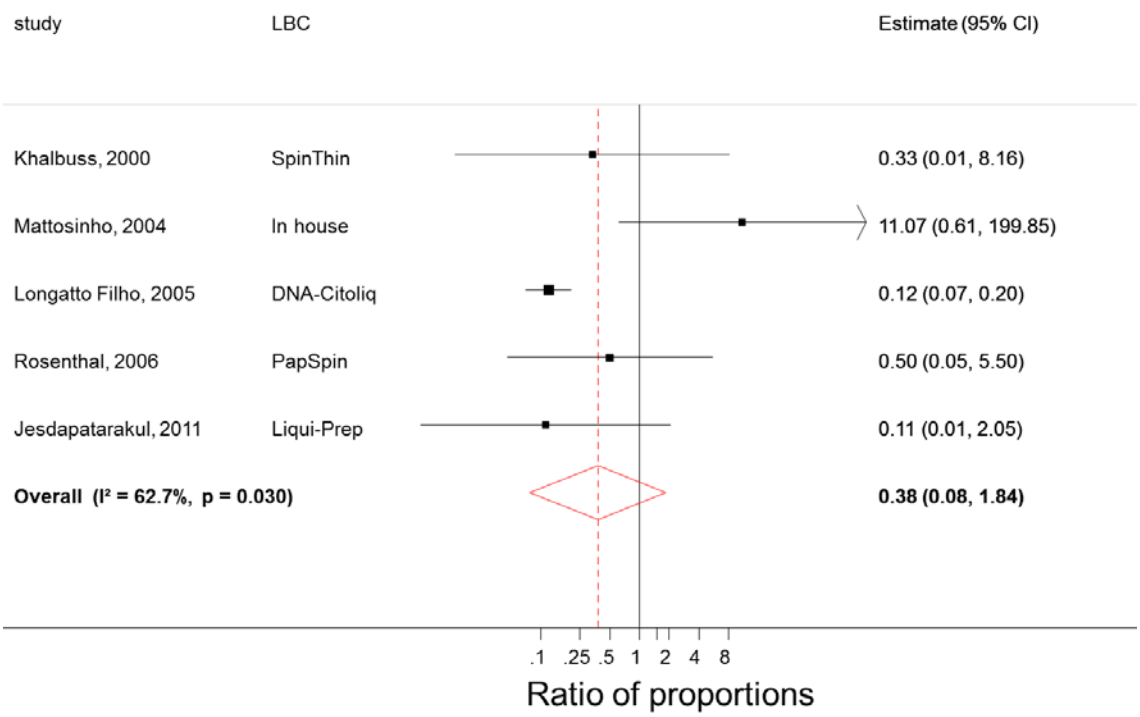


Figure 12: Ratio of the proportions of inadequate cervical cell samples of liquid based cytology (LBC) versus conventional Pap smear (CP). Only studies where a split-sample design was used and where the applied LBC system was neither ThinPrep® nor SurePath® (CytoRich® or AutoCyte®).

1.3.3.2. Comparison of the adequacy of LBC and conventional Pap smears in direct-to vial studies

In direct-to-vial studies, the average of unsatisfactory specimen was 0.9% (95% CI: 0.6%-1.2%, Table 7), 0.3% (95% CI: 0.2%-0.4%, Table 8) and 0.2% (95% CI: 0.0%-0.5%, Table 9), when ThinPrep, Autocyte or other systems were used for the preparation of an LBC specimen, respectively, whereas in the these proportions for CP in the comparative studies were 1.8% (95% CI: 0.9%-2.9%, Figure 12), 1.7% (95% CI: 1.1%-2.5%, Figure 14), 0.2% (95% CI: 0.0%-0.5%, Figure 16). Both, ThinPrep (RR=0.67 [95% CI: 0.47-0.97], Figure 13) and AutoCyte specimen 0.21 [95% CI: 0.14-0.31, Figure 15] were less unsatisfactory than CP. However, the reduction was larger when Autocyte was used compared to ThinPrep. The reduction in inadequate specimen in LBC prepared with other systems was not significant (RR=0.47 [95% CI: 0.16-1.41], Figure 17).

Table 7. Proportion of inadequate cervical cell samples using liquid-based cytology (LBC) or the conventional Pap smear (CP), and ratio of these proportions. Only studies where a direct-to-vial design was used and where the applied LBC system was ThinPrep®.

	Thinprep®		CP		Ratio LBC/CP	
Study	Proportion	95% CI	Proportion	95% CI	Ratio	95% CI
Weintraub, 1997[90]	0.3%	0.2% - 0.4%	0.7%	0.6% - 0.9%	0.39	0.27 - 0.55
Bolick, 1998[91]	0.3%	0.2% - 0.4%	1.1%	1.0% - 1.2%	0.27	0.19 - 0.39
Dupree, 1998[92]	3.8%	3.5% - 4.1%	2.0%	1.8% - 2.2%	1.89	1.68 - 2.12
Johnson, 1998[93]	0.3%	0.1% - 1.0%	0.4%	0.2% - 1.0%	0.83	0.19 - 3.71
Papillo, 1998[94]	0.4%	0.3% - 0.5%	0.2%	0.2% - 0.3%	1.63	1.04 - 2.56
Quddus, 1998[95]	0.5%	0.4% - 0.6%	0.7%	0.6% - 0.8%	0.72	0.56 - 0.92
Carpenter, 1999[96]	0.3%	0.2% - 0.6%	0.6%	0.4% - 0.9%	0.49	0.22 - 1.07
Diaz-Rosario, 1999[97]	0.7%	0.6% - 0.7%	0.2%	0.2% - 0.3%	3.05	2.54 - 3.67
Guidos, 1999[98]	0.4%	0.3% - 0.6%	1.2%	0.9% - 1.5%	0.37	0.26 - 0.55
Yeoh, 1999[99]	0.6%	0.5% - 0.7%	1.4%	1.1% - 1.7%	0.41	0.31 - 0.55
Ferris, 2000[100]	1.2%	0.7% - 2.1%	3.8%	3.0% - 4.7%	0.32	0.17 - 0.58
Weintraub, 2000[101]	0.2%	0.2% - 0.2%	0.3%	0.2% - 0.3%	0.73	0.57 - 0.94
Obwegeser, 2001[102]	1.4%	0.8% - 2.3%	0.1%	0.0% - 0.6%	13.88	1.83 - 105.32
Baker, 2002[103]	0.8%	0.6% - 1.1%	0.7%	0.5% - 0.9%	1.17	0.76 - 1.81
Bentz-a, 2002[104]	0.9%	0.8% - 1.0%	0.5%	0.4% - 0.6%	1.76	1.46 - 2.12
Bentz-b, 2002[104]	1.3%	1.1% - 1.5%	1.0%	0.8% - 1.2%	1.33	1.08 - 1.65
SCSP, 2002[105]	1.9%	1.7% - 2.1%	8.0%	7.6% - 8.4%	0.23	0.21 - 0.26
Cheung, 2003[106]	0.3%	0.3% - 0.4%	0.5%	0.5% - 0.5%	0.66	0.60 - 0.74
Hoelund, 2003[107]	3.0%	2.7% - 3.3%	8.8%	8.5% - 9.1%	0.34	0.31 - 0.38
Kruger, 2003[108]	2.6%	0.5% - 13.2%	2.4%	0.4% - 12.3%	1.08	0.07 - 16.63
Moss, 2003[109]	2.0%	1.8% - 2.1%	9.0%	8.8% - 9.2%	0.22	0.20 - 0.24
Negri, 2003[110]	1.0%	0.2% - 5.5%	5.4%	2.5% - 11.3%	0.19	0.02 - 1.51
Welsh Pilot, 2003[111]	0.5%	0.5% - 0.6%	7.3%	7.1% - 7.5%	0.07	0.06 - 0.08
Confortini, 2004[33]	0.7%	0.2% - 2.4%	0.3%	0.1% - 1.9%	2.00	0.18 - 21.94
Hodgson, 2005[112]	1.1%	1.1% - 1.2%	0.8%	0.7% - 0.8%	1.46	1.32 - 1.62
Duggan, 2006[113]	0.4%	0.2% - 0.8%	0.5%	0.3% - 0.8%	0.97	0.40 - 2.32
Ronco, 2006[114]	2.6%	2.4% - 2.8%	4.1%	3.9% - 4.4%	0.63	0.57 - 0.70
Schledermann, 2006[115]	2.9%	2.7% - 3.1%	7.9%	7.6% - 8.2%	0.36	0.34 - 0.39
Taylor, 2006 [37]	2.2%	1.7% - 2.8%	0.8%	0.5% - 1.2%	2.85	1.72 - 4.72
Nance, 2007[116]	1.9%	1.8% - 2.0%	13.6%	13.4% - 13.8%	0.14	0.13 - 0.15
Williams, 2006[117]	0.6%	0.6% - 0.7%	0.3%	0.3% - 0.3%	2.13	1.88 - 2.43
Strander, 2007[44]	0.3%	0.2% - 0.5%	0.7%	0.6% - 0.9%	0.43	0.24 - 0.76
Maccallini, 2008[118]	1.3%	1.0% - 1.7%	4.3%	3.7% - 5.0%	0.30	0.23 - 0.41

Siebers, 2009[42]	0.3%	0.3% - 0.4%	1.1%	1.0% - 1.2%	0.30	0.25 - 0.36
Klug, 2013[43]	0.3%	0.2% - 0.4%	0.0%	0.0% - 0.1%	7.18	2.55 - 20.19
Overall	**0.9%**	**0.6% - 1.2%**	**1.8%**	**0.9% - 2.9%**	**0.67**	**0.47 – 0.97**

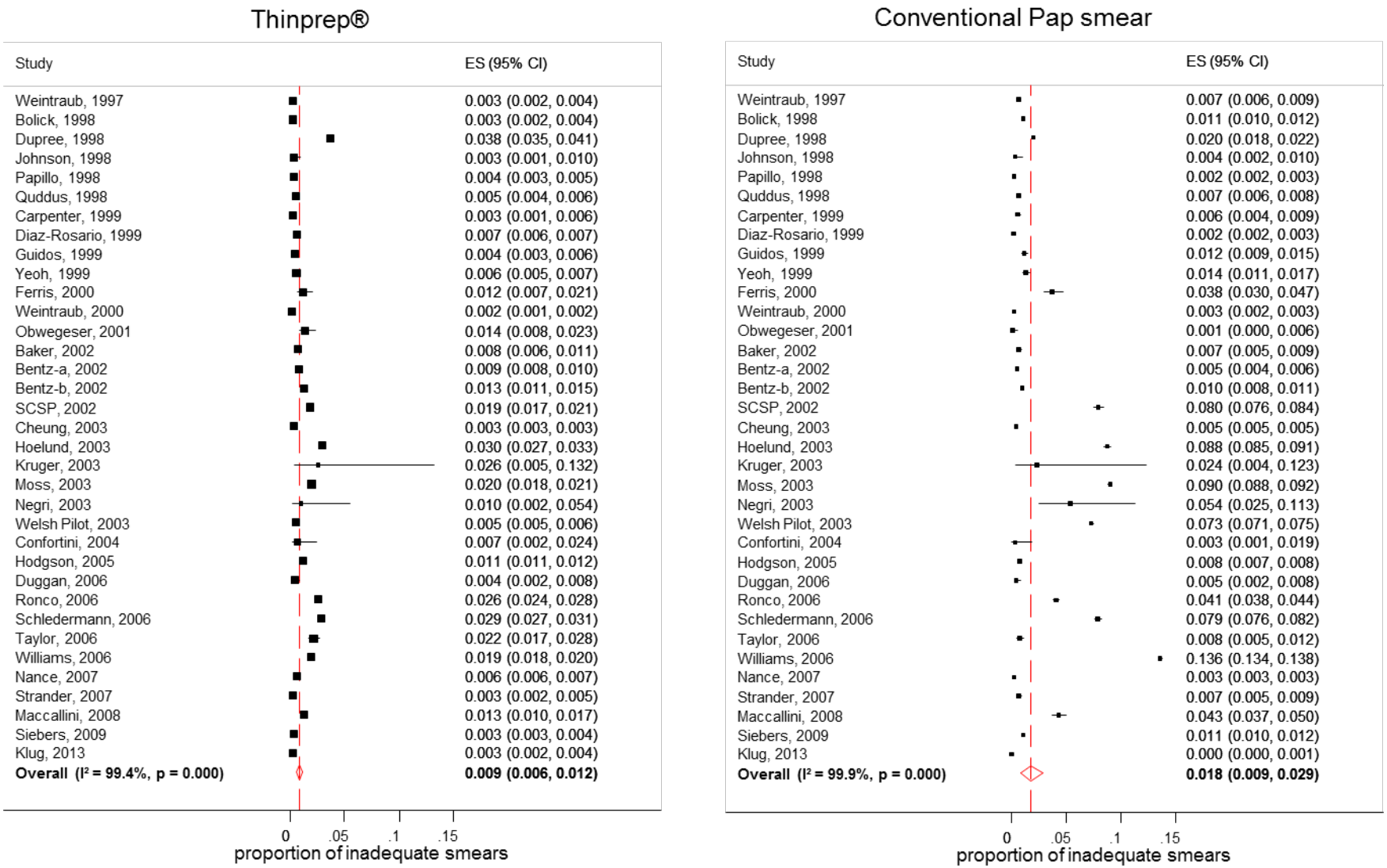


Figure 13: Proportion of inadequate cervical cell samples using ThinPrep® (left) or the conventional Pap smear (right), in studies with a direct-to-vial design.

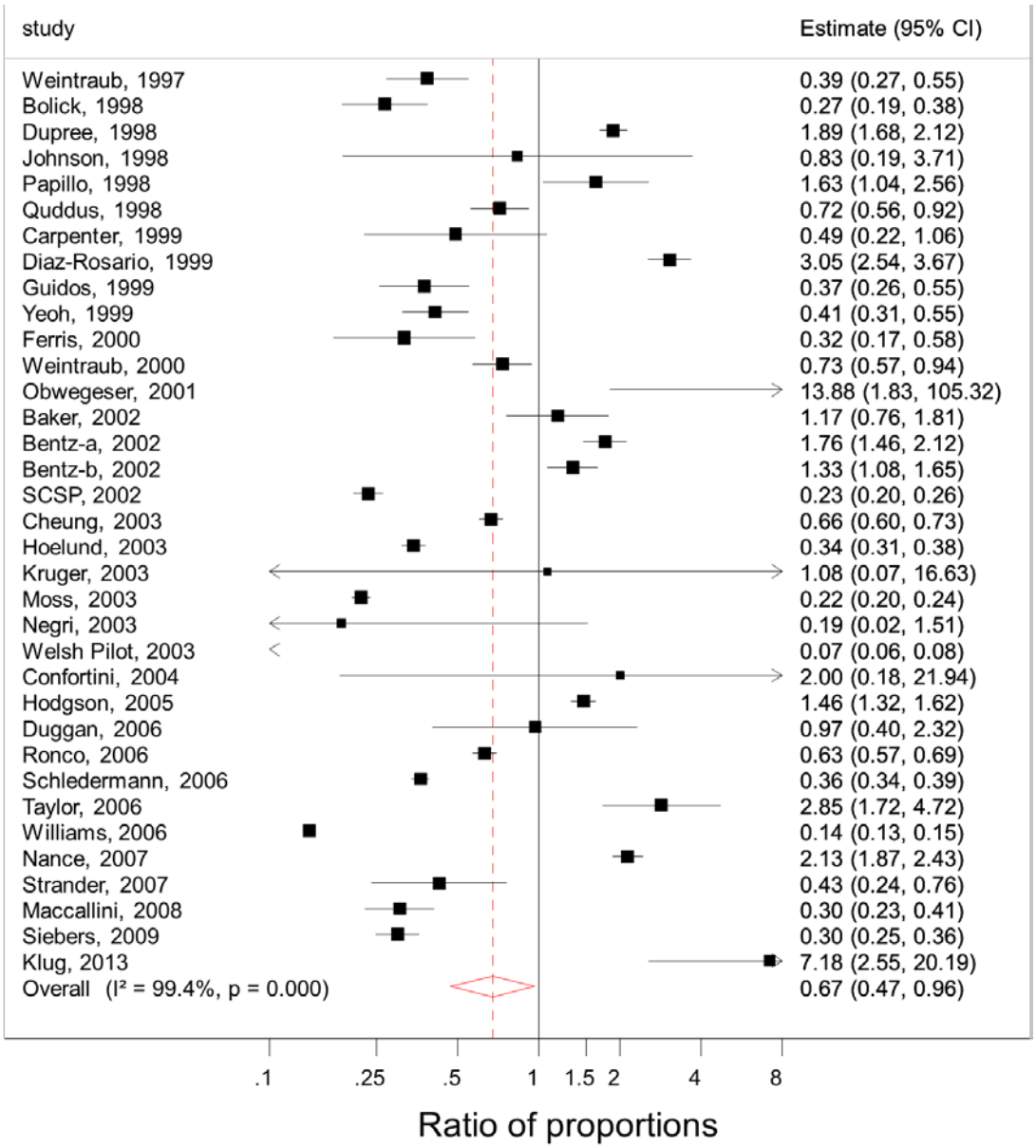


Figure 14. Ratio of the proportions of inadequate cervical cell samples of liquid based cytology (LBC) versus conventional Pap smear (CP). Only studies where a direct-to-vial design was used and where the applied LBC system was ThinPrep®.

Table 8. Proportion of inadequate cervical cell samples using liquid-based cytology (LBC) or the conventional Pap smear (CP), and ratio of these proportions. Only studies where a direct-to-vial design was used and where the applied LBC system was SurePath® (CytoRich® or AutoCyte®).

	SurePath®		**CP**		**Ratio LBC/CP**	
Study	Proportion	95% CI	Proportion	95% CI	Ratio	95% CI
Cheuvront, 1998[119]	1.5%	1.1% - 2.1%	2.0%	1.9% - 2.2%	0.75	0.52 - 1.07
Vassilakos, 1998[120]	0.4%	0.3% - 0.5%	2.0%	1.8% - 2.2%	0.21	0.17 - 0.25
Saurel, 1999[121]	0.0%	0.0% - 0.0%	1.4%	1.3% - 1.5%	0.02	0.02 - 0.03
Vassilakos, 1999[122]	0.1%	0.1% - 0.2%	1.5%	1.4% - 1.6%	0.10	0.08 - 0.12
Tench, 2000[123]	0.4%	0.2% - 0.8%	2.9%	2.6% - 3.3%	0.14	0.07 - 0.27
Marino, 2001[124]	0.1%	0.1% - 0.2%	0.3%	0.2% - 0.5%	0.29	0.15 - 0.57
Royer, 2001[125]	0.5%	0.4% - 0.7%	0.8%	0.6% - 1.0%	0.65	0.45 - 0.93
Day, 2002[126]	0.1%	0.1% - 0.2%	4.1%	3.9% - 4.2%	0.03	0.02 - 0.05
Moss, 2003[109]	0.9%	0.8% - 1.1%	9.1%	8.9% - 9.4%	0.10	0.09 - 0.12
Colgan, 2004[127]	0.2%	0.2% - 0.3%	0.6%	0.6% - 0.6%	0.41	0.38 - 0.45
Fremont-S, 2004[2]	0.2%	0.2% - 0.3%	0.5%	0.5% - 0.6%	0.42	0.34 - 0.51
Sass, 2004[128]	0.2%	0.1% - 0.3%	0.9%	0.6% - 1.1%	0.19	0.10 - 0.34
Longatto-F., 2005[129]	0.6%	0.2% - 2.3%	1.1%	0.9% - 1.3%	0.58	0.14 - 2.34
Kirschner, 2006[130]	0.3%	0.3% - 0.4%	2.3%	2.2% - 2.4%	0.15	0.13 - 0.17
Nance, 2007[116]	0.2%	0.2% - 0.2%	0.3%	0.3% - 0.3%	0.57	0.47 - 0.68
Stark, 2007[131]	0.2%	0.2% - 0.3%	0.6%	0.5% - 0.6%	0.39	0.35 - 0.45
Sykes, 2008[38]	2.7%	1.5% - 4.6%	9.1%	6.7% - 12.1%	0.29	0.16 - 0.55
Beerman, 2009[132]	0.1%	0.1% - 0.2%	0.9%	0.8% - 0.9%	0.15	0.11 - 0.21
Overall	**0.3%**	**0.2% - 0.4%**	**1.7%**	**1.1% - 2.5%**	**0.21**	**0.14 - 0.31**

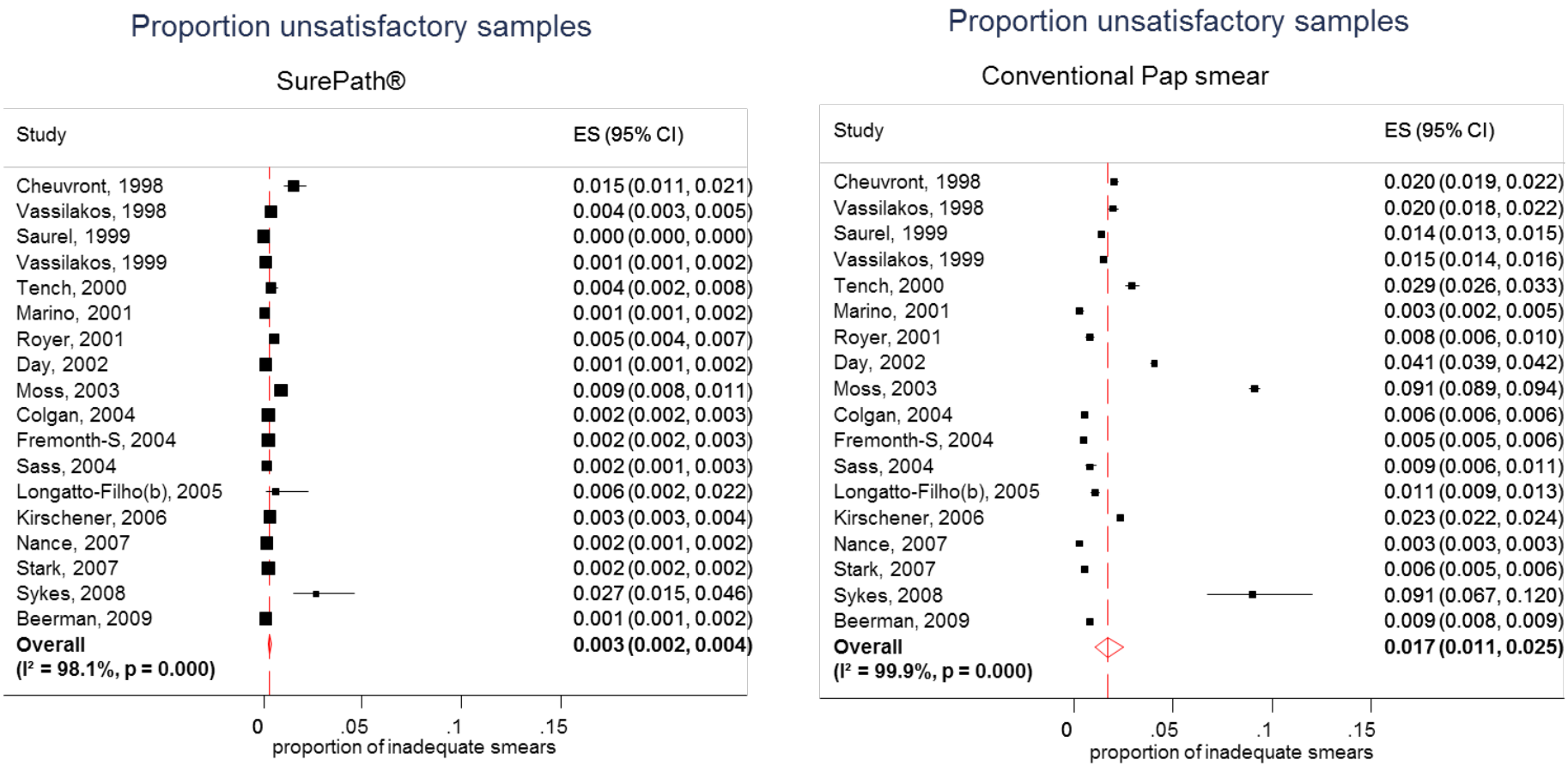


Figure 15. Proportion of inadequate cervical cell samples using SurePath® (CytoRich® or AutoCyte®, left) or the conventional Pap smear (right), in studies with a direct-to-vial design.

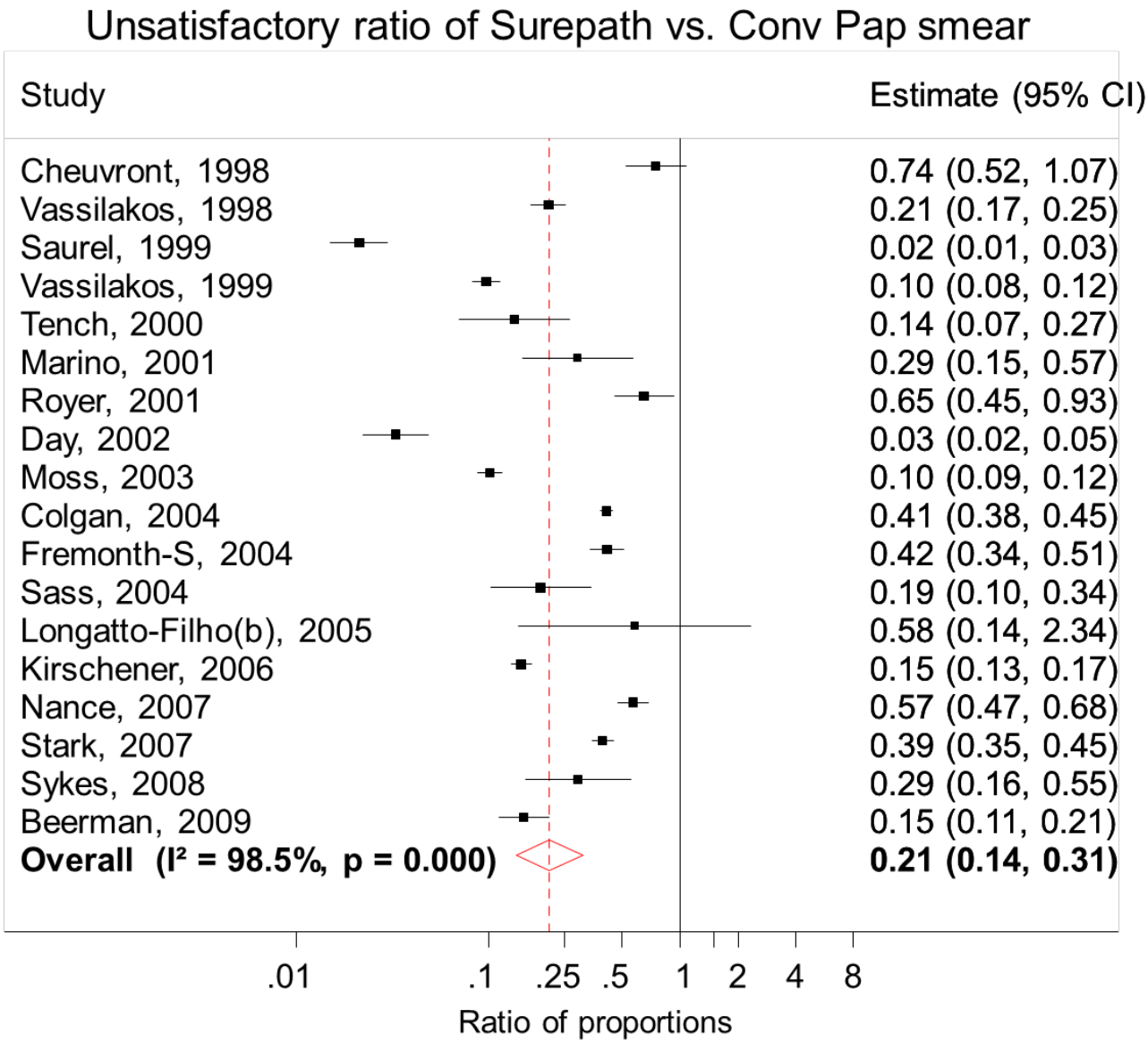


Figure 16. Ratio of the proportions of inadequate cervical cell samples of liquid based cytology (LBC) versus conventional Pap smear (CP). Only studies where a direct-to-vial design was used and where the applied LBC system was SurePath® (CytoRich® or AutoCyte®).

Table 9. Proportion of inadequate cervical cell samples using liquid-based cytology (LBC) or the conventional Pap smear (CP), and ratio of these proportions. Only studies where a direct-to-vial design was used and where the applied LBC system was neither ThinPrep® nor SurePath® (CytoRich® or AutoCyte®).

		OtherLBC		**CP**		**Ratio LBC/CP**	
Study	**LBC-system**	Proportion	95% CI	Proportion	95% CI	Ratio	95% CI
Bergeron, 2003[133]	CYTO-screen®	0.2%	0.1% - 0.3%	1.0%	0.8% - 1.2%	0.17	0.11 - 0.28
Longatto-F, 2005[129]	DNA-Citoliq®	0.6%	0.3% - 1.2%	1.1%	0.9% - 1.3%	0.55	0.27 - 1.13
Boon, 2006[134]	Pap-Spin®	0.4%	0.2% - 0.5%	0.3%	0.2% - 0.3%	0.35	0.88 - 2.07
Park, 2007[135]	Liqui-PREP®	0.0%	0.0% - 0.1%	0.1%	0.1% - 0.1%	0.37	0.15 - 0.91
Overall		**0.2%**	**0.0%- 0.5%**	**0.5%**	**0.1%- 1.1%**	**0.47**	**0.16 - 1.41**

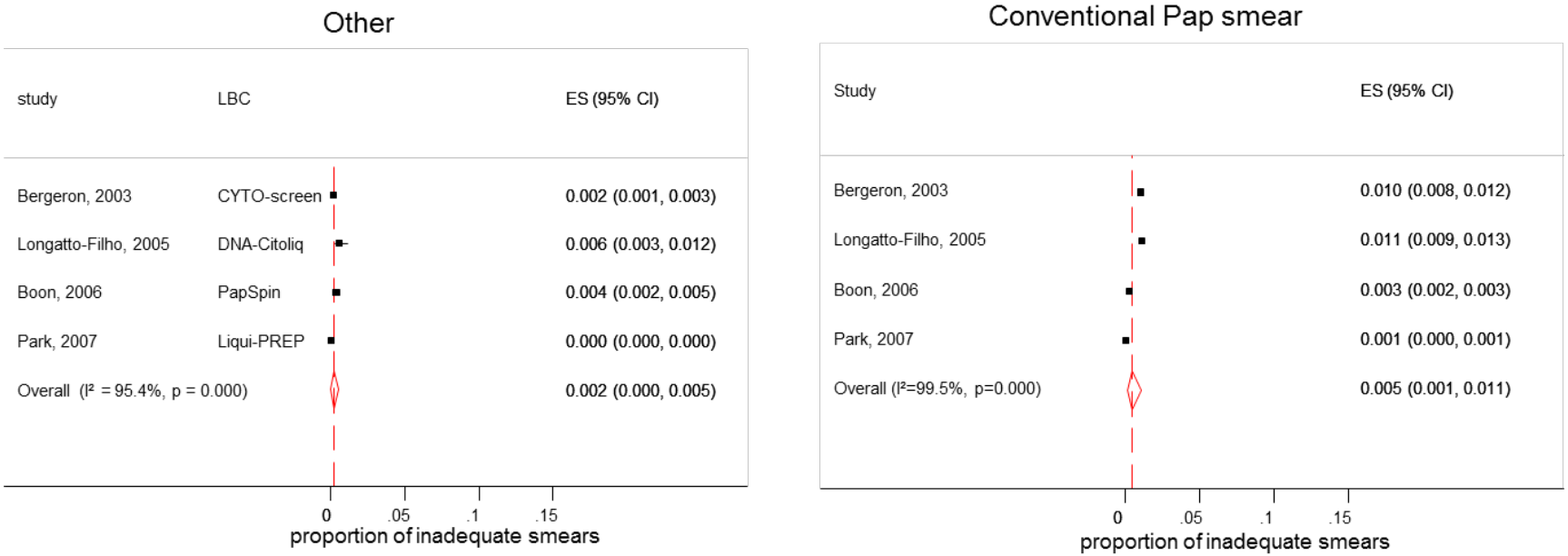


Figure 17. Proportion of inadequate cervical cell samples using different liquid based cytology (LBC) systems (left) or the conventional Pap smear (right), in studies with a direct-to-vial design.

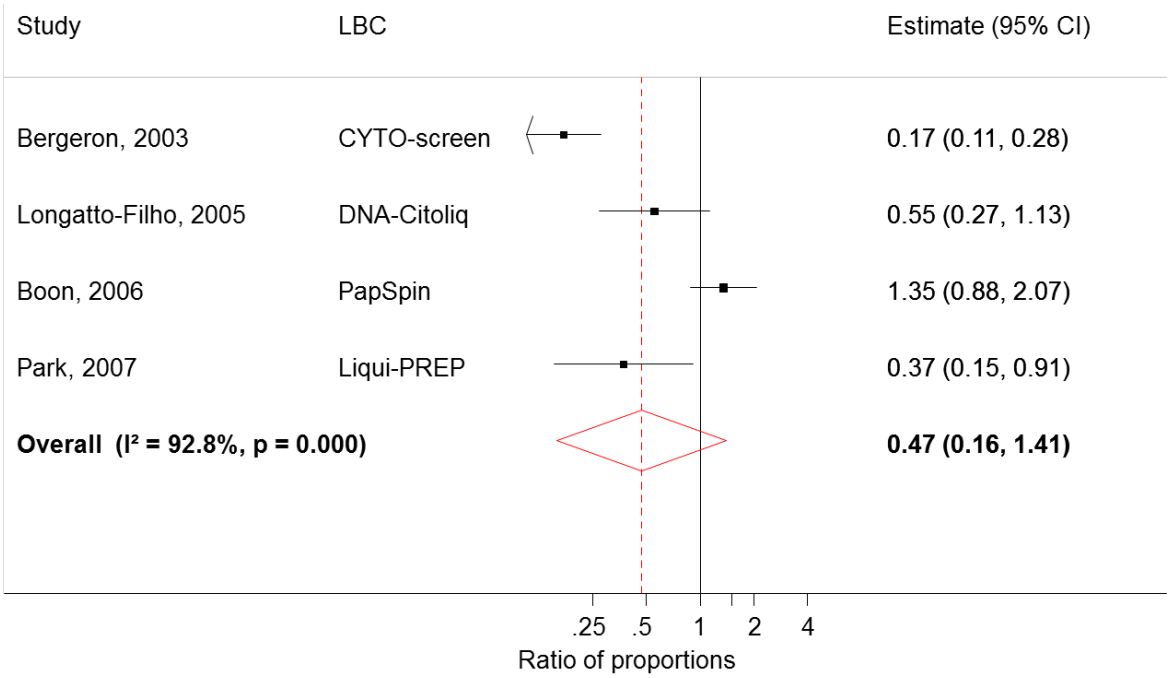


Figure 18. Ratio of the proportions of inadequate cervical cell samples of liquid based cytology (LBC) versus conventional Pap smear (CP). Only studies where a split-sample design was used and where the applied LBC system was neither ThinPrep® nor SurePath® (CytoRich® or AutoCyte®).

Table 10. Pooled proportion of unsatisfactory cervical cell specimen prepared with LBC vs CP, by LBC preparation system and by study design.

Study Design	LBC system	LBC Overall	LBC 95% CI	CP Overall	CP 95% CI	Ratio (LBC/CP) Overall	Ratio (LBC/CP) 95% CI
Split-Sample	ThinPrep®	1.6%	1.2-2.1%	2.4%	1.5-3.3%	0.74	0.54-1.02
	SurePath®	0.9%	0.5-1.4%	2.0%	1.2-3.0%	0.46	0.28-0.74
	Other	0.3%	0.0- 1.0%	1.5%	0.0-6.9%	0.38	0.08-1.84
Direct-to-vial	ThinPrep®	0.9%	0.6-1.2%	1.8%	0.9-2.9%	0.67	0.47-0.97
	SurePath®	0.3%	0.2-0.4%	1.7%	1.1-2.5%	0.21	0.14-0.31
	Other	0.2%	0.0- 0.5%	0.2%	0.0- 0.5%	0.47	0.16-1.41

1.3.3.3. Comparison of the adequacy of ThinPrep and SurePath LBCspecimen

Eight studies[48,76,77,109,116,136-138] were retrieved where the proportion of unsatisfactory specimen could be compared in specimen prepared with ThinPrep and/or SurePath (three in split-sample and five in direct-to-vial studies). In both types of studies, the ratio of unsatisfactory smears (SurePath/ThinPrep) was significantly lower than unity: RR=0.16 [95% CI=0.05-0.48 and 0.46 [95% CI: 0.30-0.69], in split-sample and direct-to-vial studies respectively, see Figure 17

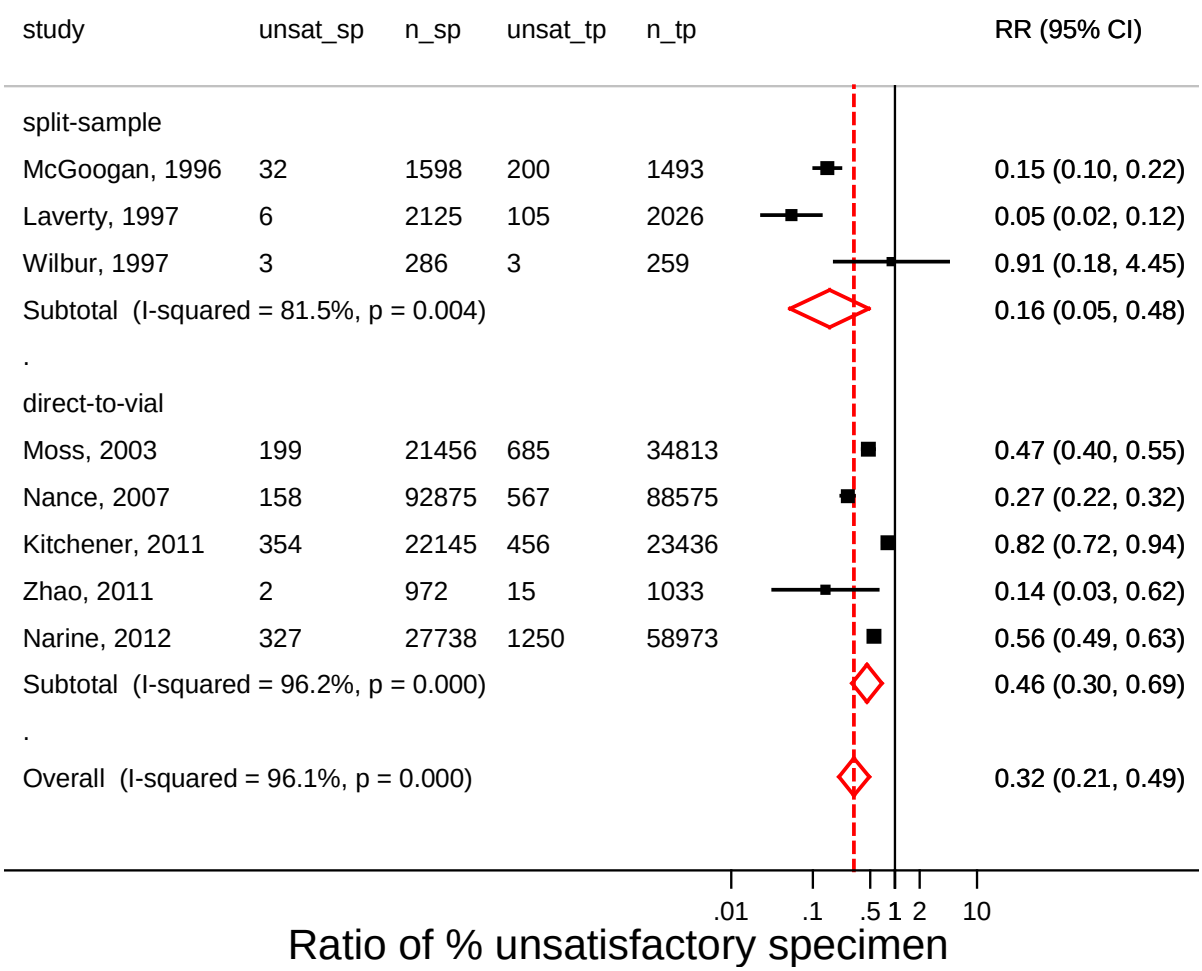


Figure 19. Ratio of the proportion of unsatisfactory LBC specimen prepared with SurePath (AutoCyto, CytoRich) versus ThinPrep.
Unsat_sp & unsat_tp: number of unsatisfactory specimen observed in Surepath and ThinPrep samples; n_sp & n_tp: number of SurePath and ThinPrep specimen.

1.3.4. Duration of microscopic interpretation

The average time needed to perform the microscopic interpretation of an LBC and CP specimen was provided in twelve studies. No meta-analytical pooling could be performed since the standard deviation, standard error or 95% CI were not reported in most of the studies. Therefore a simple average of the reading times was calculated. One study (Geyer, 1993) was excluded because of the outlying short duration (average of 60 sec) of LBC. Among the other eleven studies, the average duration of interpretation varied between 192 and 389 seconds for LBC and 270 and 450 seconds for CP.

The average reading times were 263 and 358 seconds. On average, the interpretation of an LBC required 27% less time than a CP.

Table 11. Time needed to interpret liquid-based (LBC) and conventional Pap (CP) smears and ratio of these times, by study and simple overall mean time.

		LBC			CP		Ratio
Study	Study type	Preparation type	Time (sec)	Remarks	Time (sec)	Remarks	LBC/CP
Geyer, 1993[72]	Split-sample	CytoRich	60	-	203	-	0.30
Bur, 1995[139]	Split-sample	ThinPrep-beta	192	sd=108	366	sd=198	0.52
Ferenczy, 1996[140]	Split-sample	ThinPrep-beta	195	sd=54	278	sd=73	0.70
McGoogan, 1996[48]	Split-sample	ThinPrep 2000	194	-	398	-	0.49
Cheuvront, 1998[119]**	Direct-to-vial	AutoCyte	216	-	294	-	0.73
Hoerl, 2000[141]	Direct-to-vial	ThinPrep 2000	259	-	323	-	0.80
SCSP, 2002	Direct-to-vial	ThinPrep 2000	318	-	450	-	0.71
Moss, 2003[109]	Direct-to-vial	ThinPrep 2000	398	se=0.13	434	se=0.26	0.92
Weynand, 2003[142]	Split-sample	Papspin	300	-	300	-	1.00
Confortini, 2005[34]	Split-sample	CellSlide	275	-	393	-	0.70
Dowie, 2006 [143](m)	Direct-to-vial	ThinPrep	292	sd=107	383	sd=132	0.76
Jesdapatarakul, 2011[40]	Split-sample	Liqui-Prep	258	72	324	66	0.80
Simple average of reading time***			263		358		0.73

** if performed by experienced LBC readers
μ only reading time; LBC reading time derived from 3rd survey conducted in Manchester
*** no meta-analytical pooling is performed since the sd or se usually is not reported

1.4. Conclusion

1.4.1. Accuracy for cervical precancer or cancer

On overage, liquid-based cytology is neither more sensitive nor more specific to predict presence of underlying CIN2+ than conventional cytology. This conclusion is consistent by cytological cut-off to define test positivity, study design and LBC preparation system. However, in a minority of individual situations, the use of LBC instead of CP may yield an increase in sensitivity and a decrease in specificity.

1.4.2. Specimen quality

The proportion of unsatisfactory specimen was extremely heterogeneous which can be attributed to the lack of standardized definitions of specimen adequacy or inadequacy and/or to the poor application of quality definitions. Nevertheless, the proportion of un-interpretable specimen was lower in LBC compared to CP, in particular when the SurePath system was used. The improvement in specimen quality results in a reduction of re-invitations for taking a new sample. The advantage of the lower rate of unsatisfactory samples can be substantial or small depending on the local quality of collected conventional Pap smears.

1.4.3. Duration of microscopic interpretation

The microscopic interpretation of LBC requires about one fifth to one quarter less time compared to a CP.

1.4.4. Ancillary testing

LBC offers the possibility to perform additional tests for instance for presence of high-risk HPV types or for other molecular biomarkers using the residual liquid remnant after cytological processing.

1.4.5. General conclusion

LBC is not more accurate than CP to detect cervical preancer. However, given its similar sensitivity and specificity compared to CP, and its additional advantages (improved specimen quality, shorter duration of microscopic interpretation, possibility to perform ancillary tests), LBC can be considered acceptable for cervical cancer screening. The cost price for equipment, disposables and processing and the appropriateness to perform additional

tests must be taken into account when recommending a particular LBC system in a screening programme.

1.5. References

1. Gutman S. Labeling liquid-based systems: FDA clarification. *J Reprod Med* 2000;**45**: 969-70.

2. Fremont-Smith M, Marino James, Griffin Bryan, Spencer Lynn, Bolick David. Comparison of the SurePath liquid-based Papanicolaou smear with the conventional Papanicolaou smear in a multi-site direct-to-vial study. *Cancer Cytopathology* 2004; 269-79.

3. Australian Health Technology Advisory Committee. Review of automated and semi-automated cervical screening devices. Canberra : Australian Health Technology Advisory Committee 1998. Canberra: Commonwealth of Australia; 1998. Report No.:

4. Austin RM, Ramzy I. Increased detection of epithelial cell abnormalities by liquid-based gynecologic cytology preparations. *Acta Cytol* 1998;**42**: 178-84.

5. McCrory DC, Matchar DB, Bastian L, Datta S, Hasselblad V, Hickey J, et al. Evaluation of cervical cytology. Rockville (MD), USA: AHCPR; 1999. Report No.:

6. Bishop JW, Marshall CJ, Bentz JS. New technologies in gynecologic cytology. *J Reprod Med* 2000;**45**: 701-19.

7. Payne N, Chilcott J, McGoogan E. Liquid-based cytology in cervical screening : a rapid and systematic review. *Health Technol Assess* 2000;**4**: 1-73.

8. Broadstock M. Effectiveness and cost effectiveness of automates and semi-automated cervical cancer screening devices. Christchurch, New Zealand: NZHTA; 2000. Report No.:

9. Bernstein SJ, Sanchez-Ramos L, Ndubisi B. Liquid-based cervical cytologic smear study and conventional papanicolaou smears. A metaanalysis of prospective studies comparing cytologic diagnosis and sample adequacy. *Am J Obstet Gynecol* 2001;**185**: 308-17.

10. Sulik SM, Kroeger K, Schultz JK, et al. Are Fluid-based Cytologies Superior to the Conventional Papanicolaou Test? A Systematic Review. *J Fam Pract* 2001;**50**: 1040-6.

11. Moseley RP, Paget S. Liquid-based cytology: is this the way forward for cervical screening? *Cytopathology* 2002;**13**: 71-82.

12. MSAC. Liquid based cytology for cervical screening. Canberra: Commonwealth of Australia; 2002 Aug. Report No.:

13. ANAES. Guidelines for follow-up of patients having an abnormal Pap smear. Update 2002. (in French). Paris: ANAES; 2002. Report No.:

14. Noorani HZ, Brown A, Skidmore B, Stuart GCE. Liquid-based cytology and human papillomavirus testing in cervical cancer screening. Ottawa, Ontario, Canada: CCOHTA (Canadian Coordinating Office for Health Technology Assessment); 2003. Report No.:

15. Karnon J, Peters J, Platt J, Chilcott J, McGoogan E. Assessment report: liquid-based cytology in cervical screening: an updated rapid and systematic review. Sheffield, UK: The National Institute for Clinical Excellence (NICE); 2003. Report No.:

16. Randall P, Farquhar C, Marjoribanks J. Liquid-based cytology in cervical screening: a systematic review. A report for the National Screening Unit of the Ministry of Health, April 2004. Auckland: Department of Obstetrics and Gynaecology, University of Auckland, New Zealand.; 2004. Report No.:

17. Abulafia O, Pezzullo JC, Sherer DM. Performance of ThinPrep liquid-based cervical cytology in comparison with conventionally prepared Papanicolaou smears: a quantitative survey. *Gynecol Oncol* 2003;**90**: 137-44.

18. Davey E, Barratt A, Irwig L, et al. Effect of study design and quality on unsatisfactory rates, cytology classifications, and accuracy in liquid-based versus conventional cervical cytology: a systematic review. *Lancet* 2006;**367**: 122-32.

19. Arbyn M, Anttila A, Jordan J, et al. European Guidelines for Quality Assurance in Cervical Cancer Screening. Second Edition - Summary Document. *Ann Oncol* 2010;**21**: 448-58.

20. Arbyn M, Bergeron C, Klinkhamer P, et al. Liquid compared with conventional cervical cytology: a systematic review and meta-analysis. *Obstet Gynecol* 2008;**111**: 167-77.

21. Arbyn M, Ronco G, Anttila A, et al. Evidence regarding HPV testing in secondary prevention of cervical cancer. *Vaccine* 2012;**30 Suppl 5**: F88-F99.

22. Arbyn M, Roelens J, Cuschieri K, et al. The APTIMA HPV assay versus the Hybrid Capture II test in triage of women with ASC-US or LSIL cervical cytology: a meta-analysis of the diagnostic accuracy. *Int J Cancer* 2013;**132**: 101-8.

23. Roelens J, Reuschenbach M, von Knebel-Doeberitz M, et al. p16INK4a immunocytochemistry versus HPV testing for triage of women with minor cytological abnormalities: A systematic review and meta-analysis. *Cancer* 2012;**120**: 294-307.

24. Verdoodt F, Szarewski A, Halfon P, Cuschieri K, Arbyn M. Triage of women with minor abnormal cervical cytology: Meta-analysis of the accuracy of an assay targeting messenger ribonucleic acid of 5 high-risk human papillomavirus types. *Cancer Cytopathol* 2013;**121**: 675-87.

25. Arbyn M, Roelens J, Simoens C, et al. Human papillomavirus testing versus repeat cytology for triage of minor cytological cervical lesions. *Cochrane Database Syst Rev* 2013;**3**: 1-201.

26. Arbyn M, Ronco G, Cuzick J, Wentzensen N, Castle PE. How to evaluate emerging technologies in cervical cancer screening? *Int J Cancer* 2009;**125**: 2489-96.

27. Schatzkin A, Connor RJ, Taylor PR, Bunnag B. Comparing new and old screening tests when a reference procedure cannot be performed on all screenees. Example of automated cytometry for early detection of cervical cancer. *Am J Epidemiol* 1987;**125**: 672-8.

28. Harris R, Bradburn M, Deeks J, et al. metan: fixed- and random-effects meta-analysis. *Stata Journal* 2008;**8**: 3-28.

29. Dersimonian R, Laird NM. Meta-analysis in clinical trials. *Controlled Clin Trials* 1986;**7**: 177-88.

30. Ferenczy A, Franco EL, Arseneau J, Wright TC, Richart RM. Diagnostic performance of Hybrid Capture human papillomavirus deoxyribonucleic acid assay combined with liquid-based cytologic study. *Am J Obstet Gynecol* 1996;**175**: 651-6.

31. Bergeron C, Bishop J, Lemarie A, et al. Accuracy of Thin-Layer Cytology in Patients Undergoing Cervical Cone Biopsy. *Acta Cytol* 2001;**45**: 519-24.

32. Coste J, Cochand-Priollet B, de Cremoux P, et al. Cross sectional study of conventional cervical smear, monolayer cytology, and human papillomavirus DNA testing for cervical cancer screening. *BMJ* 2003;**326**: 733-6.

33. Confortini M, Bulgaresi P, Cariaggi MP, et al. Comparing conventional and liquid-based smears from a consecutive series of 297 subjects referred to colposcopy assessment. *Cytopathology* 2004;**15**: 168-9.

34. Confortini M, Carozzi F, Cortecchia S, et al. Technical evaluation of the new thin layer device cellslidetrade mark (Menarini Diagnostics). *Diagn Cytopathol* 2005;**33**: 387-93.

35. Hussein T, Desai M, Tomlinson A, Kitchener HC. The comparative diagnostic accuracy of conventional and liquid-based cytology in a colposcopic setting. *BJOG* 2005;**112**: 1542-6.

36. Longatto-Filho A, Pereira SM, Di Loreto C, et al. DCS liquid-based system is more effective than conventional smears to diagnosis of cervical lesions: study in high-risk population with biopsy-based confirmation. *Gynecol Oncol* 2005;**97**: 497-500.

37. Taylor S, Kuhn L, Dupree W, et al. Direct comparison of liquid-based and conventional cytology in a South African screening trial. *Int J Cancer* 2006;**118**: 957-62.

38. Sykes PH, Harker DY, Miller A, et al. A randomised comparison of SurePath liquid-based cytology and conventional smear cytology in a colposcopy clinic setting. *BJOG* 2008;**115**: 1375-81.

39. Angstetra D, Tait T, Tan J, Symonds I. Should liquid-based cytology be performed prior to colposcopy? A comparison of the accuracy, unsatisfactory rates and cost in a tertiary referral setting. *Aust N Z J Obstet Gynaecol* 2009;**49**: 681-4.

40. Jesdapatarakul S, Tangjitgamol S, Nguansangiam S, Manusirivithaya S. Liqui-Prep(R) versus conventional Papanicolaou smear to detect cervical cells abnormality by split-sample technique: A randomized double-blind controlled trial. *Diagn Cytopathol* 2011;**32**: 22-7.

41. Ronco G, Cuzick J, Pierotti P, et al. Accuracy of liquid based versus conventional cytology: overall results of new technologies for cervical cancer screening: randomised controlled trial. *BMJ* 2007;**335**: 28-*.

42. Siebers AG, Klinkhamer PJ, Grefte JMM, et al. Comparison of liquid-based cytology with conventional cytology for detection of cervical cancer precursors: a randomized controlled trial. *JAMA* 2009;**302**: 1757-64.

43. Klug SJ, Neis KJ, Harlfinger W, et al. A Randomized Trial Comparing Conventional Cytology to Liquid-Based Cytology and Computer Assistance. *Int J Cancer* 2013;**132**: 2849-57.

44. Strander B, Andersson-Ellstrom A, Milsom I, Radberg T, Ryd W. Liquid-based cytology versus conventional Papanicolaou smear in an organized screening program : a prospective randomized study. *Cancer* 2007;**111**: 285-91.

45. Hutchinson ML, Cassin CM, Ball HG. The efficacy of an automated preparation device for cervical cytology. *Am J Clin Pathol* 1991;**96**: 300-5.

46. Wilbur DC, Cibas ES, Merrir S, et al. ThinPrep processor clinical trials demonstrate an increased detection rate of abnormal cervical cytology specimens. *Am J Clin Pathol* 1994;**101**: 209-14.

47. Laverty CRA, Thurloe JK, Redman NL, Farnsworth A. An Australian trial of Thinprep: a new cytopreparatory technique. *Cytopathology* 1995;**6**: 140-8.

48. McGoogan E, Reith A. Would monolayers provide more representative samples and improved preparations for cervical screening ? *Acta Cytol* 1996;**40**: 107-19.

49. Wilbur DC, Dubeshter B, Angel C, Atkison KM. Use of thin-layer preparations for gynecologic smears with emphasis on the cytomorphology of high-grade intraepithelial lesions and carcinomas. *Diagn Cytopathol* 1996;**14**: 201-11.

50. Lee KR, Ashfaq R, Birdsong GG, et al. Comparison of conventional Papanicolaou smears and a fluid-based, thin-layer system for cervical cancer screening. *Obstet Gynecol* 1997;**90**: 278-84.

51. Roberts JM, Gurley AM, Thurloe JK, Bowditch R, Laverty CRA. Evaluation of the thinprep pap test as an adjunct to the conventional pap smear. *MJA* 1997;**167**: 466-9.

52. Corkill M, Knapp D, Hutchinson ML. Improved accuracy for cervical cytology with the ThinPrep method and the endocervican brush-spatula collection procedure. *J Low Genit Tract Dis* 1998;**2**: 12-6.

53. Boman F, Farré I, Farine MO, et al. Pourquoi nous préférons la technique en couche mince aux frottis cervico-utérins conventionnels. *Clin Exp Path* 1999;**47**: 81-7.

54. Hutchinson ML, Zahniser DJ, Schiffman MA, et al. Utility of liquid-based cytology for cervical carcinoma screening: results of population-based study conducted in a region of Costa Rica with a high incidende of cervical carcinoma. *Cancer* 1999;**87**: 48-55.

55. McGoogan E. The ThinPrep test provides improved adequacy rates for routine cervical cytopathology. *Cytopathology* 1999;**10**: 2.

56. Shield PW, Nolan GR, Phillips GE, Cummings MC. Improving cervical cytology screening in a remote, high risk population. *MJA* 1999;**170**: 255-8.

57. Wang T-Y, Chen HS, Yang YC, Tsou M-C. Comparison of fluid-based, Thin-Layer processing and conventional papanicolaou methods for uterine cervical cytology. *J Formos Med Assoc* 1999;**98**: 500-5.

58. Monsonego J, Autillo Touati A, Bergeron C, et al. Liquid-based cytology for primary cervical cancer screening: a multi-centre study. *Br J Cancer* 2001;**84**: 360-6.

59. Park IA, Lee SN, Chae SW, et al. Comparing the Accuracy of ThinPrep Pap Tests and Conventional Papanicolaou Smears on the Basis of the Histologic Diagnosis: A Clinical Study of Women with Cervical Abnormalities. *Acta Cytol* 2001;**45**: 525-31.

60. Armstrong GP, Moriarty HT, Krajewska I. An in-house evaluation of the ThinPrep® split-sample Pap test. *Aust J Med Science* 2002;**23**: 162-73.

61. Biscotti CV, O'Brien DL, Gero MA, et al. Thin-layer Pap test vs. conventional Pap smear: analysis of 400 split samples. *J Reprod Med* 2002;**47**: 9-13.

62. Grace A, McBrearty P, Troost S, et al. Comparative study: conventional cervical and ThinPrep Pap tests in a routine clinical setting. *Cytopathology* 2002;**13**: 200-5.

63. Luthra UK, Chishti M, Dey P, et al. Performance of monolayered cervical smears in a gynecology outpatient setting in Kuwait. *Acta Cytol* 2002;**46**: 303-10.

64. Ring M, Bolger N, O'Donnell M, et al. Evaluation of liquid-based cytology in cervical screening of high-risk populations: a split study of colposcopy and genito-urinary medicine populations. *Cytopathology* 2002;**13**: 152-9.

65. Malle D, Pateinakis P, Chakka E, Destouni C. Experience with a thin-layer, liquid-based cervical cytologic screening method. *Acta Cytol* 2003;**47**: 129-34.

66. Tuncer ZS, Basaran M, Sezgin Y, Firat P, Mocan KG. Clinical results of a split sample liquid-based cytology (ThinPrep) study of 4,322 patients in a Turkish institution. *Eur J Gynaecol Oncol* 2005;**26**: 646-8.

67. Davey E, d'Assuncao J, Irwig L, et al. Accuracy of reading liquid based cytology slides using the ThinPrep Imager compared with conventional cytology: prospective study. *BMJ* 2007;**335**: 31-*.

68. Rahimi S, Carnovale-Scalzo C, Marani C, et al. Comparison of conventional Papanicolaou smears and fluid-based, thin-layer cytology with colposcopic biopsy control in central Italy: A consecutive sampling study of 461 cases. *Diagn Cytopathol* 2008;**37**: 1-3.

69. Treacy A, Reynolds J, Kay EW, Leader M, Grace A. Has the ThinPrep method of cervical screening maintained its improvement over conventional smears in terms of specimen adequacy? *Diagn Cytopathol* 2009;**37**: 239-40.

70. Halford JA, Batty T, Boost T, et al. Comparison of the sensitivity of conventional cytology and the ThinPrep Imaging System for 1,083 biopsy confirmed high-grade squamous lesions. *Diagn Cytopathol* 2010;**38**: 318-26.

71. Jensen OM, Whelan S. Planning a cancer registry, ... In: Jensen OM, Parkin DM, MacLennan R, Muir CS, Skeet RG, editors. Cancer registration: principles and methods. 1st ed. 1991. p. 26-158.

72. Geyer JW, Hancock F, Carrico C, Kirkpatrick M. Preliminary evaluation of Cyto-Rich : an improved automated cytology preparation. *Diagn Cytopathol* 1993;**9**: 417-22.

73. Sprenger E, Schwarzmann P, Kirkpatrick M, et al. The false negative rate in cervical cytology. Comparison of monolayers to conventional smears. *Acta Cytol* 1996;**40**: 81-9.

74. Vassilakos P, Cossali D, Albe X, et al. Efficacy of monolayer preparation for cervical cytology. Emphasis on suboptimal specimen. *Acta Cytol* 1996;**40**: 496-500.

75. Bishop JW. Comparison of the CytoRich system with conventional cervical cytology. Preliminary data on 2,032 cases from a clinical trial site. *Acta Cytol* 1997;**41**: 15-23.

76. Laverty CR, Farnsworth A, Thurloe JK, Grieves A, Bowditch R. Evaluation of the CytoRich slide preparation process. *Anal Quant Cytol Histol* 1997;**19**: 239-45.

77. Wilbur DC, Facik MS, Rutkowski MA, Mulford DK, Atkison KM. Clinical trials of the CytoRich specimen-preparation device for cervical cytology. Preliminary results. *Acta Cytol* 1997;**41**: 24-9.

78. Bishop JW, Bigner SH, Olgan TJ, et al. Multicenter masked evaluation of autocyte PREP thin layers with matched conventional smears. *Acta Cytol* 1998;**42**: 189-97.

79. Howell LP, Davis RL, Belk TI, Agdigos R, Lowe J. The AutoCyte preparation system for gynecologic cytology. *Acta Cytol* 1998;**42**: 171-7.

80. Kunz J, Rondez R, Yoshizaki Ch, et al. Vergleich konventioneller PAP-Abstriche mit Dünnschicht-Präparaten (Liquid-Based PAP-Test) und Korrelation der zytopathologischen Befunde mit dem HPV-Status nach Hybrid Capture System. *Schweiz Rundsch Med Prax* 1998;**87**: 1434-40.

81. Stevens MW, Nespolon WW, Milne AJ, Rowland R. Evaluation of the CytoRich technique for cervical smears. *Diagn Cytopathol* 1998;**18**: 236-42.

82. Minge L, Fleming M, VanGeem T, Bishop JW. AutoCyte Prep system vs. conventional cervical cytology. Comparison based on 2156 cases. *J Reprod Med* 2000;**45**: 179-84.

83. Hessling JJ, Raso DS, Schiffer B, et al. Effectiveness of thin-layer preparations vs. conventional Pap smears in a blinded, split-sample study. *J Reprod Med* 2001;**46**: 880-6.

84. Robyr R, Nazeer S, Vassilakos P, et al. Feasibility of cytology-based cervical cancer screening in rural Cameroon. *Acta Cytol* 2002;**46**: 1110-6.

85. Harkness CB, Theofrastous JP, Ibrahim SN, Galvin SL, Lawrence HC. Papanicolaou and thin-layer cervical cytology with colposcopic biopsy control. A comparison. *J Reprod Med* 2003;**48**: 681-6.

86. Bowditch RC, Clarke JM, Baird PJ, Greenberg ML. Results of an Australian trial using SurePath liquid-based cervical cytology with FocalPoint computer-assisted screening technology. *Diagn Cytopathol* 2012;**40**: 1093-9.

87. Khalbuss WE, Rudomina D, Kauff ND, Chuang L, Melamed MR. SpinThin, a simple, inexpensive technique for preparation of thin-layer cervical cytology from liquid-based specimens. Data on 791 cases. *Cancer* 2000;**90**: 135-42.

88. Mattosinho De Castro Ferraz G, Nicolau SM, Stavale JN, et al. Cervical biopsy-based comparison of a new liquid-based thin-layer preparation with conventional Pap smears. *Diagn Cytopathol* 2004;**30**: 220-6.

89. Rosenthal DL, Geddes S, Trimble CL, Carson KA, Alli PM. The PapSpin: A reasonable alternative to other, more expensive liquid-based papanicolaou tests. *Cancer* 2006;**108**: 137-43.

90. Weintraub J. The coming revolution in cervical cytology: a pathologist's guide for the clinician. *RGO* 1997;**5**: 1-6.

91. Bolick DR, Hellman DJ. Laboratory implementation and efficacy assessment of the ThinPrep cervical cancer screening system. *Acta Cytol* 1998;**42**: 209-13.

92. Dupree WB, Suprun HZ, Beckwith DG, Shane JJ, Lucente V. The promise and risk of a new technology: the Lehigh Valley Hospital's experience with liquid-based cervical cytology. *Cancer* 1998;**84**: 202-7.

93. Johnson JE, Jones HW, Conrad KA, Huff BC. Increased rate of SIL detection with excellent biopsy correlation after implementation of direct-to-vial Thinprep liquid-based preparation of cervicovaginal specimens at a university medical centre. *Acta Cytol* 1998;**42**: 1242-3.

94. Papillo JL, Zarka MA, St.John TL. Evaluation of the ThinPrep Pap test in clinical practice. A seven-month, 16.314-case experience in Northern Vermont. *Acta Cytol* 1998;**42**: 203-8.

95. Quddus MR, Xu B, Sung CJ, Boardman L, Lauchlan SC. Cytohisto correlations support the observation of increased detection of squamous intraepithelial lesions by the ThinPrep process. *Acta Cytol* 1998;**42**: 1243.

96. Carpenter AB, Davey DD. ThinPrep Pap Test: performance and biopsy follow-up in a university hospital. *Cancer* 1999;**87**: 105-12.

97. Diaz-Rosario LA, Kabawat SE. Performance of a fluid-based, thin-layer papanicolaou smear method in the clinical setting of an independent laboratory and an outpatient screening population in New England. *Arch Pathol Lab Med* 1999;**123**: 817-21.

98. Guidos BJ, Selvaggi SM. Use of Thin Prep Pap Test™ in clinical practice. *Diagn Cytopathol* 1999;**20**: 70-3.

99. Yeoh GP, Chan KW, Lauder I, Lam MB. Evaluation of the ThinPrep papanicolaou test in clinical practice: 6-month study of 16541 cases with histological correlation in 220 cases. *HKMJ* 1999;**5**: 233-9.

100. Ferris DG, Heidemann NL, Litaker MS, Crosby JH, Macfee MS. The efficacy of liquid-based cervical cytology using direct-to-vial sample collection. *J Fam Pract* 2000;**49**: 1005-11.

101. Weintraub J, Morabia A. Efficacy of a liquid-based thin layer method for cervical cancer screening in a population with a low incidence of cervical cancer. *Diagn Cytopathol* 2000;**22**: 52-9.

102. Obwegeser JH, Brack S. Does liquid-based technology really improve detection of cervical neoplasia? A Prospective, randomized trial comparing the ThinPrep Pap test with the conventional Pap test, including follow-up of HSIL cases. *Acta Cytol* 2001;**45**: 709-14.

103. Baker JJ. Conventional and liquid-based cervicovaginal cytology: a comparison study with clinical and histologic follow-up. *Diagn Cytopathol* 2002;**27**: 185-8.

104. Bentz JS, Rowe LR, Gopez EV, Marshall CJ. The unsatisfactory ThinPrep Pap Test: missed opportunity for disease detection? *Am J Clin Pathol* 2002;**117**: 457-63.

105. Quality Assurance Reference Centre for the Trent NHSCSP. Do borderline nuclear changes in gynaecological cytology constitute a reliable reporting category? *Cytopathology* 2002;**13**: 220-31.

106. Cheung ANY, Szeto EF, Leung BSY, Khoo U-S, Ng AWY. Liquid-based cytology and conventional cervical smears: a comparison study in an Asian screening population. *Cancer Cytopathol* 2003;**99**: 331-5.

107. Hoelund B. Implementation of liquid-based cytology in the screening programme against cervical cancer in the County of Funen, Denmark, and status for the first year. *Cytopathology* 2003;**14**: 269-74.

108. Kruger J, Dunton CJ, Sewell C, Cardonick E. Randomized pilot study comparing rates of endocervical cell recovery between conventional Pap smears and liquid-based cytology in a pregnant population. *J Low Genit Tract Dis* 2003;**7**: 101-3.

109. Moss SM, Gray A, Legood R, Henstock E. Evaluation of HPV/LBC. Cervical screening pilot studies. First report to the Department of Health on evaluation of LBC (December 2002). Institute of Cancer Research (Sutton); Institute of Health Sciences (Oxford); 2003 Jan. Report No.:

110. Negri G, Menia E, Egarter-Vigl E, Vittadello F, Mian C. ThinPrep versus conventional Papanicolaou smear in the cytologic follow-up of women with equivocal cervical smears. *Cancer* 2003;**99**: 342-5.

111. Cervical Cancer Screening Wales. Cervical Cancer Screening Wales. Liquid Based Cytology - Pilot Project: Project Report. Cardiff: Velindre/NHS Trust; NHS/CYMRU Wales; 2003. Report No.:

112. Hodgson W, Kaplan KJ, Rodriguez M, et al. The impact of converting to liquid-based cervical cytology in a military population. *Gynecol Oncol* 2005;**99**: 422-6.

113. Duggan MA, Khalil M, Brasher PM, Nation JG. Comparative study of the ThinPrep Pap test and conventional cytology results in a Canadian cohort. *Cytopathology* 2006;**17**: 73-81.

114. Ronco G, Segnan N, Giorgi-Rossi P, et al. Human Papillomavirus testing and liquid-based cytology in primary cervical screening: results at recruitment from the New Technologies for Cervical Cancer randomized controlled trial. *J Natl Cancer Inst* 2006;**98**: 765-74.

115. Schledermann D, Ejersbo D, Hoelund B. Improvement of diagnostic accuracy and screening conditions with liquid-based cytology. *Diagn Cytopathol* 2006;**34**: 780-5.

116. Nance KV. Evolution of Pap testing at a community hospital-A ten year experience. *Diagn Cytopathol* 2007;**35**: 148-53.

117. Williams AR. Liquid-based cytology and conventional smears compared over two 12-month periods. *Cytopathology* 2006;**17**: 82-5.

118. Maccallini V, Angeloni C, Caraceni D, et al. Comparison of the conventional cervical smear and liquid-based cytology: results of a controlled, prospective study in the Abruzzo Region of Italy. *Acta Cytol* 2008;**52**: 568-74.

119. Cheuvront DA, Elston RJ, Bishop JW. Effect of a Thin-Layer Preparation System on Workload in a Cytology Laboratory. *Lab Med* 1998;**29**: 174-9.

120. Vassilakos P, Griffin S, Megevand E, Campana A. CytoRich liquid-based cervical cytologic test. Screening results in a routine cytopathology service. *Acta Cytol* 1998;**42**: 198-202.

121. Saurel J, Rabreau M, Landi M, et al. Dépistage cytologique du cancer du col utérin par prélèvements en milieu liquide (CytoRich). Etude préliminaire d'une série de 111 292 patientes. *Contracept Fertil Sex* 1999;**27**: 853-7.

122. Vassilakos P, Saurel J, Rondez R. Direct-to-vial use of the AutoCyte PREP liquid-based preparation for cervical-vaginal specimens in three European laboratories. *Acta Cytol* 1999;**43**: 65-8.

123. Tench W. Preliminary assessment of the autoCyte PREP. Direct-to-vial performance. *J Reprod Med* 2000;**45**: 912-6.

124. Marino JF, Fremont-Smith M. Direct-to-vial experience with AutoCyte PREP in a small New England regional cytology practice. *J Reprod Med* 2001;**46**: 353-8.

125. Royer MC, Smith KL. Comparison of AutoCyte PREP adequacy to the conventional Pap test. *Acta Cytol* 2001;**45**: 827.

126. Day SJ, Deszo EL, Freund KM. Dual sampling of the endocervix and its impact on AutoCyte prep endocervical adequacy. *Am J Clin Pathol* 2002;**118**: 41-6.

127. Colgan TJ, McLachlin CM, Cotterchio M, et al. Results of the implementation of liquid-based cytology-SurePath in the Ontario screening program. *Cancer* 2004;**102**: 362-7.

128. Sass MA. Use of a liquid-based, thin-layer Pap test in a community hospital. Impact on cytology performance and productivity. *Acta Cytol* 2004;**48**: 17-22.

129. Longatto-Filho A, Maeda MY, Erzen M, et al. Conventional Pap smear and liquid-based cytology as screening tools in low-resource settings in Latin America: experience of the Latin American screening study. *Acta Cytol* 2005;**49**: 500-6.

130. Kirschner B, Simonsen K, Junge J. Comparison of conventional Papanicolaou smear and SurePath liquid-based cytology in the Copenhagen population screening programme for cervical cancer. *Cytopathology* 2006;**17**: 187-94.

131. Stark A. Rejected and unsatisfactory specimens: A comparative study of liquid-based (SurePath) preparations and conventional pap smears for cervicovaginal screening. *Lab Med* 2007;**35**: 729-33.

132. Beerman H, van Dorst EB, Kuenen-Boumeester V, Hogendoorn PC. Superior performance of liquid-based versus conventional cytology in a population-based cervical cancer screening program. *Gynecol Oncol* 2009;**112**: 572-6.

133. Bergeron C, Fagnani F. Performance of a new, liquid-based cervical screening technique in the clinical setting of a large French laboratory. *Acta Cytol* 2003;**47**: 753-61.

134. Boon ME, Rijkaart DC, Ouwerkerk-Noordam E, Korporaal H. Dutch solutions for liquid-based cytology: analysis of unsatisfactory slides and HPV testing of equivocal cytology. *Diagn Cytopathol* 2006;**34**: 644-8.

135. Park J, Jung EH, Kim C, Choi YH. Direct-to-vial comparison of a new liquid-based cytology system, liqui-PREP versus the conventional pap smear. *Diagn Cytopathol* 2007;**35**: 488-92.

136. Kitchener HC, Blanks R, Dunn G, et al. Automation-assisted versus manual reading of cervical cytology (MAVARIC): a randomised controlled trial. *Lancet Oncol* 2011;**12**: 56-64.

137. Zhao FH, Hu SY, Bian JJ, et al. Comparison of ThinPrep and SurePath liquid-based cytology and subsequent human papillomavirus DNA testing in China. *Cancer Cytopathol* 2011;**119**: 387-94.

138. Fontaine D, Narine N, Naugler C. Unsatisfactory rates vary between cervical cytology samples prepared using ThinPrep and SurePath platforms: a review and meta-analysis. *BMJ Open* 2012;**2**: e000847.

139. Bur M, Knowles K, Pekow P, Corral O, Donovan J. Comparison of ThinPrep preparations with conventional cervicovaginal smears - practical considerations. *Acta Cytol* 1995;**39**: 631-42.

140. Ferenczy A, Robitaille J, Franco EL, et al. Conventional cervical cytologic smears vs. ThinPrep smears. A paired comparison study on cervical cytology. *Acta Cytol* 1996;**40**: 1136-42.

141. Hoerl HD, Shalkham JE, Cheung K, et al. Screening Parameters for ThinPrep and Conventional Gynecologic Cytology via Automated Monitoring. *Acta Cytol* 2000;**44**: 618-24.

142. Weynand B, Berliere M, Haumont E, et al. A new, liquid-based cytology technique. *Acta Cytol* 2003;**47**: 149-53.

143. Dowie R, Stoykova B, Crawford D, et al. Liquid-based cytology can improve efficiency of cervical smear readers: evidence from timing surveys in two NHS cytology laboratories. *Cytopathology* 2006;**17**: 65-72.

1.6. GRADE-Profil

Grading of Recommendations Assessment, Development and Evaluation (GRADE)

GRADE has built on previous systems to create a highly structured, transparent, and informative system for rating quality of evidence(Guyatt *et al.,* 2008b).

Steps in evidence assessment for making guidelines
1) Formulate a question
2) Identify the PICO(S) components
3) Qualify outcomes as critical, important, not important

1) Questions
What is the absolute clinical accuracy (sensitivity and specificity) to identify or exclude high-grade cervical precancer or worse (CIN2+) using liquid-based cytology or conventional cytology and what is the relative accuracy of liquid-based versus conventional cytology?

Other question
What is the percentage of cervical cell specimen that is judged as unsatisfactory for microscopic interpretation? Does the use of LBC result in less unsatisfactory specimen? Or, in other words: what is the relative

inadequacy rate of LBC compared to CP (ratio of percentages of unsatisfactory samples [LBC/CP])?

Other question
What is the average time needed to interpret an LBC specimen versus a CP? Does the interpretation of LBC require less time?

2) PICOS

P: Women attending cervical cancer screening. Women being tested for cervical cancer precursors (high-risk group) or under follow-up because of previously found cervical lesions will be included also, but will be considered as less relevant for answering the study question.
I: LBC (index test).
C: Conventional cytology (comparator test).
O: absolute sensitivity and specificity of index- and comparator tests; relative sensitivity and specificity of index versus comparator tests to detect CIN2+ and CIN3+.
S: diagnostic test accuracy studies (with complete verification with a reference standard); screening studies with different screening tests involving at least complete verification of women with one or more positive screening test results; randomised trials with different screening tests in separate study arms (these studies will only include relativity sensitivity) .

3) Importance of outcomes

Outcome:

1. Reduction of mortality from cervical cancer, (quality-adjusted) life-years gained.
2. Reduction of morbidity due to cervical cancer: incidence of cancer (Ib+).
3. Reduction of incidence of cancer (including micro-invasive cancer).
4. Reduction of incidence of CIN3 or worse disease (CIN3+).
5. Increased detection rate of CIN3+ or CIN2+.
6. Increased test positivity with increased, similar or hardly reduced positive predictive value.

Quality of evidence for each outcome in four categories;

- High: ++++
- Moderate:+++
- Low:++
- Very low: +

5 factors that lower the quality of the evidence (Guyatt *et al.*, 2008a)

1. Risk of bias due to limitations in design or execution: see CONSORT, STROBE, QUADAS
2. Inconsistency or heterogeneity: if consistency unexplained, lower quality
3. Indirectness, applicability (relevance of studies for answering the PICPO question)

4. Imprecision: number of studies, width of CI
5. Reporting bias, publication bias.

3 factors that increase the quality

1. Large effect
2. Dose effect gradient
3. Confounders (presence of confounders that have lowered the observed effect, absence of these would have increased the effect) (Guyatt *et al.,* 2008a)

4 studies

Mostly observational studies → basic evidence of low quality (++--)

Items downgrading quality of evidence		**Downgrading**
Bias, design	Low risk of bias according to Quadas and Cochrane tool for Risk of Bias assessment	No (-0)
Inconsistency	According to the test of heterogeneity (I-squared) there is substantial heterogeneity between the studies.	**Yes (-1)**
Indirectness	No.	No (-0)
Imprecision	No.	No (-0)
Publication bias, other	Not assessed.	No (-0)
Items downgrading quality of evidence		
Large effect	No.	No (+0)
Dose-effect correlation	No.	No (+0)
Confounding factors neutralising effects	No.	No (+0)°

Conclusion: evidence of very low quality (+---)

Overall Grade of the quality of evidence is assigned and this is based on the outcome with the lowest quality of evidence given that it is a critical outcome.

For the accuracy we consider the relative sensitivity (outcome CIN3+) and specificity (outcome CIN2+) as critical. The other outcomes: absolute

accuracy, relative sensitivity for CIN2+ and relative specificity for CIN3+ are considered as important.

GRADE evidence profile f

	Quality of evidence								
# studies (N)	Absence of study limitations	Consistency	Directness	Precision	Absence publication bias	Large effect	Dose-effect relation	Bias lowering effect	Comment
Outcome 1: Absolute clinical accuracy (sensitivity and specificity) to identify or exclude high-grade cervical precancer or worse (CIN2+) using liquid-based cytology or conventional cytology [CRITICAL]									
	Yes	No	Yes	Yes	Yes	No	No	No	

Outcome 2: Relative accuracy of liquid-based versus conventional cytology [CRITICAL]									
	Yes	No	Yes	Yes	Yes	No	No	No	

Formulation of recommendation: Pro/against, strong (we recommend)/weak (we suggest = conditional on). In concertation between guideline and systematic review group.
(Guyatt *et al.,* 2008a)

Factors that influence the strength of recommendation:

- Quality of evidence: by outcome and across outcomes
- Balance benefits/harms
- Values and preferences
- Resource use, costs

References

Guyatt G.H., Oxman A.D., Kunz R., Falck-Ytter Y., Vist G.E., Liberati A., & Schunemann H.J. (2008a) Going from evidence to recommendations. *BMJ* **336:** 1049-1051.

Guyatt G.H., Oxman A.D., Kunz R., Vist G.E., Falck-Ytter Y., & Schunemann H.J. (2008b) What is "quality of evidence" and why is it important to clinicians? *BMJ* **336:** 995-998.

Guyatt G.H., Oxman A.D., Vist G.E., Kunz R., Falck-Ytter Y., Alonso-Coello P., & Schunemann H.J. (2008c) GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *BMJ* **336:** 924-926.

Schunemann H.J., Oxman A.D., Brozek J., Glasziou P., Jaeschke R., Vist G.E., Williams J.W., Jr., Kunz R., Craig J., Montori V.M., Bossuyt P., & Guyatt G.H. (2008) GRADE: Grading quality of evidence and strength of recommendations for diagnostic tests and strategies. *BMJ* **336:** 1106-1110.

2. Question: Is primary screening with a biomarker better than conventional cytology or HPV-testing?

2.1. Background and rationale

Cervical cytology has been and still is, in most industrialised countries, the mainstay of secondary prevention of cervical cancer. HPV-based screening, however, may take over the role of the Pap smear as first screening test. Indeed, randomised trials have demonstrated (see chapter 4) that screening with a hrHPV DNA assay results finally in lower incidence of invasive cervical cancer compared to screening with cytology[1;2]. More specific triage markers might be needed to triage HPV-positive women and molecular markers, distinguishing progressing from regressing HPV infection, have been proposed in management algorithms making HPV-based screening more specific[3-6]. HPV testing was proposed originally as a triage method for women with minor cytological abnormalities and in a subsequent stage it was evaluated in primary screening[7;8]. In analogy, molecular markers, currently proposed as triage tools, may also be considered as alternative primary screening tests. Whereas molecular markers, were evaluated in the setting of triage of equivocal and low-grade cytological abnormalities (see chapter 6), we assess in the current chapter their performance in primary cervical cancer screening.

2.2. Responsibility

The Unit of Cancer Epidemiology (UCE, IPH-Brussels) had the task to conduct a systematic review on the performance of molecular markers and to compare the performance with cytology and HPVDNA-based cervical cancer screening. The translation of the evidence into practice recommendations will be lead by Prof. Löning and other members of Working Group 5 (Biomakers) following the GRADE process. UCE will assist WG5 in the GRADE process.

2.3. Clinical question

Is primary screening with a biomarker better than conventional cytology or HPV-testing?

2.3.1. PICOS

- P: women participating in cervical cancer screening

- I: testing with a biomarker (p16, p16/Ki-67 dual-stain, ProExC, E6/E7 mRNA, methylation markers, , or other)
- C1: cytology (conventional Pap smear, LBC)
- C2: HPV testing (HC2, GP5+/6+ PCR, or another clinically validated hrHPV DNA test)
- O: accuracy to detect underlying disease (=CIN2+,CIN3+/AIS):
    - Complete diagnostic studies: absolute and relative sensitivity and specificity, PPV, NPV, referral rate, detection rate, detection rate ratio
    - RCTs: relative sensitivity (or detection rate ratios), relative PPV, relative referral rate
- S:
    - diagnostic studies
        - all subjects receiving testing with a biomarker,
        - at least one comparator test
        - verification with the reference standard (colposcopy/biopsy)
            - all participants (accepting a negative colposcopy as free of CIN2+)
            - participants positive in at least one screening test (accepting a negative result for all screening tests as free of disease)
    - RCTs comparing screening with biomarkers with screening with one or more comparator tests.

2.3.2. Importance of outcomes

Preferentially reduction in disease (cumulative incidence of disease [CIN3+ or cervical cancer]) will be looked at. By lack of longitudinal outcomes, indicators of diagnostic accuracy will be assessed. Colposcopy referral rates and false positivity rates will be extracted and pooled where possible.

2.4. Methods

2.4.1. Retrieval of relevant studies

A systematic literature search was performed in three electronic bibliographic databases (MEDLINE, Embase, Cochrane Library) using the following search strings:

- Medline (Pubmed):

("Cervix Uteri"[Mesh] OR cervix OR cervical)
AND
("Papillomaviridae"[Mesh] OR HPV OR papillomavirus) AND (screening OR "Early Detection of cancer/epidemiology"[Mesh] OR "Early Detection of Cancer/methods"[Mesh] OR "Early Detection of Cancer/statistics and numerical data"[Mesh])
AND
(biomarker OR p16 OR Ki-67 OR mRNA OR methylation OR ProExC)

- Embase:

('uterine cervix'/exp OR cervix OR cervical)
AND
('papilloma virus'/exp OR HPV OR 'papillomavirus')
AND
('cancer screening'/exp/mj OR screening) AND biomarker

- Cochrane Library:

Cervix AND HPV AND screening

Additionally, references of relevant reviews were hand-searched.
No language or publication date restrictions were applied. Inclusion and exclusion parameters were identified prior to the evaluation of the retrieved literature. Studies were included if a biomarker test (E5/E6 mRNA, MCM2 & TOP2A, p16, p16Ki67, methylation marker, hTERC, *etc.*) was used on cervical samples from a screening population. Furthermore, golden standard verification had to be performed, at least on all participants that had a positive test result. Both studies with, and studies without comparator test were eligible. In case of double reporting of the same studies, the most comprehensive report was included. Studies were excluded if fewer than 1000 participants were included and the setting was not clearly stated in the manuscript. Eligibility of the studies was appraised by F.V. and was subsequently revised by M.A.

2.5. Results

2.5.1. Study retrieval

In total, ten studies[9-18] were judged relevant and in compliance to the PICOS (Figure 19)[19]. A list of included as well as excluded studies can be requested from the authors.

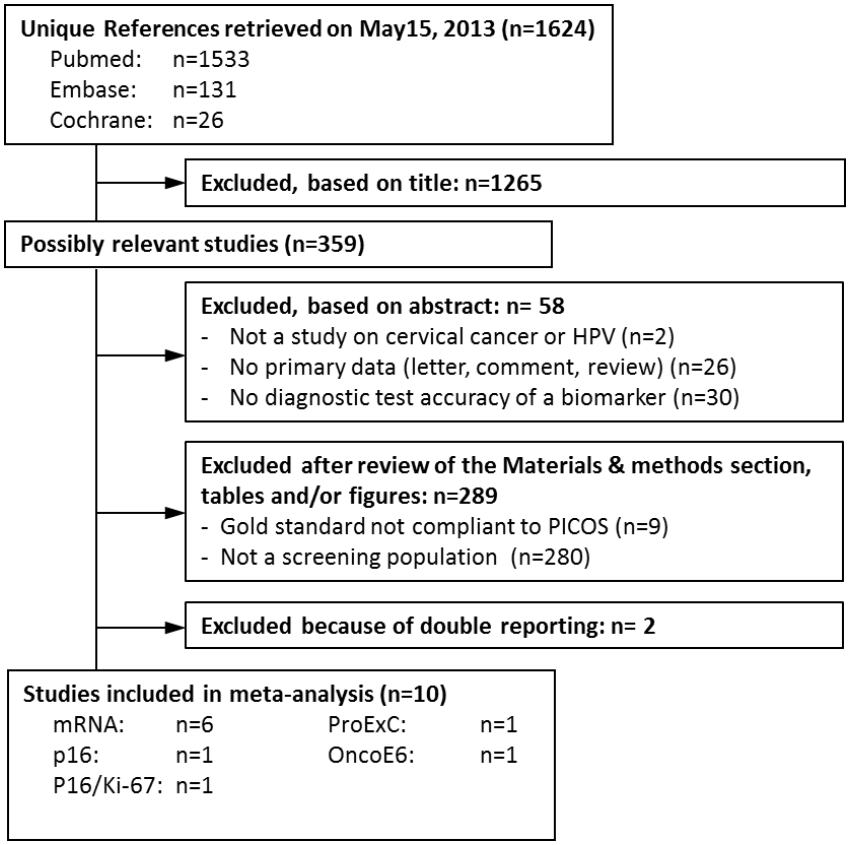


Figure 20: Prisma flow chart for the retrieval of studies.

2.5.2. Study characteristics of included studies

The majority of included studies used an assay detecting E6/E7-mRNA of five [10;15] or more[10;11;13-16] HPV types . Two studies assessed overexpression of the p16 ^{INK4a} protein by an anti-p16 ELISA assay [9], or by a double immunostaining on p16 and Ki-67[17]. Two other studies detected TOP2A and MCM2 proteins (ProEx C, BD Diagnostics–TriPath, Burlington, NC, USA)[12], and E6 proteins (OncoE6, Arbor Vita Corporation, Fremont, CA, USA)[18] of HPV 16,18 and 45, respectively. All studies provided data for hrHPV-DNA testing as a comparator, and all but two studies[14;18] performed cytology as well. In three studies, all participants received gold standard verification[10;12;14]. In the other studies, gold standard verification was performed on participants with at least one positive screening test[11;16;17], in some cases in combination with a subset of participants with all negative test results[9;13;18]. In one study, verification was only performed in patients with abnormal cytology[15].

Of the ten included studies, the population & study characteristics, and the test characteristics are listed in Table 12 and Table 13, respectively. Details on the gold standard and the outcome are presented in Table 14.

Table 12: Population and study characteristics of included studies.

Study	Study location	Study period	Study design	Study population	Inclusion criteria	Exclusion criteria	Study size
Balasubra., 2009	2 Planned Parenthood clinics, Washington, USA	Oct 2005 – Nov 2007	Concomitant testing with 3 tests: p16INK4a ELISA, cPap, HC2. Women with at least one positive test, and a random subset of women with normal screening test results, were invited to return for colposcopy and biopsy.	Women attending the clinic for routine Pap screening. Median age: 23y Range: 18-50y	Age range 18-50y	Virginity, pregnancy, chronically immunocompromised, prior treatments for CIN	1583
Hovland, 2010	3 gynaecological clinics in Bukavu, democratic Republic of Congo. The study was set up in a high-risk area.	Nov 2003 – Dec 2003	Concomitant testing with 4 g tests: Pretect HPV Proofer (5 types, 9 types), cPAP, LBC, PCR GP5+/6+ - EIA RLB. All women received gold standard verification.	Women attending gynaecological clinics in a high-risk area(mixed population: screening and obstetrical/ gynaecological problems). Age range: 25-60y	Not specified.	Pregnancy, severe gynaecological bleeding, previous hysterectomy, and age <25y or >60y	313
Wu, 2010	City of Shenzhen, Guangdong Province, in southern China	Not documented	The Shenzhen Cervical Cancer Screening Trial I (SHENCCAST I). Concomitant testing with 3 tests: APTIMA, LBC, HC2. Women with at least one positive screening result were asked to come back for gold standard verification.	2098 women from a medically underserved neighborhood in the Luohu district of Shenzhen were recruited into the study. Mean age: 35y	Age range: 25-59y	Pregnancy, had cervical cancer screening <3y, prior hysterectomy, prior pelvic radiation, age <25y or >59y	2015
Depuydt, 2011	9 gynecological practices in Flanders (Belgium).	Aug 2005 - Feb 2007	Prospective, colposcopy controlled study. After exclusion, 2,905 women underwent concomitant teststing with ProExC, LBC, hrHPV DNA testing (RT-PCR), and colposcopy verification.	Women undergoing routine screening. Median age: 43y Age range: 18–84y	Not specified.	Pregnancy, history of cervical disease	2905

Study	Study location	Study period	Study design	Study population	Inclusion criteria	Exclusion criteria	Study size
Monsonego, 2011	17 private gynecology practices in Paris, France	Apr 2008 – Feb 2009	Concomitant testing with APTIMA, LBC, HC2. Colposcopy was performed on women with at least one positive screening result and a random subset of screen-negative women. Accuracy values were adjusted for verification bias (missing-at-random assumption).	Women who came to the practice for their annual screening exam (In France, cervical cancer screening is recommended every 3 years, but most often is conducted every 1.5–2 years at the physician's discretion.)	Age range: 20-65y Willing to participate	Having had a total hysterectomy, pregnancy, abnormal cytology (≤ 6m)	4429
Ratnam, 2011	- Screening population: two study centers, Canada - Colposcopy referral cases: five tertiary care centers in five provinces across Canada.	Not documented	The study population was comprised of a random sample of colposcopy-referred women, and a subset of routinely screened women. Two screening tests in the screening population: APTIMA, HC2. Gold standard verification was performed on all participants.	(Pop. 1) women who were routinely screened. (Pop. 2) a random sample of women referred to colposcopy. Age range: 16-81y	Pop. 1: Not specified. Pop. 2: Women either newly diagnosed with abnormal Pap cytology of any grade; or those with a history of abnormal cytology who were being followed up in colposcopy clinics as per the routine standard of care	Not specified	1373

Study	Study location	Study period	Study design	Study population	Inclusion criteria	Exclusion criteria	Study size
Cuzick, 2013	St. Mary's Hospital, London. All tests were carried out in the Centre for cancer Prevention, except for PreTect HPV-Proofer, which was carried out at The Doctors Laboratory.	Not documented	Residual material was used from liquid-based cytology PreservCyt samples from 6000 women who attended for routine screening (3 or 5 yearly, depending on age). Samples were linked to concurrent cytology results and any histology within 6 months of an abnormal smear. The following screening tests were performed on the residual samples: Pretect HPV-Proofer, APTIMA, cPAP, LBC, HC2, Cobas 4800, Abbott, Viper BD Partial verification bias is present, as HPV-test results did not influence whether women received the gold standard.	6000 women who attended for a routine 3 or 5 yearly (depending on age) screening smear	Not specified.	Unsatisfactory cytology samples were excluded; there were no other inclusion/exclusion criteria.	6000
Ikenberg, 2013	Belgium, France, Germany, Italy, and Spain	Not documented	Women undergoing routine cytology-based cervical cancer screening were enrolled in gynecologist practices and hospital-based screening centers. All women received Pap cytology, p16/Ki-67 dual-stained cytology (split-sample), and HPV testing with HC2. Participants with at least one positive screen test were referred to colposcopy (unless only hrHPV+ and age <30). Biopsies were taken if clinically indicated. Cases where no biopsies were taken and negative colposcopic examination was considered negative for disease.	Women of 18 years and older undergoing routine cytology-based cervical cancer screening.	Not specified.	Pregnancy, hysterectomy.	25577

Study	Study location	Study period	Study design	Study population	Inclusion criteria	Exclusion criteria	Study size
Nieves, 2013	Michoacan, Mexico	Feb 2009 - Apr 2009	Mexican Cervical Cancer Screening Study II (MECCS II). First women took a self-sample. Next, a nurse/physician obtained 2 direct cervical samples. Three concomitant tests: APTIMA LBC and HC2. Women with at least one positive test result received the gold standard.	Mexican unscreened or underscreened women, recruited via posted public community announcements, radio and television advertisement, local meetings within communities/villages, and promotion through family welfare offices.	Women without history of Pap smear screening or knowledge of their Pap results within the last 3 years. Non pregnant. Age range:30 – 50y Residing in the state of Michoacan, Mexico	pregnancy, history of Pap smear or knowledge of Pap results (≤3y), hysterectomy, prior pelvic radiation, age <30y or >50y	2049
Zhao, 2013	Yangcheng/Xinmi/Tonggu, China	Oct 2010 – Jun 2011	All women, aged 25–65, living in the chosen village were invited to participate. Two cervical specimens were collected, the first into a dry tube for OncoE6™ testing and the second into *care*HPV™ Collection Medium For HPV testing; finally VIA was performed. Women who had at least one positive screen test and a random sample of test negative women were referred to Colposcopy.	Chinese underscreened women.	Women physically able to undergo routine cervical cancer screening; provided informed consent.	History of cervical cancer, pregnancy, hysterectomy, not married and reported never having had sexual intercourse.	7421

Table 13: Test characteristics of the included studies.

Study	Biomarker	Comparator tests	Sampling procedure	Test cut-off
Balasubra., 2009	**p16:** $p16^{INK4a}$ ELISA (original + enhanced)	- LBC - HC2	- p16: Ayre's spatula to obtain cells from the transformation zone; cytobrush for endocervical cells. Both were rinsed and retained in the TM until laboratory processing. - LBC: Ectocervical and endocervical samples collected within Ayre's spatula and cytobrush in one clinic, and a cervix broom in	Pap: ASC-US+ P16INK4a: ≥ 8pg/ml; ≥6 pg/ml HC2: ≥1 RLU

Study	Biomarker	Comparator tests	Sampling procedure	Test cut-off
			the other clinic. Devices were rinsed and tapped in Surepath TM. - HC2: Dacron swab for ectocervical and endocervical sample collection The swab was placed in Standard Transport Medium (Digene).	
Hovland, 2010	**mRNA 5 types:** Pretect HPV-Proofer **mRNA >5 types:** NASBA	- cPAP - LBC - GP5+/6+ PCR-EIA RLB	PAP smears, PreservCyt vials and histological material were stored at room temperature and shipped to Norway and Sweden for preparation and interpretation by experienced cytotechnicians and pathologists.	Cyto: ASC-US+, LSIL+, HSIL+
Wu, 2010	**mRNA >5 types:** APTIMA	- LBC (Surepath) - HC2	One cervical specimen was collected and placed in SurePath liquid (BD Diagnostics, Sparks, Md) for cytology, and another specimen was placed in PreservCyt, from which 4 mL was used for testing by HC2 and 1 mL was transferred into TM for testing with APTIMA.	LBC: ASC-US+ HC2: ≥1 RLU
Depuydt, 2011	**ProExC:** BD ProExC Immu- nohistochemistry kit	- LBC (SurePath) - RT-PCR (in house test)	Cervical cells were collected using the Cervex-Brush (Rovers). After collection, the brush head of the sampling device was deposited directly into the vial containing the ethanol-based BD SurePath Preservative Fluid. For ProExC, all slides were screened by experienced cytotechnologists. The staining of the 2,905 slides was performed in 97 batches, with negative and positive controls included in each batch.	LBC: ASC-US+, LSIL+ PCR: hrHPV (HPV 16, 18, 31, 33, 35, 39, 45, 51, 52, 52, 56, 58,59 and 68) ProExC: Moderate-to-intense brown nuclear staining observed in atypical epithelial cells
Monsonego, 2011	**mRNA >5 types:** APTIMA	- LBC (ThinPrep) - HC2	One cervical sample from each patient was collected by the gynecologist during a routine gynecological examination. Cervical samples were collected from the TZ using a Cervex-Brush which was rinsed into PreservCyt medium. Remark: An independent external reviewer blindly double-read the cytology samples with abnormal cytology results and a random selected group of women (14%) with normal LBC samples and negative HPV tests (adjudicated cytology). Final analyses were based on the adjudicated cytology results. A high (10.5%) rate of unsatisfactory ThinPrep results was due to poor celullarity, thick preparations and obscuring blood/debris, and the high proportion of post-menopausal women (20%) with poor cervical cellularity (these results were not included in the data analyses).	LBC: ASC-US+, LSIL+ HC2: ≥1 RLU APTIMA: ≥ 1.0 S/CO
Ratnam, 2011	**mRNA >5 types:** APTIMA	- HC2	Upon enrollment, a single cervical specimen was collected from all participants using the Cervex broom-type and suspended into PreservCyt TM.	HC2: ≥1 RLU APTIMA: >0.5 S/CO
Cuzick, 2013	**mRNA 5 types:**	- Cytology (cPAP	Not documented.	HC2: ≥1 RLU

Study	Biomarker	Comparator tests	Sampling procedure	Test cut-off
	Pretect HPV-Proofer **mRNA >5 types:** APTIMA	or LBC) - HC2, Cobas 4800, Abbott, Viper BD		Abbott: hrHPV (14 types) APTIMA: 100 genomic copies/reaction
Ikenberg, 2013	**P16/Ki-67:** CINtec Plus	- Cytology (cPAP or LBC - HC2	A first cervical sample was collected for Pap cytology testing using broom-type or brush/spatula sampling devices. A split sample was prepared for p16/Ki-67 staining. A second cervical sample was taken from all study participants using the DNAPAP Cervical Sampler for HPV testing.	Cyto: ASCUS HPV: 1 RLU P16/Ki-67: presence of double immunoreactive cells
Nieves, 2013	**mRNA >5 types:** APTIMA	- LBC (Thinprep) - HC2	First women took a self-sample. Next, a nurse or a physician obtained 2 direct cervical samples in alternating order based on study ID number, using either a ''broom'' sampler (Rovers Cervex-Brush; Rovers Medical Devices, Oss, the Netherlands) placed in PreservCyt, or direct to Gen-Probe TM using the Gen-Probe cervical brush.	Cyto: ASCUS HPV: Not documented
Zhao, 2013	**E6-protein:** OncoE6	- HC2 - careHPV	Women were given instructions on how to self-collect a vaginal specimen; the procedure was completed in private room. Next, women underwent a routine pelvic exam by female clinicians, at which time two cervical specimens were collected, the first into a dry tube for OncoE6™ testing and the second into *care*HPV™ Collection Medium for HR-HPV DNA testing; finally visual inspection after 5% acetic acid (VIA) was done and results recorded.	HPV: 1 RLU OncoE6: appearance of ≥1 test lines

Abbreviations: ASCUS/-US, atypical squamous cells of undetermined significance; cPAP, conventional Pap smear; EIA, enzyme immune-assay; HC2, Hybrid Capture 2; hrHPV, high-risk human papillomavirus; HSIL, high-grade squamous intra-epithelial lesions; LBC, Liquid-based cytology; LSIL, low-grade squamous intra-epithelial lesions; RLB, reverse line blot; RLU, relative light units; NASBA, nucleic acid sequence based amplification; RT-PCR, real-time polymerase chain reaction; TM, transport medium;

Table 14: Gold standard and outcome characteristics.

Study	Gold Standard	Criteria for gold standard application	Outcome	Masking of screeners, colposcpist	Delay between screening test and Gold Standard
Balasubra., 2009	Ectocervical biopsies were obtained from areas of the cervix appearing abnormal on colposcopic examination. An ectocervical biopsy was done at the 12 o'clock position if no lesion was visible during colposcopy. Endocervical curettage was done in case of inadequate colposcopy, lesions extending into the endocervical canal, normal colposcopy but HSIL at screening cytology, or AGC at cytology.	Women were invited to return for diagnostic testing by colpo-scopy and biopsy if they had at least one positive screening test result. A random subset of women with normal screening test results was invited for colpo-scopy and biopsy.	CIN3+	Not documented	Median: 45 days Range:9-225 days)
Hovland, 2010	Colposcopy and colpo-directed biopsy or random biopsy at 12 o'clock position. ECC was performed in case the TZ was not visible. Colposcopy diagnoses were verified by two physicians. Biopsy samples and ECC material were immediately transferred into PreservCyt vials. Histological examinations were carried out by an experienced pathologist in Norway.	All participants.	CIN2+	In general, all laboratory testing procedures were performed blindly without knowledge of other test results.	Screen tests and gold standard were performed during the same visit.
Wu, 2010	Colposcopy by quadrant: a biopsy of any area suggestive of abnormalities was performed in each quadrant. A biopsy of normal quadrants was performed at 2-, 4-, 8-, or 10-o'clock position at the squamo-columnar junction. Then all patients had endocervical curettage. All exocervical biopsies were performed with the 2-mm Preventive Oncology International biopsy instrument.	Participants who had atypical squamous cells of undetermined significance (ASCUS) or greater cytologic diagnosis and/or were HPV positive on either assay were asked to return to the clinic to undergo colposcopy and biopsy (return rate: 80.6%).	CIN2+ CIN3+	Individuals performing APTIMA and HC2 testing were blinded to the other HPV assay results and to cytologic and histologic results. Individuals performing cytologic or histologic examinations were blinded to both HPV assay results.	Not documented
Depuydt, 2011	All colposcopies at baseline were performed according to a standardized protocol by 9 different gynecologists. First a 5% acetic acid solution was applied to assist in identifying undifferentiated epithelia or inflammation as	All participants.	CIN2+ CIN3+	Cytology screening was performed without knowledge of the HPV status.	For 84 participants, biopsies were taken during the screening visit (because of aceto white lesions during

Study	Gold Standard	Criteria for gold standard application	Outcome	Masking of screeners, colposcpist	Delay between screening test and Gold Standard
	well as true CIN. After the application of the acetic acid, a judgment of the transformation zone was made. After visual inspection a cervical smear was taken with the Cervex-Brush. Biopsies were only taken from all acetowhite lesions after smear taking. Those with positive colposcopic examinations received immediate cervical biopsy with subsequent histologic analysis and were given treatment as required. All other subjects were followed up based on their HPV and cytology results and either received a second colposcopy and biopsy or were monitored based on recommended Belgian followup guidelines.			Not documented.	colposcopy, 13 CIN2+). For 183 participants, biopsies were taken during follow-up visits within a 24-month period.
Monsonego, 2011	Colposcopy was performed at each clinic according to standard operating procedures. Per protocol, all women with abnormal colposcopy were to receive at least one biopsy from the most severe area and a minimum of one biopsy from each quadrant of atypical TZ. For women with normal colposcopy (and not in the random control group), two biopsies were performed at 12 and 6 o'clock of TZ. No biopsy was performed in women from the control group with normal colposcopy. An independent (international) reviewer re-examined all biopsies. In all discrepant cases, the final diagnosis was the consensus of three pathologists. No CIN was diagnosed in the random control subset.	- participants with at least one positive screening test - a random subset (14%) of women with normal screening test results	CIN2+ CIN3+	Cytopathologists were blinded to the HPV test results. Histopathologists were blinded to HPV test results, but not to cyto-logy results for safety reasons.	Not documented.
Ratnam, 2011	Participating OB/GYN specialists carried out colposcopy, and if warranted cervical biopsy, on the day of patient enrollment as per the standard of care.	All participants.	CIN2+	Researchers and technologists performing tests were blinded to results obtained in the other tests and also to	Gold standard was performed at the enrolment visit. In some cases, biopsies were taken in subsequent follow-up visits, and in

Study	Gold Standard	Criteria for gold standard application	Outcome	Masking of screeners, colposcpist	Delay between screening test and Gold Standard
				cytology, colposcopy and histology results. Pathologists were blinded to HPV-test results	such instances, histology results on biopsies taken no later than six months following enrollment were included in the analysis.
Cuzick, 2013	All results are presented based on the local histopathology and the highest grade of abnormality seen in the biopsy or treatment specimen was used.	- participants with an abnormal cytology diagnosis	CIN2+ CIN3+	HPV-testing was performed a posteriori. Therefore histo and cytopathologists were unaware of HPV-test results. It is not documented whether technologists performing the HPV-test were blinded to the results of cytology and histology.	Any histologic diagnosis performed within 6m of an abnormal smear was included.
Ikenberg, 2013	Colposcopy. Biopsies were taken if clinically indicated. Cases where no biopsies were taken during the colposcopic examination were considered negative for disease. All local histology results were verified by an independent quality control (QC) review board, blinded to all study results. All CIN2+ cases and cases with discrepant results between local pathologists and QC reviewe were subjected to an extended QC review.	Participants with at least one positive screen test were referred to colposcopy (unless only hrHPV+ and age <30).	CIN2+	All cytotechnologists/ pathologists were informed about patient's age but blinded to all other study results. Colposcopists were aware of Pap cytology and HPV test results but blinded to any dual-stained cytology results. Members of the QC review board were blinded to all study results.	Not documented
Nieves, 2013	Acetic acid 4% was applied to the cervix and colposcopy was performed according to the POI microbiopsy protocol of directed and random biopsies.	Those women who tested positive for hrHPV (on any assay) or had ASCUS+ were recalled for gold standard verification.	CIN3+ CIN2+	Testing (HC2 & APTIMA) was performed in 2 different locations by technicians who had no knowledge of cytology	Not documented

Study	Gold Standard	Criteria for gold standard application	Outcome	Masking of screeners, colposcpist	Delay between screening test and Gold Standard
	All pathology specimens were processed in Mexico and read by a local pathologist and 2 gynecologic pathologists who traveled to Mexico from Cleveland Clinic. Immunostaining for p16 was done on all available CIN1 blocks, as well as all CIN2 and CIN3 specimens after transporting the specimens back to Cleveland Clinic.			results.	
Zhao, 2013	All colposcopically detected abnormalities were biopsied. If the colposcopic examination showed no lesion in a quadrant, a random biopsy was obtained at the squamocolumnar junction in that quadrant at 2, 4, 8, or 10 o'clock. An ECC was performed after the cervical biopsies. All initial biopsy diagnoses of CIN2+ were independently reviewed by an expert US pathologist for confirmation. Additional sections of all initial biopsy diagnoses of CIN2+ were cut and tested for p16INK4a by immune-histochemistry.	Women who tested positive for any of the six screening tests performed (3 tests on clinician-collected and self-collected specimens) were referred to colposcopy, and approximately a 10% random sample of screen-negative women.	CIN2+ CIN3+	Not documented.	Not documented.

Abbreviations: ASCUS, atypical squamous cells of undetermined significance; CIN1, cervical intra-epithelial neoplasia grade one; CIN2+, cervical intra-epithelial neoplasia grade two or worse; CIN3+, cervical intra-epithelial neoplasia grade three or worse; DNA, deoxiribonucleic acid; ECC, endocervical curettage ; HC2, Hybrid Capture 2; hrHPV, high risk human papillomavirus; TZ, transformation zone;

2.5.3. Quality assessment of included studies

The quality of included studies was evaluated using the QUADAS-2 tool[20] and is summarized in Table 15.

Overall, studies scored well for the majority of QUADAS-items, except for the items 'withdrawals explained' (F4), and 'uninterpretable results reported' (F5 and F6) which were often not documented. Blinding of test and gold standard results was in some cases not assured or not sufficiently documented. In the study of Balasubramanian et al.[9], the cut-off for test-positivity of the $p16^{INK4a}$ assay was determined after analysis of the samples. The retrospective study of Cuzick et al., suffered from partial verification since the decision to perform gold standard verification was based on the cytological results only.

Table 15: Evaluation of the quality of each included study according to the QUADAS-2 check list[20].

	Risk of Bias												
	Patient Selection		Screening Test		Reference test			Flow & Timing					
Author, Year	**P1**	**P2**	**T1**	**T2**	**R1**	**R2**	**R3**	**F1**	**F2**	**F3**	**F4**	**F5**	**F6**
Balasubra., 2009	Y	Y	N	?	Y	?	Y	Y	Y	Y	N	N	Y
Hovland, 2010	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	N	N	N
Wu, 2010	Y	Y	?	Y	Y	Y	Y	?	Y	Y	Y	N	N
Depuydt, 2011	Y	Y	Y	?	Y	?	?	Y*	Y	Y	Y	N	N
Monsonego, 2011	Y	Y	Y	Y	Y	N	Y	Y	Y	Y	Y	Y	Y
Ratnam, 2011	Y	?	Y	Y	Y	Y	Y	Y°	Y	Y	N	Y	Y
Cuzick, 2013	Y	Y	Y	?	N	Y	Y	Y°	N	Y	Y	Y	N
Ikenberg, 2013	Y	Y	Y	Y	Y	Y	Y	?	Y	Y	Y	N	Y
Nieves, 2013	Y	Y	?	Y	Y	N	Y	?	Y	Y	Y	Y	Y
Zhao, 2013	Y	Y	Y	?	Y	?	Y	?	Y	Y	Y	N	N

QUADAS items: (P1) acceptable enrolment method, (P2) inappropriate exclusions avoided, (T1) pre-specified test cut-off, (T2) results of index and comparator tests blinded towards each other and reference test, (R1) acceptable reference test, (R2) results of reference test blinded towards index and comparator tests, (R3) incorporation bias avoided, (F1) acceptable delay between triage tests and reference test, (F2) partial verification avoided, (F3) differential verification avoided, (F4) withdrawals explained, (F5) uninterpretable results reported for tests, (F6) uninterpretable results reported for reference test. Each quality item is judged with: Y (fulfilled, green), ? (unclear, yellow), N (not fulfilled, red). *follow-up of 24m, °follow-up of 6 months

2.5.4. Absolute accuracy of testing with a biomarker in primary screening

The absolute sensitivity and specificity estimates for CIN2+ or CIN 3+ are listed for all included studies in

Table 16. Because the ProEx C and p16INK4a-ELISA assays were each evaluated in just one study, only accuracy values for the mRNA assays were pooled (Table 17).

Using a >5-type mRNA assay, the pooled sensitivity to detect CIN2+ and CIN3+ was 95% (95% CI = 87-98%) and 99% (95% CI = 78-100%), respectively. The specificity to exclude CIN2+ was 92% (95% CI = 90-93%) (see Table 6, Figure 2-3). Using a mRNA-assay that detects 5 HPV-types, pooling of two studies resulted in a sensitivity and specificity of 75% (95% CI = 63-87%) and 96% (95% CI = 95-96%), for the outcome CIN2+ (see Table 17, Figure 20, Figure 21, Figure 22). The sensitivity of ProExC for CIN2+ (78%) was similar to that of the 5-type mRNA assay whereas the sensitivity of ProcExC CIN3+ (92%) was higher. The p16 ELISA test at the intended cut-offs showed very low sensitivity and moderate specificity. A high sensitivity (96%) could be reached with revised a posteriori interpretation, but this was accompanied with a very substantial loss in specificity.

Cyto-immunochemistry with anti-p16/Ki67 showed a sensitivity of 86% and 89% for respectivty CIN2+ and CIN3+ with a high specificity (95% for excluding CIN2+). The OncoTect test (identifying E6 protein) showed poor sensitivity (42% for CIN2+ and 54% for CIN3+) but very high specificity (99%).

Table 16: Absolute sensitivity and specificity (95% CI) of testing with a biomarker to detect CIN2+ and CIN3+, separately for each included study.

Study	Sensitivity		Specificity	
	CIN2+	CIN3+	CIN2+	CIN3+
mRNA >5 types (APTIMA)				
Hovland, 2010†	0.94 (0.70-1.00)	-	0.94 (0.91-0.96)	-
Wu, 2010	1.00 (0.87-1.00)	1.00 (0.78-1.00)	0.91 (0.90-0.92)	0.91 (0.89-0.92)
Monsonego, 2011	0.92 (0.85-0.97)	0.96 (0.81-1.00)	0.92 (0.91-0.92)	0.90 (0.89-0.91)
Ratnam, 2011	1.00 (0.59-1.00)	-	0.88 (0.87-0.90)	-
Cuzick, 2013	0.98 (0.87-1.00)	1.00 (0.82-1.00)	0.90 (0.89-0.91)	0.90 (0.89-0.91)
Nieves, 2013	0.80 (0.64-0.91)	1.00 (0.79-1.00)	0.94 (0.93-0.95)	0.93 (0.92-0.94)
mRNA 5 types (Pretect HPV-Proofer)				
Hovland, 2010	0.81 (0.54-0.96)	-	0.97 (0.94-0.98)	-
Cuzick, 2013	0.73 (0.56-0.85)	0.69 (0.41-0.89)	0.95 (0.95-0.96)	0.95 (0.94-0.96)
TOP2A/MCM2 (ProExC)				
Depuydt, 2011	0.78 (0.64-0.89)	0.92 (0.74-0.99)	0.91 (0.90-0.92)	0.91 (0.90-0.92)
p16^{INK4a} (p16^{INK4a} ELISA)				
Balasubr., 2009 (orig., 6pg)*	-	0.38 (0.26-0.50)	-	0.77 (0.75-0.79)
Balasubr., 2009 (orig., 8pg)*	-	0.29 (0.19-0.41)	-	0.90 (0.89-0.92)
Balasubr., 2009 (enhan., 6pg)*	-	0.96 (0.88-0.99)	-	0.67 (0.64-0.69)
Balasubr., 2009 (enhan, 8pg)*	-	0.83 (0.72-0.91)	-	0.83 (0.81-0.84)
P16/Ki-67 (CINtec Plus)				
Ikenberg, 2013	0.86 (0.80-0.91)	0.87 (0.79-0.93)	0.95 (0.95-0.95)	0.95 (0.95-0.95)
E6 (OncoE6)				
Zhao, 2013	0.42 (0.34-0.51)	0.54 (0.43-0.64)	0.99 (0.99-0.99)	0.99 (0.99-0.99)

†In Hovland 2010, NASBA methodology was used to detect mRNA of 9 hrHPV types. *In Balasubramanian 2009, two methods were performed, the original p16^{INK4a} ELISA (orig.), and an enhanced version of this method (enhan.). For each method, two positivity cut-offs were applied (6 pg/ml and 8 pg/ml).

Abbreviations: CI, confidence interval; CIN2+, cervical intra-epithelial neoplasia grade two or worse; CIN3+, cervical intra-epithelial neoplasia grade three or worse; MCM2, minichromosome maintenance protein-2; mRNA, messenger ribonucleic acid; p16^{INK4a}, cyclin-dependent kinase inhibitor 2A; TOP2A, topoisomerase II-alpha.

Table 17: Pooled absolute sensitivity and specificity of using mRNA testing to detect CIN2+ and CIN3+.

Biomarker	Number of studies		Sensitivity (95% CI)		Specificity (95%CI)	
	CIN2+	CIN3+	CIN2+	CIN3+	CIN2+	CIN3+
mRNA >5types*	6	4	0.95 (0.87-0.98)	0.99 (0.78-1.00)	0.92 (0.90-0.93)	0.91 (0.90-0.92)
mRNA 5types$	2	1	0.75 (0.63-0.87)	0.69 (0.41-0.89)	0.96 (0.95-0.96)	0.95 (0.94-0.96)

Abbreviations: CI, confidence interval; CIN2+, cervical intra-epithelial neoplasia grade two or worse; CIN3+, cervical intra-epithelial neoplasia grade three or worse. * joint meta-analytical pooling of sensitivity and specificity with binormal model using metadas; $ separate pooling of sensitivity and specificity using metan.

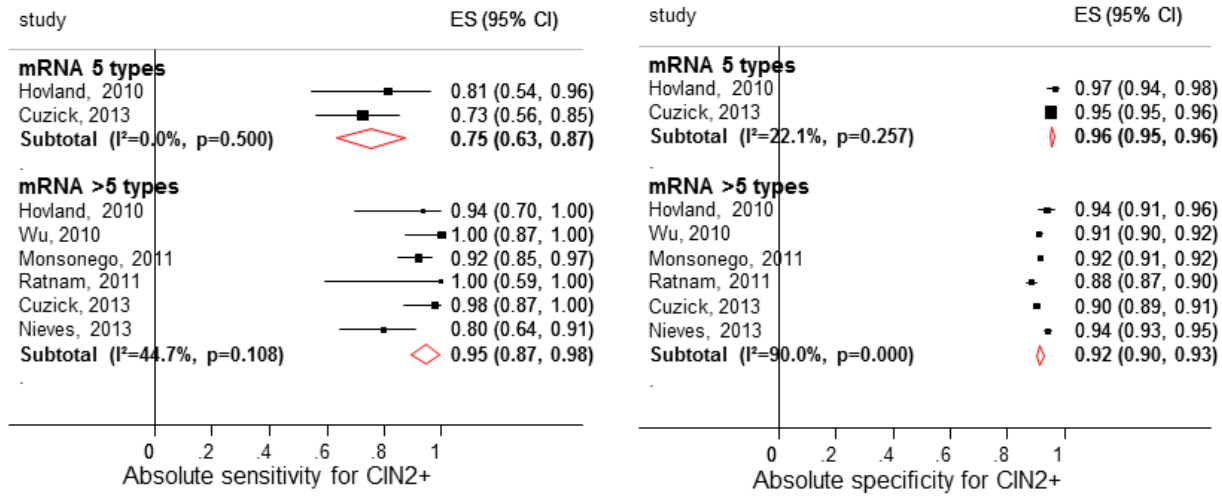


Figure 21: Absolute sensitivity (left) and specificity (right) of mRNA testing to detect CIN2+ in a screening population.

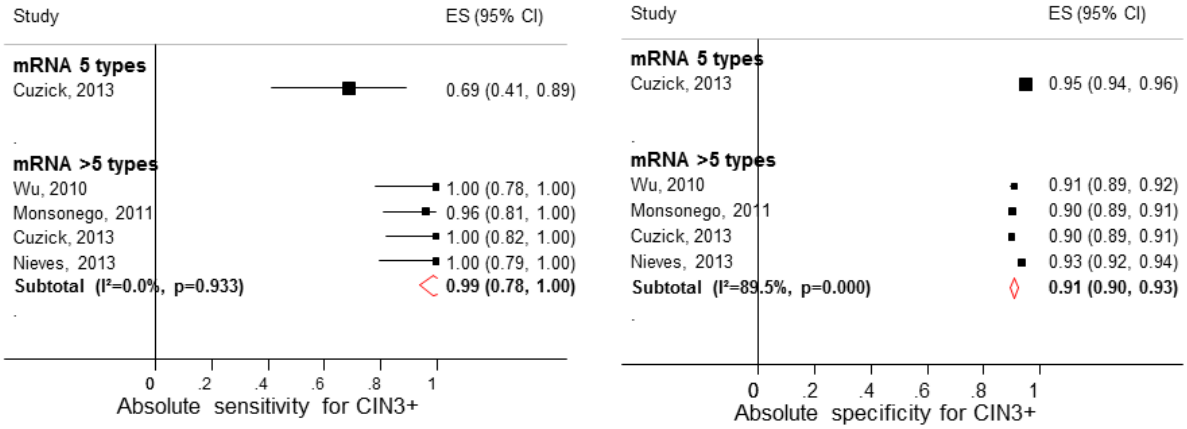


Figure 22: Absolute sensitivity (left) and specificity (right) of mRNA testing to detect CIN3+ in a screening population.

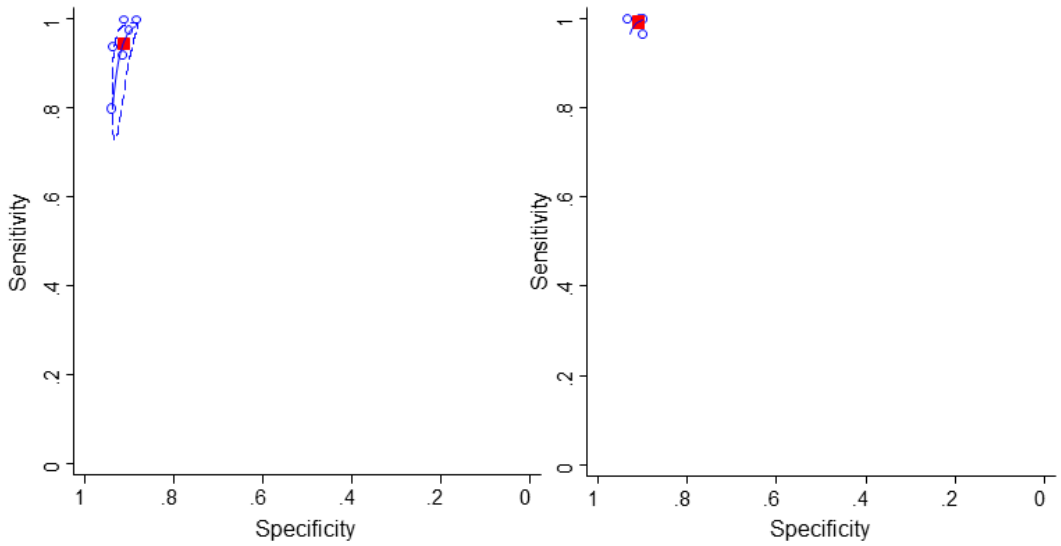


Figure 23: Meta-analysis of the accuracy of mRNA (>5 types) testing for the detection of CIN2+ (left) and CIN3+ (right) in primary cervical cancer screening. Hollow circles represent individual studies, the full curved line is the summary ROC curve, the filled red square is the pooled accuracy value surrounded by the dotted 95% confidence ellips. To detect CIN3+, no confidence ellips is shown due to the little number of studies (n=4).

2.5.5. Relative accuracy of testing with a biomarker versus HPV-DNA testing

The relative accuracy estimates of the biomarkers compared to hrHPV-DNA testing are listed for each study in Table 18. The pooled relative accuracy is documented in Table 19. Due to the limited number of available studies, only the results for both mRNA assays were pooled.

The pooled sensitivity of the >5-type mRNA assay was similar to that of hrHPV-DNA testing, both to detect CIN2+ (ratio: 0.99, 95% CI=0.95-1.04) and CIN 3+ (ratio: 1.01, 95% CI=0.98-1.04). A significantly improved specificity of the former compared to the latter assay was observed for CIN2+ (ratio: 1.05, 95% CI=1.03-1.07). The mRNA assay detecting five HPV types demonstrated a significantly lower sensitivity (ratio: 0.77, 95% CI=0.65-0.90), but higher specificity (ratio: 1.12, 95% CI=1.10-1.13) for the outcome CIN2+, compared to HPV-DNA testing.

ProExC was 18% (95% CI=4-31%) less sensitive for CIN2+ but not significantly less sensitive for CIN3+ (ratio:0.92, 95% CI=0.80-1.06) than testing with RT-PCR for hrHPV types. ProExC was also more specific in excluding CIN2+ (ratio: 1.07, 95% CI=1.05-1.10). Screening with p16/Ki67 immunostaining was 10% (95% CI=4-16%) less sensitive but 6% (95% CI=5-6%) more specific for the outcome CIN2+ than with HC2. Identification of the E6 oncoproteint detected only halve of CIN2+ or CIN3+ but was substantially more specific than hrhPV testing with HC2 (ratio: 1.14, 95% CI= 1.13-1.15) or careHPV (ratio: 1.13, 95% CI: 1.12-1.14).

Table 18: Relative sensitivity and specificity of testing with a biomarker compared to HPV-DNA testing, for the outcome CIN2+ and CIN3+.

Study	Comparator test	Relative sensitivity		Relative specificity	
		CIN2+	CIN3+	CIN2+	CIN3+
mRNA >5 types (APTIMA)					
Hovland, 2010†	GP 5+/6+ PCR- EIA	0.94 (0.79-1.11)	-	1.09 (1.04-1.16)	-
Wu, 2010	HC2	1.12 (0.97-1.30)	1.07 (0.89-1.28)	1.08 (1.05-1.10)	1.08 (1.05-1.10)
Monsonego, 2011	HC2	0.95 (0.89-1.01)	1.00 (0.90-1.11)	1.06 (1.05-1.08)	1.06 (1.05-1.08)
Ratnam, 2011	HC2	1.00	-	1.04 (1.01-1.07)	-
Cuzick, 2013	HC2	1.00 (0.93-1.07)	1.00	1.06 (1.04-1.07)	1.06 (1.04-1.07)
	COBAS4800	1.00 (0.93-1.07)	1.00	1.07 (1.05-1.08)	1.07 (1.05-1.08)
	Abbott	1.03 (0.94-1.12)	1.05 (0.91-1.22)	1.04 (1.02-1.05)	1.04 (1.02-1.05)
	Viper BD	1.00 (0.93-1.07)	1.00	1.07 (1.06-1.09)	1.07 (1.06-1.09)
Nieves, 2013	HC2	0.99 (0.80-1.23)	1.00	1.01 (1.00-1.03)	1.01 (1.00-1.03)
mRNA 5 types (Pretect HPV-Proofer)					
Hovland, 2010	GP 5+/6+ PCR- EIA	0.82 (0.63-1.06)	-	1.13 (1.07-1.18)	-
Cuzick, 2013	HC2	0.74 (0.61-0.91)	0.69 (0.50-0.97)	1.12 (1.10-1.13)	1.12 (1.10-1.13)
	COBAS4800	0.74 (0.61-0.91)	0.69 (0.50-0.97)	1.13 (1.11-1.14)	1.13 (1.11-1.14)
	Abbott	0.76 (0.62-0.94)	0.73 (0.51-1.03)	1.09 (1.08-1.11)	1.09 (1.08-1.11)
	Viper BD	0.74 (0.61-	0.69 (0.50-0.97)	1.13 (1.12-1.14)	1.13 (1.12-

		0.91)			1.15)
TOP2A/MCM2 (ProExC)					
Depuydt, 2011	RT-PCR	0.82 (0.69-0.96)	0.92 (0.80-1.06)	1.07 (1.05-1.10)	1.08 (1.06-1.10)
p16^{INK4a} (p16^{INK4a} ELISA)					
Balasubr., 2009 (orig., 6pg)*	HC2	-	0.65 (0.45-0.94)	-	1.11 (1.06-1.15)
Balasubr., 2009 (orig., 8pg)*	HC2	-	0.50 (0.33-0.76)	-	1.29 (1.25-1.34)
Balasubr., 2009 (enhan., 6pg)*	HC2	-	1.65 (1.34-2.03)	-	0.96 (0.91-1.00)
Balasubr., 2009 (enhan, 8pg)*	HC2	-	1.42 (1.13-1.79)	-	1.18 (1.14-1.23)
P16/Ki-67 (CINtec Plus)					
Ikenberg, 2013	HC2	0.90 (0.84-0.96)	-	1.06 (1.05-1.06)	-
E6 (OncoE6)					
Zhao, 2013	HC2	0.44 (0.36-0.54)	0.55 (0.46-0.67)	1.14 (1.13-1.15)	1.14 (1.13-1.15)
Zhao, 2013	CareHPV	0.44 (0.36-0.54)	0.55 (0.46-0.67)	1.13 (1.12-1.14)	1.14 (1.13-1.15)

Red and green font respectively indicate significantly lower and higher accuracy of testing with a biomarker compared to testing with a HPV-DNA test. †In Hovland 2010, NASBA methodology was used to detect mRNA of 9 hrHPV types. *In Balasubramanian 2009, two methods were performed, the original p16^{INK4a} ELISA (orig.), and an enhanced version of this method (enhan.). For each method, two positivity cut-offs were applied (6 pg/ml and 8 pg/ml).

Abbreviations: CI, confidence interval; CIN2+, cervical intra-epithelial neoplasia grade two or worse; CIN3+, cervical intra-epithelial neoplasia grade three or worse; DNA, deoxiribonucleic acid; HPV, human papillomavirus; MCM2, minichromosome maintenance protein-2; mRNA, messenger ribonucleic acid; p16^{INK4a}, cyclin-dependent kinase inhibitor 2A; TOP2A, topoisomerase II-alpha.

Table 19: Pooled relative sensitivity and specificity of mRNA testing compared to HPV-DNA testing to detect CIN2+ and CIN3+.

	Number of studies (test combination)		Sensitivity ratio (95% CI)		Specificity ratio (95%CI)	
	CIN2+	CIN3+	CIN2+	CIN3+	CIN2+	CIN3+
mRNA versus validated HPV-DNA testing						
mRNA >5types	6	4	0.99 (0.95-1.04)	1.01 (0.98-1.04)	1.05 (1.03-1.07)	1.05 (1.02-1.07)
mRNA 5types	2	1	0.77 (0.65-0.90)	0.69 (0.50-0.97)	1.12 (1.10-1.13)	1.12 (1.10-1.13)

Only comparisons with the validated HC2 assay[9;11;13-16] or PCR Gp5+/6+ assay[10] are included. Red and green font respectively indicate significantly lower and higher accuracy of testing with a biomarker compared to testing with a HPV-DNA test.

Abbreviations: CI, confidence interval; CIN2+, cervical intra-epithelial neoplasia grade two or worse; CIN3+, cervical intra-epithelial neoplasia grade three or worse; DNA, deoxiribonucleic acid; HPV, human papillomavirus; mRNA, messenger ribonucleic acid.

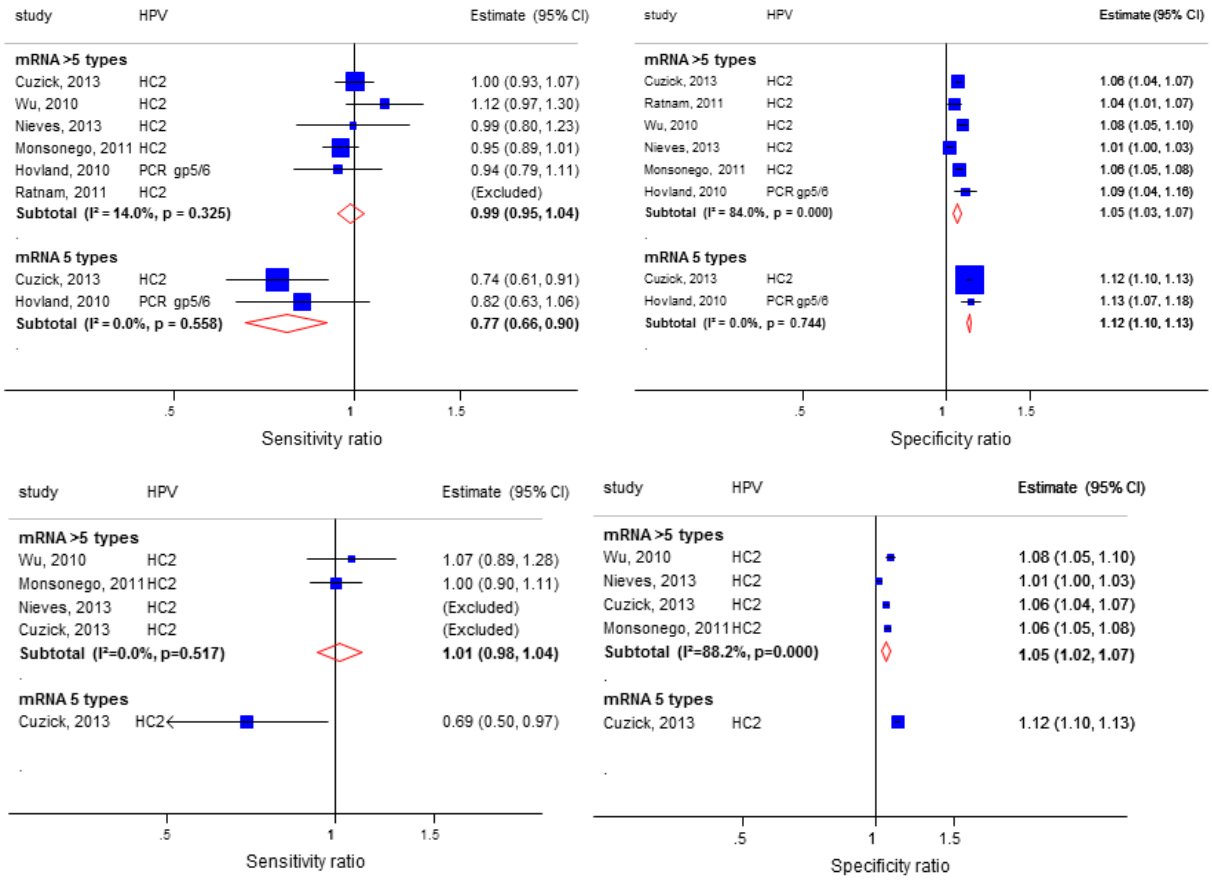


Figure 24: Relative sensitivity (left) and specificity (right) of testing with a mRNA test compared to testing with a validated HPV-DNA test to detect CIN2+ (top) and CIN3+ (bottom). p=test for inter-study heterogeneity, I^2=the percentage of total variations across studies due to inter-study

heterogeneity. Abbreviations: HC2, Hybrid Captury-2; PCR gp5/6, Polymerase Chain Reaction using consensus primers GP5+/6+.

2.5.6. Relative accuracy of testing with a biomarker versus cytology

5.6.1. Comparator: cytology at cutoff ASC-US+

The relative accuracy of testing with a biomarker versus cytology at cut-off ASC-US and LSIL for the outcome CIN2+ and CIN3+ is documented in Table 20 for each study. The pooled accuracy values for >5-type mRNA testing and for 5-type mRNA testing was are presented in Table 21.

Compared to LBC at cut-off ASC-US, detecting the mRNA of more than five types demonstrated a non-significantly higher sensitivity (ratio: 1.14, 95% CI=0.88-1.49) and similar specificity (ratio: 1.01, 95% CI=0.94-1.08) for the outcome CIN2+, compared to cytology-testing at cut-off ASC-US. For the outcome CIN3+, the >5-type mRNA assay demonstrated an improved sensitivity (ratio: 1.21, 95% CI=1.03-1.42) and similar specificity (ratio: 0.98, 95 % CI=0.94-1.01). The 5-type mRNA assay demonstrated similar sensitivity and specificity as cytology testing for CIN2+, but a lower sensitivity (ratio: 0.69, 95% CI=0.50-0.97) to detect CIN3+.

Compared to ASC-US+ cytology, ProExC and p16/Ki67 staining were significantly more sensitive (ratio: 1.33, 95% CI=1.00-1.78; and 1.21, 95% CI=1.08-1.35, respectively) whereas ProEx was less specific (ratio: 0.97, 95% CI=0.95-0.98) but p16/Ki67 staining was as specific (ratio: 1.00, 95% CI=0.99-1.00).

5.6.2. Comparator: cytology at cutoff LSIL+

Compared to cytology at cut-off LSIL, >5-type mRNA testing showed a substantially higher sensitivity (although non-significant) for CIN2+ (ratio: 1.32, 95% CI=0.97-1.81) and CIN3+ (ratio: 1.25, 95% CI=0.77-2.03) whereas specificity was significantly lower (ratio: 0.95, 95% CI=0.91-0.98 for CIN2+; ratio: 0.94, 95% CI=0.90-0.98 for CIN3+).

The sensitivity of 5-type mRNA testing was not significantly different from cytology (ratio: 1.00, 95% CI=0.60-1.67 for CIN2+; ratio: 0.73, 95% CI=0.51-1.03 for CIN3+) but the specificity was slightly but significantly lower (ratio: 0.96, 95% CI=0.96-0.99 for CIN2+; ratio: 0.97, 95% CI=0.96-0.98).

ProExC was substantially and significantly more sensitive (ratio: 1.57, 95% CI=1.13-2.17 for CIN2+; ratio: 1.92, 95% CI=1.25-2.93 for CIN3+) but less specific (ratio: 0.93, 95% CI=0.92-0.94 for CIN2+ and CIN3+) than LSIL+ cytology.

Table 20: Relative sensitivities and specificities (95% CI) of testing with a biomarker compared to cytology at cut-off ASC-US and LSIL, for the outcome CIN2+ and CIN3+.

Study	Comparator test	Relative sensitivity		Relative specificity	
		CIN2+	CIN3+	CIN2+	CIN3+
Cut-off ASC-US					
mRNA >5 types (APTIMA)					
Hovland, 2010†	LBC	1.28 (0.92-1.78)	-	0.97 (0.94-1.01)	-
	cPAP	1.14 (0.96-2.06)	-	0.98 (0.94-1.01)	-
Wu, 2010	LBC	1.49 (1.13-1.95)	1.48 (1.02-2.13)	0.96 (0.94-0.97)	0.95 (0.94-0.97)
Monsonego, 2011	LBC	1.33(1.15-1.53)	1.30 (1.03-1.64)	1.00 (0.99-1.01)	0.99 (0.98-1.01)
Cuzick, 2013	LBC	0.98 (0.91-1.04)	1.00	0.95 (0.94-0.96)	0.95 (0.94-0.96)
Nieves, 2013		1.06 (0.84-1.34)	1.14 (0.92-1.41)	1.12 (1.10-1.15)	0.99 (0.98-1.01)
mRNA 5 types (Pretect HPV-Proofer)					
Hovland, 2010	LBC	1.11 (0.75-1.63)	-	1.00 (0.97-1.03)	-
	cPAP	1.22 (0.79-1.87)	-	1.00 (0.97-1.04)	-
Cuzick, 2013	LBC	0.73 (0.60-0.88)	0.69 (0.50-0.97)	1.00 (0.99-1.01)	1.00 (0.99-1.01)
TOP2A/MCM2 (ProExC)					
Depuydt, 2011	LBC	1.33 (1.00-1.78)	1.77 (1.19-2.62)	0.97 (0.95-0.98)	0.97 (0.95-0.98)
P16/Ki-67 (CINtec Plus)					
Ikenberg, 2013	cPAP/LBC	1.21 (1.08-1.35)	1.18 (1.02-1.35)	1.00 (0.99-1.00)	1.00 (0.99-1.00)
Cut-off LSIL					
mRNA >5 types (APTIMA)					
Hovland, 2010†	LBC	1.56 (1.01-2.41)	-	0.96 (0.93-1.00)	-
	cPAP	1.56 (1.01-2.41)	-	0.96 (0.93-1.00)	-
Monsonego, 2011	LBC	1.48 (1.26-1.74)	1.53 (1.13-2.06)	0.96 (0.95-0.97)	0.96 (0.95-0.97)
Cuzick, 2013	LBC	1.08 (0.97-1.21)	1.05 (0.91-1.22)	0.92 (0.91-0.93)	0.92 (0.91-0.93)
mRNA 5 types (Pretect HPV-Proofer)					
Hovland, 2010	LBC	1.35 (0.84-2.18)	-	0.99 (0.96-1.02)	-
	cPAP	1.35 (0.84-	-	0.99 (0.96-	-

Study	Comparator test	Relative sensitivity		Relative specificity	
		CIN2+	CIN3+	CIN2+	CIN3+
Cuzick, 2013	LBC	2.18) 0.81 (0.65-1.00)	0.73 (0.51-1.03)	1.02) 0.97 (0.96-0.98)	0.97 (0.96-0.98)
TOP2A/MCM2 (ProExC)					
Depuydt, 2011	LBC	1.57 (1.13-2.17)	1.92 (1.25-2.93)	0.93 (0.92-0.94)	0.93 (0.92-0.94)

Red and green font respectively indicate significantly lower and higher accuracy of testing with a biomarker compared to testing with a HPV-DNA test. †In Hovland 2010, NASBA methodology was used to detect mRNA of 9 hrHPV types.

Abbreviations: ASC-US, atypical squamous cells of undetermined significance; CI, confidence interval; CIN2+, cervical intra-epithelial neoplasia grade two or worse; CIN3+, cervical intra-epithelial neoplasia grade three or worse; cPap, conventional Papanicolaou smear; LBC, liquid based cytology; LSIL, low-grade squamous intra-epithelial lesions; MCM2, minichromosome maintenance protein-2; mRNA, messenger ribonucleic acid; TOP2A, topoisomerase II-alpha.

Table 21: Pooled relative sensitivity and specificity (95% CI) of mRNA testing compared to LBC testing at cut-off ASC-US and LSIL, to detect CIN2+ and CIN3+.

	Number of studies (test combination)		**Sensitivity ratio** (95% CI)		**Specificity ratio** (95%CI)	
	CIN2+	**CIN3+**	**CIN2+**	**CIN3+**	**CIN2+**	**CIN3+**
mRNA versus LBC (ASC-US)						
mRNA >5types	4	3	1.14 (0.88-1.49)	1.21 (1.03-1.42)	1.01 (0.94-1.08)	0.98 (0.94-1.01)
mRNA 5types	2	1	0.87 (0.58-1.31)	0.69 (0.50-0.97)	1.00 (0.99-1.01)	1.00 (0.99-1.01)
mRNA versus LBC (LSIL)						
mRNA >5types	3	2	1.32 (0.97-1.81)	1.25 (0.77-2.03)	0.95 (0.91-0.98)	0.94 (0.90-0.98)
mRNA 5types	2	1	1.00 (0.60-1.67)	0.73 (0.51-1.03)	0.96 (0.96-0.99)	0.97 (0.96-0.98)

Red and green font respectively indicate significantly lower and higher accuracy of testing with a biomarker compared to testing with a HPV-DNA test. Abbreviations: ASC-US+, atypical squamous cells of undetermined significance or worse; CI, confidence interval; CIN2+, cervical intra-epithelial neoplasia grade two or worse; CIN3+, cervical intra-epithelial neoplasia grade three or worse; LSIL+, low grade squamous intra-epithelial lesions or worse; mRNA, messenger ribonucleic acid.

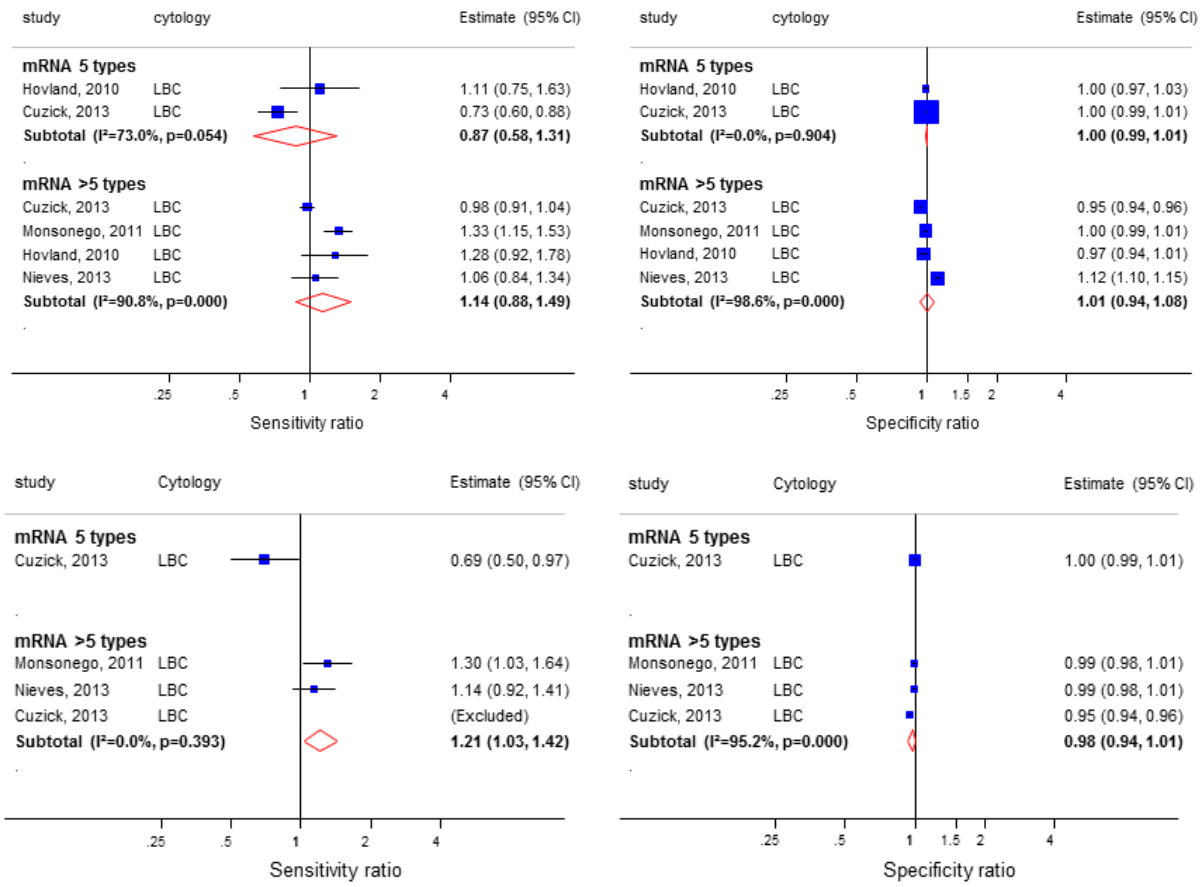


Figure 25: Relative sensitivity (left) and specificity (right) of testing with a mRNA test compared to testing with liquid based cytology (ASC-US+) to detect CIN2+ (top) and CIN3+ (bottom). p=test for inter-study heterogeneity, I^2=the percentage of total variations across studies due to inter-study heterogeneity. Abbreviations: ASC-US+, Atypical Squamous Cells of Undetermined Significance or worse; CIN2+, Cervical Intraepithelial Neoplasia grade 2 or worse; CIN3+, Cervical Intraepithelial Neoplasia grade 3 or worse; LBC, Liquid Based Cytology; mRNA, messenger ribonucleic acid.

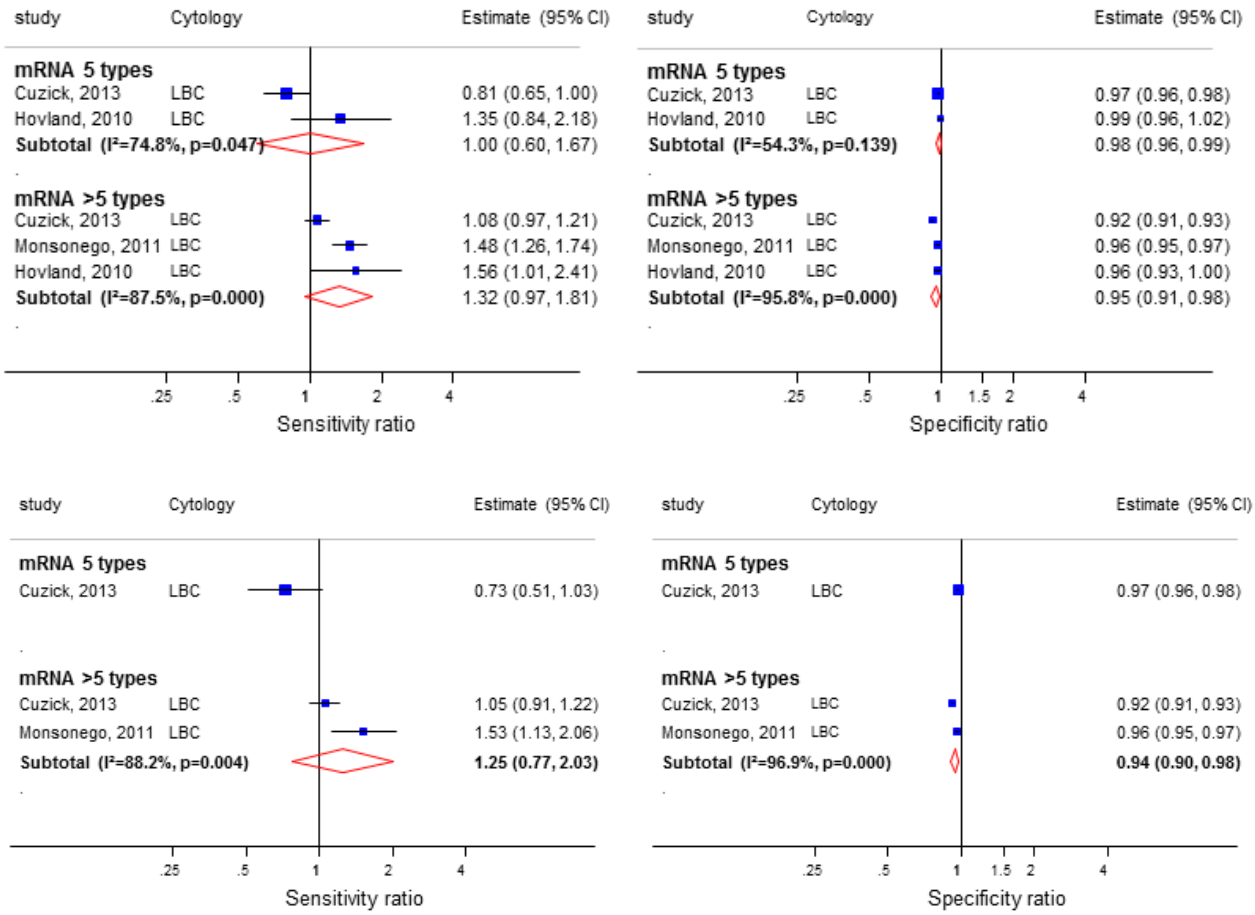


Figure 26: Relative sensitivity (left) and specificity (right) of testing with a mRNA test compared to testing with liquid based cytology (LSIL+) to detect CIN2+ (top) and CIN3+ (bottom). p=test for inter-study heterogeneity, I^2=the percentage of total variations across studies due to inter-study heterogeneity. Abbreviations: CIN2+, Cervical Intraepithelial Neoplasia grade 2 or worse; CIN3+, Cervical Intraepithelial Neoplasia grade 3 or worse; LBC, Liquid Based Cytology; LSIL+, Low-grade Squamous Intra-epithelial Lesions or worse; mRNA, messenger ribonucleic acid.

2.6. Discussion and interpretation

Only for screening with mRNA, multiple studies could be pooled. mRNA testing for 9-14 types is as sensitive as screening with hrHPV-DNA tests and slightly more specific for identifying underlying high-grade CIN. Testing for transcripts of only 5 HPV types usually was less sensitive but substantially more specific than screening with hrHPV DNA tests. TOP2A/MCM2 or p16/Ki67 immunocytochemistry was substantially and slightly less sensitive than hrHPV DNA testing but both markers showed similarly increased specificity. Staining for E6 was poorly sensitive in finding CIN3+ but was very specific in excluding high-grade CIN.

Given the lower sensitivity and the higher specificity of most of the evaluated markers (5-type mRNA testing and protein markers) than hrHPV-DNA screening, triage of hrHPV-positive women identified through HPV-based screening may be a more appropriate application of these biomarkers than in primary screening. However screening for E6/E7 mRNA of 14 hrHPVtypes (APTIMA) was not accompanied with a loss in sensitivity when compared to hrHPV-DNA screening. Moreover this 14- type mRNA test shows a small but significant gain in specificity. Nonetheless, the demonstration over a long period of at least five years, a similarly low cumulative incidence among APTIMA-negative women as among hrHPV-DNA negative women is not yet demonstrated. This longitudinal outcome should preferentially be documented before recommending APTIMA for primary screening.

Screening with biomarkers often was more sensitive than cervical cytology in detecting high-grade CIN and slightly less specific. However, given the very well documented higher performance of hrHPV-DNA-based screening compared to cervical cytology (randomised evidence for lower incidence of invasive cervical cancer), the comparison of screening with biomarkers with cervical cytology is less relevant.

In conclusion, no evidence is available today to recommend screening with HPV-mRNA markers or protein markers in primary screening. The APTIMA shows good cross-sectional accuracy for cervical precancer but more data demonstrating the longitudinal safety are needed.

2.7. References

(1) Arbyn M, Ronco G, Anttila A, Meijer CJLM, Poljak M, Ogilvie G et al. Evidence regarding HPV testing in secondary prevention of cervical cancer. *Vaccine* 2012; 30 Suppl 5:F88-F99.

(2) Ronco G, Dillner J, Elfstrom KM, Tunesi S, Snijders PJ, Arbyn M et al. Efficacy of HPV-based screening for prevention of invasive cervical cancer: follow-up of four European randomised controlled trials. *Lancet* 2013;in-press.

(3) Roelens J, Reuschenbach M, von Knebel-Doeberitz M, Wentzensen N, Bergeron C, Arbyn M. p16INK4a immunocytochemistry versus HPV testing for triage of women with minor cytological abnormalities: A systematic review and meta-analysis. *Cancer* 2012; 120(5):294-307.

(4) Verdoodt F, Szarewski A, Halfon P, Cuschieri K, Arbyn M. Triage of women with minor abnormal cervical cytology: Meta-analysis of the accuracy of an assay targeting messenger ribonucleic acid of 5 high-risk human papillomavirus types. *Cancer Cytopathol* 2013;in-press.

(5) Arbyn M, Roelens J, Cuschieri K, Cuzick J, Szarewski A, Ratnam S et al. The APTIMA HPV assay versus the Hybrid Capture II test in triage of women with ASC-US or LSIL cervical cytology: a meta-analysis of the diagnostic accuracy. *Int J Cancer* 2013; 132(1):101-108.

(6) Bergeron C, Ordi J, Schmidt D, Trunk MJ, Keller T, Ridder R. Conjunctive p16INK4a testing significantly increases accuracy in diagnosing high-grade cervical intraepithelial neoplasia. *Am J Clin Pathol* 2010; 133(3):395-406.

(7) Solomon D, Schiffman MA. Have we resolved how to triage equivocal cervical cytology? *J Natl Cancer Inst* 2004; 96(4):250-251.

(8) Arbyn M, Roelens J, Simoens C, Buntinx F, Paraskevaidis E, Martin-Hirsch PPL et al. Human papillomavirus testing versus repeat cytology for triage of minor cytological cervical lesions. *Cochrane Database Syst Rev* 2013; 3(CD008054):1-201.

(9) Balasubramanian A, Hughes J, Mao C, Ridder R, Herkert M, Kiviat NB et al. Evaluation of an ELISA for p16INK4a as a screening test for cervical cancer. *Cancer Epidemiol Biomarkers Prev* 2009; 18(11):3008-3017.

(10) Hovland S, Arbyn M, Lie AK, Ryd W, Borge B, Berle EJ et al. A comprehensive evaluation of the accuracy of cervical pre-cancer detection methods in a high-risk area in East Congo. *Br J Cancer* 2010; 102(6):957-965.

(11) Wu R, Belinson SE, Du H, Na W, Qu X, Wu R et al. Human papillomavirus messenger RNA assay for cervical cancer screening: the Shenzhen Cervical Cancer Screening Trial I. *Int J Gynecol Cancer* 2010; 20(8):1411-1414.

(12) Depuydt CE, Makar AP, Ruymbeke MJ, Benoy IH, Vereecken AJ, Bogers JJ. BD-ProExC as Adjunct Molecular Marker for Improved Detection of CIN2+ after HPV Primary Screening. *Cancer Epidemiol Biomarkers Prev* 2011; 20(4):628-637.

(13) Monsonego J, Hudgens MG, Zerat L, Zerat JC, Syrjanen K, Halfon P et al. Evaluation of oncogenic human papillomavirus RNA and DNA tests with liquid based cytology in primary cervical cancer screening (The FASE study). *Int J Cancer* 2011; 129(3):691-701.

(14) Ratnam S, Coutlee F, Fontaine D, Bentley J, Escott N, Ghatage P et al. Aptima HPV E6/E7 mRNA test is as sensitive as Hybrid Capture 2 Assay but more specific at detecting cervical precancer and cancer. *J Clin Microbiol* 2011; 49(2):557-564.

(15) Cuzick J, Cadman L, Mesher D, Austin J, Ashdown-Barr L, Ho L et al. Comparing the performance of six human papillomavirus tests in a screening population. *Br J Cancer* 2013; 108(4):908-913.

(16) Nieves L, Enerson CL, Belinson S, Brainard J, Chiesa-Vottero A, Nagore N et al. Primary cervical cancer screening and triage using an mRNA human papillomavirus assay and visual inspection. *Int J Gynecol Cancer* 2013; 23(3):513-518.

(17) Ikenberg H, Bergeron C, Schmidt D, Griesser H, Alameda F, Angeloni C et al. Screening for cervical cancer precursors with p16/Ki-67 dual-stained cytology: results of the PALMS study. *J Natl Cancer Inst* 2013;in-press.

(18) Zhao FH, Jeronimo JA, Qiao YL, Schweizer J, Chen W, Valdez M et al. An Evaluation of Novel, Lower-Cost Molecular Screening Tests for Human Papillomavirus in Rural China. *Cancer Prev Res* 2013;1-41.

(19) Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Ann Intern Med* 2009; 151(4):264-9, W64.

(20) 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):529-536.

2.8. GRADE-Profil

Authors: M. Arbyn, M. Jentschke

GRADE has built on previous systems to create a highly structured, transparent, and informative system for rating quality of evidence(Guyatt *et al.,* 2008b).

Steps in evidence assessment for making guidelines
1) Formulate a question
2) Identify the PICO(S) components
3) Qualify outcomes as critical, important, not important
~~Scaling of Critical, Important but not critical, Limited.~~

1) Questions
Is primary screening with a biomarker better than conventional cytology or HPV-testing?

2) PICOS

- P: women participating in cervical cancer screening
- I: testing with a biomarker (p16, p16/Ki-67 dual-stain, ProExC, E6/E7 mRNA, methylation markers, , or other)
- C1: cytology (conventional Pap smear, LBC)
- C2: HPV testing (HC2, GP5+/6+ PCR, or another clinically validated hrHPV DNA test)
- O: accuracy to detect underlying disease (=CIN2+,CIN3+/AIS):
    - Complete diagnostic studies: absolute and relative sensitivity and specificity, PPV, NPV, referral rate, detection rate, detection rate ratio
    - RCTs: relative sensitivity (or detection rate ratios), relative PPV, relative referral rate
- S:
    - diagnostic studies
        - all subjects receiving testing with a biomarker,
        - at least one comparator test
        - verification with the reference standard (colposcopy/biopsy)
            - all participants (accepting a negative colposcopy as free of CIN2+)
            - participants positive in at least one screening test (accepting a negative result for all screening tests as free of disease)

- RCTs comparing screening with biomarkers with screening with one or more comparator tests.

3) Importance of outcomes

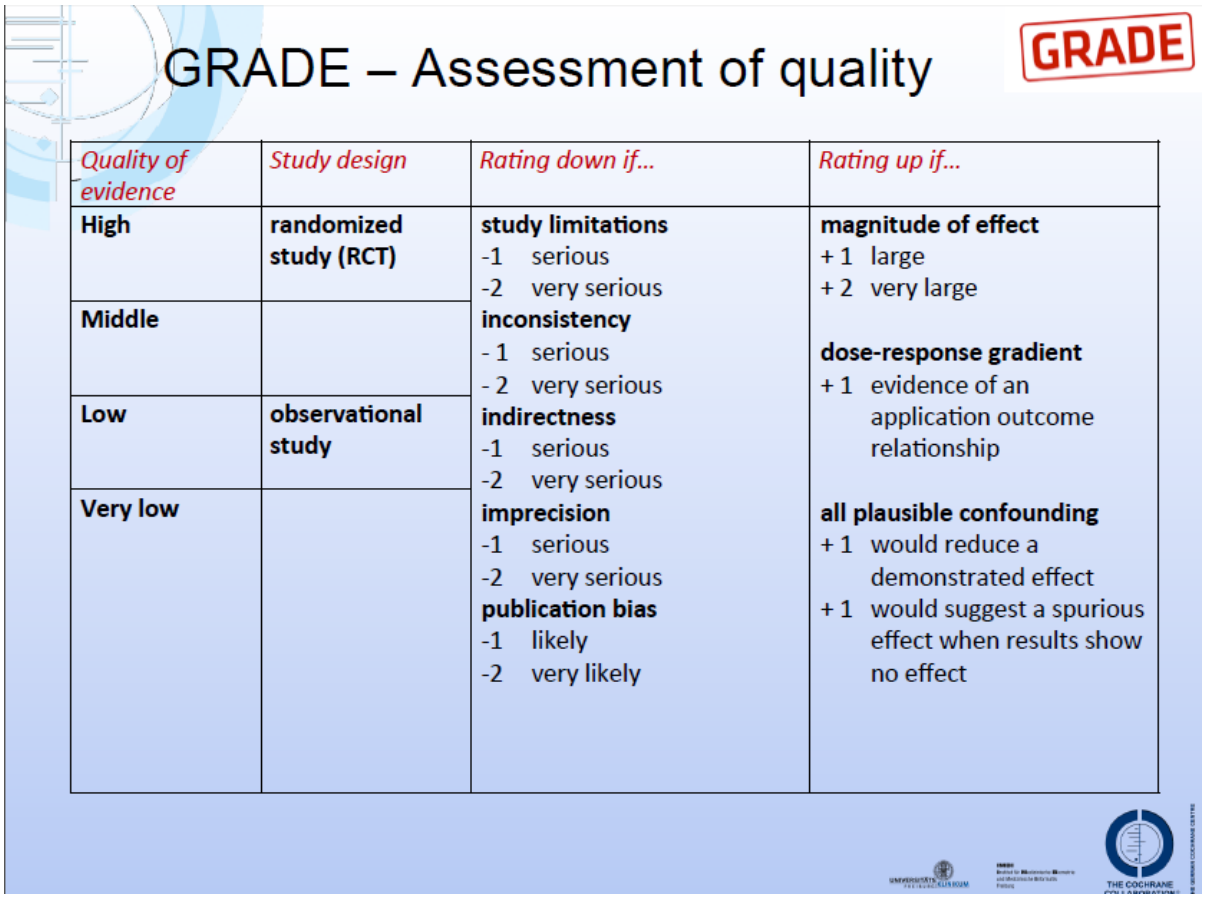
Outcome:

7. Reduction of mortality from cervical cancer, (quality-adjusted) life-years gained.
8. Reduction of morbidity due to cervical cancer: incidence of cancer (Ib+).
9. Reduction of incidence of cancer (including micro-invasive cancer).
10. Reduction of incidence of CIN3 or worse disease (CIN3+).
11. Increased detection rate of CIN3+ or CIN2+.
12. Increased test positivity with increased, similar or hardly reduced positive predictive value.

Preferentially reduction in disease (cumulative incidence of disease [CIN3+ or cervical cancer] will be looked at. By lack of longitudinal outcomes, indicators of diagnostic accuracy will be assessed. Colposcopy referral rates and false positivity rates will be extracted and pooled where possible.

4) Quality of evidence for each outcome in four categories;

- High: ++++
- Moderate: +++
- Low: ++
- Very low: +



The quality of included studies was evaluated using the QUADAS-2 tool[20] and is summarized in Table 15.

Overall, studies scored well for the majority of QUADAS-items, except for the items 'withdrawals explained' (F4), and 'uninterpretable results reported' (F5 and F6) which were often not documented. Blinding of test and gold standard results was in some cases not assured or not sufficiently documented. In the study of Balasubramanian et al.[9], the cut-off for test-positivity of the $p16^{INK4a}$ assay was determined after analysis of the samples. The retrospective study of Cuzick et al., suffered from partial verification since the decision to perform gold standard verification was based on the cytological results only.

Table 22: Evaluation of the quality of each included study according to the QUADAS-2 check list[20].

	Risk of Bias												
	Patient Selection		Screening Test		Reference test			Flow & Timing					
Author, Year	P1	P2	T1	T2	R1	R2	R3	F1	F2	F3	F4	F5	F6
Balasubra., 2009	Y	Y	N	?	Y	?	Y	Y	Y	Y	N	N	Y

	Risk of Bias												
	Patient Selection		Screening Test		Reference test			Flow & Timing					
Author, Year	P1	P2	T1	T2	R1	R2	R3	F1	F2	F3	F4	F5	F6
Hovland, 2010	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	N	N	N
Wu, 2010	Y	Y	?	Y	Y	Y	Y	?	Y	Y	Y	N	N
Depuydt, 2011	Y	Y	Y	?	Y	?	?	Y*	Y	Y	Y	N	N
Monsonego, 2011	Y	Y	Y	Y	Y	N	Y	Y	Y	Y	Y	Y	Y
Ratnam, 2011	Y	?	Y	Y	Y	Y	Y	Y°	Y	Y	N	Y	Y
Cuzick, 2013	Y	Y	Y	?	N	Y	Y	Y°	N	Y	Y	Y	N
Ikenberg, 2013	Y	Y	Y	Y	Y	Y	Y	?	Y	Y	Y	N	Y
Nieves, 2013	Y	Y	?	Y	Y	N	Y	?	Y	Y	Y	Y	Y
Zhao, 2013	Y	Y	Y	?	Y	?	Y	?	Y	Y	Y	N	N

QUADAS items: (P1) acceptable enrolment method, (P2) inappropriate exclusions avoided, (T1) pre-specified test cut-off, (T2) results of index and comparator tests blinded towards each other and reference test, (R1) acceptable reference test, (R2) results of reference test blinded towards index and comparator tests, (R3) incorporation bias avoided, (F1) acceptable delay between triage tests and reference test, (F2) partial verification avoided, (F3) differential verification avoided, (F4) withdrawals explained, (F5) uninterpretable results reported for tests, (F6) uninterpretable results reported for reference test. Each quality item is judged with: Y (fulfilled, green), ? (unclear, yellow), N (not fulfilled, red). *follow-up of 24m, °follow-up of 6 months

5 factors that lower the quality of the evidence (Guyatt *et al.*, 2008a)

6. Risk of bias due to limitations in design or execution: see CONSORT, STROBE, QUADAS
7. Inconsistency or heterogeneity: if consistency unexplained, lower quality
8. Indirectness, applicability (relevance of studies for answering the PICPO question)
9. Imprecision: number of studies, width of CI
10. Reporting bias, publication bias.

3 factors that increase the quality

1. Large effect
2. Dose effect gradient
3. Confounders (presence of confounders that have lowered the observed effect, absence of these would have increased the effect) (Guyatt *et al.*, 2008a)

Items downgrading quality of evidence		Downgrading
Bias, design	The QUADAS assessment generally provided a **good** scoring of the majority of studies. The QUADAS issues did not influence study outcomes significantly.	No (-0)

Inconsistency	No	No (-0)
Indirectness	No	No (-0)
Imprecision	Using a >5-type mRNA assay, the pooled sensitivity to detect CIN2+ and CIN3+ was 95% (95% CI = 87-98%) and 99% (95% CI = 78-100%). The specificity to exclude CIN2+ was 92% (95% CI = 90-93%). ProEx C and p16INK4a-ELISA assays were each evaluated in just one study Compared to LBC at cut-off ASC-US, detecting the mRNA of more than five types demonstrated a non-significantly higher sensitivity (ratio: 1.14, 95% CI=0.88-1.49) and similar specificity (ratio: 1.01, 95% CI=0.94-1.08) for the outcome CIN2+, compared to cytology-testing at cut-off ASC-US. For the outcome CIN3+, the >5-type mRNA assay demonstrated an improved sensitivity (ratio: 1.21, 95% CI=1.03-1.42) and similar specificity (ratio: 0.98, 95 % CI=0.94-1.01). The 5-type mRNA assay demonstrated similar sensitivity and specificity as cytology testing for CIN2+, but a lower sensitivity (ratio: 0.69, 95% CI=0.50-0.97) to detect CIN3+. Compared to ASC-US+ cytology, ProExC and p16/Ki67 staining were significantly more sensitive (ratio: 1.33, 95% CI=1.00-1.78; and 1.21, 95% CI=1.08-1.35, respectively) whereas ProEx was less specific (ratio: 0.97, 95% CI=0.95-0.98) but p16/Ki67 staining was as specific (ratio: 1.00, 95% CI=0.99-1.00). Compared to cytology at cut-off LSIL, >5-type mRNA testing showed a substantially higher sensitivity (although non-significant) for CIN2+ (ratio: 1.32, 95% CI=0.97-1.81) and CIN3+ (ratio: 1.25, 95% CI=0.77-2.03) whereas specificity was significantly lower (ratio: 0.95, 95% CI=0.91-0.98 for CIN2+; ratio: 0.94, 95% CI=0.90-0.98	Yes (-1), at least regarding ProEx C and p16INK4a-ELISA

	for CIN3+). The sensitivity of 5-type mRNA testing was not significantly different from cytology (ratio: 1.00, 95% CI=0.60-1.67 for CIN2+; ratio: 0.73, 95% CI=0.51-1.03 for CIN3+) but the specificity was slightly but significantly lower (ratio: 0.96, 95% CI=0.96-0.99 for CIN2+; ratio: 0.97, 95% CI=0.96-0.98). ProExC was substantially and significantly more sensitive (ratio: 1.57, 95% CI=1.13-2.17 for CIN2+; ratio: 1.92, 95% CI=1.25-2.93 for CIN3+) but less specific (ratio: 0.93, 95% CI=0.92-0.94 for CIN2+ and CIN3+) than LSIL+ cytology.	
Publication bias, other	No information	No (-0)
Items upgrading quality of evidence		
Large effect	No	No (+0)
Dose-effect correlation	No	No (+0)
Confounding factors neutralising effects	No	No (+0)˚

<u>Conclusion: evidence of low quality.</u>

Overall Grade of the quality of evidence is assigned and this is based on the outcome with the lowest quality of evidence given that it is a critical outcome.

For the accuracy we consider the relative sensitivity (outcome CIN3+) and specificity (outcome CIN2+) as critical. The other outcomes: absolute accuracy, relative sensitivity for CIN2+ and relative specificity for CIN3+ are considered as important.

Table 23 GRADE evidence profile

	Quality of evidence								
# studies (N)	Absence of study limitati	Consistency	Directness (outcome, representativity	Precision	Absence publication bias	Large effect	Dose-effect relation	Bias lowering effect	Comment

	ons		Germany)						
Outcome 1: Relative accuracy of testing with a biomarker versus HPV-DNA testing [CRITICAL]									
10	Yes	Yes	Yes	No	Yes	No	No	No	Low

Outcome 2: Relative accuracy of testing with a biomarker versus cytology [CRITICAL]									
10	Yes	Yes	Yes	No	Yes	No	No	No	Low

Summary of findings

Table 24 Pooled relative sensitivity and specificity of mRNA testing compared to HPV-DNA testing to detect CIN2+ and CIN3+.

	Number of studies (test combination)		Sensitivity ratio (95% CI)		Specificity ratio (95%CI)	
	CIN2+	CIN3+	CIN2+	CIN3+	CIN2+	CIN3+
mRNA versus validated HPV-DNA testing						
mRNA >5types	6	4	0.99 (0.95-1.04)	1.01 (0.98-1.04)	1.05 (1.03-1.07)	1.05 (1.02-1.07)
mRNA 5types	2	1	0.77 (0.65-0.90)	0.69 (0.50-0.97)	1.12 (1.10-1.13)	1.12 (1.10-1.13)

Table 25 Pooled relative sensitivity and specificity (95% CI) of mRNA testing compared to LBC testing at cut-off ASC-US and LSIL, to detect CIN2+ and CIN3+.

	Number of studies (test combination)		Sensitivity ratio (95% CI)		Specificity ratio (95%CI)	
	CIN2+	CIN3+	CIN2+	CIN3+	CIN2+	CIN3+
mRNA versus LBC (ASC-US)						
mRNA >5types	4	3	1.14 (0.88-1.49)	1.21 (1.03-1.42)	1.01 (0.94-1.08)	0.98 (0.94-1.01)
mRNA 5types	2	1	0.87 (0.58-1.31)	0.69 (0.50-0.97)	1.00 (0.99-1.01)	1.00 (0.99-1.01)
mRNA versus LBC (LSIL)						
mRNA >5types	3	2	1.32 (0.97-1.81)	1.25 (0.77-2.03)	0.95 (0.91-0.98)	0.94 (0.90-0.98)
mRNA 5types	2	1	1.00 (0.60-1.67)	0.73 (0.51-1.03)	0.96 (0.96-0.99)	0.97 (0.96-0.98)

References

Guyatt G.H., Oxman A.D., Kunz R., Falck-Ytter Y., Vist G.E., Liberati A., & Schunemann H.J. (2008a) Going from evidence to recommendations. *BMJ* **336:** 1049-1051.

Guyatt G.H., Oxman A.D., Kunz R., Vist G.E., Falck-Ytter Y., & Schunemann H.J. (2008b) What is "quality of evidence" and why is it important to clinicians? *BMJ* **336:** 995-998.

Guyatt G.H., Oxman A.D., Vist G.E., Kunz R., Falck-Ytter Y., Alonso-Coello P., & Schunemann H.J. (2008c) GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *BMJ* **336:** 924-926.

Schunemann H.J., Oxman A.D., Brozek J., Glasziou P., Jaeschke R., Vist G.E., Williams J.W., Jr., Kunz R., Craig J., Montori V.M., Bossuyt P., & Guyatt G.H. (2008) GRADE: Grading quality of evidence and strength of recommendations for diagnostic tests and strategies. *BMJ* **336:** 1106-1110.

3. Question: Triage of women with a positive HPV-test at screening.

Authors: Verdoodt Freija, Arbyn Marc

Unit of Cancer Epidemiology
Wetenschappelijk Instituut Volksgezondheid (WIV)
Juliette Wytsmanstraat 14
1050 Brussel

3.1. Introduction

Meta-analyses discussed in deliverable 1.1. have demonstrated improved sensitivity of high-risk HPV (hrHPV)DNA-based testing, compared to cytology in primary screening with respect to detection of cervical precancer which subsequently results in a lower incidence of CIN3 and cancer observed after a first screening round (reviewed in Arbyn, 2012 Vaccine[1]; and confirmed by a pooled analysis[2]). However, the higher sensitivity for CIN2+ and CIN3+, is associated with a drop in specificity, which results in a decreased cross-sectional positive predictive value (PPV) and may lead to unnecessary follow-up of screen-positive women and over-management of patients. Consequently, the triage of women with a positive hrHPV-DNA test constitutes an important clinical issue to address.

3.2. Materials and methods

3.2.1. Clinical question

In this report, diverse triage methods are evaluated that can be used to manage women with a positive hrHPV-DNA test at screening. The clinical question is: "What is the best test or combination of tests which results in the highest sensitivity for progressive cervical precancer at the lowest burden of follow-up?" The PICOS[3,4] elements linked to this clinical question are listed in Box1.

+ PICOS

P: women participating in virological screening for cervical cancer, having a positive hrHPV-DNA test result

I: reflex testing with biomarkers (HPV genotyping, hrHPV-mRNA testing, p16, p16/KI67, other markers) and repetition of hrHPV-DNA testing, cytology and/or or combinations thereof

C: reflex cytology triage at cut-off ASC-US

O: cross-sectional and longitudinal accuracy to detect histologically identified disease (=CIN2+,CIN3+/AIS, and cervical cancer)
triage test positivity rate, referral rate for colposcopy, PPV for CIN2+ & CIN3+, risk of CIN2+, CIN3+ and cancer after negative triage testing

S: - follow-up of randomised trials comparing cytology, with HPV-based screening and applying different follow-up algorithms
- complete diagnostic studies (all subjects tested with triage method and verification with the reference standard (colposcopy/biopsy))
- cohort studies applying at least two alternative triage algorithms involving verification with the reference standard if one or more positive triage test result

BOX 1: PICOS-elements

3.2.2. Literature Search

A systematic literature search was performed in two electronic bibliographic databases (MEDLINE and EMBASE) on January 7, 2014, using the following search strings:

cervix[tw] OR "cervix uteri"[MeSH terms] OR cervical[tw] OR cervicovagin*) AND (cancer OR carcinoma OR neoplas* OR dysplas* OR squamous OR CIN[tw] OR CINII*[tw] OR CIN2*[tw] OR CINIII*[tw] OR CIN3[tw] OR SIL[tw] OR HSIL[tw] OR H-SIL[tw] OR LSIL[tw] OR L-SIL OR ASCUS[tw] OR “ASC US”[tw] OR “ASC H”[tw]))
AND
(HPV OR "human papillomavirus" OR papillomavirus infections[MeSH Terms])
AND
(("Early Detection of Cancer(("[Mesh] OR triage[MeSH Terms] OR triage[tw] OR management[tw] OR follow*up[tw]) AND (HPV*pos* [tw] OR “HPV+” OR test*positiv* OR screen*positiv* OR infection))
AND
(efficiency[tw] OR efficacy[tw] OR diagnostic[tw] OR accuracy[tw] OR “diagnostic test accuracy” OR sensitivity[tw] OR specificity[tw] OR "Sensitivity and Specificity"[Mesh:NoExp] OR PPV OR NPV OR “predictive value”)

Studies were eligible if (1) cross-sectional and/or longitudinal triage data were available for women with a positive hrHPV screening test, and (2) verification with the golden standard (colposcopy and targeted biopsy, possibly completed

with random biopsies and/or endocervical curettage) was performed on all women or women with at least one positive triage test.
Triage methods consisting of a one-step strategy or a two-step strategy were eligible. Each triage step could consist of a single test, or combined testing with two assays using an 'AND' (both tests positive) or an 'OR' (at least one test positive) approach.

3.2.3. Statistical analysis

Where possible (sufficient studies), the pooled absolute sensitivity and specificity of triage tests were estimated jointly using *metandi*, a procedure in STATA, based on the bivariate normal model for the logit transforms of sensitivity and specificity taking the intrinsic correlation between true and false-positivity rates and the variability between studies into account [5,6].
When insufficient studies were available (<=4) absolute relative sensitivity and specificity ratios were computed independently using the STATA procedures metaprop[2] and metan[7], respectively. In this case, overall pooled measures, with 95% confidence intervals were calculated using random effects models[8]. The statistical heterogeneity was assessed by the p-value for heterogeneity (following a chi2 distribution) as well as by the I^2 statistic, which measures the variation across studies that is due to inter-study heterogeneity.
Anticipating on scarcity of data (triage scenarios only assessed in one study), we also considered estimating the absolute accuracy of a given triage strategy by using the absolute sensitivity and specificity of the reference triage strategy (reflex cytology at cut-off ASC-US+) multiplied by the relative sensitivity and specificity of a given strategy, as assessed from a bivariate normal model with triage strategy as a covariate, using the SAS macro metadas [9].

$$\text{Sensitivity}_{\text{strategy X}} = \text{pooled Sensitivity}_{\text{ref *}} \text{ modeled Relative Sensitivity}_{\text{x versus ref}}$$
$$\text{Specificity}_{\text{strategy X}} = \text{pooled Specificity}_{\text{ref *}} \text{ modeled Relative Specificity}_{\text{x versus refS}}.$$

This method is built on the general finding that ranges of variability on relative accuracy are smaller than on absolute accuracy.

3.3. Results

3.3.1. Literature retrieval

For the analysis presented here, we included data from controlled trials conducted in population-based, organized screening programs. Based on this criterion, seven large trials were identified, which incorporated virological

[2] Metaprop is a statistical procedure in STATA developed at the Unit of Cancer Epidemiology (IPH Brussels) to pool proportions based on binomial distributions.

testing in primary screening. These seven trials comprised six European (NTCC[10-13], ARTISTIC[14], SWEDESCREEN[15], VUSA[16], POBASCAM[17], and PUBLIC HEALTH TRIAL FINLAND[18]) and one American trial (ATHENA[19]). Since the data for the Italian NTCC trial, were separated in four reports[10-13], a total of ten reports were found eligible, containing accuracy data for diverse triage strategies in the management of women with a positive primary screening hrHPV-DNA test. The process of literature retrieval and study selection are shown in Figure 26.

The study characteristics of the included reports are listed in Table 26. Cross-sectional triage data were extracted for NTCC, SWEDESCREEN, ATHENA, and PUBLIC HEALTH TRIAL FINLAND. Longitudinal data were extracted for NTCC, ARTISTIC, VUSA and POBASCAM, comprising three, three, two and four years of follow-up, respectively. Five studies[10-13,19] had a complete design, referring all hrHPV-positive women to verification with the golden standard, while in the other five studies an incomplete design was applied, which means that only triage positive women were submitted to the golden standard.

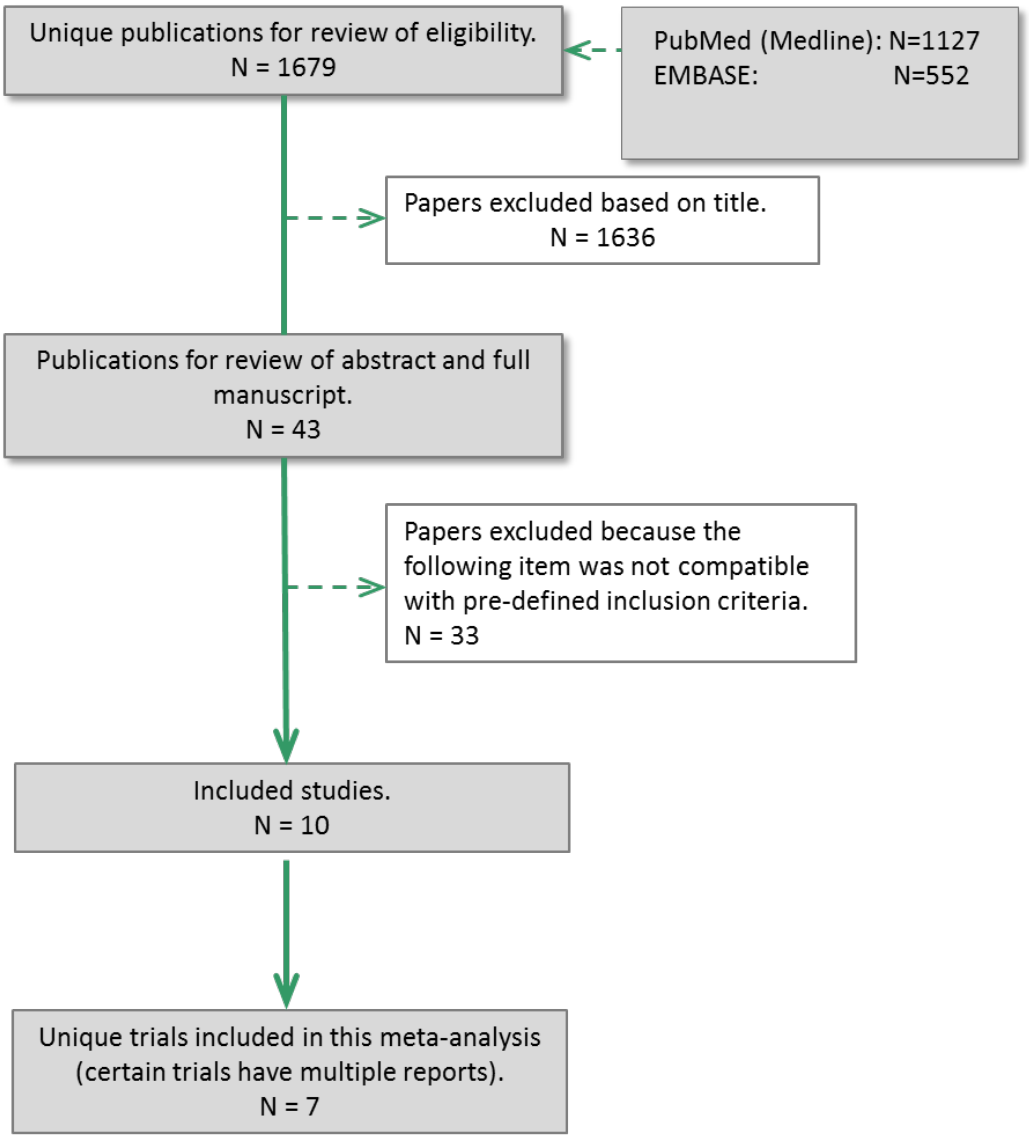


Figure 27: PRISMA flow chart for literature retrieval

Table 26: Study characteristics of the included studies.

Trial name	**Study**	**Country**	**Population**	**Follow-up**	**Gold standard**	**Triage tests**
ARTISTIC (round 1)	Kitchener 2009[14]	UK	Screening population. 20-64y	36m	Colposcopy + targeted biopsies. - If LSIL+, immediate verification. If ASC-US or LSIL, repeat cytology (6m, 12m) and verification if LSIL+. If <ASC-US, repeat HPV testing (12m, 24m) and verification if positive.	LBC (ThP) HC2
ATHENA	Castle 2011[19]	USA	Screening population. ≥25y	4m	Colposcopy + targeted biopsies or ECC. - All participants.	LBC (ThP) Linear Array Cobas(HPV1618)
NTCC	Ronco 2006[10,11]	Italy	Screening population. 25-34y[11] 35-60y[10]	6m	Colposcopy + targeted biopsies. If ≥ 35 y: all participants. <35y: colposcopy referral if ASCUS+; repeat testing if cyto-/HPV+ & referral if 2nd testing (cyto/HPV) showed a + result.	LBC (ThP)
NTCC-2	Carozzi 2008[12]	Italy	Screening population. 25-60y	6m	Colposcopy + targeted biopsy. - All participants.	p16
NTCC-2	Carozzi 2013[13]	Italy	Screening population. 25-60y	36m	Colposcopy + targeted biopsy. - All participants.	p16
POBASCAM	Dijkstra 2013[17]	The Netherlands	Screening population.	48m	Colposcopy + targeted biopsies.	CP PCR (GP5+/6+)

			29-61y		- If HSIL+, immediate verification. If <HSIL, repeat HPV and cytology (6m, 18m) and verification if ASC-US+ at 6m, or ASC-US+ and/or hrHPV+ at 18m. If <ASC-US,	and RLB
PUBLIC HEALTH TRIAL FINLAND	Leinonen 2013[18]	Finland	Screening population. 25-65y	12m	Colposcopy + targeted biopsies. - If LSIL+, immediate verification. If <LSIL, repeat (12, 24m).	CP PCR Luminex
SWEDESCREEN	Naucler 2009[15]	Sweden	Screening population. 32-38y	20m	Colposcopy + targeted biopsies or 2 random biopsies. - If ASC-US+, immediate verification or repeat cytology (depending on local practices. If <ASC-US, repeat HPV (12m) and verification if type-specific persistence.	CP PCR (GP5+/6+)
VUSA	Rijkaart 2012[16]	The Netherlands	Screening population. 30-60y	24m	Colposcopy + targeted biopsies. - If ASC-US+, immediate verification. If <ASC-US, repeat HPV and cytology (12m, 24m) and verification if ASC-US+ at 12m, or ASC-US+ and/or HPV+ at 24m.	CP PCR (GP5+/6+) and RLB

Abbreviations: ASC-US+, atypical squamous cells of undetermined significance; CP, conventional Pap smear; ECC, endocervical curettage; HSIL, high-grade squamous intraepithelial lesions or worse; HC2, Hybrid Capture

2 assay; LBC, liquid based cytology; LSIL+, low-grade squamous intraepithelial lesions or worse; PCR, polymerase chain reaction; RLB, reverse line blotting; ThP, ThinPrep.

3.3.2. Absolute accuracy of cytology and/or hrHPV-DNA based triage algorithms

Diverse triage algorithms were available in the included studies, ranging from one-step to two-step triage strategies with diverse methods such as cytology, repeat hrHPV testing, HPV genotyping, and/or p16 cytoimmunochemistry. In most studies, the available data was detailed enough enabling extraction of absolute and uncorrected values for true-positives (TP) and –negatives (TN) , and false-positives (FP) and –negatives (FN) for all or a subset of triage algorithms. Some studies only allowed extraction of a corrected accuracy measures (e.g. sensitivity, specificity, PPV, NPV, etc.) which were adjusted for non-compliance to the study protocol. Meta-analytic pooling was performed using the available values for TP, FN, FP, and TN.

3.3.2.1. Triage with reflex cytology (cut-off ASC-US+)

Eight studies contained uncorrected absolute numbers of true- and false-positive and negative results for one-step triage with reflex cytology at cut-off ASC-US+[10,11,14-19]. Two reports by Ronco et al. (NTCC-1)[10,11] were combined for women between 35-60y and women below 35y, respectively. The pooled sensitivity and specificity for reflex cytology at cut-off ASC-US+ to detect CIN2+ was 79.5% (95% CI: 65.2-90.8%) and 79.1% (95% CI: 73.0-84.6%), respectively (Figure 28 and Figure 29). To detect CIN3+, the pooled sensitivity and specificity of reflex cytology at cut-off ASC-US+ was 82.0% (95% CI: 66.9-93.4%) and 72.3% (95% CI: 67.0-77.3%), respectively (Figure 27).

In Table 27, the absolute accuracy measures for the different triage algorithms are listed.

Accuracy data for the addition of a second triage step after six months to manage women who had a negative cytology triage test at baseline, were available only for the POBASCAM trial[17]. In POBASCAM, the sensitivity of cytological triage at ASC-US+ for CIN3+ at baseline was 82% and by adding a second triage, sensitivity increased to 96%, 100% and 100% for, with the 2nd triage test being ASC-US+ cytology, hrHPV testing, or ASC-US+ cytology with hrHPV DNA testing. At the same time, the specificity decreased from 72% to 57%, 30%, or 28%. The PPV decreased from 35% to 23%, the NPV increased 97.5% to 100% and the referral rate increased from 31% to 75%. (see Table 27). Considering the outcome CIN2+, the sensitivity increased from 75% (baseline ASC-US+ triage) to 93%, 99.5% or 99.7%, whereas the specificity decreased from 88% to 80%, 44% or 41%, by adding the one of the considered 2nd triage tests (see Table 27).

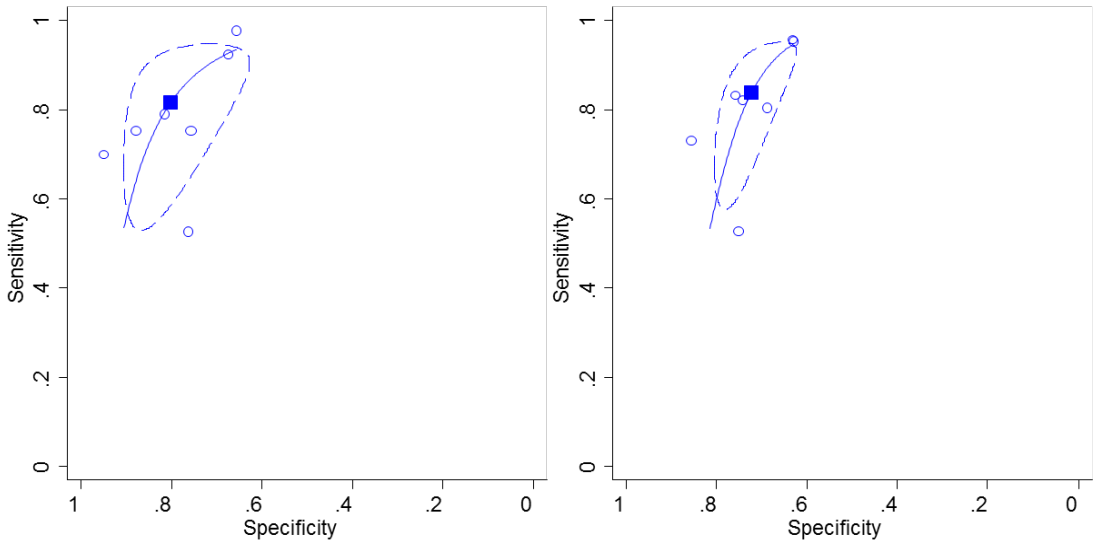


Figure 28: sROC plots of the sensitvity as a function of the specificity of reflex cytology at cut-off ASC-US to detect CIN2+ (left) and CIN3+ (right) in the triage of women with a positive hrHPV-DNA screening test.

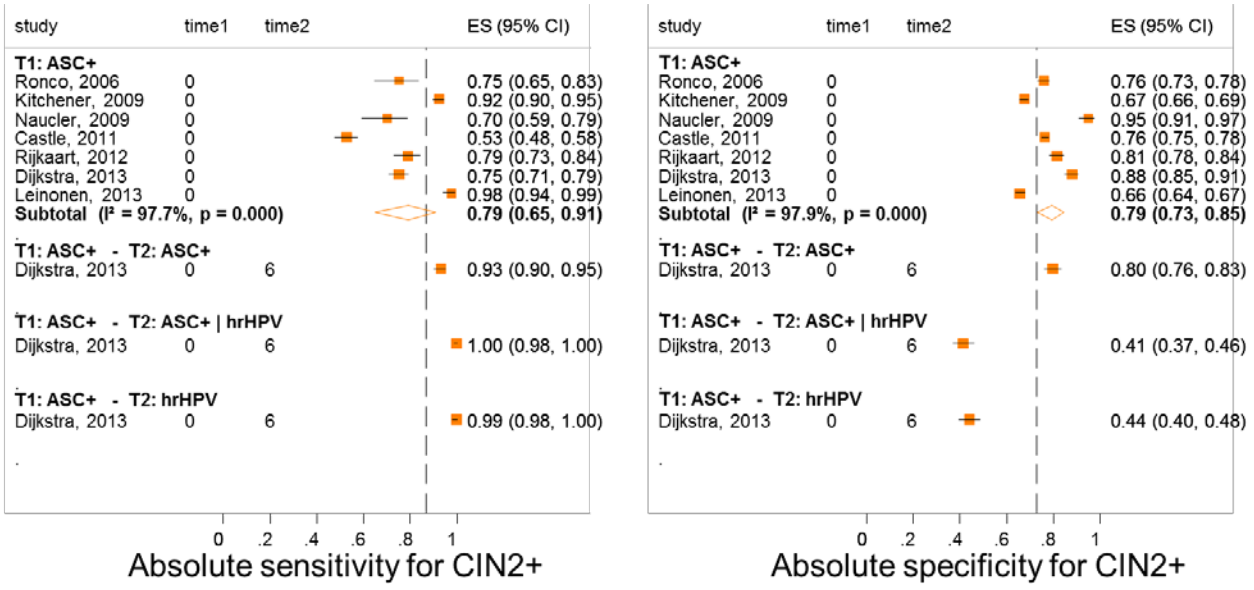


Figure 29: Meta-analysis of the absolute sensitivity and specificity to detect CIN2+ of four triage algorithms with reflex cytology (cut-off: ASC-US) as first triage. Time1 and time2 correspond to the timing of the triage step (in months). Abbreviations: |, 'OR'; ASC+, atypical squamous cells or worse; CI, confidence interval; I², percentage of total variation across studies due to heterogeneity; hrHPV, high-risk human papillomavirus; p, test for inter-study heterogeneity.

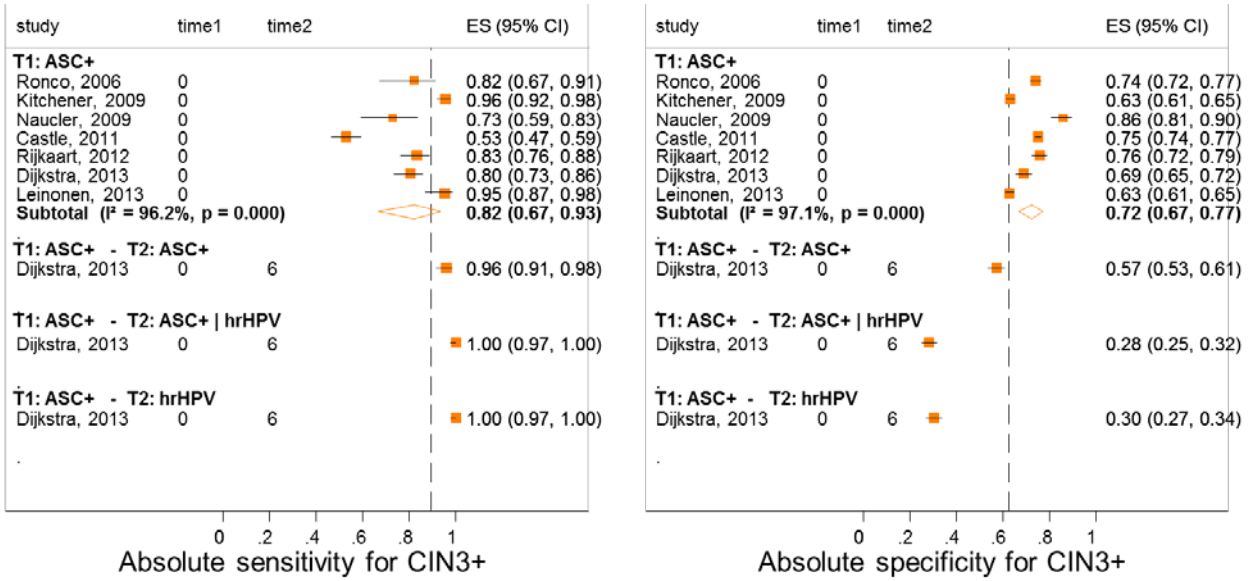


Figure 30: Meta-analysis of the absolute sensitivity and specificity to detect CIN3+ of four triage algorithms with reflex cytology (cut-off: ASC-US) as first triage. Time1 and time2 correspond to the timing of the triage step (in months). Abbreviations: |, 'OR'; ASC+, atypical squamous cells or worse; CI, confidence interval; I², percentage of total variation across studies due to heterogeneity; hrHPV, high-risk human papillomavirus; p, test for inter-study heterogeneity.

Table 27: Absolute sensitivity, specificity, positive and negative predictive values, and referral rate for triage with reflex cytology at cutoff ASC-US+ combined or not with second triage step among women with a positive hrHPV test.

Triage1	Triage2	Number of studies	Outcome	Sensitivity % (95% CI)	Specificity % (95% CI)	PPV % (95% CI)	NPV % (95% CI)	Referral rate % (95% CI)
ASC-US+		**7**[3]	CIN3+	82.0 (66.9-93.4)	72.3 (67.0-77.3)	22.9 (13.7-33.6)	97.5 (95.0-99.1)	32.0 (27.0-37.3)
ASC-US+	ASC-US+	1	CIN3+	95.9 (91.4-98.1)	57.2 (53.5-60.8)	32.2 (28.0-36.7)	98.5 (96.8-99.3)	52.1 (48.8-55.5)
ASC-US+	hrHPV	1	CIN3+	100 (97.5-100)	30.2 (26.9-33.7)	23.3 (20.2-26.8)	100 (98.2-100)	75.1 (72.0-77.9)
ASC-US+	ASC-US+ | hrHPV	1	CIN3+	100 (97.5-100)	28.2 (25.0-31.7)	22.8 (19.7-26.2)	100 (98.1-100)	76.7 (73.7-79.4)
ASC-US+		7	CIN2+	79.5 (65.2-90.8)	79.1 (73.0-84.6)	44.8 (27.2-63.1)	94.3 (89.4-97.8)	32.0 (27.0-37.3)
ASC-US+	ASC-US+	1	CIN2+	93.0 (89.9-95.2)	79.8 (75.9-83.2)	78.2 (74.1-81.8)	93.6 (90.8-95.6)	52.1 (48.8-55.5)
ASC-US+	hrHPV	1	CIN2+	99.5 (98.1-99.9)	44.0 (39.6-48.5)	58.1 (54.2-61.9)	99.1 (96.6-99.7)	75.1 (72.0-77.9)
ASC-US+	ASC-US+ | hrHPV	1	CIN2+	99.7 (98.5-100)	41.3 (36.9-45.7)	57.0 (53.2-60.8)	99.5 (97.2-99.9)	76.7 (73.7-79.4)

[33] For comparison in Dijkstra et al 2013: the accuracy measure of triage with ASC-US+ cytology for CIN3+ SE=80%, SP=69%, PPV=35%, NPV=94%, referral rate=31%; and for CIN2+ SE=75%, SP=88%, PPV=83%, NPV=82%.

3.3.2.2. Triage with a combination of reflex cytology (cut-off ASC-US+) and HPV16 or HPV1618 genotyping

From the POBASCAM and ATHENA trials, data could be extracted on reflex triage with a combination of cytology (cut-off ASC-US+) and HPV16 or HPV1618 genotyping, either using an 'OR' approach (one or both tests positive) [17,19] or an 'AND' approach (both tests positive)[19]. In Table 28, the absolute accuracy measures for the different triage algorithms are listed.

Based on these two studies, the pooled sensitivity and specificity of reflex triage with cytology (ASC-US+) or HPV1618 genotyping to detect CIN2+ was 85.7% (95% CI: 61.9-98.9%) and 62.1 (95% CI: 55.1-68.9%), respectively (Figure 30). To detect CIN3+, the pooled sensitivity and specificity were 89.6% (95% CI: 64.5-100.0%) and 52.8% (95% CI: 42.6-62.8%), respectively (Figure 31). Adding a second triage step using cytology (ASC-US+) or a combination of cytology and hrHPV-DNA testing resulted in a 13-14% gain in sensitivity for CIN2+ (98.7% [95%CI=96.9-99.4%] and 100% [95% CI=99.0-100%, respectively], and a ~2% gain in NPV (98.2% [95% CI=95.9-99.2%] and 100% [95% CI=97.4-100%], respectively). Referral rate increased up to 67.0% (95% CI= 63.8-70.1%) and 82.9% (95% CI=80.2-85.3%) when the second triage step was cytology or combined cytology-hrHPV respectively. Given the large contrast between ATHENA and POBASCAM, an intra-study (POBASCAM) comparison is appropriate. In POBASCAM, reflex triage with ASC-US+ cytology and HPV1618 genotyping reached a sensitivity for CIN2+ of 94.1% (95% CI: 91.2-96.1%). Adding a second triage step using cytology (ASC-US+) or a combination of cytology and hrHPV-DNA testing resulted in a 5-6% gain in sensitivity for CIN2+ (reaching 99% and 100%, respectively), and a 5-6% gain in NPV (98% and 100%, respectively). Referral rate increased up to 67% and 83% when the second triage step was cytology or combined cytology-hrHPV respectively. The accuracy of reflex ASC-US+ combined with HPV1618 genotyping reached a high sensitivity for CIN3+ (97%), adding a 2nd triage step resulted in a sensitivity of 99-100% but this resulted in a specificity loss of 18-37%.

In a triage algorithm where both cytology (ASC-US+) and the HPV1618 genotyping test had to be positive, the sensitivity dropped considerably (30.0% [95% CI=25.6-34.8%] for CIN2+, and 34.1%[95% CI=28.6-40.2%] for CIN3+), while the specificity increased (92.9% [95% CI=91.9-93.7%] for CIN2+, and 92.3% [95% CI=91.3-93.1%] for CIN3+).

The ATHENA trial contained data on accuracy of triage with a combination of cytology (ASC-US+) and HPV16 genotyping[19]. Compared to related triage algorithms that use HPV1618 genotyping, sensitivity was 3% lower using the 'OR' approach, and 4-5% lower using the 'AND' approach.

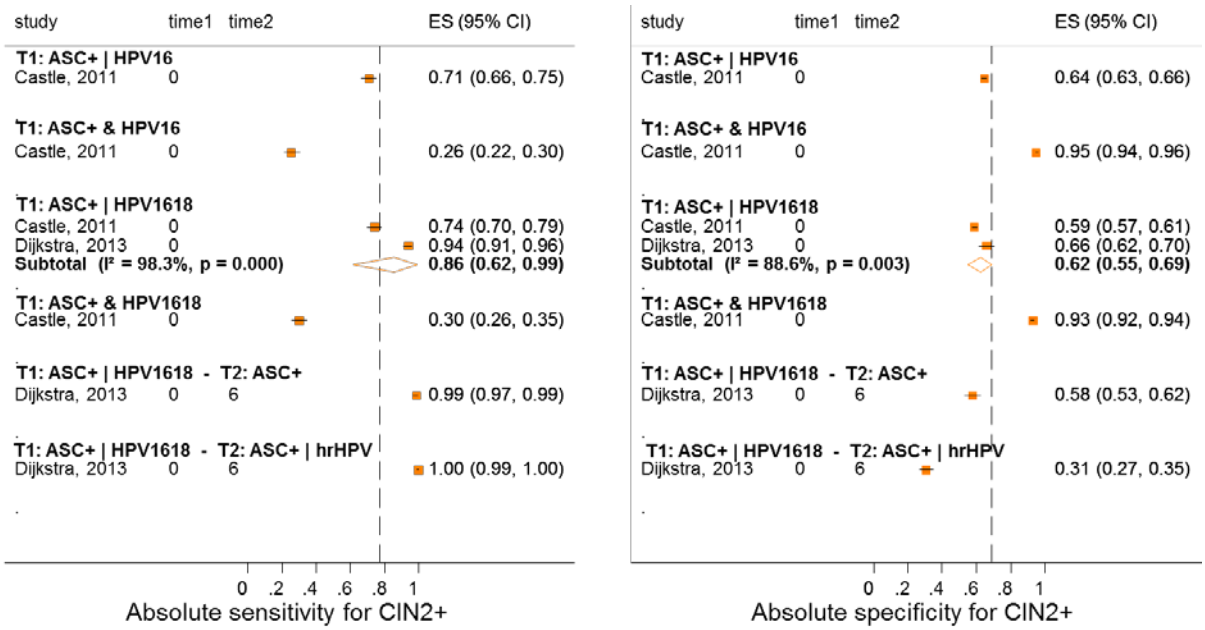


Figure 31: Meta-analysis of the absolute sensitivity and specificity to detect CIN2+ of six triage algorithms with a combination of reflex cytology (cut-off: ASC-US+) and HPV16 or HPV1618 genotyping as first triage. Time1 and time2 correspond to the timing of the triage steps (in months). Abbreviations: |, 'OR'; &, 'AND', ASC+, atypical squamous cells or worse; CI, confidence interval; I^2, percentage of total variation across studies due to heterogeneity; hrHPV, high-risk human papillomavirus; p, test for inter-study heterogeneity.

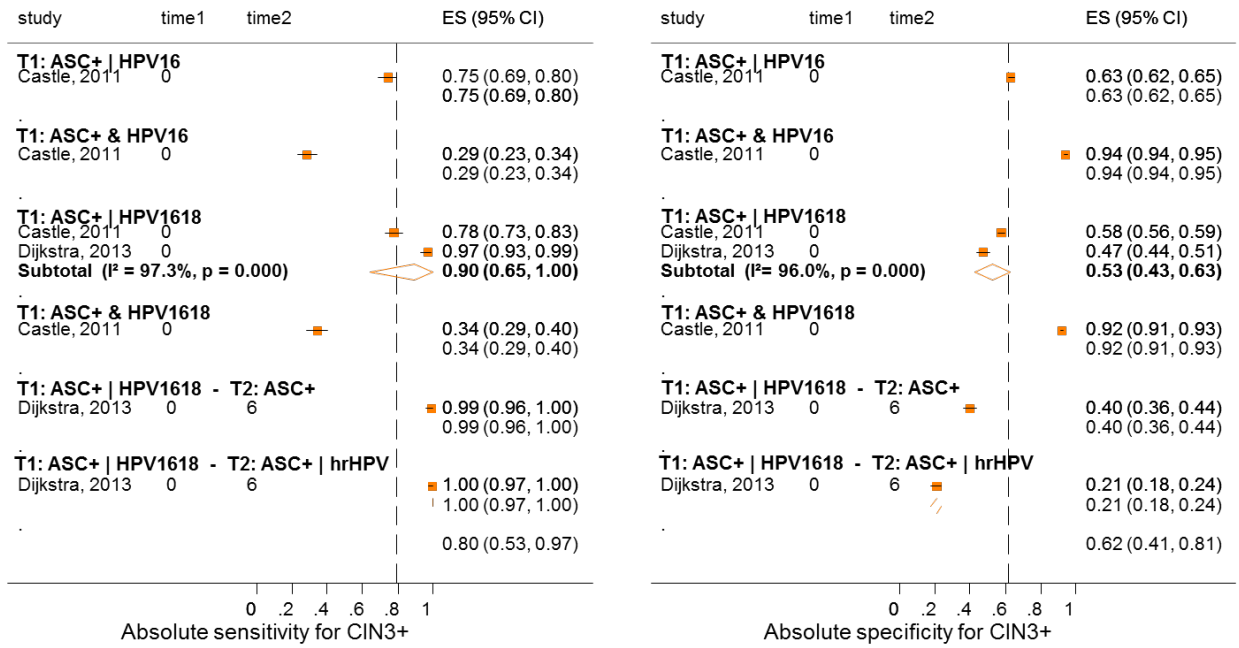


Figure 32: Meta-analysis of the absolute sensitivity and specificity to detect CIN3+ of six triage algorithms with a combination of reflex cytology (cut-off: ASC+) and HPV16 or HPV1618 genotyping as first triage. Time1 and time2

correspond to the timing of the triage step (in months). Abbreviations: |, 'OR'; &, 'AND', ASC+, atypical squamous cells or worse; CI, confidence interval; I^2, percentage of total variation across studies due to heterogeneity; hrHPV, high-risk human papillomavirus; p, test for inter-study heterogeneity.

Table 28: Absolute sensitivity, specificity, positive and negative predictive values, and referral rate of reflex cytology (ASC-US+) combined with HPV16 or HPV1618 genotyping to triage women with a positive hrHPV test.

Triage1	Triage2	Number of studies	Outcome	Sensitivity % (95% CI)	Specificity % (95% CI)	PPV % (95% CI)	NPV % (95% CI)	Referral rate % (95% CI)
ASC-US+|HPV1618		2[4]	CIN3+	89.6 (64.5-100)	52.8 (42.6-62.8)	19.7 (6.9-37.0)	97.9 (96.1-99.2)	52.6 (37.3-67.5)
ASC-US+&HPV1618		1	CIN3+	34.1 (28.6-40.2)	92.3 (91.3-93.1)	25.5 (21.2-30.4)	94.8 (93.9-95.5)	9.6 (8.7-10.6)
ASC-US+|HPV1618	ASC-US+	1	CIN3+	99.3 (96.3-99.9)	39.8 (36.3-43.5)	25.9 (22.5-29.7)	99.6 (98.0-99.9)	67.0 (63.8-70.1)
ASC-US+|HPV1618	ASC-US+|hrHPV	1	CIN3+	100 (97.5-100)	20.8 (17.9-23.9)	21.1 (18.3-24.3)	100 (97.4-100)	82.9 (80.2-85.3)
ASC-US+|HPV16		1	CIN3+	74.6 (68.9-79.6)	63.3 (61.7-65.0)	13.6 (11.9-15.5)	97.0 (96.2-97.6)	39.4 (37.8-41.0)
ASC-US+&HPV16		1	CIN3+	28.6 (23.3-34.4)	94.4 (93.6-95.2)	28.5 (23.3-34.3)	94.5 (93.6-95.2)	7.2 (6.4-8.1)
ASC-US+|HPV1618		2[3]	CIN2+	85.7 (61.9-98.9)	62.1 (55.1-68.9)	42.0 (3.2-88.8)	94.6 (93.2-95.9)	52.6 (37.3-67.5)
ASC-US+&HPV1618		1	CIN2+	30.0 (25.6-34.8)	92.9 (91.9-93.7)	33.8 (29.0-39.0)	91.6 (90.6-92.5)	9.6 (8.7-10.6)
ASC-US+|HPV1618	ASC-US+	1	CIN2+	98.7 (96.9-99.4)	57.7 (53.2-62.0)	64.6 (60.5-68.4)	98.2 (95.9-99.2)	67.0 (63.8-70.1)
ASC-US+|HPV1618	ASC-US+|hrHPV	1	CIN2+	100 (99.0-100)	30.5 (26.6-34.8)	52.9 (49.2-56.6)	100 (97.4-100)	82.9 (80.2-85.3)
ASC-US+|HPV16		1	CIN2+	71.1 (66.3-75.4)	64.4 (62.7-66.1)	19.6 (17.6-21.7)	94.8 (93.8-95.7)	39.4 (37.8-41.0)
ASC-US+&HPV16		1	CIN2+	25.8 (21.6-30.4)	95.0 (94.2-95.7)	38.7 (32.9-44.9)	91.3 (90.3-92.2)	7.2 (6.4-8.1)

[4] In Dijkstra 2013, the performance parameters of triage with reflex testing with ASC-US+ cytology and HPV1616 testing for CIN2+ were: SE=94.1%, SP=65.9%, PPV=68.3%, NPV=93.4% and referral rate=60.4%; and for CIN3+: SE=97.3%, SP=47.4%, PPV=28.2%, NPV=98.8%

3.3.2.3. Triage with reflex cytology (cut-off LSIL+) and combinations with HPV16 or HPV1618 genotyping.

Four studies were identified which provided accuracy estimates for reflex triage with LSIL+ cytology, whereas for reflex LSIL+ triage combined with genotyping for HPV16 or HPV1618 only the ATHENA trial provided accuracy data.

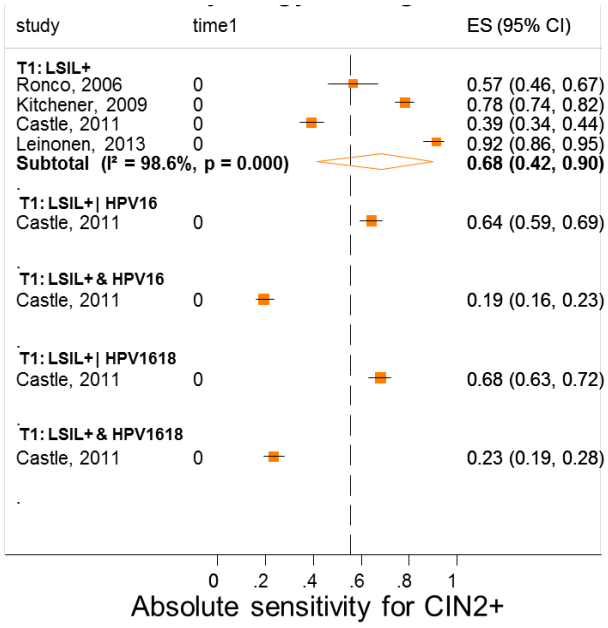
The pooled sensitivity and specificity to detect CIN2+ was 68.4% (95% CI: 41.5-90.0%) and 86.8% (95% CI: 83.4-89.8%), respectively (Figure 32). The pooled sensitivity and specificity to detect CIN3+ was 70.4% (95% CI: 42.9-91.7%) and 84.3% (95% CI: 80.1-88.1%), respectively (Figure 33).

In Table 29, the absolute accuracy measures for the different triage algorithms are listed.

Comparing the pooled accuracy estimates for LSIL+ triage with those for LSIL+ combined with HPV1618 genotyping resulted in only a small gain in sensitivity for CIN3+ (from 70% to 72%)) but resulted in a drop in specificity (from 84% to 65%) and a doubled referral rate (from 18% to 38% (Table 29). In contrast, if both tests had to be positive, the referral rate dropped considerably (6.2% [95% CI=5.4-7.0%] for CIN3+) , but a significant lower sensitivity (27.4% [95% CI: 22.2-33.2%] for CIN3+) combined with a lower NPV (94.4% [95% CI: 93.6-95.2%] for CIN3+) was observed.

Given the high variability among the four studies that contributed data for LSIL+ triage, an intra-study (ATHENA) comparison might be more appropriate to appreciate the change in accuracy for adding HPV1618 genotyping to reflex LSIL+ cytology. This intra-study comparison shows a substantial gain in sensitivity (29% for CIN2+, 32% for CIN3+) and drop in specificity (21% for CIN2+ and CIN3+).

When comparing triage algorithms that use HPV16 genotyping versus related triage algorithms that use HPV1618 genotyping, the former results in approximately 5% sensitivity loss, but a 7% gain in specificity.



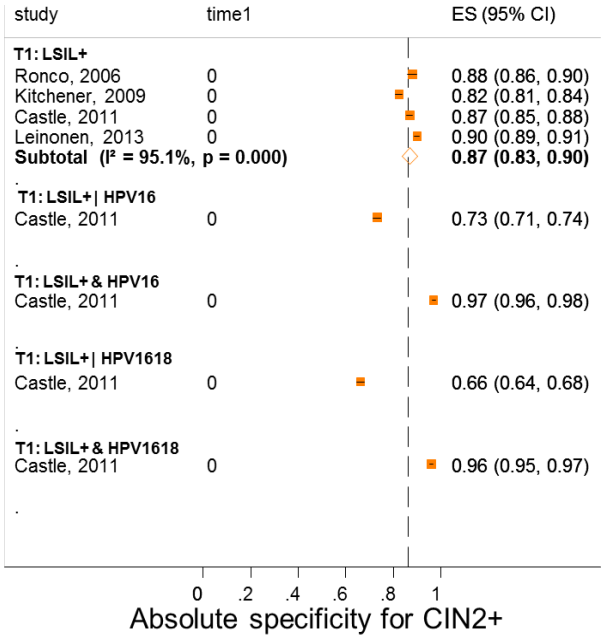
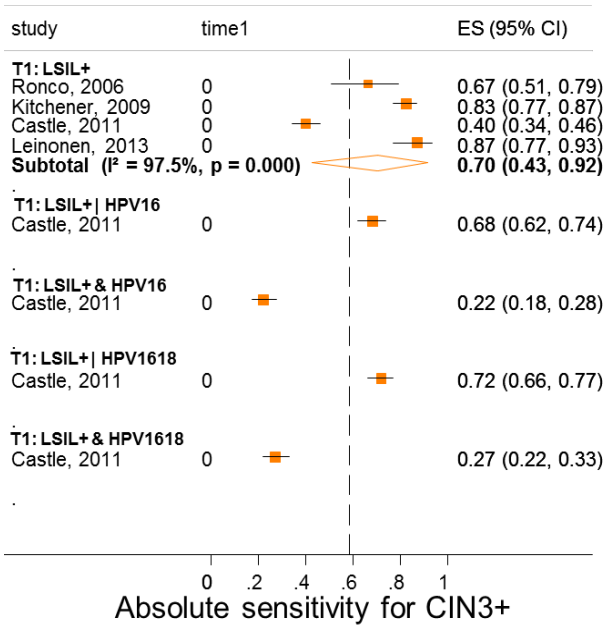


Figure 33: Meta-analysis of the absolute sensitivity and specificity to detect CIN2+ of five triage algorithms with reflex cytology (cut-off: LSIL+) and HPV16 or HPV1618 genotyping. Time1 corresponds to the timing of the triage step (in months). Abbreviations: |, 'OR'; CI, confidence interval; I², percentage of total variation across studies due to heterogeneity; hrHPV, high-risk human papillomavirus; LSIL+, low-grade squamous intraepithelial lesions; p, test for inter-study heterogeneity.



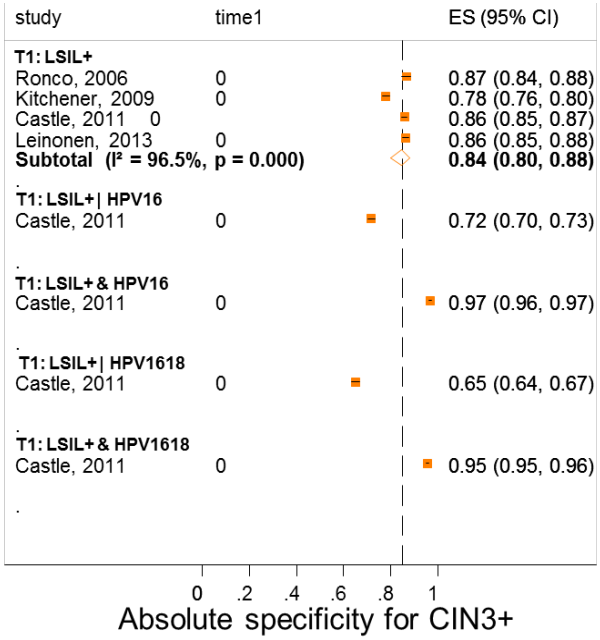


Figure 34: Meta-analysis of the absolute sensitivity and specificity to detect CIN3+ of five triage algorithms with reflex cytology (cut-off: LSIL+) and HPV16 or HPV1618 genotyping. Time1 corresponds to the timing of the triage step (in

months). Abbreviations: |, 'OR'; CI, confidence interval; I^2, percentage of total variation across studies due to heterogeneity; hrHPV, high-risk human papillomavirus; LSIL+, low-grade squamous intraepithelial lesions; p, test for inter-study heterogeneity.

Table 29: Absolute sensitivity, specificity, positive and negative predictive values, and referral rate of reflex cytology (LSIL+) and combinations with HPV16 or HPV1618 genotyping to triage women with a positive hrHPV test.

Triage1	Number of studies	Outcome	Sensitivity % (95% CI)	Specificity % (95% CI)	PPV % (95% CI)	NPV % (95% CI)	Referral rate % (95% CI)
LSIL+	**4**[5]	**CIN3+**	**70.4 (42.9-91.7)**	**84.3 (80.1-88.1)**	**17.7 (12.9-23.1)**	**98.2 (95.5-99.7)**	**18.2 (13.1-23.9)**
LSIL+ | HPV1618	1	CIN3+	72.2 (66.4-77.4)	65.2 (63.5-66.8)	13.9 (12.1-15.8)	96.8 (96.0-97.5)	37.5 (35.9-39.1)
LSIL+ & HPV1618	1	CIN3+	27.4 (22.2-33.2)	95.5 (94.7-96.1)	31.9 (26.1-38.4)	94.4 (93.6-95.2)	6.2 (5.4-7.0)
LSIL+ | HPV16	1	CIN3+	68.3 (62.3-73.7)	71.8 (70.2-73.3)	15.8 (13.7-18.1)	96.7 (95.9-97.3)	31.1 (29.6-32.6)
LSIL+ & HPV16	1	CIN3+	22.2 (17.5-27.8)	96.6 (95.9-97.1)	33.3 (26.6-40.8)	94.1 (93.3-94.9)	4.8 (4.1-5.6)
LSIL+	**4**[4]	**CIN2+**	**68.4 (41.5-90.0)**	**86.8 (83.4-89.8)**	**33.4 (24.6-42.8)**	**96.4 (92.2-99.0)**	**18.2 (13.1-23.9)**
LSIL+ | HPV1618	1	CIN2+	67.9 (63.0-72.4)	66.2 (64.5-67.8)	19.6 (17.6-21.9)	94.4 (93.4-95.3)	37.5 (35.9-39.1)
LSIL+ & HPV1618	1	CIN2+	23.2 (19.2-27.7)	95.9 (95.1-96.5)	40.7 (34.4-47.4)	91.1 (90.1-92.0)	6.2 (5.4-7.0)
LSIL+ | HPV16	1	CIN2+	64.2 (59.3-68.9)	72.9 (71.3-74.5)	22.4 (20.0-25.0)	94.4 (93.4-95.2)	31.1 (29.6-32.6)
LSIL+ & HPV16	1	CIN2+	19.2 (15.6-23.5)	97.0 (96.3-97.5)	43.5 (36.2-51.0)	90.8 (89.8-91.7)	4.8 (4.1-5.6)

[5] In Castle 2011, the performance parameters of triage with reflex testing with LSIL+ cytology for CIN2+ were: SE=39.2%, SP=86.7%, PPV=26.4%, NPV=92.1% and referral rate=16.1%; and for CIN3+: SE=40.1%, SP=85.8%, PPV=17.9%, NPV=94.9%

3.3.2.4. Joint variation of sensitivity and specificity of 3 triage strategies

In Figure 34, the variation of the sensitivity and specificity of three sensitive triage strategies are displayed in ROC space: 1) reflex cytology (ASCUS+) triage (green); 2) reflex cytology (ASCUS+) triage (green) completed with 2nd cytology (ASCUS+) triage if reflex triage was negative (blue); 3) reflex triage with cytology and HPV1618 genotyping (yellow). The accuracy measures are based on estimates adjusted for non-compliance derived from the published papers, and therefore do not always correspond to values that were calculated based on the number of absolute true- and false-positives and –negatives, unadjusted for follow-up compliance (see section 3.2.1- 3.2.4). By lack of absolute values of the adjusted accuracy parameters, no statistical inference can be made. For each displayed triage scenario, the simple average sensitivity and specificity was computed allowing a rough estimation of the pooled accuracy measures. Within the framework of the COHEAHR project, funded by the 7th Framework Programme of DG Research of the EU, authors are being contacted to obtain non-available absolute numbers which subsequently will be used for a formal meta-analysis of the accuracy to predict the outcomes of alternative triage scenarios adjusted for compliance to follow-up.

Figure 34 shows a gain in sensitivity (+10-15%) for CIN2 by repeating cytology (at ASC-US+) at a subsequent triage visit at 6-12 months, whereas the loss in specificity is limited. Adding HPV1618 genotyping to reflex cytology triage yields a small gain in average sensitivity for a considerable loss in specificity.

The reader must be warned that these estimates are very rough.

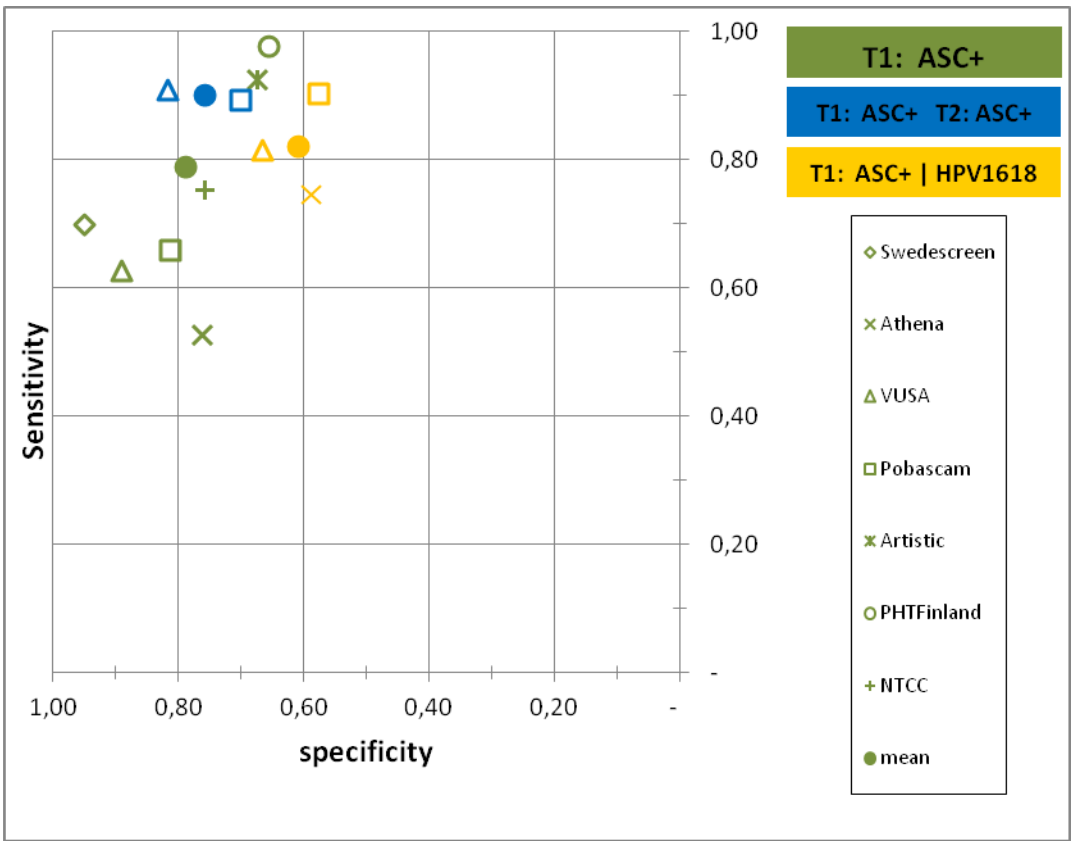


Figure 35: Sensitivity and specificity of three scenarios to triage hrHPV+ women: 1) reflex cytology (ASCUS+) triage (green); 2) reflex cytology (ASCUS+) triage (green) completed with 2nd cytology (ASCUS+) triage if reflex triage was negative; 3) reflex triage with cytology and HPV1618 genotyping. Filled symbols represent average values, other symbols represent values from individual studies.

3.3.3. Relative accuracy of cytology and/or hrHPV-DNA based triage algorithms versus reflex cytology (ASC-US+)

The relative accuracy of the different triage algorithms were compared with reflex cytology at cut-off ASC-US+.

Data on the triage with combined cytology (ASC-US+) or HPV1618 genotyping, versus cytology (ASC-US+) alone was available in the POBASCAM and ATHENA trial [17,19] (Figure 35, Figure 36). When either cytology or the HPV1618 genotyping test had to be positive, a significantly higher sensitivity (ratio: 1.32 [95% CI=1.16-1.51] for CIN2+) but lower specificity (ratio:0.77 [95% CI=0.74-0.79] for CIN2+) was observed, compared to cytology testing alone. Results were similar for CIN3+ (sensitivity ratio: 1.33[95% CI=1.06-1.68], and specificity ratio: 0.73 [95% CI=0.66-0.82]).

Four studies allowed comparison of reflex cytology triage at cut-of LSIL+ versus ASC-US+[10,11,14,18,19] ((Figure 35, Figure 36). Using LSIL+ as cut-off resulted in a 16% drop in sensitivity (ratio: 0.84 [95% CI=0.74-0.95] for CIN2+), but a 22% increase in specificity (ratio: 1.22 [95% CI=1.12-1.33] for CIN2+).

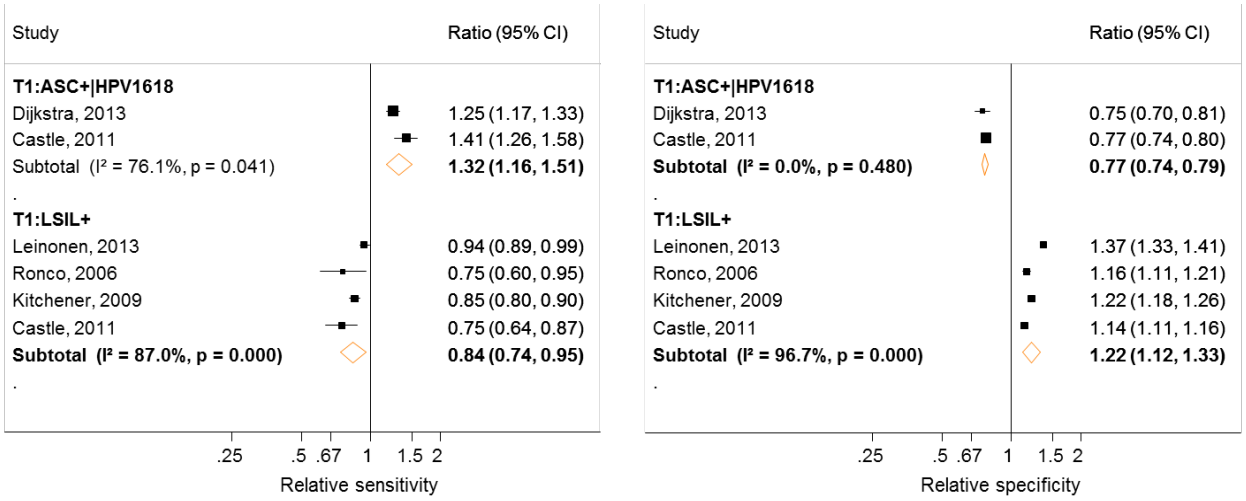


Figure 36: Relative sensitivity (left) and specificity (right) of two scenarios compared to reflex cytology at cutoff ASC-US+ to detect CIN2+ in women with a positive hrHPV DNA screening test (restricted to scenarios where a pooling from at least 2 studies was possible).

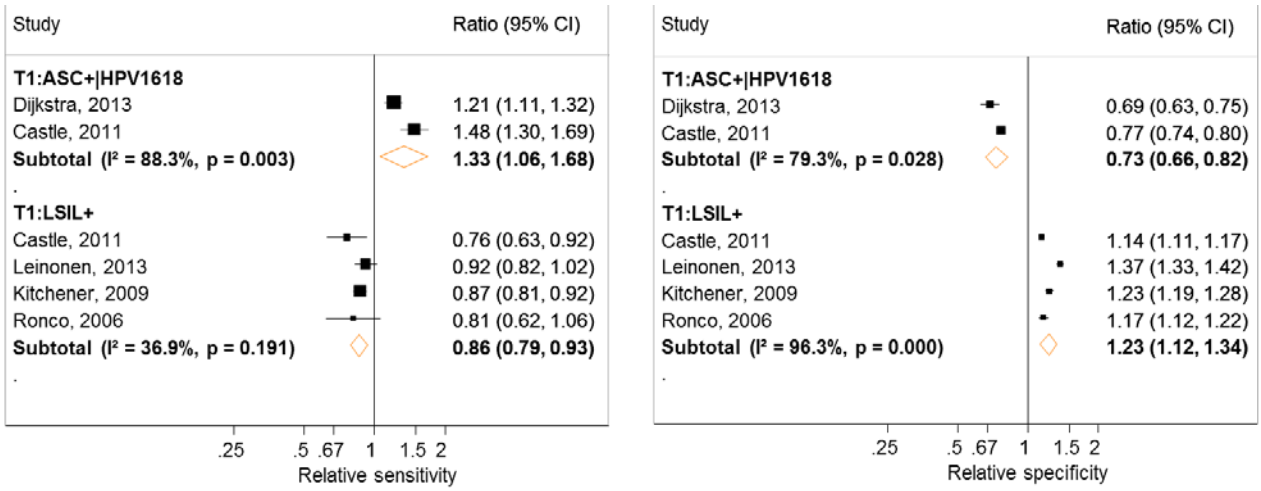


Figure 37: Relative sensitivity (left) and specificity (right) of two scenarios compared to reflex cytology at cutoff ASC-US+ to detect CIN3+ in women with a positive hrHPV DNA screening test. (restricted to scenarios where a pooling from at least 2 studies was possible).

For most triage algorithms, data were available for only one study. Comparisons were separated in two groups containing algorithms with an increased sensitivity (Figure 37, Figure 38) and those with sensitivity loss (Figure 39**,** Figure 40), compared to reflex cytology at cut-off ASC-US+.

Recalling women with normal reflex cytology for a second triage test with cytology at cut-off ASC-US+ after 6 months (T1:ASC-US+, T2:ASC-US+), resulted in a significantly increased sensitivity (ratio: 1.24 [95% CI= 1.16-1.32] for CIN2+), but a drop in specificity (ratio: 0.91 [95% CI= 0.86-0.96] for CIN2+) in the study of Dijkstra and colleagues[17]. Using a hrHPV-DNA assay in the second triage step further increased sensitivity, but specificity was halved compared to reflex cytology (ASC-US+) alone.

The largest gain in sensitivity was observed when reflex cytology was combined with HPV16 or HPV1618 genotyping and only one of both assays had to be positive (sensitivity ratio: 1.35 [95% CI=1.20-1.51]and 1.41 [95% CI=1.26-1.58], respectively)[19]. This however was linked with a significant drop in specificity (ratio: 0.84 [95% CI=0.82-0.87] and 0.77 [95% CI=0.74-0.80], respectively).

Reflex triage with HPV1618 genotyping was as sensitive (ratio: 0.99 [95% CI=0.86-1.13] for CIN2+) and as specific (ratio: 0.99 [95% CI=0.96-1.02] for CIN2+) as reflex cytology at cut-off ASC-US. Reflex triage with HPV16 genotyping was less sensitive (ratio: 0.84 [95% CI=0.72-0.97] for CIN2+) but

more specific (ratio: 1.09 [95% CI=1.06-1.12] for CIN2+) compared to reflex cytology.
Triage algorithms using a higher cytology cut-off (LSIL+ or HSIL+) or where both cytology and a genotyping test had to be positive resulted in significantly lower sensitivities compared to reflex cytology.

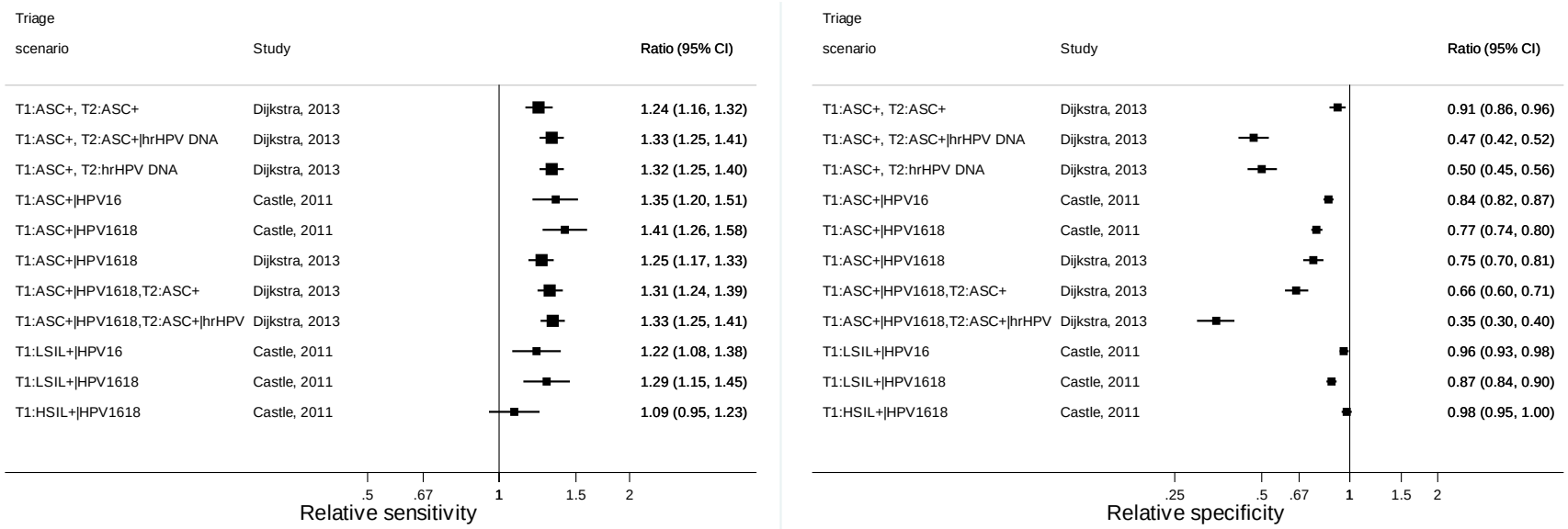


Figure 38: Relative sensitivity (left) and specificity (right) of different scenarios compared to reflex cytology at cutoff ASC-US+ to detect CIN2+ in women with a positive hrHPV DNA screening test. (restricted to scenarios being more sensitive than the comparator triage test)

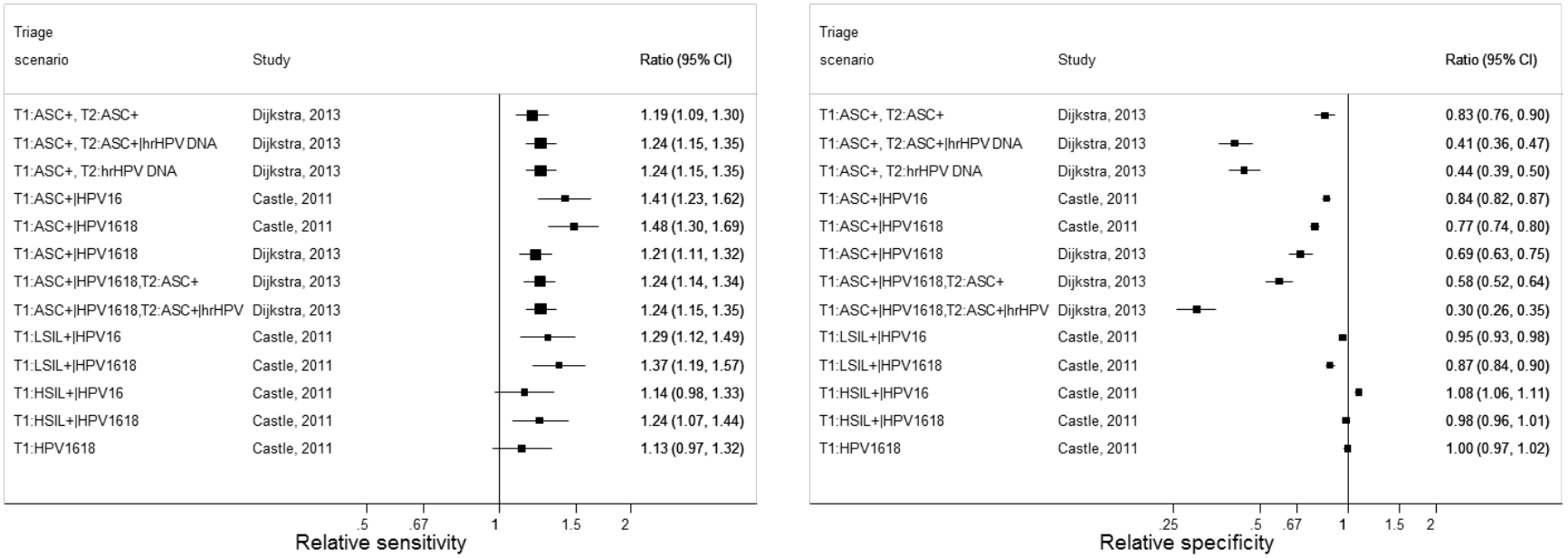


Figure 39: Relative sensitivity (left) and specificity (right) of different scenarios compared to reflex cytology at cutoff ASC-US+ to detect CIN3+ in women with a positive hrHPV DNA screening test. (restricted to scenarios being more sensitive than the comparator triage test)

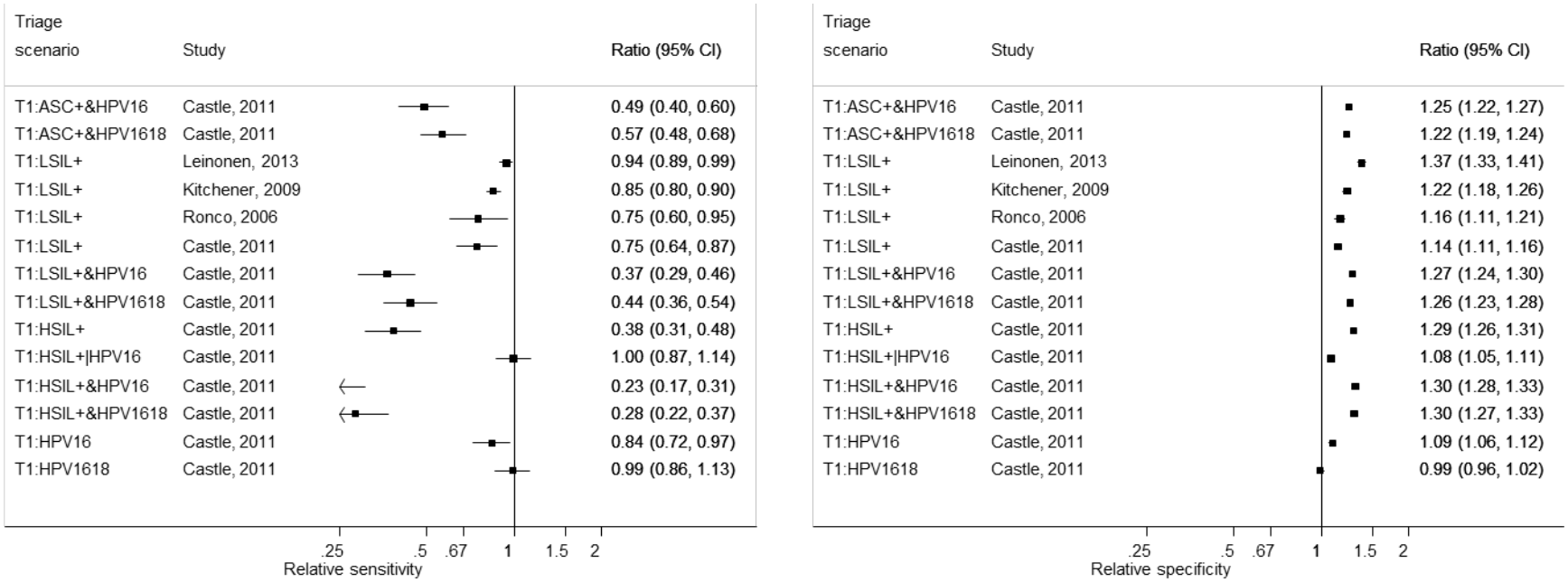


Figure 40: Relative sensitivity (left) and specificity (right) of different scenarios compared to reflex cytology at cutoff ASC-US+ to detect CIN2+ in women with a positive hrHPV-DNA screening test. (restricted to scenarios being less or as sensitive compared to reflex cytology at cutoff ASC-US+).

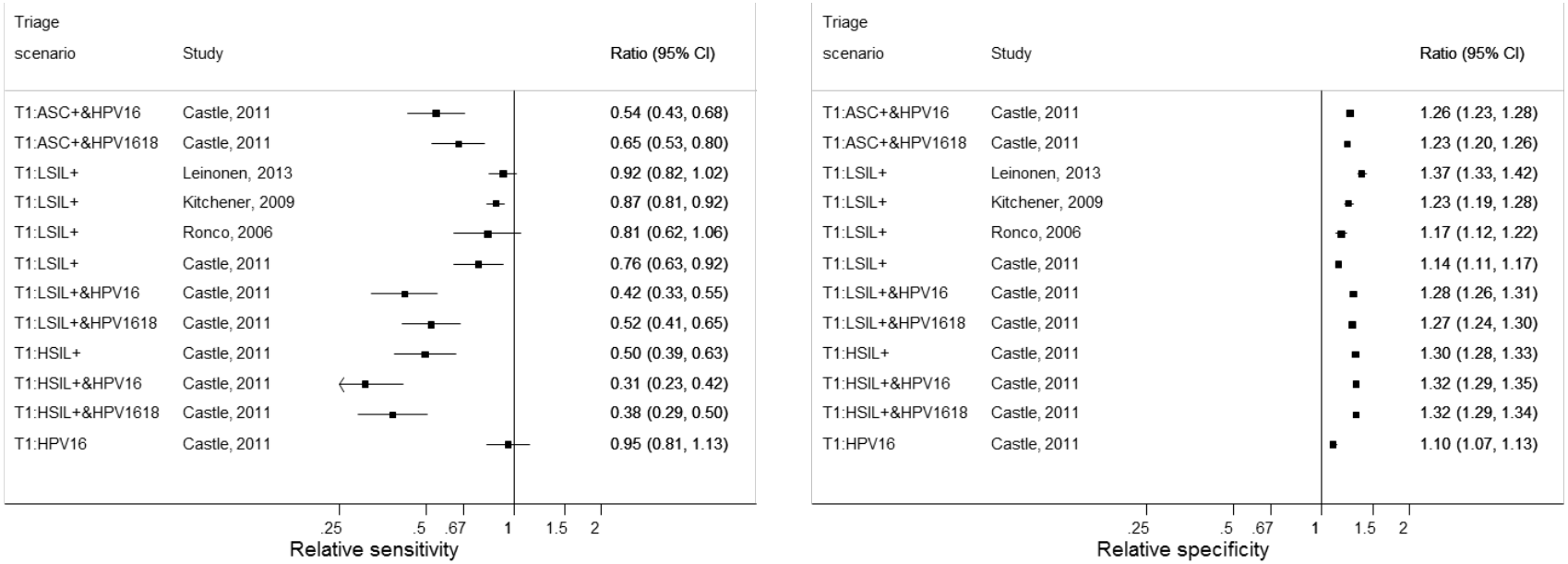


Figure 41: Relative sensitivity (left) and specificity (right) of different scenarios compared to reflex cytology at cutoff ASC-US+ to detect CIN3+ in women with a positive hrHPV DNA screening test. (restricted to scenarios being less or as sensitive compared to reflex cytology at cutoff ASC-US+).

3.3.4. p16^{INK4a} immuno-cytochemistry

One of the seven retrieved RCT's (the phase-2 study of NTCC) contained data on the use of a molecular biomarker (overexpression of p16^{INK4a}) in the triage of hrHPV-positive women. Two reports provided the cross-sectional[12] and longitudinal[13] outcomes of triage based on p16-immunocytochemistry**.** The cross-sectional sensitivity and specificity of reflex triage with p16^{INK4a}, was 88.0% (95% CI=79.6-93.9) and 60.6% (95% CI=57.5-63.6%), respectively, for CIN2+, and 90.5% (95% CI= 77.4-97.3%) and 58.4% (95% CI= 55.5-61.4%), respectively, for CIN3+[12].

The longitudinal sensitivity for p16^{INK4a} assessed over three years was lower than the cross-sectional sensitivity regarding detection of CIN2+ (88.0% vs. 79.0%) and CIN3+ (90.5% vs 84.6%), indicating disease development in women with a negative reflex p16^{INK4a} test [20].

Table 30: Absolute sensitivity and specificity for reflex p16^{INK4a} triage for women with a positive hrHPV test.

Study		Test (cut-off)	Outcome	Absolute Accuracy Sensitivity (95% CI)	Absolute Accuracy Specificity (95% CI)
Carozzi 2008	Cross-sectional	p16^{INK4a} (1 stained cell)*	CIN2+	88.0% (79.6-93.9)	60.6% (57.5-63.6)
Carozzi 2008	Cross-sectional	p16^{INK4a} (1 stained cell)*	CIN3+	90.5% (77.4-97.3)	58.4% (55.5-61.4)
Carozzi 2013	Longitudinal§	p16^{INK4a} (1 stained cell)	CIN2+	79.0% (71.4-85.4)	62.6% (59.0-66.0)
Carozzi 2013	Longitudinal§	p16^{INK4a} (1 stained cell)	CIN3+	84.6% (73.5-92.4)	59.1% (55.6-62.4)

* the cut-off with the best sensitivity. §cumulative disease after three years of follow-up.

By lack of an intra-study comparator, no relative accuracy could be derived for p16-based triage in NTCC-2. However, by comparing with absolute accuracy measures for ASC-US+ triage from NTCC-1 (cross-sectional sensitivity of 75.3% [95% CI: 64.5-84.2] for CIN2+, 82.1% [95% CI: 66.5-92.5%] for CIN3+; specificity of 75.8% [95% CI: 73.1-78.3%] for CIN2+, 74.1% [95% CI: 71.5-76.6%]), we may also obtain credible relative accuracy. Allowing for this inter-NTCC comparison,

we can conclude that p16-based triage was 1.17 times (95% CI: 1.01-1.35) more sensitive to detect CIN2+ and 1.10 times (95% CI: 0.92-1.32) to detect CIN3+ compared to simple cytology triage at cut-off ASC-US. The specificity of p16 triage was significantly lower than ASC-US+ triage: ratio of 0.80 (95% CI: 0.75-0.85) and 0.79 (95% CI: 0.74-0.84), considering outcomes CIN2+ and CIN3+, respectively.

Using p16^{INK4a}/Ki-67 detection to manage hrHPV-positive/cytology-negative women

Another study was identified containing data on triage with p16^{INK4a}/Ki-67 double staining of women who were hrHPV-positive but had normal cytology at primary screening [21].

In this study the p16^{INK4a}/Ki-67 triage had a sensitivity and specificity of 91.9% (95% CI=78.1-98.3) and 80.9% (95% CI= 76.7-84.7) for CIN2+. To detect CIN3+, triage with p16^{INK4a}/Ki-67 resulted in a sensitivity and specificity of 96.4% (95% CI=81.7-99.9) and 79.6% (95% CI=75.3-83.5).

3.3.5. Risk of CIN3+ in hrHPV-positive women with positive or negative triage test results

Sensitivity and specificity are test characteristics reflecting the capacity to identify diseased subjects by a positive test result and non-diseased subjects by a negative test result. These are test characteristics which are typically not influenced by disease prevalence. Therefore, in systematic reviews and meta-analyses, sensitivity and specificity are the test measures that are pooled to synthesize knowledge on test performance.

However, patients, clinicians, and decision makers defining policies for good clinical practice, are in the first place interested in the probability of disease when a test is positive (positive predictive value: PPV) and the risk of disease when a test is negative (complement of the negative predictive value: 1-NPV=cNPN). The PPV provides information on the risk of underlying pre-cancer and consequently on the efficiency of referral for further management. The inverse of the PPV (1/PPV) corresponds with the number needed to refer [colposcopy/biopsy] to find 1 case of cervical pre-cancer. The NPV provides assurance on the safety that a women does not have (pre)-cancer and will have a very low risk to develop (pre-)cancer by the next screening round.

Below, we computed the predictive values for a plausible series of background risks of CIN3+ (possible pretest probabilities) which are relevant for the settings where the evaluated tests will possibly be used. The predictive values, computed for a given setting/area, allow decision making regarding the use of a test in this setting/area. The risk of underlying pre-cancer or cancer (CIN3+) should be sufficiently low in case of a negative screen test result to reassure women and to refer them back to the normal screening schedule[22]. Whereas the risk of CIN3+ should be sufficiently high if the screening test is positive (=PPV). If the PPV is not high enough a triage test is needed.

We considered the following range of background risks of cervical pre-cancer or cancer among women with a positive hrHPV DNA test at screening:

- Low: 5%
- Intermediate: 9% (corresponding to the average cumulative risk of CIN3+ among hrHPV DNA+ women)
- High: 15%.

The low and high estimates correspond with rounded low and high risks observed in the aforementioned screening trials.

We accepted the following cutoffs for the measures of efficiency (PPV) and safety (cNPV), considering prevalent CIN3+ as targeted prevalent disease:

- PPV: >10%
- cNPV: <1%.

In addition, the following cut-offs for longitudinal PPV and cNPV over a period of five years after the screening test were accepted.

- PPV_{long}: >20%
- $cNPV_{long}$: <1%.

The risk or post-test probability of CIN3+ after a positive or negative result of a given triage scenario was computed from:
1) absolute accuracy of the reference triage (reflex cytology at ASC-US+);
2) relative accuracy of the given scenario estimated using a binormal model;
3) assumed underlying low, intermediate and high-risk of CIN3+.
The results are shown in Table 31.

In nearly all triage scenario's and back ground risk situations, more than 10% of triage-positive women will have or will develop CIN3+. Exceptions are some very sensitive two-step triage scenarios (3,4,9 & 10) in low-risk situations (T1:ASC-US+, T2:ASC-US+ | hrHPV; T1:ASC-US+, T2:hrHPV; T1:ASC-US+|HPV1618, T2:ASC-US+; T1:ASC-US+|HPV1619, T2:ASC-US+ | hrHPV).
All two-step scenarios, in low and intermediate risk situations, resulted in a post-test probability <1% when the triage test is negative. In a low-risk situation, also negative reflex-cytology combined with negative HPV1618 genotyping or a negative p16 test is accompanied with <1% post-test probabilities.
In a high-risk situation only two-step triage scenarios (3,4,9 & 10) are associated with <1% post-test probabilities. These last four scenarios are the only which are both efficient (PPV>10%) and safe (cNPV<1%).
In a low-risk situation, two-step triage with reflex-cytology at baseline and at 6-12 months later looks a good triage method (both triage criteria fulfilled, outcome documented over >=36 months and referral rate to colposcopy of only 39%). However, loss to follow-up should be taken into account when triage involves more visits. Avoiding the necessity for repeat testing reduces the risk of loss to follow-up. In the two Dutch trials, the compliance with follow-up after six and twelve months was ~60% and ~75%, repectivey[23,24]. Other studies have also demonstrated considerable loss to follow-up at repeat testing, particularly after normal cytology. Therefore more sensitive one step reflex-triage scenarios are interesting as well, such as T1: ASC-US+ combined with HPV1618

genotyping, which results always in an good PPV (≥10%) in low- and intermediate risk situation and an acceptable cNPV in low- risk situation.

Table 31. Number of true and false-positives among 1000 women with a positive hrHPV test at screening and triaged with one of 23 different scenarios, positive predictive value (PPV=risk of CIN3+ if triage-positive), 1/PPV (number needed to refer to find 1 case of CIN3+), negative predictive value (NPV) and the complement of NPV (cNPV=1-NPV=risk of CIN3+ if triage-negative) estimated for three situations of underlying background risk of CIN3+ (risk=5%, 9% and 15%).

	Triage1	Triage2	Background risk	Sensitivity	Specificity	Useful referrals TP	Missed cases FN	Unnecessary referrals FP	True reassurance TN	PPV	1/PPV	NPV	cNPV	test+	Criteria fulfilled
1	**ASC-US+**		**0.05**	**0.81**	**0.71**	**40**	**10**	**277**	**673**	**0.13**	**7.93**	**0.99**	**0.015**	**0.32**	
2	ASC-US+	ASC-US+	0.05	0.95	0.64	47	3	340	610	0.12	8.23	1.00	0.005	0.39	x
3	ASC-US+	ASC-US+ | hrHPV	0.05	1.00	0.35	50	0	622	328	0.07	13.44	1.00	0.000	0.67	
4	ASC-US+	hrHPV	0.05	1.00	0.37	50	0	601	349	0.08	13.02	1.00	0.000	0.65	
5	**ASC-US+|HPV16**		**0.05**	**0.91**	**0.58**	**46**	**4**	**399**	**551**	**0.10**	**9.67**	**0.99**	**0.007**	**0.45**	x
6	ASC-US+&HPV16		0.05	0.59	0.93	30	20	65	885	0.32	3.17	0.98	0.022	0.10	
7	ASC-US+|HPV1618		0.05	0.93	0.53	47	3	445	505	0.10	10.47	0.99	0.006	0.49	x
8	ASC-US+&HPV1618		0.05	0.66	0.91	33	17	85	865	0.28	3.58	0.98	0.019	0.12	
9	ASC-US+|HPV1618	ASC-US+	0.05	0.99	0.44	50	0	530	420	0.09	11.60	1.00	0.000	0.58	
10	ASC-US+|HPV1618	ASC-US+ | hrHPV	0.05	1.00	0.24	50	0	724	226	0.06	15.48	1.00	0.000	0.77	
11	**LSIL+**		**0.05**	**0.71**	**0.85**	**36**	**14**	**142**	**808**	**0.20**	**4.94**	**0.98**	**0.017**	**0.18**	
12	LSIL+ | HPV16		0.05	0.89	0.68	44	6	302	648	0.13	7.86	0.99	0.009	0.35	x
13	LSIL+ & HPV16		0.05	0.51	0.96	26	24	37	913	0.41	2.42	0.97	0.026	0.06	
14	LSIL+ | HPV1618		0.05	0.91	0.61	45	5	369	581	0.11	9.20	0.99	0.009	0.41	x
15	LSIL+ & HPV1618		0.05	0.58	0.95	29	21	49	901	0.37	2.69	0.98	0.023	0.08	
16	**HSIL+**		**0.05**	**0.56**	**0.98**	28	22	24	926	**0.54**	**1.86**	**0.98**	**0.023**	**0.05**	
17	HSIL+ | HPV16		0.05	0.85	0.78	42	8	210	740	0.17	6.00	0.99	0.011	0.25	
18	HSIL+ & HPV16		0.05	0.41	0.99	21	29	10	940	0.68	1.48	0.97	0.030	0.03	
19	HSIL+ | HPV1618		0.05	0.87	0.69	44	6	291	659	0.13	7.61	0.99	0.009	0.34	x
20	HSIL+ &		0.05	0.48	0.99	24	26	12	938	0.67	1.50	0.97	0.027	0.04	

	Triage1	Triage2	Background risk	Sensitivity	Specificity	Useful referrals TP	Missed cases FN	Unnecessary referrals FP	True reassurance TN	PPV	1/PPV	NPV	cNPV	test+	Criteria fulfilled
	HPV1618														
21	**HPV16**		**0.05**	**0.79**	**0.79**	39	11	198	752	**0.16**	**6.08**	**0.99**	**0.014**	**0.24**	
22	**HPV1618**		**0.05**	**0.84**	**0.70**	42	8	280	670	**0.13**	**7.67**	**0.99**	**0.012**	**0.32**	
23	p16		0.05	0.89	0.56	44	6	418	532	0.10	10.50	0.99	0.011	0.46	x
1	**ASC-US+**		**0.09**	**0.81**	**0.71**	**75**	**18**	**264**	**643**	**0.22**	**4.52**	**0.97**	**0.027**	**0.34**	
2	ASC-US+	ASC-US+	0.09	0.95	0.64	88	5	325	582	0.21	4.69	0.99	0.009	0.41	x
3	ASC-US+	ASC-US+ | hrHPV	0.09	1.00	0.35	93	0	594	313	0.14	7.39	1.00	0.000	0.69	x
4	ASC-US+	hrHPV	0.09	1.00	0.37	93	0	573	334	0.14	7.16	1.00	0.000	0.67	x
5	**ASC-US+|HPV16**		**0.09**	**0.91**	**0.58**	**85**	**8**	**381**	**526**	**0.18**	**5.48**	**0.99**	**0.015**	**0.47**	
6	ASC-US+&HPV16		0.09	0.59	0.93	55	38	62	845	0.47	2.13	0.96	0.043	0.12	
7	ASC-US+|HPV1618		0.09	0.93	0.53	87	6	425	482	0.17	5.89	0.99	0.012	0.51	
8	ASC-US+&HPV1618		0.09	0.66	0.91	61	32	81	826	0.43	2.33	0.96	0.037	0.14	
9	ASC-US+|HPV1618	ASC-US+	0.09	0.99	0.44	92	1	506	401	0.15	6.50	1.00	0.002	0.60	x
10	ASC-US+|HPV1618	ASC-US+ | hrHPV	0.09	1.00	0.24	93	0	691	216	0.12	8.43	1.00	0.000	0.78	x
11	**LSIL+**		**0.09**	**0.71**	**0.85**	**66**	**27**	**136**	**771**	**0.33**	**3.06**	**0.97**	**0.034**	**0.20**	
12	LSIL+ | HPV16		0.09	0.89	0.68	83	10	289	618	0.22	4.48	0.98	0.016	0.37	
13	LSIL+ & HPV16		0.09	0.51	0.96	48	45	36	871	0.57	1.75	0.95	0.049	0.08	
14	LSIL+ | HPV1618		0.09	0.91	0.61	84	9	353	554	0.19	5.20	0.98	0.016	0.44	
15	LSIL+ & HPV1618		0.09	0.58	0.95	54	39	47	860	0.53	1.87	0.96	0.043	0.10	
16	**HSIL+**		**0.09**	**0.56**	**0.98**	**52**	**41**	**22**	**885**	**0.70**	**1.42**	**0.96**	**0.044**	**0.07**	
17	HSIL+ | HPV16		0.09	0.85	0.78	79	14	200	707	0.28	3.53	0.98	0.019	0.28	
18	HSIL+ & HPV16		0.09	0.41	0.99	38	55	10	897	0.79	1.26	0.94	0.058	0.05	
19	HSIL+ | HPV1618		0.09	0.87	0.69	81	12	277	630	0.23	4.42	0.98	0.019	0.36	
20	HSIL+ & HPV1618		0.09	0.48	0.99	44	49	11	896	0.80	1.25	0.95	0.052	0.06	

	Triage1	Triage2	Background risk	Sensitivity	Specificity	Useful referrals TP	Missed cases FN	Unnecessary referrals FP	True reassurance TN	PPV	1/PPV	NPV	cNPV	test+	Criteria fulfilled
21	**HPV16**		**0.09**	**0.79**	**0.79**	**73**	**20**	**189**	**718**	**0.28**	**3.59**	**0.97**	**0.027**	**0.26**	
22	**HPV1618**		**0.09**	**0.84**	**0.70**	**78**	**15**	**268**	**639**	**0.23**	**4.44**	**0.98**	**0.023**	**0.35**	
23	p16		0.09	0.89	0.56	80	10	400	510	0.17	6.00	0.98	0.019	0.48	
1	ASC-US+		0.15	0.81	0.71	121	29	248	602	0.33	3.05	0.95	0.046	0.37	
2	ASC-US+	ASC-US+	0.15	0.95	0.64	142	8	304	546	0.32	3.14	0.99	0.014	0.45	
3	ASC-US+	ASC-US+ | hrHPV	0.15	1.00	0.35	150	0	556	294	0.21	4.71	1.00	0.000	0.71	x
4	ASC-US+	hrHPV	0.15	1.00	0.37	150	0	537	313	0.22	4.58	1.00	0.000	0.69	x
5	ASC-US+|HPV16		0.15	0.91	0.58	137	13	357	493	0.28	3.61	0.97	0.026	0.49	
6	ASC-US+&HPV16		0.15	0.59	0.93	89	61	58	792	0.61	1.65	0.93	0.072	0.15	
7	ASC-US+|HPV1618		0.15	0.93	0.53	140	10	398	452	0.26	3.84	0.98	0.022	0.54	
8	ASC-US+&HPV1618		0.15	0.66	0.91	98	52	76	774	0.56	1.78	0.94	0.063	0.17	
9	ASC-US+|HPV1618	ASC-US+	0.15	0.99	0.44	149	1	475	375	0.24	4.19	1.00	0.003	0.62	x
10	ASC-US+|HPV1618	ASC-US+ | hrHPV	0.15	1.00	0.24	150	0	647	203	0.19	5.31	1.00	0.000	0.80	x
11	LSIL+		0.15	0.71	0.85	107	43	127	723	0.46	2.19	0.94	0.056	0.23	
12	LSIL+ | HPV16		0.15	0.89	0.68	133	17	271	579	0.33	3.04	0.97	0.029	0.40	
13	LSIL+ & HPV16		0.15	0.51	0.96	77	73	33	817	0.70	1.43	0.92	0.082	0.11	
14	LSIL+ | HPV1618		0.15	0.91	0.61	136	14	330	520	0.29	3.43	0.97	0.026	0.47	
15	LSIL+ & HPV1618		0.15	0.58	0.95	87	63	44	806	0.66	1.51	0.93	0.072	0.13	
16	HSIL+		0.15	0.56	0.98	84	66	21	829	0.80	1.25	0.93	0.074	0.11	
17	HSIL+ | HPV16		0.15	0.85	0.78	127	23	188	662	0.40	2.48	0.97	0.034	0.32	
18	HSIL+ & HPV16		0.15	0.41	0.99	62	88	9	841	0.87	1.15	0.91	0.095	0.07	
19	HSIL+ | HPV1618		0.15	0.87	0.69	131	19	260	590	0.34	2.98	0.97	0.031	0.39	
20	HSIL+ & HPV1618		0.15	0.48	0.99	72	78	11	839	0.87	1.15	0.91	0.085	0.08	
21	HPV16		0.15	0.79	0.79	118	32	177	673	0.40	2.50	0.95	0.045	0.30	

	Triage1	Triage2	Background risk	Sensitivity	Specificity	Useful referrals TP	Missed cases FN	Unnecessary referrals FP	True reassurance TN	PPV	1/PPV	NPV	cNPV	test+	Criteria fulfilled
22	HPV1618		0.15	0.84	0.70	126	24	251	599	0.33	2.99	0.96	0.039	0.38	
23	p16		0.15	0.89	0.56	133	17	374	476	0.26	3.81	0.97	0.034	0.51	

3.4. Discussion

In the near future, screening for cervical cancer will likely shift from cytological to virological screening. However, the optimal management of women with a hrHPV infection remains an imperative issue to solve, since hrHPV testing has a lower cross-sectional specificity compared to cytology[1]. As a consequence, the triage of hrHPV positive women is needed to limit the burden of follow-up and to avoid over-diagnosis and over-treatment as much as possible.

Different triage options nested in large screening trials using an hrHPV assay as a primary screening test, enabled us to assess the accuracy of diverse strategies to manage hrHPV-positive women.

A two-step triage scenario with twice cytology at cutoff ASC-US+ (strategy 2 in Table 31) offers a good balance of efficiency (4 to 9 referrals to detect one CIN3+, ~40% of referral) and safety (risk of CIN3+ in triage-negative women of 0.5% to 0.9%). If the background risk is higher (>=15%), the safety becomes borderline (risk of CIN3+ in next 3-5 years of 1.4%). In the Netherlands, this scenario has been chosen for the future HPV-based screening policy, which will be applied the whole country in 2016. The safety of strategy 2 can be increased by adding HPV16 or HPV1618 genotyping and/or hrHPV testing, or by replacing cytology with a repeat hrHPV test. In these scenarios, safety criteria are obviously fulfilled, even when the background risk is high, but they are accompanied by a substantially increased referral rate (67% to 71%).

Two-step scenarios are characterised by a certain degree of drop-out of women under follow-up. Where this drop-out is important, more sensitive reflex triage scenarios could be favoured which involve reflex cytology combined with HPV1618 genotyping (scenarios 7 and 14). However, these scenarios do not reach the safety criterion when the back ground risk is intermediate or high.

Limits and strengths of the review

Because of time constraints the current review was restricted to large population-based trials comparing HPV-based with cytology-based screening. A more comprehensive literature review is currently being done. However, it is expected that the main bulk of useful information on triage of HPV+ women may be included in the studies retrieved in this review.

Timing of outcome is often limited to a few months after observation of the hrHPV screen test result. Outcomes from studies comprising up to 3-4 years of surveillance provide more useful information (see Table 26) than those

with only 3-6 months of follow-up. Unfortunately, no results were available for 5 years of follow-up or more.

Many scenarios of triage are documented in few and often even in only one study. Moreover, the inter-study heterogeneity in the absolute accuracy values observed in multiple studies assessing a particular scenario, often was large. However, by assessing the relative accuracy, variability was reduced and therefore, absolute accuracies predicting the outcomes were based on the product of the accuracy of the reference triage scenario (reflex cytology at ASC-US documented in eight reports) * relative accuracy of a given scenario compared with this reference.

Not all relevant triage information reported in secondary publications of the screening trials could be included in a formal meta-analysis since only proportions or rates were reported with different assumptions applied for adjustment for follow-up compliance. Adjustment for incomplete compliance could not be assessed statistically since it requires availability of the absolute data. Requesting data from authors will be done within COCEAHR, but cannot yet be included in the current review.

The definition of criteria for good triage scenario's (PPV>10% and cNPV<1%, considering CIN3+ as outcome) are arbitrary and depend on length of duration of follow-up. The choice was based on conventions agreed among certain experts. International consensus building on these criteria may be needed and should involve policy makers, clinical experts, systematic reviewers, health economists and patient organisations. Criteria for the outcome cancer may be preferred but these criteria would today not be verifiable. Nevertheless, incidence of invasive cancer according to triage policy and compliance with this policy should be target of monitoring based in systematic linkages with screening and cancer registries.

The future evaluation of multiple step triage scenarios should include the proportion of CIN3+ identified at each successive step beyond the baseline step and the proportion of drop-out at each additional follow-up visit in order to assess to overall cumulative sensitivity and safety compared to one-step scenarios.

Other markers may be useful in triage of hrHPV-positive women as well (in particular, double immune-staining for p16 and Ki67, hypermethylation profiles, expression of oncoproteins such as E6 and E7, chromosomal aberrations, viral mRNA testing, evolution of type specific viral load) and may provide alternatives for the triage scenarios considered in this review. Some

publications are expected to become available in the near future and should be included in updated reviews as soon as possible.

In the current review, triage with reflex cytology and repeat cytology appeared to be an acceptable scenario. However, it should be mentioned that the quality of cytology in the field may be more heterogeneous than in the trials included in this review. Triage with objective bio-markers could reduce this variability.

3.5. References

1. Arbyn M, Ronco G, Anttila A, et al. Evidence regarding HPV testing in secondary prevention of cervical cancer. *Vaccine* 2012;**30 Suppl 5**: F88-F99.

2. Ronco G, Dillner J, Elfstrom KM, et al. Efficacy of HPV-based screening for prevention of invasive cervical cancer: follow-up of four European randomised controlled trials. *Lancet* 2013; in-press.

3. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Med* 2009;**6**: e1000097.

4. Diagnostic Test Accuracy Working Group. Handbook for Diagnostic Test Accuracy Reviews. Available from: http://srdta.cochrane.org/handbook-dta-reviews. 2011.

5. Harbord RM, Deeks JJ, Egger M, Whiting P, Sterne JA. A unification of models for meta-analysis of diagnostic accuracy studies. *Biostatistics* 2007;**8**: 239-51.

6. Harbord RM, Whiting P. metandi: Meta-analysis of diagnostic accuracy using hierarchical logistic regression. *The Stata Journal* 2009;**9**: 211-29.

7. Harris R, Bradburn M, Deeks J, et al. metan: fixed- and random-effects meta-analysis. *Stata Journal* 2008;**8**: 3-28.

8. Dersimonian R, Laird NM. Meta-analysis in clinical trials. *Controlled Clin Trials* 1986;**7**: 177-88.

9. Takwoingi Y, Diagnostic Test AccuracyWorking Group. METADAS: A SAS macro for meta-analysis of diagnostic accuracy studies. http://srdta cochrane org/sites/srdta cochrane org/files/uploads/MetaDAS%20Quick%20Reference%20v1 3%20May%202012 pdf 2009:-*.

10. Ronco G, Segnan N, Giorgi-Rossi P, et al. Human Papillomavirus testing and liquid-based cytology in primary cervical screening: results at recruitment from the New Technologies for Cervical Cancer randomized controlled trial. *J Natl Cancer Inst* 2006;**98**: 765-74.

11. Ronco G, Giorgi-Rossi P, Carozzi F, et al. Human papillomavirus testing and liquid-based cytology in primary screening of women younger than 35 years: results at recruitment for a randomised controlled trial. *Lancet Oncol* 2006;**7**: 547-55.

12. Carozzi F, Confortini M, Palma PD, et al. Use of p16-INK4A overexpression to increase the specificity of human papillomavirus

testing: a nested substudy of the NTCC randomised controlled trial. *Lancet Oncol* 2008;**9**: 937-45.

13. Carozzi F, Gillio-Tos A, Confortini M, et al. Risk of high-grade cervical intraepithelial neoplasia during follow-up in HPV-positive women according to baseline p16-INK4A results: a prospective analysis of a nested substudy of the NTCC randomised controlled trial. *The Lancet Oncology* 2013;**14**: 168-76.

14. Kitchener HC, Almonte M, Thomson C, et al. HPV testing in combination with liquid-based cytology in primary cervical screening (ARTISTIC): a randomised controlled trial. *Lancet Oncol* 2009;**10**: 672-82.

15. Naucler P, Ryd W, Tornberg S, et al. Efficacy of HPV DNA testing with cytology triage and/or repeat HPV DNA testing in primary cervical cancer screening. *J Natl Cancer Inst* 2009;**101**: 88-98.

16. Rijkaart DC, Berkhof J, van Kemenade FJ, et al. Evaluation of 14 triage strategies for HPV DNA-positive women in population-based cervical screening. *Int J Cancer* 2012;**130**: 602-10.

17. Dijkstra M, Van ND, Rijkaart D, et al. Primary hrHPV DNA testing in Cervical Cancer screening: how to manage screen positive women? A POBASCAM Trial sub study. *Cancer Epidemiol Biomarkers Prev* 2013; *.

18. Leinonen MK, Anttila A, Malila N, et al. Type- and age-specific distribution of human papillomavirus in women attending cervical cancer screening in Finland. *Br J Cancer* 2013; in-press.

19. Castle PE, Stoler MH, Wright TC, Jr., et al. Performance of carcinogenic human papillomavirus (HPV) testing and HPV16 or HPV18 genotyping for cervical cancer screening of women aged 25 years and older: a subanalysis of the ATHENA study. *Lancet Oncol* 2011;**12**: 880-90.

20. Carozzi F, Gillio-Tos A, Confortini M, et al. Risk of high-grade cervical intraepithelial neoplasia during follow-up in HPV-positive women according to baseline p16-INK4A results: a prospective analysis of a nested substudy of the NTCC randomised controlled trial. *Lancet Oncol* 2012.

21. Petry KU, Schmidt D, Scherbring S, et al. Triaging Pap cytology negative, HPV positive cervical cancer screening results with p16/Ki-67 Dual-stained cytology. *Gynecol Oncol* 2011.

22. Arbyn M, Verdoodt F, Snijders PJF, et al. Accuracy of human papillomavirus testing on self-collected versus clinician-collected samples: a meta-analysis. *Lancet Oncol* 2014; pre-published.

23. Bulkmans N, Berkhof J, Rozendaal L, et al. Human papillomavirus DNA testing for the detection of cervical intraepithelial neoplasia grade 3 and

cancer: 5-year follow-up of a randomised controlled implementation trial. *Lancet* 2007;**370**: 796-802.

24. Rijkaart DC, Berkhof J, Rozendaal L, et al. Human papillomavirus testing for the detection of high-grade cervical intraepithelial neoplasia and cancer: final results of the POBASCAM randomised controlled trial. *Lancet Oncol* 2012;**13**: 78-88.

3.6. GRADE-Profil

Authors: F. Verdoodt, M. Arbyn, M. Jentschke

Steps in evidence assessment for making guidelines
1) Formulate a question
2) Identify the PICO(S) components
3) Qualify outcomes as critical, important, not important

1) Questions

What is the best test or combination of tests which results in the highest sensitivity for progressive cervical precancer at the lowest burden of follow-up?

2) PICOS

P: women participating in virological screening for cervical cancer, having a positive hrHPV-DNA test result

I: reflex testing with biomarkers (HPV genotyping, hrHPV-mRNA testing, p16, p16/KI67, other markers) and repetition of hrHPV-DNA testing, cytology and/or or combinations thereof

C: reflex cytology triage at cut-off ASC-US

O: cross-sectional and longitudinal accuracy to detect histologically identified disease (=CIN2+,CIN3+/AIS, and cervical cancer)
triage test positivity rate, referral rate for colposcopy, PPV for CIN2+ & CIN3+, risk of CIN2+, CIN3+ and cancer after negative triage testing

S: - follow-up of randomised trials comparing cytology, with HPV-based screening and applying different follow-up algorithms
- complete diagnostic studies (all subjects tested with triage method and verification with the reference standard (colposcopy/biopsy))
- cohort studies applying at least two alternative triage algorithms involving verification with the reference standard if one or more positive triage test result

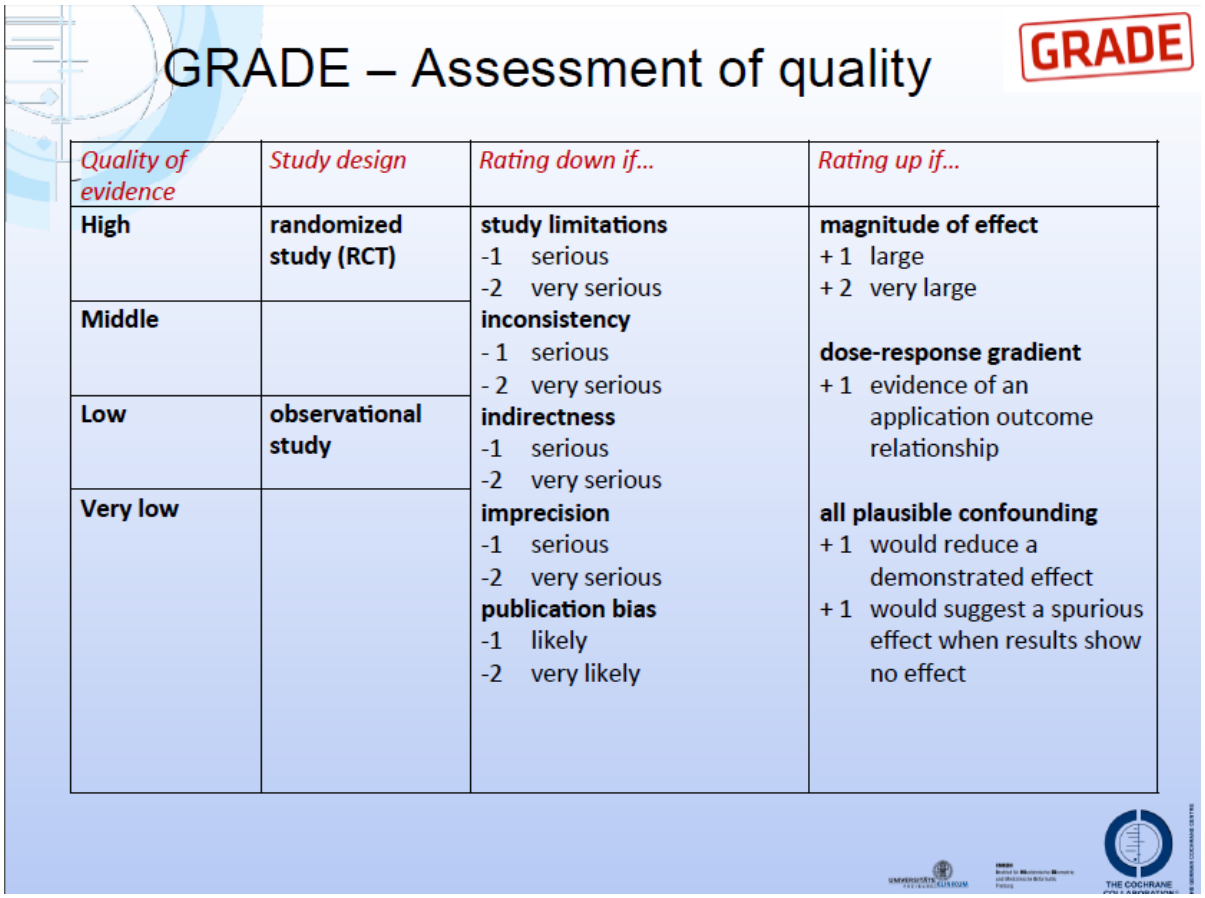
3) Importance of outcomes

Outcome:

13. Reduction of mortality from cervical cancer, (quality-adjusted) life-years gained.
14. Reduction of morbidity due to cervical cancer: incidence of cancer (Ib+).
15. Reduction of incidence of cancer (including micro-invasive cancer).
16. Reduction of incidence of CIN3 or worse disease (CIN3+).
17. Increased detection rate of CIN3+ or CIN2+.
18. Increased test positivity with increased, similar or hardly reduced positive predictive value.

4) Quality of evidence for each outcome in four categories;

- High: ++++
- Moderate: +++
- Low: ++
- Very low: +



GRADE – Assessment of quality

GRADE

Quality of evidence	*Study design*	*Rating down if…*	*Rating up if…*
High	**randomized study (RCT)**	**study limitations** -1 serious -2 very serious **inconsistency** - 1 serious - 2 very serious **indirectness** -1 serious -2 very serious **imprecision** -1 serious -2 very serious **publication bias** -1 likely -2 very likely	**magnitude of effect** + 1 large + 2 very large **dose-response gradient** + 1 evidence of an application outcome relationship **all plausible confounding** + 1 would reduce a demonstrated effect + 1 would suggest a spurious effect when results show no effect
Middle			
Low	**observational study**		
Very low			

<u>Included studies:</u>

(NTCC[10-13], ARTISTIC[14], SWEDESCREEN[15], VUSA[16], POBASCAM[17], and PUBLIC HEALTH TRIAL FINLAND[18]) and one American trial (ATHENA[19]).

➔ Large RCT's, therefore <u>high-quality evidence</u>

<u>Quadas items:</u>

Patient Sellection:	P1	acceptable sampling method	1 = Yes
	P2	inappropriate exclusions avoided	0 = No
Triage Test	T1	pre-specified cut-off	9 = Unclear

T2 results of other tests blinded when interpreting triage test

Reference Standard R1 acceptable reference test

R2 results of triage tests blinded when interpreting the reference standard

R3 Incorporation avoided

Flow & Timing F1 acceptable delay between tests

F2 partial verification avoided

F3 differential verification avoided

F4 withdrawals explained

F5 uninterpretable results reported

a) triage test
b) reference test

Concerns of Applicability:

A1 domain 1: Patient Selection h = high risk of bias

A2 domain 2: Index test L = low risk

A3 domain 3: Reference standard ? = unclear

	Risk of Bias													Concerns of Applicability		
	Patient Selection		Triage test		Reference test			Flow & Timing								
Author, Year	P1	P2	T1	T2	R1	R2	R3	F1	F2	F3	F4	F5	F6	Patient Selection	Index & comparator test	Reference test
Ronco 2006	Y	Y	Y	Y	Y	N	Y	Y	Y	Y	Y	Y	Y	Low	Low	High
Carozzi 2008	Y	Y	Y	Y	Y	N	Y	Y	Y	Y	Y	Y	N	Low	Low	High
Kitchener 2009	Y	Y	Y	Y	Y	N	Y	Y	N	Y	Y	Y	Y	Low	Low	High
Naucler 2009	Y	Y	Y	Y	Y	Y	Y	Y	N	Y	Y	Y	Y	Low	Low	Low
Castle 2011	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Low	Low	Low
Rijkaart 2013	Y	Y	Y	?	Y	?	Y	Y	?	Y	Y	Y	N	Low	Moderate	Moderate
Carozzi 2013	Y	Y	Y	Y	Y	N¥	Y	Y	Y	Y	Y	Y	N	Low	Low	High
Dijkstra 2013	Y	Y	Y	?	Y	?	Y	Y	?	Y	Y	Y	N	Low	Moderate	Moderate
Leinonen 2013	Y	Y	Y	?	Y	?	Y	Y	?	Y	Y	Y	N	Low	Moderate	Moderate

5 factors that lower the quality of the evidence (Guyatt *et al.*, 2008a)

11. Risk of bias due to limitations in design or execution: see CONSORT, STROBE, QUADAS
12. Inconsistency or heterogeneity: if consistency unexplained, lower quality
13. Indirectness, applicability (relevance of studies for answering the PICPO question)
14. Imprecision: number of studies, width of CI
15. Reporting bias, publication bias.

3 factors that increase the quality

4. Large effect
5. Dose effect gradient
6. Confounders (presence of confounders that have lowered the observed effect, absence of these would have increased the effect) (Guyatt *et al.*, 2008a)

Items downgrading quality of evidence		**Downgrading**
Bias, design		No (-0)
Inconsistency		No (-0)
Indirectness		No (-0)
Imprecision		No (-0)
Publication bias, other		No (-0)
Items upgrading quality of evidence		
Large effect	No	No (+0)
Dose-effect correlation	No	No (+0)
Confounding factors neutralising effects	No	No (+0)°

Conclusion: evidence of high quality. (regarding cytology triage).

Overall Grade of the quality of evidence is assigned and this is based on the outcome with the lowest quality of evidence given that it is a critical outcome.

Table 32 GRADE evidence profile

	Quality of evidence								
# studies (N)	Absence of study limitations	Consistency	Directness (outcome, representativity Germany)	Precision	Absence publication bias	Large effect	Dose-effect relation	Bias lowering effect	Comment

Outcome 1: Triage of women with a positive HPV-test at screening									
9	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**	**No**	**No**	**No**	**High**

References

Guyatt G.H., Oxman A.D., Kunz R., Falck-Ytter Y., Vist G.E., Liberati A., & Schunemann H.J. (2008a) Going from evidence to recommendations. *BMJ* **336:** 1049-1051.

Guyatt G.H., Oxman A.D., Kunz R., Vist G.E., Falck-Ytter Y., & Schunemann H.J. (2008b) What is "quality of evidence" and why is it important to clinicians? *BMJ* **336:** 995-998.

Guyatt G.H., Oxman A.D., Vist G.E., Kunz R., Falck-Ytter Y., Alonso-Coello P., & Schunemann H.J. (2008c) GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *BMJ* **336:** 924-926.

Schunemann H.J., Oxman A.D., Brozek J., Glasziou P., Jaeschke R., Vist G.E., Williams J.W., Jr., Kunz R., Craig J., Montori V.M., Bossuyt P., & Guyatt G.H. (2008) GRADE: Grading quality of evidence and strength of recommendations for diagnostic tests and strategies. *BMJ* **336:** 1106-1110.

4. Question: Accuracy of tests used to triage women with minor abnormal cervical cytology?

Unit of Cancer Epidemiology
Wetenschappelijk Instituut Volksgezondheid (WIV)
Juliette Wytsmanstraat 14
1050 Brussel

4.1. Introduction

Cervical cytology has been and still is, in most industrialised countries, the mainstay of secondary prevention of cervical cancer. Women with cytological lesions require further follow-up or treatment, or both, depending on the severity of the lesion. Women with high-grade cytological lesions should be referred immediately for further diagnostic work-up(1;2). However, more options are considered for management of women with minor cytologic lesions (3-7).

The term minor cytological abnormalities encompass the following categories:

- atypical squamous cells of undetermined significance (ASCUS defined according to TBS-1988(8) and ASC-US defined according to TBS-2001(9)),
- atypical glandular cells (AGUS defined according to TBS-1988(8) and AGC defined according to TBS-2001(9)),
- atypical squamous cells where a high-grade squamous intraepithelial abnormality cannot be excluded (ASC-H) (8;9)
- low-grade intraepithelial lesions (LSIL)(8;9).

Until recently, follow-up recommendations for women with ASC-US and LSIL varied from conservative repeat cytology (10-12), to immediate referral for colposcopy and biopsy (13-15). Although most women with an ASC-US or LSIL smear result do not have clinically significant disease, a substantial proportion of them do have histopathologically-confirmed high-grade cervical intra-epithelial neoplasia (CIN)(16-18). From a population of women screened in the USA, it was estimated that one third of CIN were discovered on follow-up of a previous smear with ASC-US(18). According to previous meta-analyses, the absolute risk of underlying high-grade CIN (grade II or III or worse [CIN2/3+]), among women with ASC-US, is on average, 9-10 % for CIN2+ and 4-5 % for CIN3+(19). For women with LSIL, these risks are about 1.5 to 2 times as high(19). The risks are 10 to 30 times higher than for women with normal cytology.

An appropriate triage method should identify those women that have, or will develop, a cervical cancer precursor. At the same time, an accurate triage should reduce the risk of over-diagnosis and overtreatment, including adverse obstetric outcomes associated with excision of CIN lesions(20;21). Given the evidence concerning the etiological role of high-risk human papillomavirus (hrHPV) infections in the development of cervical cancer and its precursors (22-25), HPV testing has been proposed as an alternative triage method to distinguish between women with minor cytological lesions who need referral for colposcopy, and those who can be referred back to the normal screening schedule(3;17).
In this report, an update is made of previous meta-analyses and a Cochrane review which address the accuracy of high-risk HPV testing by the Hybrid Capture assay (HC2), other hrHPV tests and repeat cytology(19;26-28).

4.2. Materials and Methods

4.2.1. Clinical Questions

What is the clinical accuracy (sensitivity and specificity) to identify or exclude high-grade cervical precancer or worse (CIN2+, CIN3+, AIS+) using hrHPV testing or biomarkers among women with minor abnormal cytology (ASC-US, LSIL, ASCH or AGC.
How compares this sensitivity and specificity with that of repeat cytolology?

4.2.2. Literature Retrieval

Previously, meta-analyses on the use of various hrHPV tests(27;29-31) and the use of HC2 versus cytology(27;28) to triage women with minor cytological abnormalities have been performed and published by the Unit of Cancer Epidemiology. To update these systematic reviews, the electronic databases Medline, EMBASE, and CENTRAL were searched for more recent studies, using the search string shown in Box 1 which was, for each database, adapted to the relevant syntax.
We refer to the published reviews for details on the strategy for literature retrieval. In general, studies were eligible if (1) cross-sectional and/or longitudinal triage data were available for women with a cytological diagnosis of ASC-US, LSIL, ASC-H or AGC and (2) verification with the golden standard (colposcopy and targeted biopsy, possibly completed with random biopsies and/or endocervical curettage) was performed on all women or women with a positive triage test in case of randomised trials.

((cervix OR cervical) AND (cancer OR carcinoma OR neoplas* OR dysplas* OR CIN OR SIL) OR (cervix neoplasm[mesh])) AND (HPV OR (HPV AND DNA) OR (HPV AND viral) OR "human papillomavirus" OR hybrid capture OR HC2-assay OR HC2 OR HC-2 OR HCII OR HC-II) AND (triage OR management OR "follow up")

Box 1: Search string for literature retrieval.

4.2.3. Outcome measures

The following outcome measures were assessed, separately for the triage of ASC-US and LSIL:

- disease rate (CIN2+ or CIN3+)
- absolute sensitivity and specificity of HC2 (cut-off: 1.00 RLU/CO) and cytology (cut-off: ASC-US) to detect CIN2+ or CIN3+
- absolute sensitivity and specificity of other hrHPV tests and biomarkers to detect CIN2+ or CIN3+
- relative sensitivity and specificity HC2 (cut-off: 1.00 RLU/CO) versus cytology (cut-off: ASC-US) to detect CIN2+ or CIN3+
- relative sensitivity and specificity of other hrHPV tests and biomarkers versus HC2 (cut-off: 1.00 RLU/CO) to detect CIN2+ or CIN3+

The following outcome measures were assessed, separately for the triage of AGC and ASC-H:

- disease rate (CIN2+ or CIN3+)
- absolute accuracy of HC2 (cut-off: 1.00 RLU/CO) and cytology (cut-off: ASC-US) to detect CIN2+ or CIN3+
- relative accuracy HC2 (cut-off: 1.00 RLU/CO) versus cytology (cut-off: ASC-US) to detect CIN2+ or CIN3+

4.2.4. Statistical analysis

Absolute and relative accuracy were computed using the STATA procedures metaprop[††] and metan(32), respectively. Overall pooled measures, with 95% confidence intervals were calculated using random effects models(33). The statistical heterogeneity was assessed by the p-value for heterogeneity (following a chi2 distribution) as well as by the I^2 statistic, which measures the proportion of the variation that is due to inter-study heterogeneity.

[††] Metaprop is a statistical procedure in STATA developed at the Unit of Cancer Epidemiology (IPH Brussels) to pool proportions based on binomial distributions.

4.3. Results

4.3.1. Triage of women with minor cytological lesions

4.3.1.1. Triage of ASC-US

A previously published meta-analysis on the accuracy of HC2 in triage of ASC-US(27), was updated with eleven new studies (see Figure 41 and section 4.5.1). Based on a total of 52 studies, the pooled sensitivity and specificity of HC2 to detect CIN2+ was 94.4% (95% CI= 92.6-96.0%) and 54.9% (95% CI= 51.0-58.9%), respectively (Figure 42, Figure 43,
Table 33). Compared to cytology, HC2 was 27% more sensitive (relative sensitivity: 1.27 [95% CI: 1.16-1.39]) and not less specific (relative specificity: 0.99 [95% CI= 0.97-1.03]) (Table 34).
Overall, the prevalence of CIN2+ and CIN3+ among women with ASC-US was 14.2% (95% CI=11.6-16.9%) for CIN2+, and 7.8% (95% CI=5.6-10.3%), respectively.

Table 33: Pooled absolute accuracy of HC2 and cytology in the triage of women with ASC-US

TEST	CIN2+			CIN3+		
	Number of studies	**Absolute sensitivity** (95% CI)	**Absolute specificity** (95% CI)	**Number of studies**	**Absolute sensitivity** (95% CI)	**Absolute specificity** (95% CI)
HC2[$]	52	94.4 (92.6-96.0)	54.9 (51.0-58.9)	24	98.1 (96.1-99.5)	50.3 (44.0-56.5)
Cytology*	16	73.4 (65.6-80.6)	63.6 (56.1-70.9)	4	83.4 (73.1-91.9)	67.3 (41.2-88.7)

[$] cut-off: 1.00 RLU/CO, * cut-off: ASC-US

Table 34: Pooled relative accuracy of HC2 versus cytology in the triage of women with ASC-US[‡‡]

Test comparison	CIN2+			CIN3+		
	Number of studies	**Relative sensitivity** (95% CI)	**Relative specificity** (95% CI)	**Number of studies**	**Relative sensitivity** (95% CI)	**Relative specificity** (95% CI)
HC2[$] vs. cytology*	10	1.27 (1.16-1.39)	0.99 (0.97-1.03)	4	1.14 (1.06-1.22)	0.99 (0.89-1.09)

[$] cut-off: 1.00 RLU/CO, * cut-off: ASC-US

[‡‡] No new studies comparing triage with HC2 and repeat cytology were identified than those included in a recent Cochrane review(28). Therefore the results are derived from the Cochrane review where a binormal model for pooling of accuracy data was used.

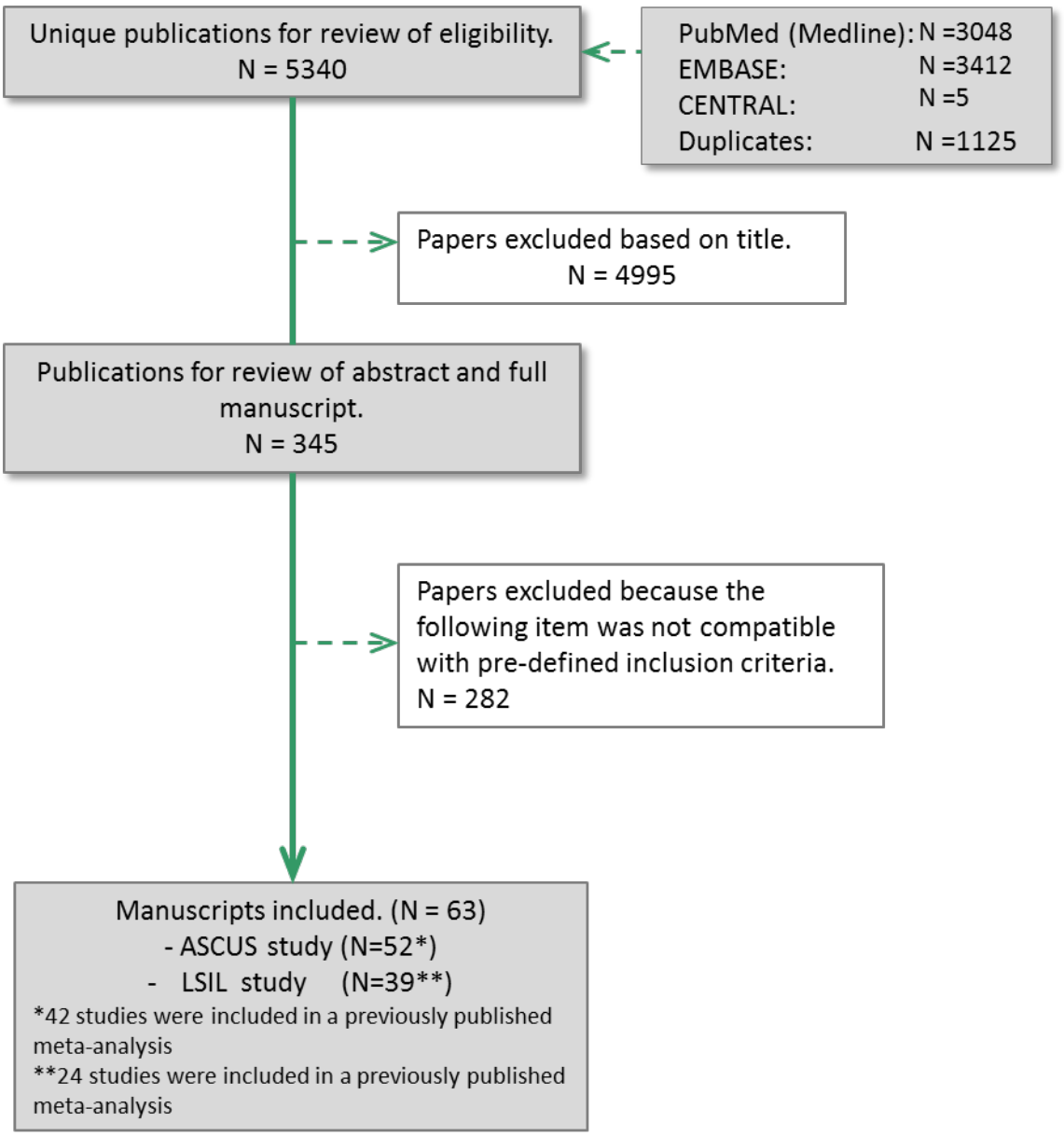


Figure 42: PRISMA flow chart for the retrieval of triage of women with ASCUS/LSIL with HC2 studies

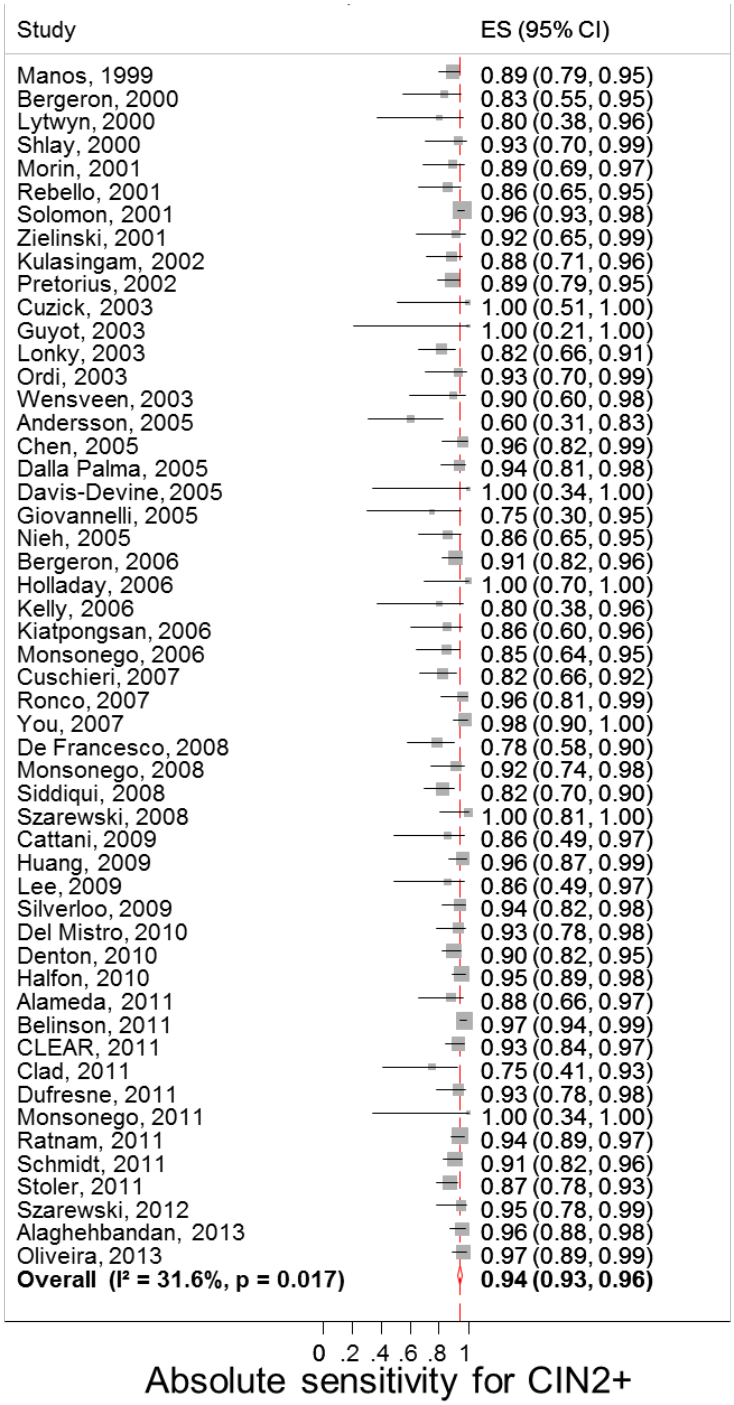


Figure 43. Absolute sensitivity of HC2 (cut-off: RLU>1) to detect CIN2+ in triage of ASC-US.[§§]

[§§] Data from the CLEAR trial were downloaded from the file submitted by the manufacturer for FDA approval: Gen-Probe. APTIMA HPV Assay. Available from: http://www.accessdata.fda.gov/cdrh_docs/pdf10/P100042c.pdf. 2011:-502170 rev. A. 502170 rev. A. Accessed November 30, 2011.

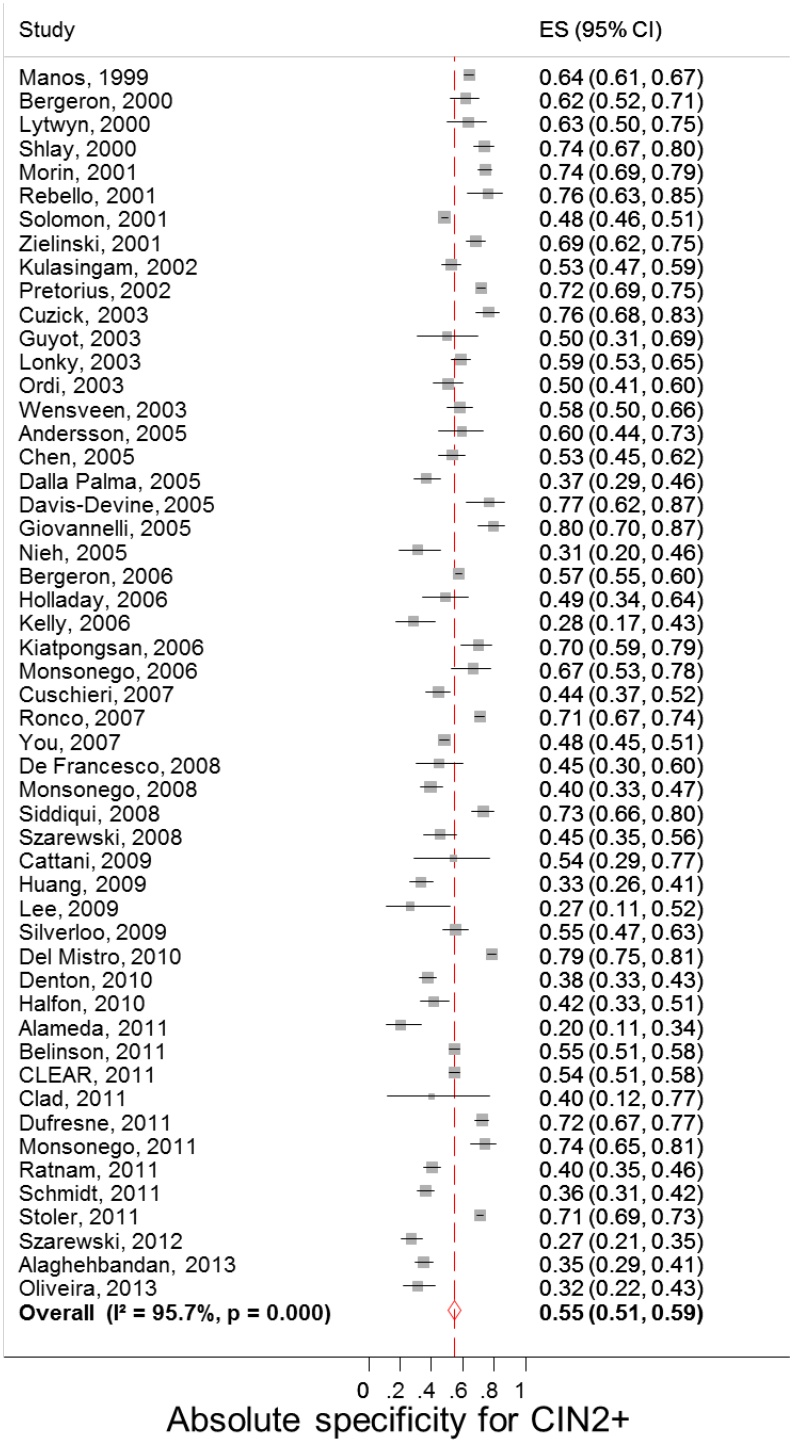


Figure 44. Absolute specificity of HC2 (cut-off: RLU>1) to detect CIN2+ in triage of ASC-US.

The absolute and relative accuracy to detect CIN2+ and CIN3+ of different HPV tests was pooled and compared to HC2 (Table 35 and Table 36). A list of included studies is shown in the appendix (see 5.1).

To detect CIN2+, *Linear Array, Cobas-4800, and ProExC* were found to have similar accuracy as HC2. The pooled sensitivity of the 5-type mRNA-assay, *Pretect HPV Proofer*, was significantly lower than that of HC2 (sensitivity ratio: 0.83 [95% CI=0.77-0.89]), but a significantly improved specificity was observed (specificity ratio: 1.95 [95% CI=1.68-2.25]). In the same way, *type-specific testing for HPV16 or HPV16/18* was less sensitive (ratio: 0.58 [95% CI=0.51-0.66] and 0.65 [95% CI=0.57-0.74], respectively) but more specific (ratio: 2.04 [95% CI=1.70-2.44] and 1.92 [95% CI=1.61-2.29], respectively) than testing for 13 hrHPV types with HC2.
Similar sensitivity combined with significantly improved specificity was also observed for *Abbott RT-PCR* (specificity ratio: 1.18 [95% CI=1.00-1.03]), *Papillocheck* (specificity ratio: 1.44 [95% CI=1.10-1.89]), *Cervista* (specificity ratio: 1.15 [95% CI=1.06-1.24]), *APTIMA* (specificity ratio: 1.15 [95% CI=1.10-1.21]), and *p16^{INK4a}* (specificity ratio: 1.80 [95% CI=1.38-2.34]). *Amplicor* was observed to have a similar sensitivity (ratio: 0.98 [95% CI=0.92-1.05]) but a significantly lower specificity (ratio: 0.87 [95% CI=0.79-0.95]).

Table 35. Pooled absolute and relative (compared to HC2) sensitivity and specificity to detect CIN2+ in the triage of ASC-US[***].

	Absolute accuracy			Relative accuracy		
Test	**Number of studies**	**Pooled sensitivity**	**Pooled specificity**	**Number of studies**	**Pooled sensitivity**	**Pooled specificity**
HC2	52	94.4 (92.6-96.0)	54.9 (51.0-58.9)	-	-	-
Amplicor	6	91.1 (86.4-95.0)	47.2 (37.3-57.3)	4	0.98 (0.92-1.05)	0.87 (0.79-0.95)
Abbott PCR	4~	94.6 (89.4-98.3)	42.2 (34.8-49.9)	4	0.98 (0.93-1.03)^	1.18 (1.00-1.03)
Linear Array	11	94.8 (90.5-98.0)	42.4 (34.2-50.7)	5	1.02 (0.99-1.06)	0.90 (0.79-1.03)
Papillocheck[¥]	1	96.1 (92.4-99.9)	60.2 (51.0-69.4)	1	1.01 (0.95-1.07)	1.44 (1.10-1.89)
Cervista[¥]	3	95.9 (93.1-98.7)	49.6 (34.6-64.6)	1	0.98 (0.95-1.02)	1.15 (1.06-1.24)
Cobas-4800	3	91.2 (85.4-95.8)	57.8 (40.4-74.3)	2	1.02 (0.94-1.11)	1.06 (0.85-1.32)

[***] Forest plots for the different tests can be provided upon request from the Unit Cancer Epidemiology, Scientific Institute of Public Health, Brussels.

Pretect	11	75.5 (66.4-83.6)	76.6	(66.0-85.9)	6	0.83	(0.77-0.89)	1.95	(1.68-2.25)
APTIMA	10	95.4 (91.6-98.4)	58.0	(49.3-66.4)	9	1.01	(0.97-1.04)^	1.15	(1.10-1.21)
HPV16	23	53.1 (48.3-57.9)	86.2	(83.1-88.9)	12[£]	0.58	(0.51-0.66)	2.04	(1.70-2.44)
HPV16/18	25	57.7 (52.2-63.1)	82.4	(79.0-85.6)	12[£]	0.65	(0.57-0.74)	1.92	(1.61-2.29)
ProExC	5	89.8 (68.2-100)	80.5	(73.9-86.3)	3	0.97	(0.66-1.42)	1.77	(0.92-3.42)
p16[INK4a]	19[$]	84.7 (77.7-90.9)	67.6	(59.2-75.4)	8	0.98	(0.91-1.06)^	1.80	(1.38-2.34)
P16/Ki-67	4	80.4 (59.7-95.5)	65.7	(51.5-78.7)	1	1.01	(0.92-1.12)	2.22	(1.89-2.62)

~The study of Wong et al., 2011(34) was excluded due to its outlying low specificity (20.8% for the Abbott PCR and 12.5% for HC2). Inclusion of Wong et al., 2011 resulted in a pooled sensitivity and specificity of 95.4% (95% CI=90.6-98.8) and 38.8% (95% CI=29.1-48.0), respectively.

[¥]No new studies were identified than those included in a recent meta-analysis(27). Therefore the results are derived from the review where a binormal model for pooling of accuracy data was used.

[£]12test comparisons from 7 studies.

[$]Denton et al. (2010)(35) reported the results of 3 independent p16[INK4a] tests (2 performed by 2 different pathologists and 1 performed by a cytotechnologist).

^continuity correction performed for studies with relative accuracy of 100% over 100%

Table 36. Pooled absolute and relative (compared to HC2) sensitivity and specificity to detect CIN3+ in the triage of ASC-US.

	Absolute accuracy			Relative accuracy		
Test	**Number of studies**	**Pooled sensitivity**	**Pooled specificity**	**Number of studies**	**Pooled sensitivity**	**Pooled specificity**
HC2	24	98.1 (96.1-99.5)	50.3 (44.0-56.5)	-	-	-
Amplicor	3	93.1 (85.6-98.3)	37.8 (26.8-49.6)	3	1.00 (0.92-1.09)^	0.86 (0.69-1.06)
Abbott PCR	4~	96.3 (93.4-98.6)	29.4 (22.0-37.4)	3	0.99 (0.92-1.06)^	1.17 (0.97-1.41)
Linear Array	7	99.2 (97.4-100)	39.3 (30.8-48.1)	4	1.03 (0.98-1.09)^	0.91 (0.79-1.06)
Papillocheck	-	-	-	-	-	-
Cervista	2	98.0 (94.6-99.9)	50.7 (35.7-65.6)	-	-	-
Cobas-4800	2	96.1 (88.8-100)	51.6 (18.2-84.2)	2	1.01 (0.93-1.11)^	1.06 (0.88-1.28)
Pretect	8	86.0 (75.1-94.6)	75.2 (61.2-86.9)	4	0.88 (0.81-0.96)	2.08 (1.77-2.46)
APTIMA	11	98.9 (96.1-100)	53.7 (44.9-62.4)	9	1.00 (0.96-1.04)^	1.15 (1.10-1.21)
HPV16	18	65.2 (58.1-72.1)	82.0 (77.6-86.1)	9[£]	0.72 (0.62-0.84)	2.13 (1.58-2.88)
HPV16/18	19	71.5 (63.5-	78.5 (74.0-	9[£]	0.76 (0.66-	2.06 (1.55-

		79.0)	82.6)		0.88)	2.74)
ProExC	3	100 (95.5-100)	78.0 (58.6-92.7)	1	1.00 (0.87-1.15)^	0.95 (0.81-1.11)
p16^{INK4a}	10[$]	85.3 (76.6-92.6)	62.2 (59.2-65.2)	4	1.00 (0.91-1.10)^	1.71 (1.36-2.15)
P16/Ki-67	1	75.0 (30.1-95.4)	61.1 (51.9-69.6)	-	-	-

~The study of Wong et al., 2011(34) was excluded due to its low specificity. Inclusion of Wong et al., 2011 resulted in a sensitivity and specificity of 98.9% (95% CI=92.3-100) and 37.4% (95% CI=29.4-45.7), respectively.

[£]9 test comparisons from 5 studies.

[$]Denton et al. (2010)(35) reported the results of 3 independent p16^{INK4a} tests (2 performed by 2 different pathologists and 1 performed by a cytotechnologist).

^continuity correction performed for studies with relative accuracy of 100% over 100%

4.3.2. Triage of LSIL

A previously published meta-analysis on the accuracy of HC2 in triage of LSIL(27), was updated with 15 new studies (see section 5.2). Overall, based on 39 studies the pooled absolute sensitivity of HC2 to triage women with LSIL was 98.0% (95% CI=96.8-99.1%) to detect CIN2+, and the specificity was 26.9% (95% CI=23.6-30.4). To detect CIN3+, the sensitivity reached 100% (95% CI=99.5-100%) after pooling the data of 20 studies, but the specificity was merely 24.7% (95% CI=20.4-29.3%) (Table 37, Figure 44 and Figure 45). Compared to cytology, HC2 demonstrated a 23% increase in sensitivity (ratio: 1.23 [95% CI=1.06-1.43]), but a 34% loss in specificity (ratio: 0.66 [95% CI=0.58-0.75]) to detect CIN2+ (Table 38).

Overall, the prevalence of CIN2+ and CIN3+ among women with LSIL was 21.1% (95% CI=18.0-24.5%) and 8.6% (95% CI=6.2-11.5%), respectively.

Table 37. Pooled absolute accuracy of HC2 and cytology to triage women with LSIL

TEST	CIN2+			CIN3+		
	Number of studies	**Absolute sensitivity** (95% CI)	**Absolute specificity** (95% CI)	**Number of studies**	**Absolute sensitivity** (95% CI)	**Absolute specificity** (95% CI)
HC2[$]	39	98.0 (96.8-99.1)	26.9 (23.6-30.4)	20	100 (99.5-100)	24.7 (20.4-29.3)
Cytology*	9	84.9 (70.4-95.7)	46.7 (32.0-61.8)	4	81.7 (65.1-94.3)	46.7 (32.0-61.8)

[$] cut-off: 1.00 RLU, * cut-off: ASC-US

Table 38. Pooled relative accuracy of HC2 versus cytology to triage women with LSIL[†††]

Test comparison	CIN2+			CIN3+		
	Number of studies	**Relative sensitivity** (95% CI)	**Relative specificity** (95% CI)	**Number of studies**	**Relative sensitivity** (95% CI)	**Relative specificity** (95% CI)
HC2[$] vs. cytology*	6	1.23 (1.06-1.43)	0.66 (0.58-0.75)	4	1.15 (0.89-1.38)	0.56 (0.37-0.84)

[$] cut-off: 1.00 RLU, * cut-off: ASC-US

[†††] No new studies comparing triage with HC2 and repeat cytology were identified than those included in a recent Cochrane review(28). Therefore the results are derived from the Cochrane review where a binormal model for pooling of accuracy data was used.

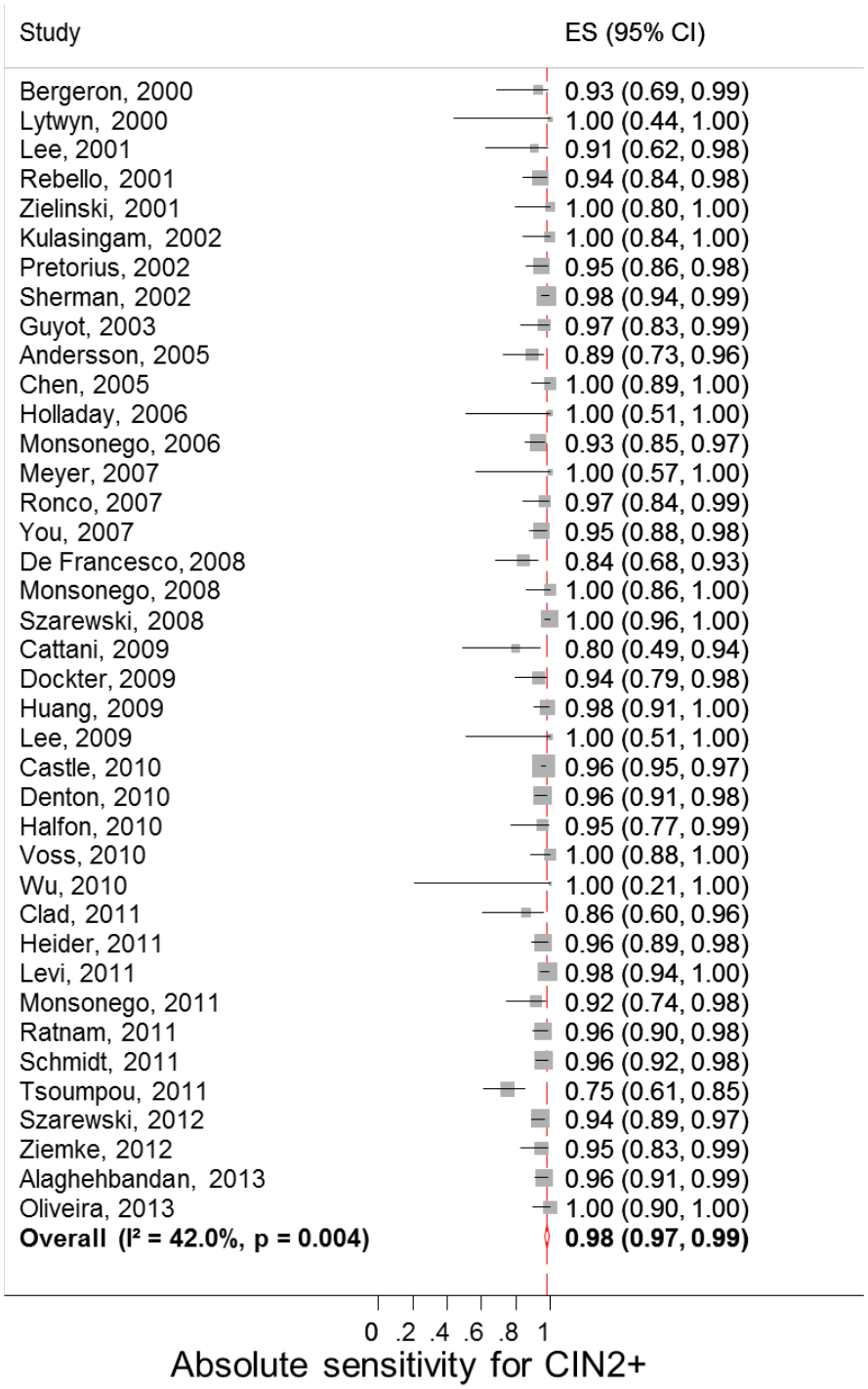


Figure 45. Absolute sensitivity of HC2 (cut-off: 1.00 RLU/CO) to detect CIN2+ in triage of LSIL.

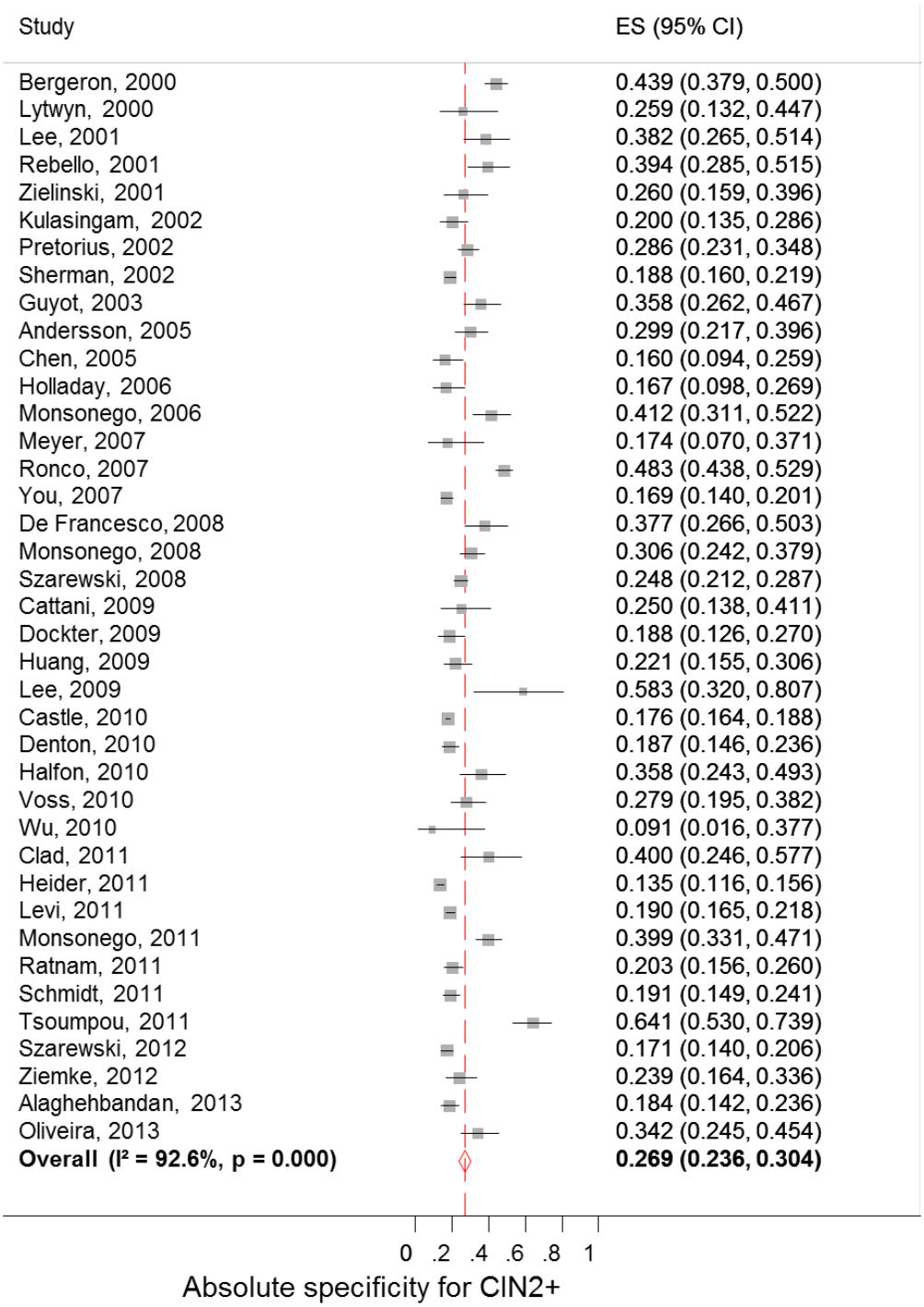


Figure 46. Absolute specificity of HC2 (cut-off: 1.00 RLU/CO) to detect CIN2+ in triage of LSIL.

The absolute and relative accuracy to detect CIN2+ and CIN3+ of different HPV tests and biomarkers was pooled and compared to HC2 (Table 39, Table 40). To detect CIN2+, the accuracy of *Linear Array, Papillocheck* and *Cobas-4800* was similar compared to the accuracy of HC2. The two mRNA-based assays both had a significantly improved specificity (ratio: 2.90 [95% CI=2.25-3.74] for Pretect HPV-Proofer, and 1.55 [95% Ci=1.39-1.73] for APTIMA). However, *Pretect HPV-Proofer*, which detects 5 types, had a significantly lower sensitivity (ratio: 0.74 [95% CI=0.69-0.78]) , while the 14-type *APTIMA* assay had a similar sensitivity (ratio: 0.97 [95% CI=0.92-1.02]) compared to HC2. A loss in sensitivity linked with an improved specificity compared to HC2 was also observed for *ProExC* (sensitivity ratio: 0.70 [95% CI=0.61-0.80] , specificity ratio: 3.26 [95% CI=2.47-4.29]), *p16^{INK4a}* (sensitivity ratio: 0.81 [95% CI=0.69-0.95] , specificity ratio: 2.21 [95% CI=1.36-3.59]), and *type-specific testing for HPV16* (sensitivity ratio: 0.57 [95% CI=0.54-0.61] , specificity ratio: 3.07 [95% CI=2.56-3.68]) and *HPV1618* (sensitivity ratio: 0.62 [95% CI=0.57-0.66] , specificity ratio: 2.85 [95% CI=2.38-3.41]). Significantly improved specificity which was not linked to a loss in sensitivity for *Abbott RT-PCR* (sensitivity ratio: 0.97 [95% CI=0.94-1.00] , specificity ratio: 1.34 [95% CI=1.10-1.62]) and *double staining with p16 and Ki-67* (sensitivity ratio: 0.94 [95% CI=0.83-1.06] , specificity ratio: 3.43 [95% CI=2.77-4.24]).

Table 39. Pooled absolute and relative (compared to HC2) sensitivity and specificity to detect CIN2+ in the triage of LSIL[‡‡‡].

Test	Absolute accuracy			Relative accuracy		
	Number of studies	Pooled sensitivity	Pooled specificity	Number of studies	Pooled relative sensitivity	Pooled relative specificity
HC2	39	98.0 (96.8-99.1)	26.9 (23.6-30.4)	-	-	-
Amplicor	4	94.1 (82.8-99.9)	29.5 (20.3-39.5)	2	0.98 (0.90-1.07)	0.80 (0.65-0.98)
Abbott PCR	4	96.3 (93.4-98.6)	29.4 (22.0-37.4)	4	0.97 (0.94-1.00)	1.34 (1.10-1.62)
Linear Array	8	99.3 (97.3-100)	26.4 (21.3-31.7)	5	1.00 (0.97-1.02)^	1.03 (0.84-1.25)
Papillocheck¥	2	94.6 (79.4-100)	32.1 (19.9-44.4)	2	1.05 (0.93-1.19)	0.90 (0.54-1.51)
Cervista¥	1	96.8 (94.0-99.6)	46.7 (40.6-52.8)	-	-	-
Cobas-4800	2	93.9 (89.4-97.3)	26.6 (17.1-37.3)	1	1.02 (0.96-1.07)	1.26 (0.98-1.62)
Pretect	10	70.1 (63.4-76.4)	73.3 (68.3-78.0)	6	0.74 (0.69-0.78)	2.90 (2.25-3.74)
APTIMA	8	96.3 (92.2-	39.6 (31.4-	7	0.97 (0.92-	1.55 (1.39-

‡‡‡ Forest plots for the different tests can be provided upon request from the Unit Cancer Epidemiology, Scientific Institute of Public Health, Brussels.

		99.2)		48.2)			1.02)		1.73)	
HPV16	22	59.8	(46.2-	82.6	(79.6-	13*	0.57	(0.54-	3.07	(2.56-
		53.5)		85.5)			0.61)		3.68)	
HPV16/18		55.1	(51.5-	78.4	(74.6-	13*	0.62	(0.57-	2.85	(2.38-
	21	58.7)		82.0)			0.66)		3.41)	
ProExC	4	68.6	(60.7-	70.4	(61.2-	1	0.70	(0.61-	3.26	(2.47-
		76.1)		78.9)			0.80)		4.29)	
p16INK4a	15$	81.3	(72.5-	59.7	(48.3-	7	0.81	(0.69-	2.21	(1.36-
		88.9)		70.5)			0.95)		3.59)	
p16INK4a/Ki67	5	90.8	(84.4-	56.5	(44.6-	2	0.94	(0.83-	3.43	(2.77-
		95.8)		68.0)			1.06)		4.24)	

¥No new studies were identified than those included in a recent meta-analysis, and therefor results are dervived from this review(27).

*13 test comparisons from 6 studies.

$Denton et al. (2010)(35) reported the results of 3 independent p16INK4a tests (2 performed by 2 different pathologists and 1 performed by a cytotechnologist).

^ continuity correction performed for studies with relative accuracy of 100% over 100%

Table 40. Pooled absolute and relative (compared to HC2) sensitivity and specificity to detect CIN3+ in the triage of LSIL.

Test	Absolute accuracy			Relative accuracy		
	Number of studies	Pooled sensitivity	Pooled specificity	Number of studies	Pooled sensitivity	Pooled specificity
HC2	20	100 (99.5-100)	24.7 (20.4-29.3)	-	-	-
Amplicor	2	97.3 (76.9-100)	22.3 (12.5-33.9)	2	1.00 (0.96-1.04)^	0.80 (0.65-0.99)
Abbott PCR	3	99.8 (96.7-100)	23.5 (16.8-31.0)	3	1.00 (0.97-1.03)^	1.33 (1.07-1.65)
Linear Array	5	100 (100-100)	22.0 (15.1-29.9)	4	1.01 (0.98-1.04)^	1.05 (0.84-1.33)
Papillocheck	-	-	-	-	-	-
Cervista	1	97.2 (92.1-99.0)	39.9 (34.5-45.4)	-	-	-
Cobas-4800	2	97.4 (84.0-100)	24.7 (14.9-36.1)	1	1.00 (0.96-1.04)^	1.21 (0.95-1.56)
Pretect	7	78.8 (68.6-87.6)	69.3 (64.2-74.3)	4	0.81 (0.75-0.88)	3.27 (2.59-4.13)
APTIMA	7	100 (99.2-100)	36.0 (28.5-43.8)	7	0.99 (0.97-1.02)^	1.47 (1.26-1.70)
HPV16	17	63.3 (58.0-68.5)	78.8 (75.3-82.0)	8*	0.67 (0.62-0.73)	3.61 (3.00-4.35)
HPV16/18	18	69.7 (63.0-76.1)	76.0 (71.9-79.9)	8*	0.70 (0.65-0.75)	3.33 (2.78-4.00)
ProExC	2	91.4 (63.0-100)	68.8 (63.0-74.2)	-	-	-
p16INK4a	8$	91.2 (83.4-97.1)	47.5 (38.4-56.7)	3	0.90 (0.79-1.02)	2.72 (2.17-3.42)

p16^{INK4a}/Ki-67	2	97.6 (90.7-100)	41.1 (27.1-55.9)	-	-	-

*8 test comparisons from 46 studies.
^continuity correction performed for studies with relative accuracy of 100% over 100%
$Denton et al. (2010)(35) reported the results of 3 independent p16^{INK4a} tests (2 performed by 2 different pathologists and 1 performed by a cytotechnologist).

4.3.2.1. Triage of ASC-H

The absolute sensitivity and specificity of HC2 in triage of women with ASC- The absolute sensitivity and specificity of HC2 in triage of women with ASC-H, derived from 19 studies (see Figure 46 for literature retrieval history and section 4.5.3), considering CIN2+ as outcome, was 93% (95% CI: 89-95) and 45% (95% CI: 41-50%), respectively (Figure 47). Merely 1 study contained data on repeat cytology at cut-off ASC-US, demonstrating a sensitivity and specificity of 38.5% (95% CI=17.7-64.5%) and 100% (95% CI=82.4-100%)[36]. Overall, 67% (95% CI=63-72%) had a positive HC2 test and 34% (95% CI: 28-40%) contained CIN2+.
The accuracy for CIN3+, was described in five studies, showing a sensitivity of 91% (95% CI: 81-96%) and a specificity of 42% (95% CI: 34-51%)[36-41].

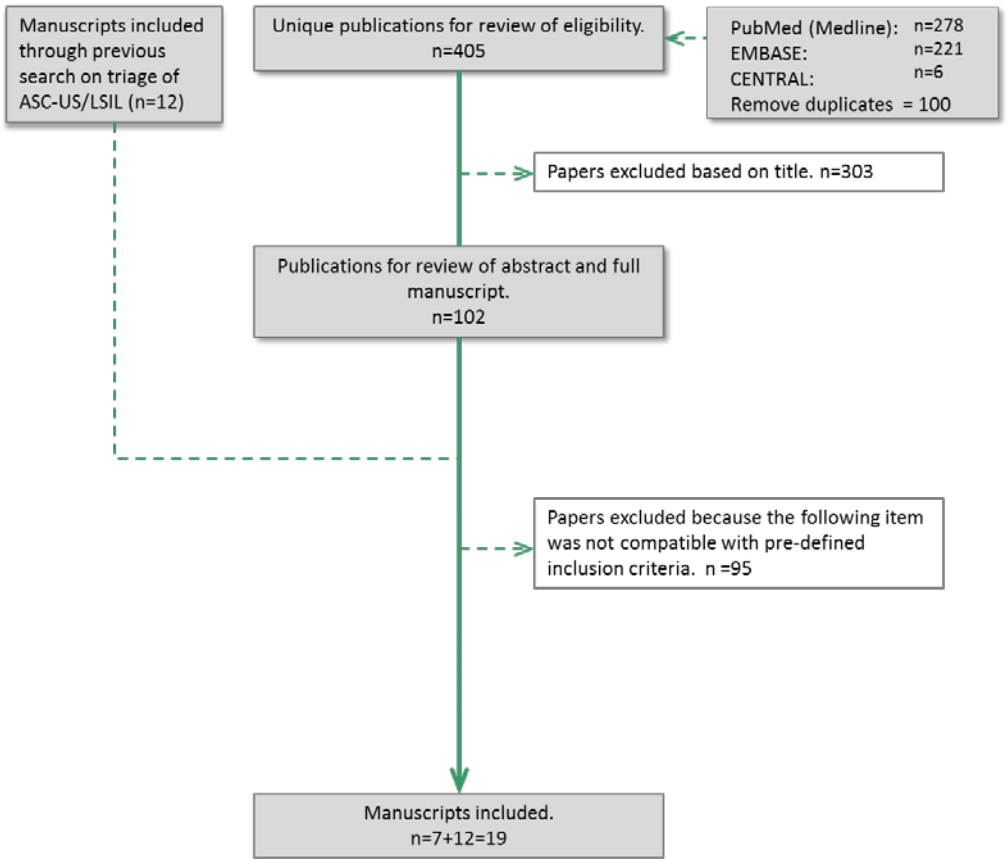


Figure 47. PRISMA flow chart for the retrieval of ASC-H studies

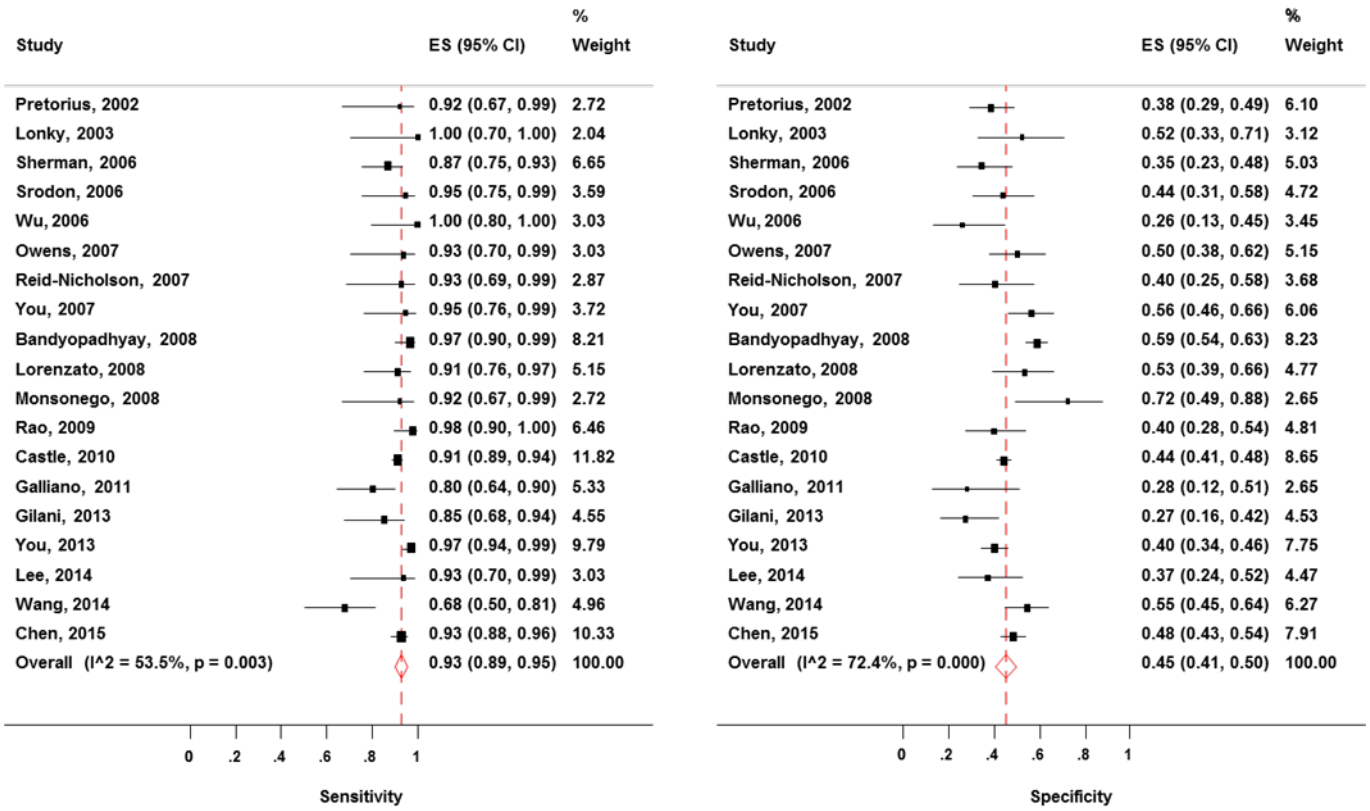


Figure 48: Absolute sensitivity (left) and specificity (right) of HC2 (cut-off: 1.00 RLU/CO) to detect CIN2+ in triage of ASC-H.

4.3.2.2. Triage of AGC

Using HC2 in the triage of women with AGC[§§§], pooled from 12 studies (see Figure 48 for literature retrieval history and section 5.4), showed high sensitivity (90.0%, 95% CI= 85.1-93.4%) for identifying underlying squamous CIN2+ or glandular AIS+[****] (Figure 49). The hrHPV-positivity (39.8%, 95% CI=32.2-47.5%) was lower than in the other categories of minor cytological abnormality, which corresponds with rather good specificity (75.1%, 95% CI= 64.8-83.2%). Accuracy, in particular the specificity, was extremely heterogeneous, which may be due to inclusion of various age groups, restriction/extension to different sub-categories of AGC and variability in the case-mix of AGC patients with different degrees of follow-up or disease verification.

One study (Castle,2010) provided age-strafied data and found that HPV+ women have 10% risk of cervical cancer but 0% of endometrial cancer, whereas hrHPV-negative women had a 11% risk of cervical cancer.

Only one study contained data on repeat cytology at cut-off ASC-US/AGC+ and demonstrated a sensitivity and specificity of 62.5% (95% CI=38.6-81.5%) and 71.2% (95% CI=62.5-78.6%), respectively[42].

[§§§] Only AGC of endocervical or unknown origin was included in the meta-analysis. Endometrial atypia was excluded. In principle, glandular intraepithelial neoplasia, endo-cervical adenocarcinima in situ and invasive endo-cervical adenocarcinima were considered as glandular disease outcomes.

[****] Adenocarcinoma in situ, or worse

The proportion of underlying or incipient CIN2+ or glandular AIS+ among women with AGC was 55.7% (95% CI=14.0-93.2%).

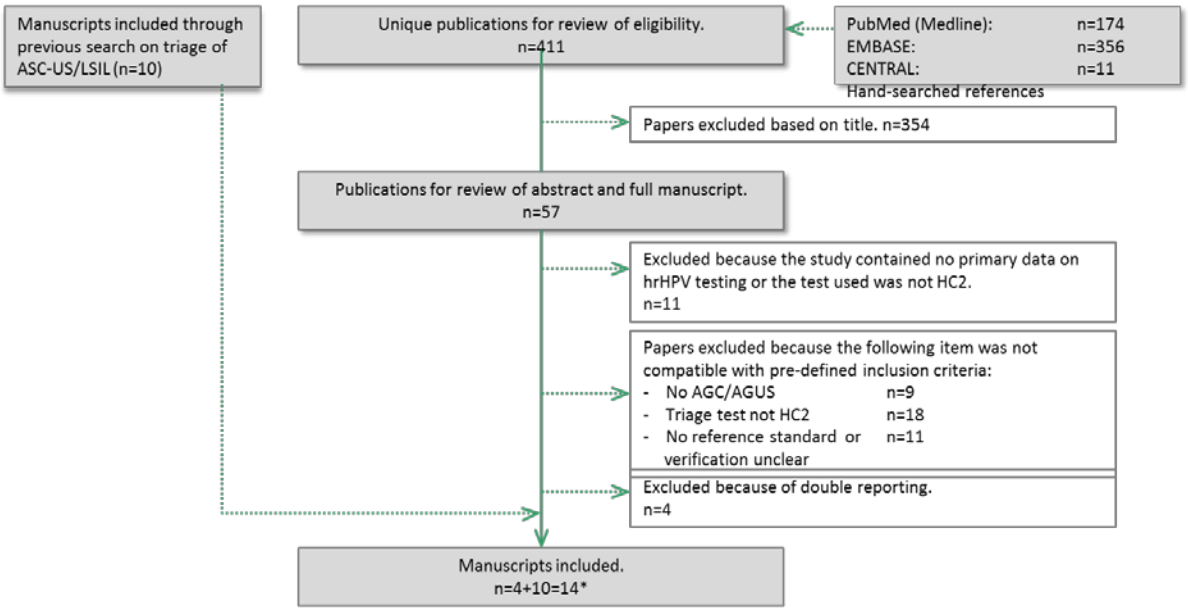


Figure 49. PRISMA flow chart for the retrieval of AGC studies

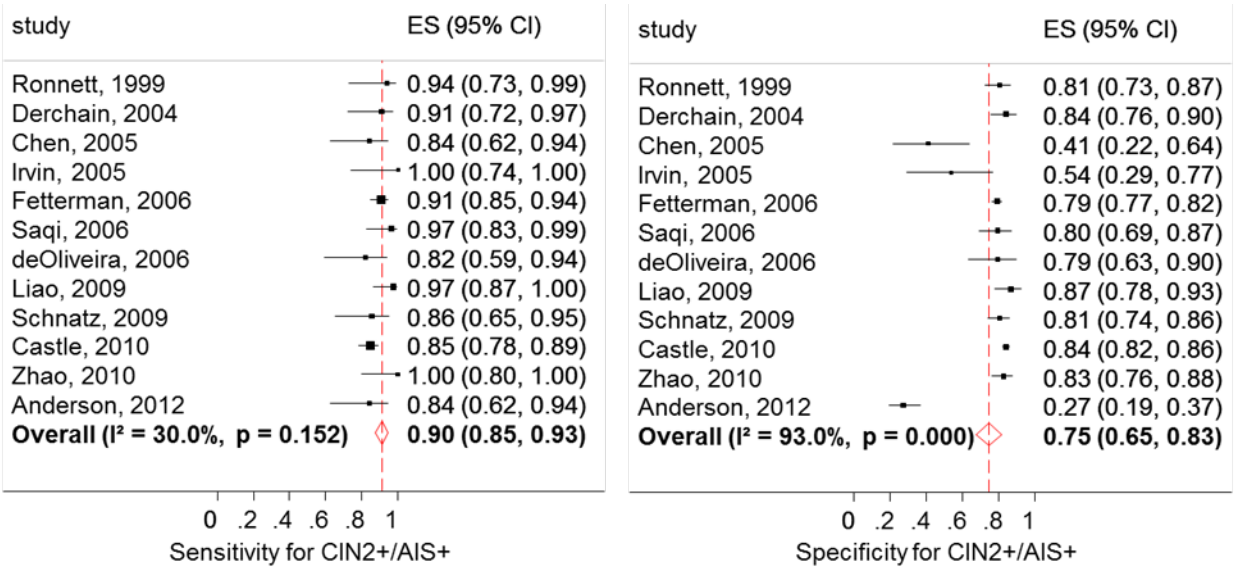


Figure 50: Absolute sensitivity (left) and specificity (right) of HC2 (cut-off: 1.00 RLU/CO) to detect CIN2+ in triage of AGC.

4.3.3. Requirements for a test usable in triage of minor cytological lesions

Preferentially, triage of screen-positive women should be performed with clinically validated tests. Experts agree on equivalency principles of minimal diagnostic test accuracy for HPV tests usable in primary screening where

HC2 is accepted as a reference comparator(see guideline of Meijer et al, Int J Cancer 2009)(42). Consistent evidence underpins the clinical utility of HC2 also in triage of women with equivocal squamous abnormalities. Therefore proven equivalency of an HPV assay relative to HC2 in screening, can be considered for use in triage of ASC-US However, specific evidence regarding the diagnostic accuracy of a new triage method to predict presence of underlying or incipient squamous or glandular cervical precancer in women with minor cytological abnormalities is preferable.

4.3.3.1. hrHPV testing in triage of ASC-US

In ASC-US triage, a new triage test should be at least as sensitive and, but preferably more specific than HC2.

→ assessing the relative accuracy of the new test versus HC2, to detect CIN2+:

- relative sensitivity: lower 95% CI bound ≥ 0.90
- relative specificity: lower 95% CI bound ≥ 0.95

Following this set of requirements the following tests are considered valid: Abbott RT-PCR, Papillocheck, Cervista, APTIMA, $p16^{INK4a}$, p16/Ki-67. It should be noted that the relative accuracy requirements for Papillocheck and Cervista are based on only one study. The Cobas-4800 assay demonstrates similar pooled sensitivity and specificity as HC2 but the precision of the specificity is not sufficient (lower CIB at 0.86).

Table 41. Pooled relative accuracy of hrHPV tests and biomarkers versus HC2 (cut-off: 1.00 RLU/CO) considering CIN2+. These tests fulfill the requirements of a triage test for women with ASC-US.

Test	Relative sensitivity	Relative specificity
Abbott PCR	0.98 (0.93-1.03)^	1.18 (1.00-1.03)
Papillocheck¥	1.01 (0.95-1.07)	1.44 (1.10-1.89)
Cervista¥	0.98 (0.95-1.02)	1.15 (1.06-1.24)
APTIMA	1.01 (0.97-1.04)^	1.15 (1.10-1.21)
$p16^{INK4a}$	0.98 (0.91-1.06)^	1.80 (1.38-2.34)
P16/Ki-67	1.01 (0.92-1.12)	2.22 (1.89-2.62)

4.3.3.2. hrHPV testing in triage of LSIL

For LSIL triage, the new test should be at least as sensitive but necessarily more specific than HC2. The requirement for higher specificity is indicated because of the very high positivity rate, and low specificity, of the latter test, in LSIL.

→ assessing the relative accuracy of the new test versus HC2, to detect CIN2+:

- relative sensitivity: lower 95% CI bound ≥ 0.90
- relative specificity: measure ≥ 1.10, and lower 95% CI bound ≥ 1.00

Following these requirements, Abbott RT-PCR and APTIMA are valid hrHPV tests in triage of women with LSIL.

Table 42. Pooled relative accuracy of hrHPV tests versus HC2 (cut-off: 1.00 RLU/CO) considering CIN2+. These tests fulfill the requirements of a triage test for women with LSIL.

Test	Relative sensitivity	Relative specificity
Abbott PCR	0.97 (0.94-1.00)	1.34 (1.10-1.62)
APTIMA	0.97 (0.92-1.02)	1.55 (1.39-1.73)

When no relative accuracy data (compared to HC2) are available, absolute accuracy verified with a good gold standard could be acceptable as well. Colposcopy and multiple biopsies of colposcopic suspicious areas and high-quality standards for colposcopy and histology, including review of biopsies are necessary criteria. Absolute accuracy data of a new HPV assay or biomarker can then be compared with the pooled absolute accuracy data for HC2 applied in triage of a given lesion in other studies. However, a lower level of evidence should be attributed if no comparator information relative to another validated triage test is available.

4.4. Interpretation

4.4.1. Triage of women with ASC-US

In the triage of ASC-US, the meta-analysis presented here, demonstrates a significantly improved sensitivity of HC2 compared to repeat cytology at cut-off ASC-US (27% and 14% increase for CIN2+ and CIN3+, respectively), which was coupled with a non-inferior specificity. These findings corroborate conclusions formulated in previous reviews which indicate a better performance of HC2 to triage women with ASC-US (improved sensitivity, similar specificity) compared to repeat cytology(26;28;43). As a result, hrHPV testing for triage of ASC-US cytology is now widely accepted in the United States and Europe.

In the report presented here, three other HPV DNA-based assays (Abbott PCR, Papillocheck, and Cervista) and one mRNA-based assay (APTIMA) were found to be appropriate triage tests for women with a diagnosis of ASC-US. The same results and conclusion were formulated in a previous meta-

analysis(27). Nonetheless, looking at the absolute accuracy of these assays, the specificity still is suboptimal (range42-60% for CIN2+, 29-54% for CIN3+), resulting in colposcopy referral for many women without disease. In comparison, triage with repeat cytology at a cut-off of ASC-US demonstrated specificity to exclude CIN2+ of 64%. However, as for HC2, the high sensitivities of Abbott PCR, Papillocheck, Cervista and APTIMA (range 95-96% for CIN2+, 96-99% for CIN3+) demonstrate the highly improved potential of hrHPV testing to identify women with cervical precancer, compared to repeat cytology at cut-off ASC-US (sensitivity 73% for CIN2+, 83% for CIN3+).
Additionally, both the single $p16^{INK4a}$ and double p16/Ki-67 immunostaining were shown to meet the criteria for better specificity without loss of sensitivity, compared to HC2.
In the past, p16 immunostaining has demonstrated promising results in terms of lowering the number of false referrals, compared to HC2(44;45). In this report and a previous meta-analysis(30), based on data from 8 studies, the specificity of $p16^{INK4a}$ to reject CIN2+ was 1.80 times higher than that of HC2, without losing sensitivity. p16/Ki-67 immunostaining demonstrated a more than doubled in specificity, which was, however, based on one study.

4.4.2. Triage of women with LSIL

In the triage of women with LSIL, HC2 demonstrated a considerable improved potential to detect CIN2+ (relative sensitivity: 1.23) compared to cytology. For CIN3+, sensitivity of HC2 was not significantly better than that of repeat cytology (cut-off ASC-US). However, a substantial loss in specificity (34% for CIN2+, 44% for CIN3+) was observed, leading to over-diagnosis and over-treatment.
A strong association has been documented between the presence of squamous intra-epithelial lesions and HPV-infection, and it is generally believed that LSIL in fact is the manifestation of a productive HPV infection with low potential of neoplastic transformation(19). In the ALTS-trial, 75-80% of women with LSIL were confirmed to harbor hrHPV types (46). In the same way, our meta-analysis demonstrated that 79.5% (95% CI=77.1-81.7%) of women with LSIL had a positive HC2 test, while underlying CIN2+ was present in 21%. Given its high positivity rate, HC2 has limited capacity to distinguish between cases with and without clinically significant disease.
Due to poor values for specificity, the American Society for Colposcopy and Cervical Pathology (ASCCP) does not recommend reflex HPV triage, but proposes to refer to colposcopy. If colposcopy and/or biopsy are normal or only reveal CIN1, an HPV test at 12 months or two repeat smears after the initial LSIL smear is recommended. In The Netherlands, Bais and Berkhof showed that delayed HPV and repeat cytology testing in patients with borderline or mild dyskaryosis after 6 and 18 months is both safe and more

cost-effective than immediate HPV triage (47;48). Postponing triage, allows viral clearance, which over a period of 6–12 months can vary from 18% to 45%(47) and therefore reduces the need for colposcopy. However, this strategy requires good compliance with follow-up recommendations.
In certain situations, the prevalence of hrHPV in LSIL is lower than average. For instance, in an Italian study, it was 55%(49) (similar to the prevalences observed in most studies among women with ASC-US). This may be due to local interpretation of cytology. In these situations, triage of LSIL with a hrHPV test may be efficient(50).

hrHPV tests, other than HC2, generally demonstrate a good sensitivity with approximately half of them meeting the requirement of a lower 95% confidence bound of at least 0.90 (relative sensitivity compared to HC2). Of the assays that met the sensitivity criterion, Abbott RT-PCR and APTIMA satisfied the requirements for specificity (relative specificity ≥1.10 and lower 95% confidence bound ≥ 1.00) as well. Testing with biomarkers or identifying mRNA of the 5 main HPV types or DNA of HPV16/18 showed substantially improved specificity but did not reach the minimal sensitivity.

4.4.3. TRIAGE OF ASC-H

hrHPV testing with HC2 shows good average sensitivity and reasonable specificity for detection of HPV-related cervical precancer or worse including glandular endocervical precancer in women with ASC-H.

4.4.4. TRIAGE OF AGC/AGUS

hrHPV testing with HC2 shows good average sensitivity and specificity for detection of HPV-related cervical precancer or worse including glandular endocervical precancer in women with cytological glandular atypia. Data on specificity were heterogeneous. In particular, in postmenopausal women, HPV testing may help in distinguishing risk for endometrial and cervical/endocervical cancer.

4.4.5. GENERAL CONCLUSION

The meta-analyses conducted in this review provide estimates of the average sensitivity and specificity of different tests which are potentially usable to triage women with minor cytological abnormalities identified through cytological cervical cancer screening. To judge on their applicability in the framework of cervical cancer screening in Germany, requires the estimation of the risk of cervical precancer or cancer in women in Germany with the

particular cytological abnormality when the test is positive or negative. A paradigm which recently was suggested in the aforementioned Cochrane Review(28) was:

1) Positive triage: a positive triage test result should be associated with an underlying risk (=PPV) of CIN3+ of >=10% or risk of CIN2+ of 20%.

2) Negative triage: the negative triage test result should be associated with an underlying risk (=1-NPV) of CIN3+ <1% or risk of CIN2+<2%.

Whether these conditions of an efficient (PPV>10%) and safe triage (1-NPV<1%) requires computing the predictive values for ranges of underlying pretest risk of precancer which are plausible for Germany.

This computation requires estimation of ranges of pretest risks (prevalence of underlying CIN2+ and CIN3+) for the different cytological which are relevant for Germany. These predictive values should be computed during the GRADE assessment.

4.5. Appendix – included studies

4.5.1. Triage ASC-US

A list of all included studies for triage of ASC-US is shown. Studies in bold, are studies that were not yet included in a previously published meta-analysis(27).

Test	Number	1st author, Journal Year; Vol: Pages
HC2		
	1	Manos, JAMA 1999; 281: 1605-1610
	2	Bergeron, Obstet Gynecol 2000; 95: 821-827
	3	Lytwyn, Can Med Ass J 2000; 19: 701-707
	4	Shlay, Gynecol Oncol 2000; 96: 410-416
	5	Morin, J Reprod Med 2001; 46: 799-805
	6	Rebello, BMJ 2001; 322: 893-894
	7	Solomon, J Natl Cancer Inst 2001; 93: 293-299
	8	Zielinski, J Pathol 2001; 195: 300-306
	9	Kulasingam, JAMA 2002; 288: 1749-1757
	10	Pretorius, J Reprod Med 2002; 47: 290-296
	11	Cuzick, Lancet 2003; 363: 1871-1876
	12	Guyot, BMC Infect Dis 2003; 3: 23-23
	13	Lonky, Obstet Gynecol 2003; 101: 481-489
	14	Ordi, Med Clin (Barc) 2003; 121: 441-445
	15	Wensveen, Acta Obstet Gynecol Scand 2003; 82: 883-889
	16	Andersson, Acta Obstet Gynecol Scand 2005; 84: 996-1000
	17	Chen , Taiw J Obstet Gynecol 2005; 44: 252-257
	18	Dalla Palma, Cytopathology 2005; 16: 22-26
	19	Davis-Devine, Am J Clin Pathol 2005; 124: 24-30
	20	Giovannelli, J Clin Virol 2005; 33: 281-286
	21	Nieh, Gynecol Oncol 2005; 97: 35-40
	22	Bergeron, Gyn Obstet Fert 2006; 34: 312-316
	23	Holladay, Cancer Cytopathol 2006; 108: 451-461
	24	Kelly, Cancer 2006; 108: 494-500
	25	Kiatpongsan, In J Gynecol Cancer 2006; 16: 262-265
	26	Monsonego, In J Gynecol Cancer 2006; 16: 591-598
	27	Cushieri, J Clin Virol 2007; 38: 14-18
	28	Ronco, Eur J Cancer 2007; 43: 476-480
	29	You, Aust NZ J Obstet Gynecol 2007; 47: 141-144
	30	De Francesco, J Virol Meth 2008; 147: 10-17
	31	Monsonego, Sex Transm Dis 2008; 35: 521-527
	32	Siddiqui, Arch Pathol Lab Med 2008; 132: 1648-1652
	33	Szarewski,CEBP 2008; 17: 3033-3043
	34	Cattani, J Clin Microbiol 2009; 47: 3895-3901
	35	Huang, J Clin Virol 2009; 45S: 19-23
	36	Lee, Int J Gynecol Cancer 2009; 19: 266-272
	37	Silverloo, Acta Obstet Gynaecol 2009; 88: 1006-1010
	38	Del Mistro, Gynecol Oncol 2010; 117: 77-88
	39	Denton, L Clin Pathol 2010; 134: 12-21
	40	**Halfon, Cancer Biomerkers 2010; 7: 133-139**

41	**Alameda, Diagn Cytopathol 2011; 39: 110-114**
42	**Belinson, Am J Clin Pathol 2011; 135: 790-795**
43	**CLEAR, http://www.accessdata.fda.gov/ 2011**
44	**Clad, J Clin Microbiol 2011; 49: 1071-1076**
45	**Dufresne, J Clin Microbiol 2011; 49: 48-53**
46	**Monsonego, Int J Cancer 2011; 129: 691-701**
47	**Ratnam, J Clin Microbiol 2011; 49: 557-564**
48	**Schmidt, Cancer Cytopathol 2011; 119: 158-166**
49	**Stoler, Am J Clin Pathol 2011; 135: 468-475**
50	**Szarewski, J Clin Microbiol 2012; 50: 1867-**
51	**Alaghehbandan, Diagn Cytopathol 2013; 41: 767-775**
52	**Oliveira, J Med Virol 2013; 85: in-press**
Amplicor	
1	Monsonego, Gynecol Oncol 2005; 99: 160-168
2	De Francesco, J Virol Meth 2008; 147: 10-17
3	Szarewski,CEBP 2008; 17: 3033-3043
4	Wentzensen, CEBP2009; 18: 1341-1349
5	Dufresne, J Clin Microbiol 2011; 49: 48-53
6	**Jakobsson, Int J STD AIDS 2012; 23: 485-489**
Abott RT PCR	
1	Szarewski,CEBP 2008; 17: 3033-3043
2	Huang, J Clin Virol 2009; 45S: 19-23
3	Halfon, J Clin Virol 2010; 48: 246-250
4	Wong, J Clin Virol 2011; 51: 136-138
5	**Szarewski, J Clin Microbiol 2012; 50: 1867-**
Linear Array	
1	Castle, J Clin Microbiol 2008; 46: 109-117
2	Fröberg, BJC 2008; 99: 563-568
3	Monsonego, Sex Transm Dis 2008; 35: 521-527
4	Szarewski,CEBP 2008; 17: 3033-3043
5	Lee, Int J Gynecol Cancer 2009; 19: 266-272
6	Halfon, J Clin Virol 2010; 47: 38-42
7	Ratnam, J Clin Microbiol 2011; 49: 557-564
8	Lapierre, J Clin Microbiol 2012; 50: 1240-1244
9	**Dona, Gynecol Oncol 2012; 126: 198-202**
10	Waldstrom, Cytopathol 2012; 23:389-395
11	**Wentzensen, Clin Cancer Res 2012; 18: 4154-4162**
PapilloCheck	
1	Halfon, J Clin Virol 2010; 47: 38-42
Cervista	
1	Wong, Cancer 2009; 115: 823-832
2	Einstein, Gynecol Oncol 2010; 118: 116-122
3	Bellinson, Am J Clin Pathol 2011; 135: 790-795
Cobas-4800	
1	Stoler, Am J Clin Pathol 2011; 135: 468-475
2	Lapierre, J Clin Microbiol 2012; 50: 1240-1244
3	**Szarewski, J Clin Microbiol 2012; 50: 1867-**
Pretect-HPV Proofer	
1	Molden, Int J Cancer 2005; 114: 973-976
2	**Andersson, Int J Oncol 2006; 29: 705-711**
3	Szarewski,CEBP 2008; 17: 3033-3043
4	Halfon, J Clin Virol 2010; 47: 177-181
5	Sorbye , J Virol Meth 2010; 169: 219-222

	6 Andersson, J Clin Microbiol 2011; 49: 3794-3799	
	7 Ratnam, J Clin Microbiol 2011; 49: 557-564	
	8 Koliopoulos, Acta Obstet Gynecol Scand 2012; 91: 794-801	
	9 Szarewski, J Clin Microbiol 2012; 50: 1867-	
	10 Alaghehbandan, Diagn Cytopathol 2013; 41: 767-775	
	11 Oliveira, J Med Virol 2013; 85: in-press	
APTIMA		
	1 Szarewski,CEBP 2008; 17: 3033-3043	
	2 Dockter, J Clin Virol 2009; 45 (S1): S55-S61	
	3 Wu, Int J Gynecol Cancer 2010; 20: 1411-1414	
	4 Clad, J Clin Microbiol 2011; 49: 1071-1076	
	5 CLEAR, http://www.accessdata.fda.gov/ 2011	
	6 Monsonego, Int J Cancer 2011; prepub: -	
	7 Ratnam, J Clin Microbiol 2011; 49: 557-564	
	8 Szarewski, J Clin Microbiol 2012; 50: 1867-	
	9 Waldstrom, Cytopathol 2012; 23:389-395	
	10 Stoler, Am J Clin Pathol 2011; 135: 468-475	
$p16^{INK4}$ [a]		
	1 Nieh, Gynecol Oncol 2005; 97: 35-40	
	2 Andersson, Int J Oncol 2006; 29: 705-711	
	3 Holladay , Cancer 2006; 108: 451-461	
	4 Meyer, Cancer 2007; 111: 83-92	
	5 Monsonego, Acta Cytol 2007; 51: 755-766	
	6 Wentzensen, Cancer 2007; 111: 58-66	
	7 Schledermann, Diagn Cytopathol 2008; 36: 453-459	
	8 Szarewski, CEBP2008; 17: 3033-3042	
	9-11 Denton, Am J Clin Pathol 2010; 134: 12-21 (3 interpretations)	
	12 Guo, Diagn Cytopathol 2010; 39: 482-488	
	13 Passamonti, Pathalogica 2010; 102: 6-11	
	14 Samarawardana, Cancer Cytopathol 2010; 118: 146-156	
	15 Sung, Diagn Cytopathol 2010; 38: 168-171	
	16 Alameda, Diagn Cytopathol 2011; 39: 119-114	
	17 Nasioutziki, Int J Gynecol Cancer 2011; 21: 79-85	
	18 Szarewski, J Clin Microbiol 2012; 50: 1867-	
	19 Loghavi, Diagn Cytopathol 2013; 41: 582-587	
P16/Ki-67		
	1 Edgerton, Diagn Cytopathol 2011; 41: 35-40	
	2 Schmidt, Cancer Cytopathol 2011; 119: 158-166	
	3 Wentzensen, Clin Cancer Res 2012; 18: 4154-4162	
	4 Loghavi, Diagn Cytopathol 2013; 41: 582-587	
HPV16 genotyping		Assay
	1 Froberg, Br J Cancer 2008; 99: 563-568	Linear array
	2 Monsonego, Int J STD AIDS 2008; 19: 385-392	Linear array
	3 Halfon, J Clin Virol 2010; 47: 177-181	Linear array
	4 **Dona, Gynecol Oncol 2012; 126: 198-202**	Linear array
	5 Lapierre, J Clin Microbiol 2012; 50: 1240-1244	Linear array
	6 **Wentzensen, Clin Cancer Res 2012; 18: 4154-4162**	Linear array
	7 **Gage, Cancer Epidemiol Biomarkers Prev 2013; 22: 1095-1101**	Linear array
	8 Guo, Mod Pathol 2008; 21: 1037-1043	Easy Chip
	9 Huang, J Clin Virol 2009; 45: S19-S23	Easy Chip

	10	Szarewski,CEBP 2008; 17: 3033-3043	Abbott
	11	Halfon, J Clin Virol 2010; 47: 177-181	Abbott
	12	**Szarewski, J Clin Microbiol 2012; 50: 1867-**	Abbott
	13	Szarewski,CEBP 2008; 17: 3033-3043	Clinical Arrays
	14	Halfon, J Clin Virol 2010; 47: 177-181	PapilloCheck
	15	Wong, Cancer 2009; 115: 823-832	Cervista
	16	**Einstein, Gynecol Oncol 2010; 118: 116-122**	Cervista
	17	**Belinson, Am J Clin Pathol 2011; 135: 790-795**	MALDI-TOF
	18	**Depuydt, Cancer Epidemiol Biomarkers Prev 2011; 20: 628-637**	E6/7 qPCR
	19	Stoler, Am J Clin Pathol 2011; 135: 468-475	COBAS-4800
	20	Lapierre, J Clin Microbiol 2012; 50: 1240-1244	COBAS-4800
	21	**Szarewski, J Clin Microbiol 2012; 50: 1867-**	COBAS-4800
	22	**Szarewski, J Clin Microbiol 2012; 50: 1867-**	BD Viper
	23	**Oliveira, J Med Virol 2013; 85: in-press**	CLART
HPV16/18 genotyping			
	1	Froberg, Br J Cancer 2008; 99: 563-568	Linear array
	2	Monsonego, Int J STD AIDS 2008; 19: 385-392	Linear array
	3	Halfon, J Clin Virol 2010; 47: 177-181	Linear array
	4	**Dona, Gynecol Oncol 2012; 126: 198-202**	Linear array
	5	Lapierre, J Clin Microbiol 2012; 50: 1240-1244	Linear array
	6	**Wentzensen, Clin Cancer Res 2012; 18: 4154-4162**	Linear array
	7	**Gage, Cancer Epidemiol Biomarkers Prev 2013; 22: 1095-1101**	Linear array
	8	Guo, Mod Pathol 2008; 21: 1037-1043	Easy Chip
	9	Huang, J Clin Virol 2009; 45: S19-S23	Easy Chip
	10	Szarewski,CEBP 2008; 17: 3033-3043	Abbott
	11	Halfon, J Clin Virol 2010; 47: 177-181	Abbott
	12	Wong, J Clin Virol 2011; 51: 136-138	Abbott
	13	**Szarewski, J Clin Microbiol 2012; 50: 1867-**	Abbott
	14	Szarewski,CEBP 2008; 17: 3033-3043	Clinical Arrays
	15	Halfon, J Clin Virol 2010; 47: 177-181	PapilloCheck
	16	Wong, Cancer 2009; 115: 823-832	Cervista
	17	**Einstein, Gynecol Oncol 2010; 118: 116-122**	Cervista
	18	**Belinson, Am J Clin Pathol 2011; 135: 790-795**	MALDI-TOF
	19	**Depuydt, Cancer Epidemiol Biomarkers Prev 2011; 20: 628-637**	E6/7 qPCR
	20	Stoler, Am J Clin Pathol 2011; 135: 468-475	COBAS-4800
	21	Lapierre, J Clin Microbiol 2012; 50: 1240-1244	COBAS-4800
	22	**Szarewski, J Clin Microbiol 2012; 50: 1867-**	COBAS-4800
	23	**Spathis, Plos one 2012; 7: e49205-**	CLART
	24	**Oliveira, J Med Virol 2013; 85: in-press**	CLART
	25	**Szarewski, J Clin Microbiol 2012; 50: 1867-**	BD Viper
ProExC			
	1	**Kelly, Cancer Cytopathol 2006; 108: 494-500**	
	2	**Siddiqui, Arch Pathol Lab Med 2008; 132: 1648-1652**	
	3	**Tambouret, Arch Pathol Lab Med 2008; 132: 918-925**	
	4	**Depuydt, Cancer Epidemiol Biomarkers Prev 2011; 20: 628-637**	
	5	**Alaghehbandan, Diagn Cytopathol 2013; 41: 767-775**	

4.5.2. Triage LSIL

A list of all included studies for triage of LSIL is shown. Studies in bold, are studies that were not yet included in a previously published meta-analysis(27).

Test	Number	1st author, Journal Year; Vol: Pages
HC2		
	1	Bergeron, Obstet Gynecol 2000; 95: 821-827
	2	Lytwyn, Can Med Ass J 2000; 19: 701-707
	3	**Lee, Arch Pathol Lab Med 2001; 125: 1453-1457**
	4	Rebello, BMJ 2001; 322: 893-894
	5	Zielinski, J Pathol 2001; 195: 300-306
	6	Kulasingam, JAMA 2002; 288: 1749-1757
	7	Pretorius, J Reprod Med 2002; 47: 290-296
	8	Sherman, J Natl Cancer Inst 2002; 94: 102-107
	9	Guyot, BMC Infect Dis 2003; 3: 23-23
	10	Andersson, Acta Obstet Gynecol Scand 2005; 84: 996-1000
	11	Chen , Taiw J Obstet Gynecol 2005; 44: 252-257
	12	Holladay, Cancer Cytopathol 2006; 108: 451-461
	13	Monsonego, In J Gynecol Cancer 2006; 16: 591-598
	14	**Meyer, Cancer 2007; 111: 83-92**
	15	Ronco, Eur J Cancer 2007; 43: 476-480
	16	You, Aust NZ J Obstet Gynecol 2007; 47: 141-144
	17	De Francesco, J Virol Meth 2008; 147: 10-17
	18	Monsonego, Sex Transm Dis 2008; 35: 521-527
	19	Szarewski,CEBP 2008; 17: 3033-3043
	20	Cattani, J Clin Microbiol 2009; 47: 3895-3901
	21	Huang, J Clin Virol 2009; 45S: 19-23
	22	Lee, Int J Gynecol Cancer 2009; 19: 266-272
	23	Castle , Obstet Gynecol 2010; 116: 76-84
	24	Denton, L Clin Pathol 2010; 134: 12-21
	25	**Halfon, Cancer Biomerkers 2010; 7: 133-139**
	26	Voss, Anal Quant Cytol Histol 2010; 32: 121-130
	27	**Wu, Int J Gynecol Cancer 2010; 20: 1411-1414**
	28	**Clad, J Clin Microbiol 2011; 49: 1071-1076**
	29	**Heider, Acta Cytol 2011; 55: 48-53**
	30	**Levi, Cancer Cytopathol 2011; 119: 228-234**
	31	**Monsonego, Int J Cancer 2011; 129: 691-701**
	32	**Ratnam, J Clin Microbiol 2011; 49: 557-564**
	33	**Schmidt, Cancer Cytopathol 2011; 119: 158-166**
	34	**Tsoumpou, Gynecol Oncol 2011; 121: 49-53**
	35	**Szarewski, J Clin Microbiol 2012; 50: 1867-**
	36	**Ziemke, Pathologe 2012; in press: 1-6**
	37	**Alaghehbandan, Diagn Cytopathol 2013; 41: 767-775**
	38	**Oliveira, J Med Virol 2013; 85: in-press**
Amplicor		
	1	Monsonego, Gynecol Oncol 2005; 99: 160-168
	2	De Francesco, J Virol Meth 2008; 147: 10-17

Test	Number	1st author, Journal Year; Vol: Pages
	3	Szarewski,CEBP 2008; 17: 3033-3043
	4	**Szarewski, J Clin Microbiol 2012; 50: 1867-**
Abott RT PCR		
	1	Szarewski,CEBP 2008; 17: 3033-3043
	2	Huang, J Clin Virol 2009; 45S: 19-23
	3	Halfon, J Clin Virol 2010; 48: 246-250
	4	**Szarewski, J Clin Microbiol 2012; 50: 1867-**
Linear Array		
	1	Fröberg, BJC 2008; 99: 563-568
	2	Monsonego, Sex Transm Dis 2008; 35: 521-527
	3	Szarewski,CEBP 2008; 17: 3033-3043
	4	Lee, Int J Gynecol Cancer 2009; 19: 266-272
	5	Halfon, J Clin Virol 2010; 47: 38-42
	6	Ratnam, J Clin Microbiol 2011; 49: 557-564
	7	**Dona, Gynecol Oncol 2012; 126: 198-202**
	8	**Wentzensen, Clin Cancer Res 2012; 18: 4154-4162**
PapilloCheck		
	1	Jones, J Clin Virol 2009; 45: 100-104
	2	Halfon, J Clin Virol 2010; 47: 38-42
Cervista		
	1	Belinson, Am J Clin Pathol 2011; 135: 790-795
COBAS-4800		
	1	**Szarewski, J Clin Microbiol 2012; 50: 1867-**
	2	**Cuzick, Int J Cancer 2013; 132: 959-966**
Pretect-HPV Proofer		
	1	Molden, Int J Cancer 2005; 114: 973-976
	2	**Andersson, Int J Oncol 2006; 29: 705-711**
	3	Szarewski,CEBP 2008; 17: 3033-3043
	4	Halfon, J Clin Virol 2010; 47: 177-181
	5	Sorbye , J Virol Meth 2010; 169: 219-222
	6	Ratnam, J Clin Microbiol 2011; 49: 557-564
	7	**Koliopoulos, Acta Obstet Gynecol Scand 2012; 91: 794-801**
	8	**Szarewski, J Clin Microbiol 2012; 50: 1867-**
	9	**Alaghehbandan, Diagn Cytopathol 2013; 41: 767-775**
	10	**Oliveira, J Med Virol 2013; 85: in-press**
APTIMA		
	1	Szarewski,CEBP 2008; 17: 3033-3043
	2	Dockter, J Clin Virol 2009; 45 (S1): S55-S61
	3	Wu, Int J Gynecol Cancer 2010; 20: 1411-1414
	4	Clad, J Clin Microbiol 2011; 49: 1071-1076
	5	Monsonego, Int J Cancer 2011; prepub: -
	6	Ratnam, J Clin Microbiol 2011; 49: 557-564
	7	**Szarewski, J Clin Microbiol 2012; 50: 1867-**
	8	**Waldstrom, Cancer Cytopathol 2013; 121: 136-145**
ProExC		
	1	**Kelly, Cancer Cytopathol 2006; 108: 494-500**
	2	**Tambouret, Arch Pathol Lab Med 2008; 132: 918-925**
	3	**Depuydt, Cancer Epidemiol Biomarkers Prev 2011; 20: 628-637**
	4	**Alaghehbandan, Diagn Cytopathol 2013; 41: 767-775**

p16^{INK4a}		
1	**Andersson, Int J Oncol 2006; 29: 705-711**	
2	Holladay , Cancer 2006; 108: 451-461	
3	Meyer, Cancer 2007; 111: 83-92	
4	Wentzensen, Cancer 2007; 111: 58-66	
5	Schledermann, Diagn Cytopathol 2008; 36: 453-459	
6	Szarewski, CEBP 2008; 17: 3033-3042	
7-9	Denton, Am J Clin Pathol 2010; 134: 12-21 (3 interpretations)	
10	Passamonti, Pathalogica 2010; 102: 6-11	
11	Samarawardana, Cancer Cytopathol 2010; 118: 146-156	
12	Nasioutziki, Int J Gynecol Cancer 2011; 21: 79-85	
13	Tsoumpou, Gynecol Oncol 2011; 121: 49-53	
14	**Szarewski, J Clin Microbiol 2012; 50: 1867-**	
15	**Loghavi, Diagn Cytopathol 2013; 41: 582-587**	
p16/Ki-67		
1	Schmidt, Cancer Cytopathol 2011; 119: 158-166	
2	**Wentzensen, Clin Cancer Res 2012; 18: 4154-4162**	
3	**Ziemke, Pathologe 2012; in press: 1-6**	
4	**Loghavi, Diagn Cytopathol 2013; 41: 582-587**	
5	**Waldstrom, Cancer Cytopathol 2013; 121: 136-145**	
HPV16 genotyping		Assay
1	Froberg, Br J Cancer 2008; 99: 563-568	Linear array
2	Monsonego, Int J STD AIDS 2008; 19: 385-392	Linear array
3	Szarewski,CEBP 2008; 17: 3033-3043	Linear array
4	Halfon, J Clin Virol 2010; 47: 177-181	Linear array
5	**Dona, Gynecol Oncol 2012; 126: 198-202**	Linear array
6	**Wentzensen, Clin Cancer Res 2012; 18: 4154-4162**	Linear array
7	**Gage, Cancer Epidemiol Biomarkers Prev 2013; 22: 1095-1101**	Linear array
8	Guo, Mod Pathol 2008; 21: 1037-1043	Easy Chip
9	Huang, J Clin Virol 2009; 45: S19-S23	Easy Chip
10	Szarewski,CEBP 2008; 17: 3033-3043	Clinical Arrays
11	Szarewski,CEBP 2008; 17: 3033-3043	Abbott
12	Halfon, J Clin Virol 2010; 47: 177-181	Abbott
13	**Szarewski, J Clin Microbiol 2012; 50: 1867-**	Abbott
14	Jones, J Clin Virol 2009; 45: 100-104	Papillocheck
15	Halfon, J Clin Virol 2010; 47: 177-181	Papillocheck
16	Jones, J Clin Virol 2009; 45: 100-104	PCR
17	**Belinson, Am J Clin Pathol 2011; 135: 790-795**	MALDI-TOF
18	**Depuydt, Cancer Epidemiol Biomarkers Prev 2011; 20: 628-637**	E6/7 qPCR
19	**Szarewski, J Clin Microbiol 2012; 50: 1867-**	COBAS-4800
20	**Cuzick, Int J Cancer 2013; 132: 959-966**	COBAS-4801
21	**Szarewski, J Clin Microbiol 2012; 50: 1867-**	BD-Viper
22	**Oliveira, J Med Virol 2013; 85: in-press**	CLART
HPV16/18 genotyping		Assay
1	Froberg, Br J Cancer 2008; 99: 563-568	Linear array
2	Monsonego, Int J STD AIDS 2008; 19: 385-392	Linear array
3	Szarewski,CEBP 2008; 17: 3033-3043	Linear array
4	Halfon, J Clin Virol 2010; 47: 177-181	Linear array
5	**Dona, Gynecol Oncol 2012; 126: 198-202**	Linear array

6	**Wentzensen, Clin Cancer Res 2012; 18: 4154-4162**	Linear array
7	**Gage, Cancer Epidemiol Biomarkers Prev 2013; 22: 1095-1101**	Linear array
8	Guo, Mod Pathol 2008; 21: 1037-1043	EasyChip HPV Blot
9	Huang, J Clin Virol 2009; 45: S19-S23	EasyChip HPV Blot
10	Szarewski,CEBP 2008; 17: 3033-3043	Clinical Arrays
11	Szarewski,CEBP 2008; 17: 3033-3044	Abbott
12	Jones, J Clin Virol 2009; 45: 100-104	Papillocheck
13	Halfon, J Clin Virol 2010; 47: 177-181	Papillocheck
14	Jones, J Clin Virol 2009; 45: 100-104	PCR
15	**Belinson, Am J Clin Pathol 2011; 135: 790-795**	MALDI-TOF
16	**Depuydt, Cancer Epidemiol Biomarkers Prev 2011; 20: 628-637**	E6/7 qPCR
17	**Szarewski, J Clin Microbiol 2012; 50: 1867-**	COBAS-4800
18	**Cuzick, Int J Cancer 2013; 132: 959-966**	COBAS-4801
19	**Spathis, Plos one 2012; 7: e49205-**	CLART
20	**Oliveira, J Med Virol 2013; 85: in-press**	CLART
21	**Szarewski, J Clin Microbiol 2012; 50: 1867-**	BD-Viper

4.5.3. Triage ASC-H

A list of included studies for triage of ASC-H is shown.

Test	Number	1st author, Journal Year; Vol: Pages
HC2		
	1	**Pretorius, J Reprod Med 2002; 47: 290-296**
	2	**Lonky, Obstet Gynecol 2003; 101: 481-489**
	3	**Liman, Cancer 2005; 105: 457-460**
	4	**Chivukula, CytoJournal 2006; 3: 1-23**
	5	**Sherman, Cancer Cytopathol 2006; 108: 298-305**
	6	**Srodon, Cancer 2006; 108: 32-38**
	7	**Wu, Diagn Cytopathol 2006; 34: 707-710**
	8	**Owens, Am J Clin Pathol 2007; 128: 398-403**
	9	**Reid-Nicholson, Diagn CytoPathol 2007; 35: 1-5**
	10	**You, Aust NZ J Obstet Gynecol 2007; 47: 141-144**
	11	**Bandyopadhyay, Arch Pathol Lab Med 2008; 132: 1874-1881**
	12	**Monsonego, Sex Transm Dis 2008; 35: 521-527**
	13	**Siddiqui, Arch Pathol Lab Med 2008; 132: 1648-1652**
	14	**Rao, J Obstet Gynecol Res 2009; 35: 503-506**
	15	**Castle , Obstet Gynecol 2010; 116: 76-84**
	16	**Monsonego, Int J Cancer 2011; 129: 691-701**

4.5.4. Triage AGC

A list of included studies for triage of AGC is shown.

Test	Number	1st author, Journal Year; Vol: Pages
HC2		
	1	**Ronnett, Hum Pathol 1999; 30: 816-825**
	2	**Derchain, Gynecol Oncol 2004; 95: 618-623**
	3	**Chen , Gynecol Oncol 2005; 99: 578-584**
	4	**Irvin, Am J Obstet & Gynecol 2005; 193: 559-567**
	5	**de Oliveira, Int J Gynecol Cancer 2006; 16: 1055-1062**
	6	**Saqi, Diagn Cytopathol 2006; 34: 235-239**
	7	**Liao, Int J Cancer 2009; 125: 2434-2440**
	8	**Schnatz, J Low Genit Tract Dis 2009; 13: 94-101**
	9	**Castle, Obstet Gynecol 2010; 115: 243-248**
	10	**Zhao, Arch Pathol Lab Med 2010; 134: 103-108**

4.6. References

(1) Wright TC, Cox JT, Massad LS, Wilkinson EJ. 2001 Consensus guidelines for the management of women with cervical cytological abnormalities. JAMA 2002;287(16):2120-9.

(2) Jordan J, Martin-Hirsch P, Arbyn M, Schenck U, Baldauf J-J, Da Silva D et al. European guidelines for management of abnormal cervical cytology, Part 2. Cytopathology 2009;20(1):5-16.

(3) Cox JT. HPV testing: is it useful in triage of minor Pap abnormalities? J Fam Pract 1998;46:121-4.

(4) Cox JT. Management of women with cervical cytology interpreted as ASC-US or as ASC-H. Clin Obstet Gynecol 2005 March;48(1):160-77.

(5) Sawaya GF. A 21-year-old woman with atypical squamous cells of undetermined significance. JAMA 2005 November 2;294(17):2210-8.

(6) Soutter WP. Management of women with mild dyskaryosis. Immediate referral to colposcopy is safer. BMJ 1994;309:591-2.

(7) Stoler MH, Schiffman MA. Interobserver reproducibility of cervical cytologic and histologic interpretations. JAMA 2001;285(11):1500-5.

(8) Lundberg GD, National Cancer Institute. The 1988 Bethesda System for Reporting Cervical/Vaginal Cytologic Diagnoses. JAMA 1989;262(7):931-4.

(9) Solomon D, Davey D, Kurman R, Moriarty A, O'Connor D, Prey M et al. The 2001 Bethesda System: terminology for reporting results of cervical cytology. JAMA 2002;287:2114-9.

(10) Robertson JH, Woodend BE, Crozier EH, Hutchinson J. Risk of cervical cancer associated with mild dyskaryosis. BMJ 1988;297:18-21.

(11) Coleman D, Day N, Douglas G, Farmery E, Lynge E, Philip J et al. European Guidelines for Quality Assurance in Cervical Cancer Screening. Europe against cancer programme. Eur J Cancer 1993;29A Suppl 4:S1-S38.

(12) Flannelly G, Anderson D, Kitchener HC, Mann EM, Campbell M, Fisher P et al. Management of women with mild and moderate cervical dyskaryosis. BMJ 1994;308:1399-403.

(13) Richart RM. Causes and management of cervical intraepithelial neoplasia. Cancer 1987;60:1951-9.

(14) Noumoff JS. Atypia in cervical cytology as a risk factor for intraepithelial neoplasia. Am J Obstet Gynecol 1987;156:628-31.

(15) Richart RM, Wright TC. Controversies in the management of low-grade cervical intraepithelial neoplasia. Cancer 1993;71(4 Suppl):1413-21.

(16) Cox JT, Lorincz AT, Schiffman MH, Sherman ME, Cullen A, Kurman RJ. Human papillomavirus testing by hybrid capture appears to be useful in triaging women with

a cytologic diagnosis of atypical squamous cells of undetermined significance. Am J Obstet Gynecol 1995;172(3):946-54.

(17) Wright TC, Sun XW, Koulos J. Comparison of management algorithms for the evaluation of women with low-grade cytologic abnormalities. Obstet Gynecol 1995 February;85(2):202-10.

(18) Kinney WK, Manos MM, Hurley LB, Ransley JE. Where's the high grade cervical neoplasia? The importance of minimally abnormal Papanicolaou diagnoses. Obstet Gynecol 1998;91:973-6.

(19) Arbyn M, Sasieni P, Meijer CJ, Clavel C, Koliopoulos G, Dillner J. Chapter 9: Clinical applications of HPV testing: a summary of meta-analyses. Vaccine 2006 August 31;24 (SUPPL. 3):S78-S89.

(20) Arbyn M, Kyrgiou M, Simoens C, Raifu AO, Koliopoulos G, Martin-Hirsch P et al. Perinatal mortality and other severe adverse pregnancy outcomes associated with treatment of cervical intraepithelial neoplasia: a meta-analysis. BMJ 2008;337:a1284.

(21) Kyrgiou M, Koliopoulos G, Martin-Hirsch P, Arbyn M, Prendiville W, Paraskevaidis E. Obstetric outcomes after conservative treatment for intra-epithelial or early invasive cervical lesions: a systematic review and meta-analysis of the literature. Lancet 2006;367(9509):489-98.

(22) zur Hausen H. Molecular pathogenesis of cancer of the cervix and its causation by specific human papillomavirus types. Curr Top Microbiol Immunol 1994;186:131-56.

(23) Walboomers JM, Jacobs MV, Manos M, 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.

(24) Bosch FX, Lorincz AT, Munoz N, Meijer CJ, Shah KV. The causal relation between human papillomavirus and cervical cancer. J Clin Pathol 2002 April;55(4):244-65.

(25) IARC Monograph Working Group. IARC Monographs on the Evaluation of Carcinogenic Risks to Humans. Vol 90: Human Papillomaviruses. Lyon: IARCPress; 2007.

(26) Arbyn M, Buntinx F, Van Ranst M, Paraskevaidis E, Martin-Hirsch P, Dillner J. Virologic versus cytologic triage of women with equivocal Pap smears: a meta-analysis of the accuracy to detect high-grade intraepithelial neoplasia. J Natl Cancer Inst 2004;96:280-93.

(27) Arbyn M, Ronco G, Anttila A, Meijer CJLM, Poljak M, Ogilvie G et al. Evidence regarding HPV testing in secondary prevention of cervical cancer. Vaccine 2012;30 Suppl 5:F88-F99.

(28) Arbyn M, Roelens J, Simoens C, Buntinx F, Paraskevaidis E, Martin-Hirsch PPL et al. Human papillomavirus testing versus repeat cytology for triage of minor cytological cervical lesions. Cochrane Database Syst Rev 2013;3(CD008054):1-201.

(29) Arbyn M, Roelens J, Cuschieri K, Cuzick J, Szarewski A, Ratnam S et al. The APTIMA HPV assay versus the Hybrid Capture II test in triage of women with ASC-US or LSIL cervical cytology: a meta-analysis of the diagnostic accuracy. Int J Cancer 2013;132(1):101-8.

(30) Roelens J, Reuschenbach M, von Knebel-Doeberitz M, Wentzensen N, Bergeron C, Arbyn M. p16INK4a immunocytochemistry versus HPV testing for triage of women

with minor cytological abnormalities: A systematic review and meta-analysis. Cancer 2012;120(5):294-307.

(31) Verdoodt F, Szarewski A, Halfon P, Cuschieri K, Arbyn M. Triage of women with minor abnormal cervical cytology: Meta-analysis of the accuracy of an assay targeting messenger ribonucleic acid of 5 high-risk human papillomavirus types. Cancer Cytopathol 2013 July 23;121(12):675-87.

(32) Harris R, Bradburn M, Deeks J, Harbord R, Altman D, Sterne J. metan: fixed- and random-effects meta-analysis. Stata Journal 2008;8(1):3-28.

(33) Dersimonian R, Laird NM. Meta-analysis in clinical trials. Controlled Clin Trials 1986;7:177-88.

(34) Wong OG, Lo CK, Szeto E, Cheung AN. Efficacy of Abbott RealTime High Risk HPV test in evaluation of atypical squamous cells of undetermined significance from an Asian screening population. J Clin Virol 2011 April 27;51(2):136-8.

(35) Denton KJ, Bergeron C, Klement P, Trunk MJ, Keller T, Ridder R. The Sensitivity and Specificity of p16INK4a Cytology vs HPV Testing for Detecting High-Grade Cervical Disease in the Triage of ASC-US and LSIL Pap Cytology Results. Am J Clin Pathol 2010 July;134(1):12-21.

(36) Monsonego J, Pollini G, Evrard MJ, Sednaoui P, Monfort L, Zerat L et al. Detection of Human Papillomavirus Genotypes Among High-Risk Women: A Comparison of Hybrid Capture and Linear Array Tests. Sex Transm Dis 2008 May;35(5):521-7.

(37) Sherman ME, Castle PE, Solomon D. Cervical cytology of atypical squamous cells-cannot exclude high-grade squamous intraepithelial lesion (ASC-H): characteristics and histologic outcomes. Cancer 2006 March 16;298-305.

(38) Siddiqui MT, Cohen C, Nassar A. Detecting high-grade cervical disease on ASC-H cytology: role of BD ProEx C and Digene Hybrid Capture II HPV DNA testing. Am J Clin Pathol 2008 November;130(5):765-70.

(39) Castle PE, Fetterman B, Thomas CJ, Shaber R, Poitras N, Lorey T et al. The age-specific relationships of abnormal cytology and human papillomavirus DNA results to the risk of cervical precancer and cancer. Obstet Gynecol 2010 July;116(1):76-84.

(40) Monsonego J, Hudgens MG, Zerat L, Zerat JC, Syrjanen K, Halfon P et al. Evaluation of oncogenic human papillomavirus RNA and DNA tests with liquid based cytology in primary cervical cancer screening (The FASE study). Int J Cancer 2011;129(3):691-701.

(41) Ronnett BM, Manos M, Ransley JE, Fetterman B, Kinney W, Hurley A et al. Atypical glandular cells of undetermined significance (AGUS) : cytopathologic features, histopathologic results, and human papillomavirus DNA detection. Hum Pathol 1999;30:816-25.

(42) Meijer CJLM, Castle PE, Hesselink AT, Franco EL, Ronco G, Arbyn M et al. Guidelines for human papillomavirus DNA test requirements for primary cervical cancer screening in women 30 years and older. Int J Cancer 2009;124(3):516-20.

(43) Cuzick J, Sasieni P, Davies P, Adams J, Normand C, Frater A et al. A systematic review of the role of human papillomavirus testing within a cervical screening programme. Health Technol Assess 1999;3:1-204.

(44) Wentzensen N, Bergeron C, Cas F, Vinokurova S, von Knebel DM. Triage of women with ASCUS and LSIL cytology: use of qualitative assessment of p16INK4a positive cells to identify patients with high-grade cervical intraepithelial neoplasia. Cancer 2007 January 1;111(1):58-66.

(45) Wentzensen N, von Knebel DM. Biomarkers in cervical cancer screening. Dis Markers 2007;23(4):315-30.

(46) Zuna RE, Wang SS, Rosenthal DL, Jeronimo J, Schiffman M, Solomon D. Determinants of human papillomavirus-negative, low-grade squamous intraepithelial lesions in the atypical squamous cells of undetermined significance/low-grade squamous intraepithelial lesions triage study (ALTS). Cancer 2005 June 30;105(5):253-62.

(47) Bais AG, Rebolj M, Snijders PJ, De Schipper FA, van der Meulen DA, Verheijen RH et al. Triage using HPV-testing in persistent borderline and mildly dyskaryotic smears: Proposal for new guidelines. Int J Cancer 2005 August 10;116(1):122-9.

(48) Berkhof J, de Bruijne MC, Zielinski GD, Bulkmans NW, Rozendaal L, Snijders PJ et al. Evaluation of cervical screening strategies with adjunct high-risk human papillomavirus testing for women with borderline or mild dyskaryosis. Int J Cancer 2006;118:1759-68.

(49) Ronco G, Cuzick J, Segnan N, Brezzi S, Carozzi F, Folicaldi S et al. HPV triage for low grade (L-SIL) cytology is appropriate for women over 35 in mass cervical cancer screening using liquid based cytology. Eur J Cancer 2007 January 11;43:476-80.

(50) Arbyn M, Martin-Hirsch P, Buntinx F, Van Ranst M, Paraskevaidis E, Dillner J. Triage of women with equivocal or low-grade cervical cytology results. A meta-analysis of the HPV test positivity rate. J Cell Mol Med 2009;13(4):648-59.

4.7. GRADE-Profil

Authors: M. Arbyn, M. Jentschke

The term minor cytological abnormalities encompass the following categories:

- atypical squamous cells of undetermined significance (ASCUS defined according to TBS-1988(8) and ASC-US defined according to TBS-2001(9)),
- atypical glandular cells (AGUS defined according to TBS-1988(8) and AGC defined according to TBS-2001(9)),
- atypical squamous cells where a high-grade squamous intraepithelial abnormality cannot be excluded (ASC-H) (8;9)
- low-grade intraepithelial lesions (LSIL)(8;9).

Steps in evidence assessment for making guidelines
1) Formulate a question
2) Identify the PICO(S) components
3) Qualify outcomes as critical, important, not important
~~Scaling of Critical, Important but not critical, Limited.~~

1) Questions
What is the clinical accuracy (sensitivity and specificity) to identify or exclude high-grade cervical precancer or worse (CIN2+, CIN3+, AIS+) using hrHPV testing or biomarkers among women with minor abnormal cytology (ASC-US, LSIL, ASCH or AGC)?
How compares this sensitivity and specificity with that of repeat cytolology?

2) PICOS

- P: women participating in cervical cancer screening with minor abnormal cytology
- I: hrHPV testing or biomarkers (p16, p16/Ki-67 dual-stain, ProExC, E6/E7 mRNA, methylation markers or other)
- C: cytology (conventional Pap smear, LBC)
- O: accuracy to detect underlying disease (=CIN2+,CIN3+/AIS):
    - Complete diagnostic studies: absolute and relative sensitivity and specificity, PPV, NPV, referral rate, detection rate, detection rate ratio
    - RCTs: relative sensitivity (or detection rate ratios), relative PPV, relative referral rate
- S:
    - diagnostic studies
        - all subjects receiving testing with a biomarker,

- at least one comparator test
- verification with the reference standard (colposcopy/biopsy)
    - all participants (accepting a negative colposcopy as free of CIN2+)
    - participants positive in at least one screening test (accepting a negative result for all screening tests as free of disease)

o RCTs comparing screening with biomarkers with screening with one or more comparator tests.

3) Importance of outcomes

Outcome:

19. Reduction of mortality from cervical cancer, (quality-adjusted) life-years gained.
20. Reduction of morbidity due to cervical cancer: incidence of cancer (Ib+).
21. Reduction of incidence of cancer (including micro-invasive cancer).
22. Reduction of incidence of CIN3 or worse disease (CIN3+).
23. Increased detection rate of CIN3+ or CIN2+.
24. Increased test positivity with increased, similar or hardly reduced positive predictive value.

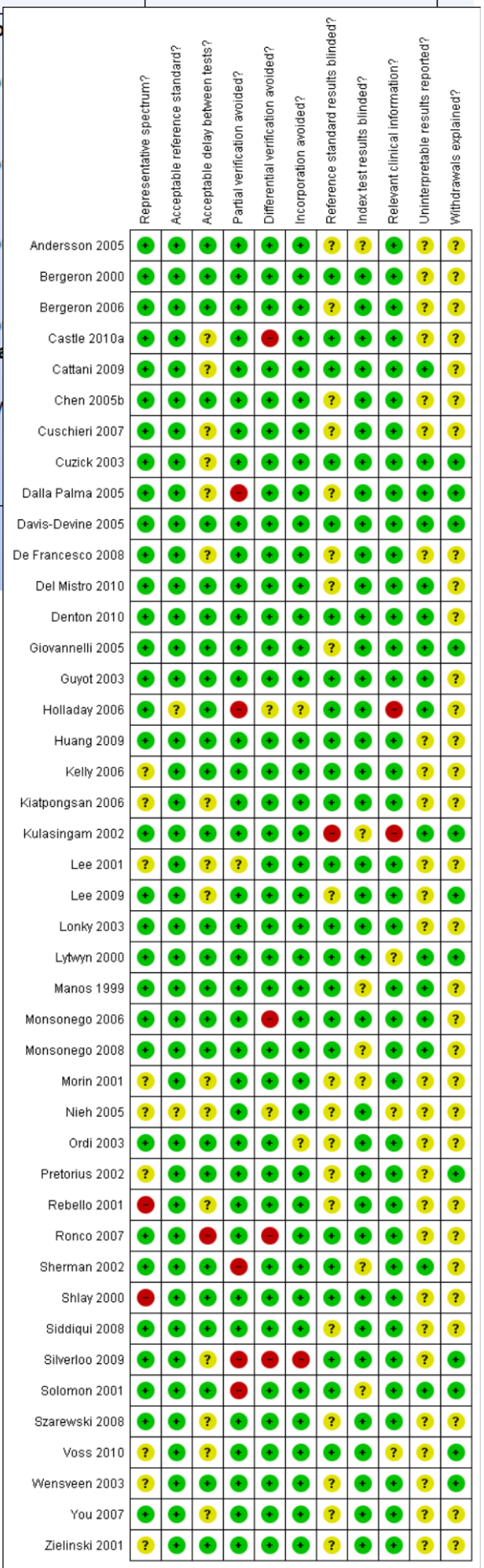
The following outcome measures were assessed, separately for the triage of ASC-US and LSIL:
- disease rate (CIN2+ or CIN3+)
- absolute sensitivity and specificity of HC2 (cut-off: 1.00 RLU/CO) and cytology (cut-off: ASC-US) to detect CIN2+ or CIN3+
- absolute sensitivity and specificity of other hrHPV tests and biomarkers to detect CIN2+ or CIN3+
- relative sensitivity and specificity HC2 (cut-off: 1.00 RLU/CO) versus cytology (cut-off: ASC-US) to detect CIN2+ or CIN3+
- relative sensitivity and specificity of other hrHPV tests and biomarkers versus HC2 (cut-off: 1.00 RLU/CO) to detect CIN2+ or CIN3+

The following outcome measures were assessed, separately for the triage of AGC and ASC-H:
- disease rate (CIN2+ or CIN3+)
- absolute accuracy of HC2 (cut-off: 1.00 RLU/CO) and cytology (cut-off: ASC-US) to detect CIN2+ or CIN3+
- relative accuracy HC2 (cut-off: 1.00 RLU/CO) versus cytology (cut-off: ASC-US) to detect CIN2+ or CIN3+

GRADE – Assessment of quality

GRADE

Quality of evidence	*Study design*	*Rating down if…*	*Rating up if…*
High	**randomized study (RCT)**	**study limitatio** -1 serious -2 very serio **inconsistency** - 1 serious - 2 very serio **indirectness** -1 serious -2 very serio **imprecision** -1 serious -2 very serio **publication bia** -1 likely -2 very likely	
Middle			
Low	**observational study**		
Very low			



4) Quality of evidence for each outcome in four categories;

- High: ++++
- Moderate: +++
- Low: ++
- Very low: +

Quality assessment was only done for the studies included in a previous systematic review (Arbyn M, Roelens J, Simoens C, Buntinx F, Paraskevaidis E, Martin-Hirsch PPL, Prendiville WJ. Human papillomavirus testing versus repeat cytology for triage of minor cytological cervical lesions. Cochrane Database of Systematic Reviews 2013, Issue 3).

The methodological quality of selected studies were assessed using the QUADAS guidelines. The effective sample size funnel plot and associated regression test of asymmetry were used to detect publication bias.

Study design: Three studies were randomised controlled trials (Lytwyn 2000; Solomon 2001; Sherman 2002; Cuzick 2003). In all other studies, a concomitant testing design was used, where enrolled women received the HPV test, a repeat smear (if done) and the reference standard.

Overall, the quality of included studies was good with average negative scores for the 11 QUADAS items varying between 0% and 1à%.

5 factors that lower the quality of the evidence (Guyatt *et al.*, 2008a)

16. Risk of bias due to limitations in design or execution: see CONSORT, STROBE, QUADAS
17. Inconsistency or heterogeneity: if consistency unexplained, lower quality
18. Indirectness, applicability (relevance of studies for answering the PICPO question)
19. Imprecision: number of studies, width of CI
20. Reporting bias, publication bias.

Figure 51 Methodological quality summary: review authors' judgements about each QUADAS item for each

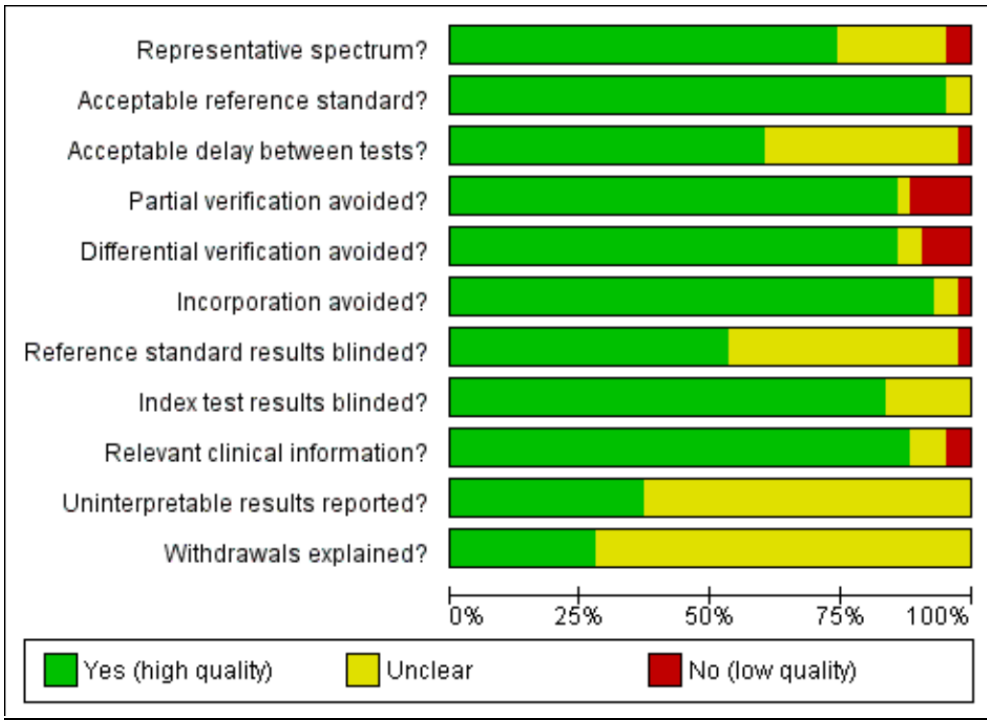


Figure 52 Methodological quality graph: review authors' judgements about each methodological quality

3 factors that increase the quality

7. Large effect
8. Dose effect gradient

9. Confounders (presence of confounders that have lowered the observed effect, absence of these would have increased the effect) (Guyatt *et al.,* 2008a)

Items downgrading quality of evidence		**Downgrading**
Bias, design	Overall, the quality of included studies was good.	No (-0)
Inconsistency		No (-0)
Indirectness		No (-0)
Imprecision	The heterogeneity analysis by covariate was performed only when the groups compared contained at least five studies in one group and at least three studies in the other group. Most often, the absolute accuracy of triage withHC2 or repeat cytology did not change significantly by covariate.	No (-0)
Publication bias, other	The findings are suggestive of a positive relationship between diagnostic accuracy and sample size. The ALTS trial was one of the best designed triage studies with high-quality disease certification. The data show the opposite of the usual publication bias where excessive accuracy in small published studies is unbalanced by nonpublished small studies with low accuracy.	Yes (-0)
Items upgrading quality of evidence		
Large effect	No	No (+0)
Dose-effect correlation	No	No (+0)
Confounding factors neutralising effects	No	No (+0)˚

<u>Conclusion: evidence of moderate quality.</u>

Overall Grade of the quality of evidence is assigned and this is based on the outcome with the lowest quality of evidence given that it is a critical outcome.

Table 43 GRADE evidence profile

	Quality of evidence	

# studies (N)	Absence of study limitations	Consistency	Directness (outcome, representativity Germany)	Precision	Absence publication bias	Large effect	Dose-effect relation	Bias lowering effect	Comment
Outcome 1: Triage of women with ASC-US									
52	Yes	Yes	Yes	Yes	No	No	No	No	Moderate

Outcome 2: Triage of women with L-SIL									
38	Yes	Yes	Yes	Yes	No	No	No	No	Moderate

References

Guyatt G.H., Oxman A.D., Kunz R., Falck-Ytter Y., Vist G.E., Liberati A., & Schunemann H.J. (2008a) Going from evidence to recommendations. *BMJ* **336:** 1049-1051.

Guyatt G.H., Oxman A.D., Kunz R., Vist G.E., Falck-Ytter Y., & Schunemann H.J. (2008b) What is "quality of evidence" and why is it important to clinicians? *BMJ* **336:** 995-998.

Guyatt G.H., Oxman A.D., Vist G.E., Kunz R., Falck-Ytter Y., Alonso-Coello P., & Schunemann H.J. (2008c) GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *BMJ* **336:** 924-926.

Schunemann H.J., Oxman A.D., Brozek J., Glasziou P., Jaeschke R., Vist G.E., Williams J.W., Jr., Kunz R., Craig J., Montori V.M., Bossuyt P., & Guyatt G.H. (2008) GRADE: Grading quality of evidence and strength of recommendations for diagnostic tests and strategies. *BMJ* **336:** 1106-1110.

5. Question: Organised versus opportunistic cervical cancer screening

Unit of Cancer Epidemiology
Wetenschappelijk Instituut Volksgezondheid (WIV)
Juliette Wytsmanstraat 14
1050 Brussels

5.1. Introduction

In 2003, the European Council including all national Ministers responsible for public health in the Member States of the European Union endorsed the scientific evidence on mass screening for three cancers (breast, colo-rectum and uterine cervix) and recommended to offer organised screening for these cancers for the whole target populations concerned in agreement with European guidelines[1]. According to the recommendation, systematic implementation of cancer screening programmes requires an organisation with a call/recall system and with quality assurance at all levels; an effective and appropriate diagnostic, treatment, and after-care service following evidence-based guidelines. Centralised data systems are also needed in running organised screening programmes. There is further guidance for implementation, registration, monitoring and evaluation, training, informing women, and on introducing novel screening tests. In many countries, the European recommendations are not yet met [2,3]. The recommended test to screen for cervical cancer was the Pap smear. However, the Council recommendation also encourages research to evaluate new screening methods using robust scientific study designs, preferentially randomised controlled trials, taking in to account public health relevant outcomes as mortality or established surrogate indicators. Further, pooling the results of the trials using meta-analyses should assess the level of evidence concerning effects of new methods.

In the framework of the preparation of the S3 Guideline Prevention of Cervical Cancer in Germany, questions were raised related to the definition and characteristics of organised cervical cancer screening, as well as on the evidence that organised screening is better than non-organised or opportunistic screening.

5.2. Materials and Methods

5.2.1. Questions to be addressed

Three questions were proposed by the Working Group Versorgungsstrukturen.

- What are the quality characteristics of organised screening "Was sind Qualitätsmerkmale eines organisierten Screenings? (Definition)"
- Is organised screening better than opportunistic screening? "Ist ein organisiertes Screeningverfahren besser geeignet als ein opportunistisches?"
- Which parameters are crucial for an effective screening programme? "Welche Parameter sind entscheidend für ein effektives Screeningprogramm?"

5.2.2. Literature Retrieval

To answer these questions, the Unit of Cancer Epidemiology updated its previous review conducted in charge of the Belgian Health Care Knowledge Centre[4] in 2006, which was updated in 2009[5]. We further used materials from the second edition of the European Guidelines on Quality Assurance in Cervical Cancer Screening[6]. Newer data sources were retrieved from the literature.

The following search string was used to retrieve data on the performance of organised versus non-organised cervical cancer screening.

((cervix OR cervical) AND cancer AND screening AND (organized OR organised OR organisation). Restricted to references published after 1st of January 2009.

Box 1: Search string for literature retrieval.

5.3. Results

5.3.1. Definition of organized screening – components of an organized screening programme

To maximise the positive impact and minimise the adverse effects, screening should be offered only in organised settings [7]. Designing a programme includes defining the screening policy: this means determining the target age group, the screening interval, choice of the screening test, and the establishment of follow-up and treatment strategies for screen-positive women. The screening policy should be defined taking into account the natural history of the disease and the variation in background risk. Moreover, a screening programme must reach high population acceptance and coverage, and assure and demonstrate good quality at all levels[8,9].

Population-based information systems need to be set up allowing continuous monitoring of screening process indicators. A legal framework should be established permitting registration of individual data and linkage between population databases, screening files, cancer and mortality registers [7,10-13]. The information system is an essential tool to run the programme, to compute the indicators of attendance, compliance, quality and impact, and to provide feed-back to involved health professionals, stake holders and health authorities.

A concern is the completeness of the recorded information of the programme. Reliable cancer registration is important. Individual-level links between population, screening, cancer registry, and treatment data are needed.

As with any public health policy, a screening programme should be designed in such a way that it can be evaluated. Key components in the monitoring and evaluation of screening are: regularly published results on the screening performance so that it is clear for the decision-makers, key personnel groups, and for the general public how well the programme is running; scientific evaluation of the effectiveness and outcomes of the screening programme based on established epidemiological methods; and ascertaining and feed-back of information about invasive cancers detected in connection or after screening[8].

Effectiveness of an organised programme is a function of the quality of its individual components. Epidemiology provides instruments that permit planning, guidance and evaluation of the entire process of a screening programme, from the organisational and administrative aspects up to assessment of the impact[14].

Organised cervical cancer screening is a multi-step process including (see Table 44):

- Identification of the target population
- Reaching women belonging to the target population
- Collection of an adequate Pap smear
- Examination of the Pap smear and reporting
- Communication of smear test results to women with a normal result.
- Recall of women with unsatisfactory/inadequate smears
- Follow-up of women with abnormal smears i.e. diagnostic procedures and treatment if needed, including a fail-safe system to make sure this actually happens

In some countries, with established opportunistic screening, among which Germany and Belgium, re-allocation of resources already used for screening activities would be sufficient to cover the entire target population within a screening interval as defined in European guidelines [2,15,16]. Different solutions can be proposed to implement organised screening (depending on the level of existing opportunistic screenign). In general, systems that have demonstrated effectiveness can be recommended but relevant cost-effectiveness aspects and aspects to minimise potential adverse effects need to be considered also. However, strategies that have demonstrated good performance in other countries are often not transposable and require careful piloting or implementation trials.

Table 44. Components of an organized cervical cancer screening programme

1	Definition of the screening policy: determination of the screening test, target age group (start & end screening) and the screening interval
2	Existence of guidelines for management of screen test positives
3	Screening and management policy in agreement with European guidelines or other high-quality evidence assessment
4	Population registry available for targeted individualised invitation
5	Existence of an invitational system (call, call-recall, recall)
6	Registration of participation in screening (not response to invitation but also opportunistic)
7	Registration of screen test results (organised and opportunistic)
8	Registration of further follow-up/management of screen-positive women (repeat tests, triage tests, colposcopy, histology, treatment)
9	Linkage with cancer registry
10	Linkage with HPV vaccination registry
11	Computation of screening performance (process and impact) indicators, production of a regular reports including the evaluation of cost-effectiveness
12	Regular feedback to health authorities, international networks, stakeholders, field workers involved in screening, health professionals and the public
13	Use of a universal system for reporting of screen test results allowing national and international comparison. For cervical cytology, a system compatible with the Bethesda system is recommended in EU guidelines{Herbert, 2007 26729 /id}.

14	Adaption of evidence-based guidelines, in agreement with EU guidelines or in agreement with new evidence regarding effectiveness, cost-effectiveness and endorsed by competent bodies in the country/region.
15	Organised piloting of new elements of the screening programme, adaptation if required, general roll-out if piloting is successful

5.3.2. Target age groups and screening interval for cytological screening

The Council of the European Union states that screening should start in the age range of 20 to 30 years. It does not define the age at which screening can be stopped neither the screening frequency. The Council Recommendation is a political basic document, which is universally accepted throughout the Union and is therefore not very detailed. However, the Advisory Committee on Cancer Prevention, which consists of cancer screening experts, have formulated recommendations including the screening frequency and the target age group. According to this committee, cytology screening should be offered at 3 to 5 year intervals up to the age 60. Depending on resources, screening can be continued beyond that age [17]. The definition of the start age should be based on the local age-specific incidence rates of cervical cancer. Three to five years before incidence starts rising from a very low level is good thumb rule (Anttila *et al.,* 2006).

The European countries have opted for quite different target age groups. Screening more frequently than every three years should be discouraged as it is only marginally more effective and is certainly not cost-effective [3]. There is no firm evidence for the optimal age to start screening. However a smear taken between 35 and 64 years of age is much more effective in detecting a progressive lesion, than a smear taken at age 20. Table 47 illustrates the impact of different screening policies on cancer incidence, based on the follow-up of women with negative smears (from IARC, 1986). There was no additional impact of starting screening at age 20 compared to starting at age 25. Evidence of a lower effect of screening below age 30 was suggested by a recent study from the UK [18]. An early start will imply treatment of many CIN which would, if untreated, never have progressed to invasive cervical cancer. Treatment of young women by excision can compromise future pregnancy outcomes [19]. A very late start will inevitably imply that some early invasive cancers are missed. A start at the age of 15 as in Luxembourg is clearly too early as the incidence of invasive cancer is virtually zero until the age of 20, and as the early start will lead to overtreatment.

Table 45. The calculated effectiveness of different screening policies. Proportionate reduction in incidence of invasive squamous cell carcinoma of the cervix uteri assuming 100% compliance (IARC, 1986).

Screening frequency	Age group	Numbers of smears per woman	Reduction in cumulative incidence (%)
Every year	20-64	45	93
Every 3 years	20-64	15	91
Every 3 years	25-64	13	90
Every 3 years	35-64	10	78
Every 5 years	20-64	9	84
Every 5 years	25-64	8	82
Every 5 years	35-64	6	70
Every 10 years	25-64	5	64

The Europe against Cancer recommendations stated also that cervical cancer screening should be offered at least every fifth year, and if resources are available, every third year. The number of unnecessary treatments increases with a large number of smears per lifetime. With limited resources, screening every fifth year with high quality and high compliance is preferable to screening every third year at a proportionally lower coverage.

5.3.3. Evidence showing better performance of organised screening

To assess the effectiveness of cervical cancer screening programmes we first describe the screening systems currently in place in the Member States in the European Union. Next, we will describe the impact of introduction of organised screening in certain countries.

5.3.4. An overview of cervical cancer screening systems in Europe

Organised screening programmes for cervical cancer exist in several countries of the European Union. The screening policies, the organisation and practices of screening vary between the countries [2,2,3,3,3,20,21]. So do their effectiveness and cost-effectiveness [14,22].

In Table 47 are summarized the major features of screening systems in use in EU Member States.

Table 46: Characteristics of cervical cancer screening systems in the 15 old Member States of the European Union (adapted from Anttila et al, 2009)[23].

Countries	Target age group (years)		Screening interval (years)	Number of smears per woman in target age	Population covered by	Organised screening
	Start	End				

					programme	
Austria	18	-	1	50	-	No
Belgium	25	64	3	14	100 %[d]	Partly (58%)[d]
Denmark	23	59	3(<50)/ 5[b]	12	100%	Nationwide
Finland	(25)30	60 (65)	5	7-9	100%	Nationwide
France	(20)25	65	3	14	5%	Partly
Germany	20	-	1	>50	90%	No
Greece	20	-	1	>50	-	No
Ireland	25	60	3(<45)/ 5[c]	10	100%	Nationwide
Italy	25	64	3	14	64%	Partly
Luxembourg	15	-	1	>50	100%	No
The Netherlands	30	60	5	7	100%	Nationwide
Portugal	25	64	3	14	-	Partly
Spain	18-35	59-65	3/5	5-15	-	Partly
Sweden	23	60	3(<50)/ 5[b]	12	100%	Nationwide
UK[a]	(20)25	(60)64	3(<50)/ 5	12	100%	Nationwide

[a] In 2003, the screening policy in England was adapted subsequent to a case-control study where screening histories were compared of women with cancer with those of age-matched women who never developed cancer [18]. The current policy is to screen women aged 25-49 every 3 years, and women aging 50-64 every 5 years. In Scotland, screening still starts at the age of 20.

[b] In Denmark and Sweden, women aged 23-50 years are currently recommended to be screened every 3 and women aged 51-60 years every 5 years.

[c] In Ireland, women aged 25-44 years are currently recommended to be screened every 3 and women aged 45-60 years every 5 years.

[d] In Belgium: opportunistic screening is completely reimbursed for all women (without age restriction), since 2013. An invitational system only exist in the Flemish Region inviting women aged 25-64 without Pap smear in the last 3 years.

Table 47. Screening coverage (proportion of women having had at least 1 Pap since screening interval, in the target age range as defined in the previous table) in the 15 old EU Member States.

Country	Coverage	Interval (years)	Source
Belgium	61%	3	Arbyn 2011; 2014[24,25]
Denmark	69%	3	Antilla, 2009[23]
Finland	>70%	5	Antilla, 2009[23]
France	60%	3	Rousseau, 2002[26]
Germany	80,60%	1	Mund, 2009[27]

	[a]		
Greece[b]	~80%	3	Simou, 2010[28]
Ireland	75%	3	www.cervicalcheck.ie
Italy	>59%	3	Antilla, 2009[23]
The Netherlands	77%	5	Antilla, 2009[23]
Portuga[b]l	66%	3	Oliveira, 2014[29]
Spain[b]	76%	3	Puig-Tintore, 2008[30]
Sweden	80%	3/5	NKCx annual report 2013[31]
UK (England)	74%	3/5	Antilla, 2009[23]

[a] Coverage estimated for 3-year interval: 80% in age group 20-40, 60% in age group 41-65.
[b] Estimated from surveys, therefore most-probably overestimated.

Table 48. National cancer screening policies in twelve new Member States of the European Union

Countries	Screening system	Target age group (years) start	end	Interval (years)
Bulgaria	Opportunistic	31	65	2
Cyprus	Opportunistic	-	-	1
Czech Republic	Partly organised (national roll out)	25	69	1
Estonia	Partly organised (national roll out)	30	59	5
Hungary	~ organised	25	64	3
Latvia	Partly organised	25	69	3
Lithuania	Partly organised (national roll out)	30	60	3
Malta	Opportunistic	-	-	-
Poland	Partly organised (national roll out)	25	59	3
Romania	Partly organised	25	65	5
Slovakia	Opportunistic	18	-	1
Slovenia	Nationwide	20	64	3

Table 49. Estimated coverage for cervical cancer screening in certain new Member States of the EU

Countries	Coverage	Screening interval (years)	Source
Estonia			
Within programme	13%	5	Veerus,

			2010[32]
Outside programme	50%		
Latvia			
Response to invitation	27%	3	Receberga, 2014[a]
Coverage (organised + opportunistic)	42%		
Poland			
Ever screened	78%	5	Spazynski, 2009[b]
Received invitation	71%		
Responded to invitation	6%		
Romania			
Cluj Region	10%	3	Anttila, 2009[23]
Slovenia	71%	3	Zakelj, 2009[c]

[a] Receberga, 2014: presented at the 39th Congress of the Nordic Federation of Societies of Obstetrics and Gynecology (NFOG): Stockholm (Sweden), 10-12 June 2014.
[b] Spaczynski, Ginekol Polska 2009
[c] Zakelj, Onkologija 2009.

Concerning the system of invitation, three major systems are used: the call system, the call-recall system and the recall system.

CALL

The call system is an invitation system where all women from the target population are drawn for invitations to the programme. For this reason, an accurate list of the population is needed. Sources of such lists vary between countries and include population registries, general practitioners medical files, electoral registers and others.

The advantages are that all women in the list have access to well-organised screening. The disadvantage is that no information is captured for opportunistic screening, where respect of quality standards cannot be verified, and that women already screened in an opportunistic screening are invited unnecessarily††††. Call systems are in place in the Finland, Italy, the Netherlands and the UK [33-36].

Non-attenders are identified by the laboratories and are reminded. In the Netherlands every smear taken in the country is recorded in the PALGA (Dutch Network and National Database for Pathology) with the reasons for the smear (programme smear, opportunistic smear, repeat smear), its result and advice on follow-up. Opportunistic smears are not paid and their frequency has therefore decreased.

†††† If women are screened opportunistically in services without quality control, organized re-invitation should not be considered as a disadvantage;

CALL-RECALL

A call recall system is an invitation system where only those women from the target population who are not recently screened are invited. Women with a recent screening history are excluded from invitation

When opportunistic screening is already widespread, some countries restrict invitations to women who have not had a smear taken within the screening interval as in Denmark [37], Sweden [38] and Slovenia. This approach is acceptable when opportunistic smears are subjected to systematic quality control to avoid ineffectiveness and inequalities. One disadvantage of this system is that unnecessary smears are taken from women at low risk, who continue to be screened at high frequency in an opportunistic setting.

In some regions of France, a call-recall system is integrated in the French health care system, where screening remains essentially opportunistic [39]. All smears are registered including the identification of the patient and the smear taker, the data of specimen collection and the result. Quality assurance procedures must be accepted by all laboratories were the tests are performed. Women who have not had a smear reimbursed by the health insurance system are sent a personal letter within three years. No reminder is sent to non-participants.

RECALL SYSTEM

In a recall system, only women who were screened before and who are due for a subsequent smear are invited.

A recall system is often run in opportunistic systems by centres that invite their clients who contacted the service before. Recall systems do not contribute in reaching non-screened populations but are useful in maintaining continued coverage among previously screened women.

5.3.5. Examples of successful introduction of organized cervical cancer screening

The main objective of screening for cancer is to reduce mortality from the disease. In cervical screening, reducing the incidence of invasive disease is also an objective as pre-cancerous lesions are detected and treated before they develop into invasive cancer. Nowadays there is strong evidence that organised cervical cancer screening can reduce incidence and mortality substantially among screened women [3,40-43]. Firm evidence comes from the Nordic countries, where the implementation of widely different policies results in sharply contrasting trends in incidence and mortality. Concerning demonstration of the effect of organised screening, particularly important are the data on time trends in the incidence of invasive cervical cancer and the mortality from cervical cancer in the Nordic countries [40,42] where reliable national data are available from the period before screening programmes were implemented.

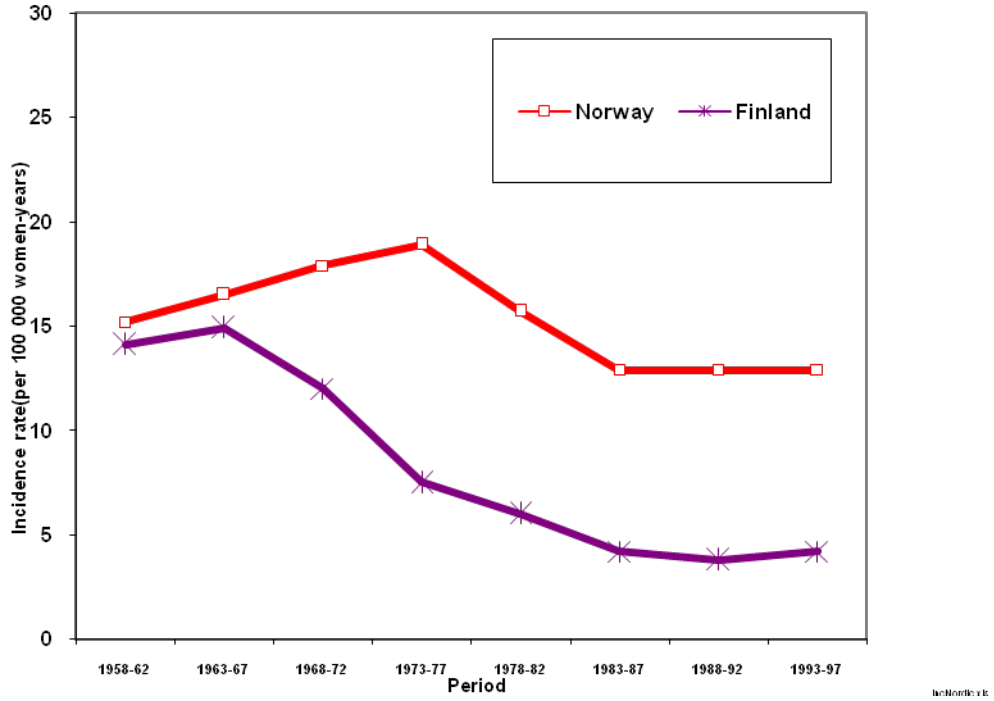


Figure 1. Age standardised incidence of cervical cancer in Finland and Norway (1958-96).

Time trends in mortality from cervical cancer in Denmark, Finland, Iceland, Norway, and Sweden since the early 1950s were investigated in relation to the extent and intensity of organised screening programmes in these countries. A clear parallelism was found between the population coverage achieved by organised screening programmes and the decline in the incidence of invasive cervical cancer.
In all five countries the cumulative mortality rates (0-74 years) fell between 1965 and 1982. In Iceland, where the nationwide programme has the widest target age range, the fall in mortality was greatest (80%). Finland and Sweden have nationwide programmes also; the mortality fell by 50% and 34%, respectively. In Denmark, where about 40% of the population is covered by organised programmes, the overall mortality fell by 25%, but in Norway, with only 5% of the population covered by organised screening, the mortality fell by only 10%. The most striking contrasts, between Finland and Norway, are shown in Figure 2.2.
These observations support the conclusion that organised screening programmes have had a major impact on the reduction in mortality from cervical cancer in the Nordic countries [42].

FINLAND
To compare the effectiveness of organised Pap smear screening with that of the spontaneous one on the incidence of invasive cervical cancer, a case-control study was conducted within the catchment's area of the Helsinki University Hospital (Helsinki, Finland). The study material consisted of 179 incident cases of invasive cervical cancer and 1,507 population controls. Data on lifetime Pap smears before the year of the cancer diagnosis were

collected using a self-administered questionnaire. The questionnaire information was obtained for 82% of the cases and 73% of the controls. The main outcome measure was the odds ratio associated with relative risk of invasive cervical cancer according to participation in organised and/or opportunistic screening compared to those with no history of screening. Non-screened women were the reference group. The odds ratio of invasive cervical cancer was 0.25 (CI 0.13- 0.48) for those who participated only in the organised screening programme, 0.57 (CI: 0.30-1.06) for those who had participated only in opportunistic screening and 0.27 (CI: 0.15-0.49) for those who had participated in both types of screening. The odds ratios were adjusted for age and the other type of the screening activity. These results indicate that the substantial decrease in the incidence of cervical cancer in Finland is mainly due to the organised mass screening [44].

UK

Although cervical cancer screening in England started in 1964, for over 20 years it failed to achieve sufficient coverage of women or an adequate follow-up of all women with cytological lesions. Near the end of the eighties it was also recognised that the incidence and even the mortality was rising among young cohorts [45]. A national call and recall system was set up in 1988. Financial incentives were first introduced with general practitioners contracts in 1990 [36]. The impact of this screening programme was assessed by trend analyses of incidence and cause-specific mortality and related to screening coverage and other indicator [46-49].

Quinn illustrated very clearly the tremendous impact of the new screening programme [46].

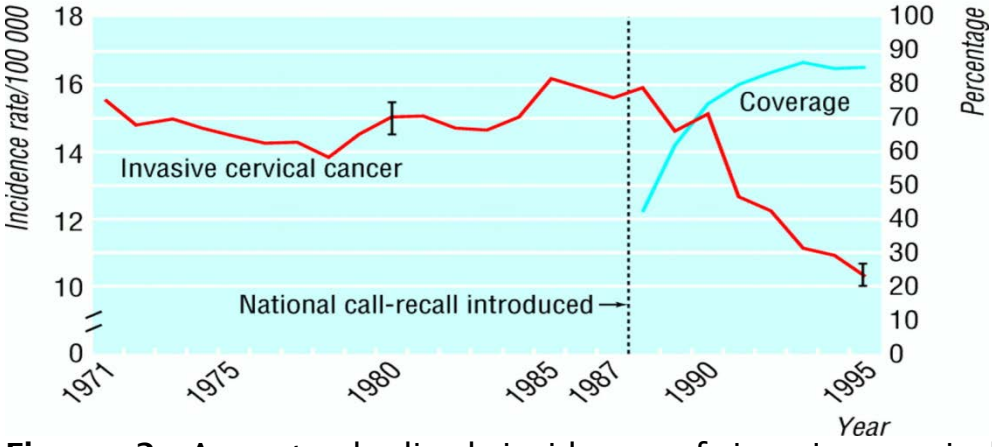


Figure 2. Age standardised incidence of invasive cervical cancer and coverage of screening, England, 1971-95 [46].

The coverage of the target group in the screening programme rose from 42% in 1988 to 85% in 1994, a level that was subsequently maintained. Coverage increased in all age groups, but particularly for older women (55 to 64 years).

Improvements in the screening programme have resulted in a 35% fall in incidence of invasive disease.

ITALY

Until recently cervical cancer screening in Italy has been mainly spontaneous, with only a few organised programmes. This has resulted in low coverage and high frequency of tests in screened women. The situation is, however, rapidly changing. In 1996 nation wide organised programmes on a regional basis, were recommended. National guidelines recommend personal invitation of women aged 25-64 years for a Pap test every third year. At the end of 1999, 34% of the Italian population 25-64 years old was included in organised programmes. Most organised programmes have a fail-safe system allowing picking up screen-positive women who skip follow-up visits. In recent years data have been collected in a standardised way by most organised programmes, allowing internal and external comparisons. An evaluation of the effectiveness of screening activity is therefore not easy. Only three cancer registries (Varese, Parma and Ragusa) have produced data for at least 10 years. They show a secular trend to a decreasing incidence. This is, probably, the result of spontaneous screening, but the proportion attributable to it is difficult to estimate [34]. In Florence, a significant trend towards a reduction in the incidence of invasive cancer was found. It was strongly associated with age-specific coverage, and thus most likely to be attributed to screening [50].

In Turin, where no trend towards a reduced incidence was present before start of organised programme in 1992, preliminary data for 1992-1995 show a very low incidence of interval cases after the first round, suggesting a high protection. The age-adjusted cervical cancer incidence ratio in 1992-98 was 0.81 (95% C.I. 0.59-1.09) for invited versus not invited women and 0.25 (95% CI 0.13-0.50) for attending versus non attending women [51].

A recent case-control study conducted in the Region of Firenze, indicated that protection against invasive cervical offered by cytology screening less than 3 years ago was elevated (OR= 0.15 (95% CI 0.07–0.31). Screening at an interval of 3 to 6 years was elevated as well in women of 40 years or older (OR=0.14; 95% CI: 0.06–0.33 but considerably lower if younger (OR= 0.45; 95% CI 0.14–1.48) [52]. There was no statistically significant protection against adeno-carcinoma.

DENMARK

In Denmark Pap smears started to be used in the late 1950s, and it has resulted in a decline over time in both cervical cancer incidence and mortality (see

Figure **3**). Nevertheless, considerable differences have been observed across Denmark in the organization of cervical cancer screening, because health care is under the responsibility of the sixteen counties. National guidelines, issued in 1986, recommended screening of women aged 23-59 years, every third year. The first cervical cancer screening pilot programme was set up in a small county in 1962, followed by the implementation of programmes in three larger counties in 1967/68. However, 30 years passed before screening was organized in the last county in 1996 [53]. In addition to the organized programmes, opportunistic screening activity expanded after 1969 when all smears started to be provided free of charge.

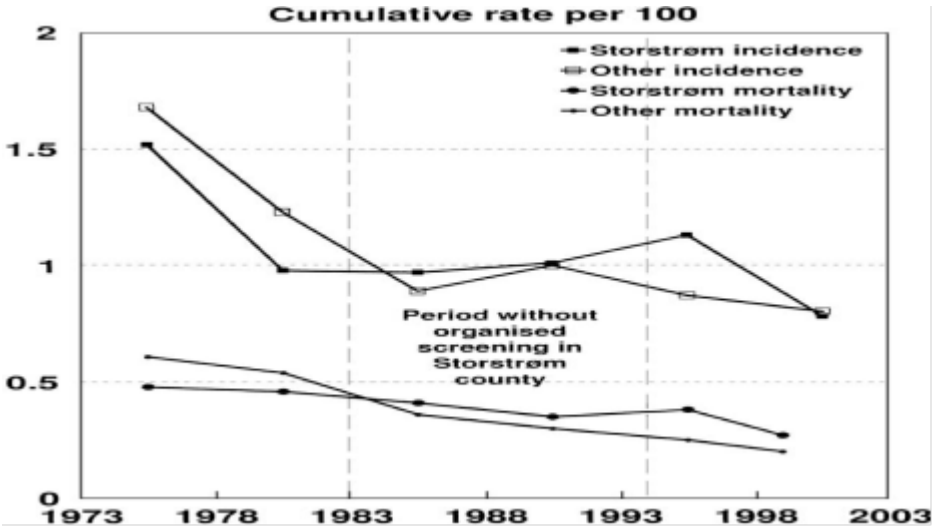


Figure 3. Cumulative rates per 100 for cervical cancer incidence and mortally 1973-2002 for women aged 30-64 for Storstrom county and other counties with long-term operating organised screening programmes in 1982 (from[53].

It is important to mention the particular development that took place in the Storstrom County when the organized programme was stopped at the end of 1982. It took another 11 years before the programme was restarted in 1994. Stopping the organized programme had a considerable impact on the screening coverage in the different age groups, where an opportunistic pattern developed after the organized programme stopped. Second, the 11 years' gap in the organized screening resulted in a statistically significant increase in incidence and mortality rates, which was observed at the restart of the organised programme.
During the interruption of organised screening in Storstrom the number of smears was higher than during the organized period before 1982. This experience shows that organization of screening should be a continued activity.
Opportunistic screening was for a long time the preferred approach of cervical cancer control in several Danish counties. The number of smears used in the opportunistic setting exceeded the number of smears needed for an organized programme, and the impact on the occurrence of invasive cancer is lower. It was estimated that close to 800 Danish women could have been spared the fate of becoming cervical cancer patients if organized screening programmes had been implemented nationwide at an earlier point in time.

THE NETHERLANDS
In 1996, the Dutch cervical cancer screening programme was restructured. The restructuring concerned the management and financing of the programme, organisation, target age ranges and interval, follow-up of abnormal results, and evaluation [35]. When comparing before (1996) and after (2003) the restructuration the most important achievements are the following [54]:

Substantial increase of the five-year coverage in the added target age groups (30-34, and 54-60) while in the old target age group (35-53 years) it remained around 80%.
Decrease of the proportion of screened women sent to follow-up from almost 19% to 3% per screening round.
Improvement of the follow-up compliance among screened women.
Shortening of the average time until a woman is either referred or rejoins the regular screening schedule
Reduction of the test positivity rate from over 10% up to approximately 2%
Reduction in the number of smears taken outside the target age group by 20% while maintaining high coverage rate. No increased interval cancer rate, in spite of less screening and lower percentage of women under follow-up was observed.

Compared to other countries with organised national programmes, the Netherlands has been successful in limiting the number of excess smears while maintaining a high coverage rate. The procedures in the Netherlands allow sorting out women with recent smears. Further, smears taken outside of the regular screening schedule are only reimbursed when the women have medical complaints. Last, Pap smear screening in the Netherlands is principally performed in the GP practices.

NORWAY
In Norway, a 20% of reduction in incidence of cervical cancer has been observed since the initiation of organised screening in 1995. This was achieved through more efficient use of Pap smears (by fixing the screening interval at 3 years) yielding lower consumption of screening examinations but also an increased population coverage. More details about the situation in Norway are given below.

5.3.6. Cost-effectiveness of different screening policies

Figure 4 shows the efficient cost-effectiveness frontier of optimal starting ages, number of scheduled examinations, and screening intervals, including cost-effectiveness of different screening policies in use in several old member states in the 1990s[36]. The costs and number of life years gained were computed assuming 100% participation of the target population, absence of excess Pap smears, average sensitivity and natural history parameters[36]. When moving toward a more intensive policy (starting at younger and ending at older age with shorter interval), the incremental cost-effectiveness ratio increased because the incremental effects rapidly diminish. Screening policies from Finland and the Netherlands were remarkably close to the efficient frontier. Screening every year starting at young adult age without upper age limit, as recommended in Austria, Germany and Luxembourg (>50 smears/lifetime), yielded a rather small gain in life years but at a cost that was dramatically high (figure 1). The costs per percentage reduction of life-years lost due to cervical cancer estimated for the German screening policy (yearly intervals, 50 smears per lifetime) are

approximately five times greater than for the Finish or Dutch policy (5-yearly screening)[44].

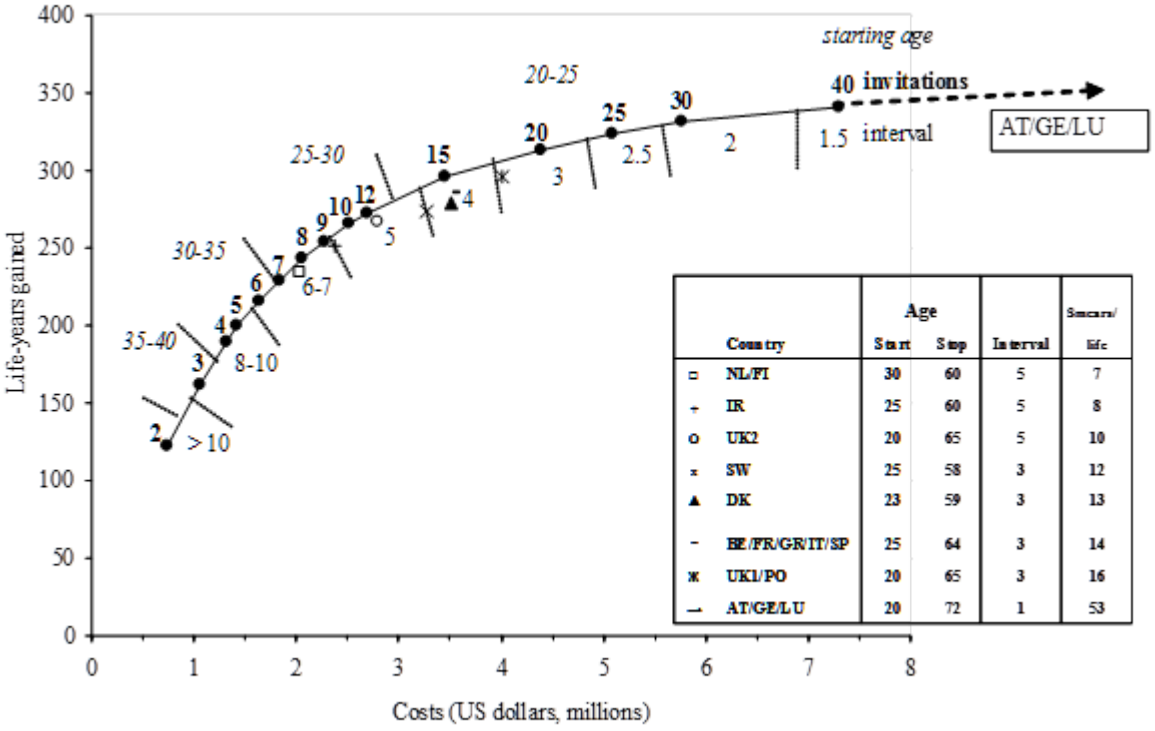


Figure 4. Schematic representation of the simulated efficient frontier showing the location of optimal starting ages, number of scheduled examinations, and screening intervals, including a comparison of the costs and effects for screening policies used in countries with a cervical screening program or program recommended in national guidelines. The starting age ranges (in years), number of invitations and screening intervals (in years) are indicated above, on, or under the curve, respectively.
The estimated life-years gained (per 1,000,000 screened women) & costs are shown for 9 screening policies in place in EU member states in the 1990s (AT=Austria, BE=Belgium, DK=Denmark, FI=Finland, FR=France, GE=Germany, GR=Greece, IR=Ireland, LU=Luxembourg, NL=the Netherlands, PO=Portugal, S=Sweden, SP=Spain, UK=United Kingdom), using a discount rate for costs & effects of 3% (adopted from Van den Akker, et al [36]).

5.3.7. Three highlighted screening programmes

In this paragraph we will discuss which parameters are essential in a well-organised cervical cancer screening programme and which determine its success and cost-effectiveness. We will illustrate their importance by highlighting aspects of three screening programmes in Europe: Norway, Sweden and the Netherlands.

5.3.7.1. The Norwegian cervical cancer screening programme

In Norway, a centralized system has been set up comprising obligatory registration of screen tests carried out in the organised programme or in an opportunistic setting. The Norwegian screening programme was introduced in 1995. It is population-based, nationwide, and recommends women of 25 to 69 years of age to have a Pap smear taken every 3 years. However in Norway, spontaneous screening activities were present since the early 1960s. Those activities brought a 50% reduction of the invasive cervical cancer incidence in women of 40-59 years old [55]. From 1990, the incidence has remained stable. Several attempts during the 1960s, 1970s and 1980s to introduce an organised screening programme in Norway failed. In that period screening was characterised by frequent testing of young women at low risk and low-coverage in among women older than 50 and among women at high risk..

The introduction of a screening programme into a population where screening is already widespread poses problems different from those when implementing a novel programme. The Norwegian challenge was therefore to try to implement an organised programme in coordination with the spontaneous screening activity. The choice was made to integrate spontaneous screening into the organised programme in order to minimise changes in the healthcare system. By establishing a Cytology Register that registered every Pap smear taken in Norway, and by linking information at the individual level, the Norwegian coordinated screening programme started posting recommendation letters in 1995 to women who had not had a Pap smear in the previous 3 years [56]. The main purpose of the coordinated screening programme was to increase coverage, especially for women older than 50.

The impact of organised screening was assessed by comparing Pap smear use, screening coverage and incidence of invasive cervical cancer in the 4 years (1992-5) before start of the programme with the two subsequent screening rounds.

After the introduction of the programme, a substantial increase in coverage was observed, particularly in the age group 50-69 years (see Figure 5). In the last 2 years studied, the incidence of invasive cancer was 22% lower than in

the period just before the start of the organised programme. The 3-year coverage in the 25-67 year age group in the period 2001-4 increased with about 7% compared to the period 1992-1995. However, this increase in coverage was accompanied by a decrease in the average number of yearly smears used (533 thousand versus 494 thousand).

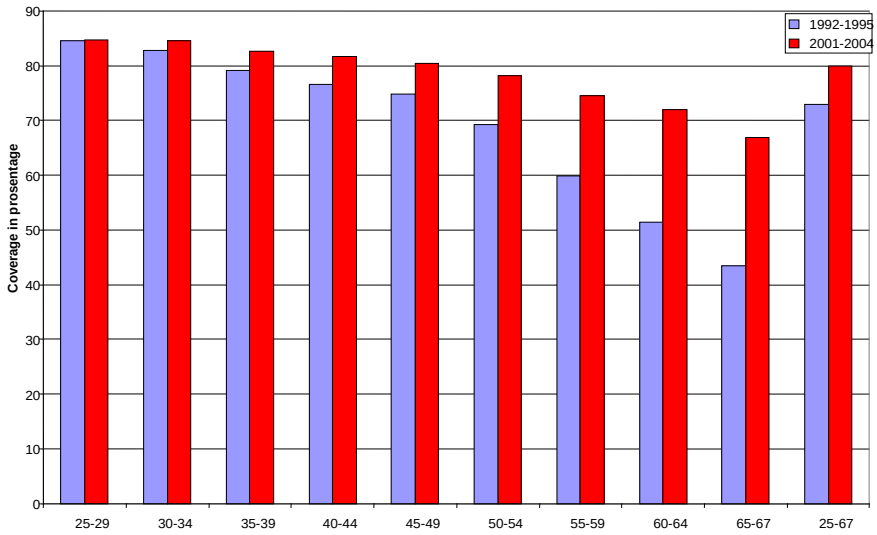


Figure 5. Coverage 1992-1995 and 2001-2004, Norway (from Jan Nygard)

The Norwegian programme has demonstrated that it is possible to mobilise resources spent in over-screening towards higher coverage and lower cancer incidence.

5.3.7.2. Decentralised screening programmes in Sweden

Sweden has a decentralised cervical cancer screening programme. Organised cervical screening was first implemented in Sweden in the mid-1960s. Pap smears are also taken outside the screening programme by gynaecologists, midwives and general practitioners [57]. Organised screening and opportunistic use of Pap smear have been in existence for several decades in Sweden. A marked decline in cervical cancer incidence could be attributed to the time point of start of screening. In the period 1959-1963, the age-standardised incidence of cervical cancer in Sweden was 20.6/100 000. Following the introduction of organised screening, there has been a regular decline and in the period 1989-1993, the standardised incidence was 10.1/100 000/year. The Swedish screening policy recommends 3-yearly Pap tests between 23 and 50 years of age and 5-yearly tests between 50 and 60 years of age.

The healthcare in Sweden is organised regionally in each county (26 in total). The different counties implemented organised screening according to the

national guidelines for cervical cancer screening issued in 1985 where it was recommended that all women between 20 and 59 years of age should be screened every third year. It was also stated that quality assurance in terms of smear usage records should be maintained and registry linkages with cancer registries be set up. The population registry is used. Every person has a personal identification number (PIN) and screening registries, cancer registries, pathology and cytology registries are all based on the PIN [38].
Sweden applied a call-recall invitation system. By a linkage with the population register and cytology registries all women who had a spontaneous smear taken in the past 18 months are sorted out and not invited for screening. The situation is heterogeneous in with respect to coverage and consumption of Pap tests. In certain counties over-screening is a substantial problem. The very high (86%) coverage of Pap smears in Stockholm has also led to a remarkable decrease in both incidence and mortality of cervical cancer. Rodvall demonstrated, in the Stockholm area, that organised screening reached high coverage among certain groups such as older and immigrant women, which are usually less covered in opportunistic settings.
The screening programme in Sweden is heterogeneous in quality. The new national guidelines seek to remedy some of the major limitations in particular by means of a national working group responsible for reviewing the programme.
It is of particular importance that registration of screening tests, in Norway and in Sweden, is compulsory and based on the national ID numbers and allowing linkages with other databases. Besides the advantages, outlined above for public health, this system offers enormous possibilities for bio-bank research.

5.3.7.3. Intensity of cervical cancer screening and impact in the Netherlands vs the USA

In a recent comparative effectiveness study, Habbema et al analysed the evolution in intensity of the use of Pap smears in the Netherlands and the USA (Figure 6) and plotted this against the trend of cervical cancer incidence and mortality (Figure 7). In the Netherlands, a well-organised screening programme has been set up since the xxx, whereas in the United States, screening for cervical cancer has been manly opportunistic. Although the consumption of Pap smears (expressed as numbers/1000 women, standardised for age) was three to four times more larger in the US compared to the Netherlands, very similar decreases in incidence (Figure 7) and in cause-specific mortality were observed. A second study is currently going which shows that in USA much more abnormal cytology results were noted and more diagnostic and therapeutic procedures were performed resulting not only in more costs but also more adverse effects.

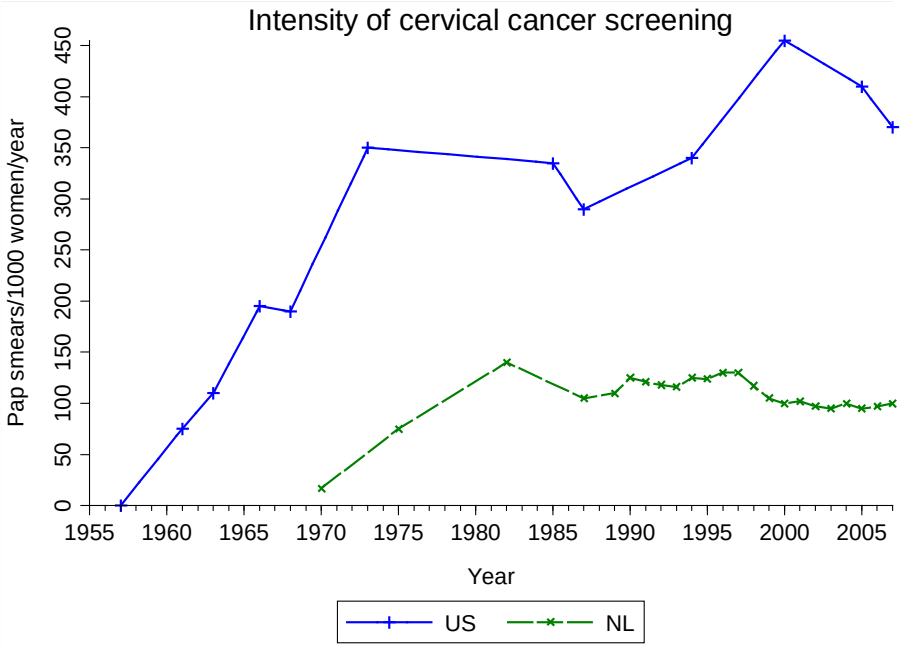


Figure 6. Evolution in the intensity of cervical cancer screening in the US and the Netherlands, expressed as the standardized prevalence of an annual Pap smear per 1,000 women women/year (using the 2000 female population in the US as reference) (Source: Habbema et al, 2012)[58].

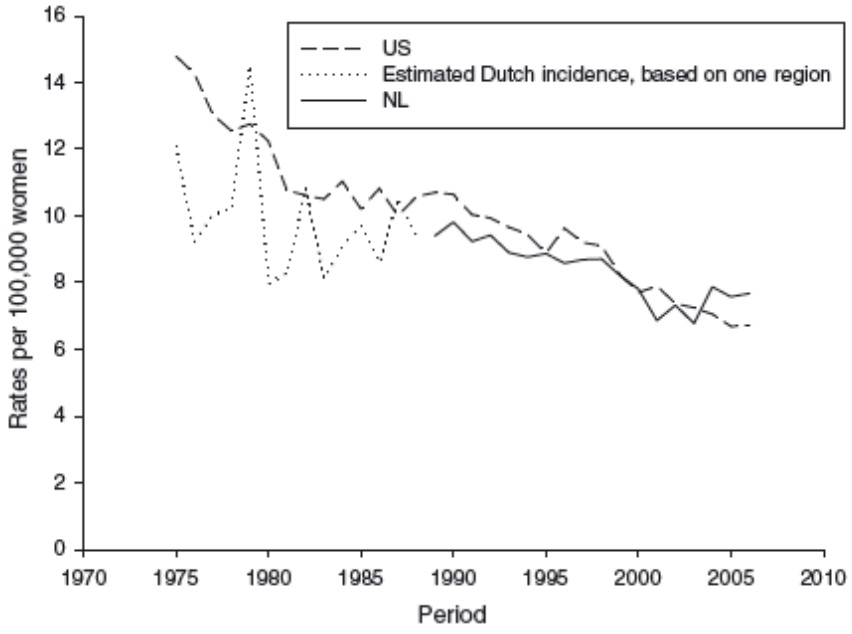


Figure 7. Trend of the standardised incidence rate of cervical cancer in the US and the Netherlands (using the US 2000 population as reference) (Source: Habbema et al, 2012)[58]

5.4. Discussion and conclusion

The European Council recommends offering organised screening for cervical cancer in agreement with evidence-based guidelines. The examples outlined above suggest that organised screening is more efficient and

Despite evidence indicating greater effectiveness and cost-effectiveness of organised screening and in spite of the European Council Recommendation, detection of cervical cancer precursors remains mainly opportunistic in a majority of the EU states, including Germany. It should be considered as a compelling responsibility for national or regional health authorities of these countries to set up organised programmes preferably extending over the whole country in agreement with current European Guidelines for Quality Assurance for Cervical Cancer Screening. Stakeholders and health professionals must understand that organised screening is not a question of economy to save resources for the public treasury but is, first of all, a question of optimising the effectiveness and minimising the adverse effects.

It should be realised that converting an established opportunistic practice into an organised programme is not easy and not always provides the expected positive result. An invitational system is often proposed as an intrinsic component of an organised programme. As an example, we can refer to findings from Belgium. The screening coverage for Pap smears can be computed precisely, by analysing health insurance databases[24,25]. In the period 2002-2006, a small increase in screening coverage[‡‡‡‡] of two percent was observed in province with a call-recall system whereas in another province where the invitation system was abandoned (East-Flanders) an increase of even four % was observed [24]. Piloting, careful planning and evaluation and health-services research, and comparative effectiveness research are strongly recommended for countries planning to switch from opportunistic to organised cervical cancer screening, a fortiori when new screening and/or triage methods will be implemented.

[‡‡‡‡] Screening coverage defined as % of women aged 25-64 with a Pap smear in the last 3 years.

5.5. References

1. Council of the European Union. Council Recommendation of 2 December 2003 on Cancer Screening. *Off J Eur Union* 2003;**878**: 34-8.

2. Anttila A, Ronco G, Clifford G, et al. Cervical cancer screening programmes and policies in 18 European countries. *Br J Cancer* 2004;**91**: 935-41.

3. IARC. Cervix Cancer Screening. IARC Handbooks of Cancer Prevention. Vol. 10. Lyon: IARCPress; 2005.

4. Hulstaert F, Arbyn M, Huybrechts M, Vinck I, Puddu M, Ramaekers D. Cervical cancer screening and HPV. Brussels: Federaal Kenniscentrum voor de Gezondheidszorg; Centre fédéral d'Expertise des Soins de Santé (Belgian Health Care Knowledge Centre); 2006. Report No.:

5. Arbyn M, Rebolj M, de Kok IM, et al. The challenges for organising cervical screening programmes in the 15 old member states of the European Union. *Eur J Cancer* 2009;**45**: 2671-8.

6. European Commission. European Guidelines for Quality Assurance in Cervical Cancer Screening. 2nd ed. Editors: Arbyn M, Anttila A, Jordan J, Ronco G, Schenck U, Segnan N et al. Luxembourg: Office for Official Publications of the European Communities; 2008.

7. The Council of the European Union. Council Recommendation of 2 December on Cancer Screening. *Off J Eur Union* 2003;**878**: 34-8.

8. Arbyn M, Anttila A, Jordan J, et al. European Guidelines for Quality Assurance in Cervical Cancer Screening. Second Edition - Summary Document. *Ann Oncol* 2010;**21**: 448-58.

9. Advisory Committee on Cancer Prevention. Recommendations on cancer screening in the European Union. Advisory Committee on Cancer Prevention. *Eur J Cancer* 2000;**36**: 1473-8.

10. Proposal for a Council Recommendation on Cancer Screening., 2003/0093 (CNS), Commission of the European Communities, (2003).

11. Storm H, Buiatti E, Hakulinen T, Ziegler H. Guidelines on confidentiality in population-based cancer registration in the European Union. Lyon; 2004. Report No.:

12. Arbyn M, Wallyn S, Van Oyen H, et al. The new privacy law in Belgium: a legal basis for organised cancer screening. *Eur J Health Law* 1999;**6**: 401-7.

13. Anttila A, Ronco G, Lynge E, Fender M, Arbyn M, Baldauf JJ, et al. Chapter 2:Epidemiological guidelines for quality assurance in cervical cancer screening. In: Commission of the European Communities, editor. European Guidelines for Quality Assurance in Cervical Cancer Screening.Luxembourg: 2006. p. 1-43.

14. Miller AB. The (in)efficiency of cervical screening in Europe. *Eur J Cancer* 2002;**38**: 321-6.

15. Arbyn M, Van Oyen H. Cost-analysis of cervical cancer screening in the Flemish Region: the spontaneous versus the organised approach. *Eur J Cancer* 1997;**33**: 15-6.

16. Hillemanns P, Soergel P, Jentschke M. Cervical cancer screening. State of discussion. *Gynakologe* 2014;**47**: 334-9.

17. Advisory Committee on Cancer Prevention, Lynge E. Recommendations on cancer screening in the European union. Advisory Committee on Cancer Prevention. *Eur J Cancer* 2000;**36**: 1473-8.

18. Sasieni P, Adams J, Cuzick J. Benefit of cervical screening at different ages: evidence from the UK audit of screening histories. *Br J Cancer* 2003;**89**: 88-93.

19. Kyrgiou M, Koliopoulos G, Martin-Hirsch P, et al. Obstetric outcomes after conservative treatment for intra-epithelial or early invasive cervical lesions: a systematic review and meta-analysis of the literature. *Lancet* 2006;**367**: 489-98.

20. Linos A, Riza E. Comparisons of cervical cancer screening programmes in the European Union. *Eur J Cancer* 2000;**36**: 2260-5.

21. Arbyn M. An overview of cervical cancer screening systems in Europe. 3rd European Congress for Colposcopy and Cervical Pathology; 2004 Jan 24; Paris 2004 p. 1-46.

22. van Ballegooijen M, van den Akker van Marle ME, Patnick J, et al. Overview of important cervical cancer screening process values in EU-countries, and tentative predictions of the corresponding effectiveness and cost-effectiveness. *Eur J Cancer* 2000;**36**: 2177-88.

23. Anttila A, von Karsa L, Aasmaa A, et al. Cervical cancer screening policies and coverage in Europe. *Eur J Cancer* 2009;**45**: 2649-58.

24. Arbyn M, Fabri V, Temmerman M, Simoens C. Attendance at cervical cancer screening and use of diagnostic and therapeutic procedures on the uterine cervix assessed from individual health insurance data (Belgium, 2002-2006). *PLoS ONE* 2014;**9**: e92615.

25. Arbyn M, Simoens C, Van Oyen H, et al. Analysis of 13 million individual patient records pertaining to Pap smears, colposcopies, biopsies and surgery on the uterine cervix (Belgium, 1996-2000). *Prev Med* 2009;**48**: 438-43.

26. Rousseau A, Bohet P, MerliŠre J, et al. Evaluation du dépistage organisé, et du dépistage individuel du cancer du col de l'utérus: utilité, des données de l'Assurance maladie. *Bull Epidemiol Hebdom* 2002;**19**: 81-4.

27. Mund M, Knöpnadel J, Schenck U, Becker N. Screening for cervical cancer in Germany. In: Description of the national situation of cervical cancer screening in the member states of the European Union. *Eur J Cancer* 2009;**45**: 2685-708.

28. Simou E, Maniadakis N, Pallis A, Foundoulakis E, Kourlaba G. Factors associated with the use of Pap smear testing in Greece. *J Women's Health* 2010;**19**: 1577-85.

29. Oliveira M, Peleteiro B, Lunet N. Cytology use for cervical cancer screening in Portugal: results from the 2005/2006 National Health Survey. *Eur J Public Health* 2014;**24**: 253-8.

30. Puig-Tintore LM, Castellsague X, Torne A, et al. Coverage and factors associated with cervical cancer screening: results from the AFRODITA study: a population-based survey in Spain. *J Low Genit Tract Dis* 2008;**12**: 82-9.

31. Dillner J. Nationellt Kvalitetsregister för Cervixcancerprevention. Svensk förening för Obstetrik och Gynekologi (SFOG) 2013

32. Veerus P, Arbyn M, Amati C, Baili P, EUROCHIP 3. Impact of implementing a nationwide cervical cancer screening programme on population coverage in Estonia. *Tumori* 2010;**96**: 524-8.

33. Anttila A, Nieminen P. Cervical cancer screening programme in Finland. *Eur J Cancer* 2000;**36**: 2209-14.

34. Segnan N, Ronco G, Ciatto S. Cervical cancer screening in Italy. *Eur J Cancer* 2000;**36**: 2235-9.

35. van Ballegooijen M, Hermens R. Cervical cancer screening in the Netherlands. *Eur J Cancer* 2000;**36**: 2244-6.

36. Patnick J. Cervical cancer screening in England. *Eur J Cancer* 2000;**36**: 2205-8.

37. Coleman D, Day N, Douglas G, et al. European Guidelines for Quality Assurance in Cervical Cancer Screening. Europe against cancer programme. *Eur J Cancer* 1993;**29A Suppl 4**: S1-S38.

38. Dillner J. Cervical cancer screening in Sweden. *Eur J Cancer* 2000;**36**: 2255-9.

39. Schaffer P, Sancho-Garnier H, Fender M, et al. Cervical cancer screening in France. *Eur J Cancer* 2000; 2215-20.

40. Hakama M. Trends in the incidence of cervical cancer in the Nordic countries. In: Magnus K, editor. Trends in Cancer Incidence.Washington: Hemisphere Publishing Corporation; 1982. p. 279-92.

41. Hakama M, Miller AB, Day NE. Screening for cancer of the uterine cervix. From the IARC Working Group on Cervical Cancer Screening and the UICC Project Group on the Evaluation of Screening Programmes for Cancer. WHO, IARC, UICC, editors. *IARC Sci.Publ.* 1-315. 1986.

42. Laara E, Day NE, Hakama M. Trends in mortality from cervical cancer in the Nordic countries: association with organised screening programmes. *Lancet* 1987;**1**: 1247-9.

43. Sankila R, Demaret E, Hakama M, Lynge E, Schouten LJ, Parkin DM. Evaluation and monitoring of screening programmes. Luxembourg: Office for Official publications of the European Communities; 2000.

44. Nieminen P, Kallio M, Anttila A, Hakama M. Organised vs. spontaneous Pap-smear screening for cervical cancer: A case-control study. *Int J Cancer* 1999;**83**: 55-8.

45. Cook GA, Draper GJ. Trends in cervical cancer and carcinoma in situ in Great Britain. *Br J Cancer* 1984;**50**: 367-75.

46. Quinn M, Babb P, Jones J. Effect of Screening on Incidence an Mortality from Cancer of the Cercix in England: Evaluation Based on Routinely Collected Statistics. *BMJ* 1999;**318**: 904-8.

47. Sasieni P, Cuzick J, Farmery E. Accelerated decline in cervical cancer mortality in England and Wales. *Lancet* 1995;**346**: 1566-7.

48. Sasieni PD, Adams J. Analysis of cervical cancer mortality and incidence data from England and Wales: evidence of a beneficial effect of screening. *J Royal Stat Soc* 2000;**163**: 191-209.

49. Peto J, Gilham C, Fletcher O, Matthews FE. The cervical cancer epidemic that screening has prevented in the UK. *Lancet* 2004;**364**: 249-56.

50. Ciatto S, Cecchini S, Iossa A, et al. Trends in cervical cancer incidence in the district of Florence. *Eur J Cancer* 1995;**31A**: 354-5.

51. Ronco G, Pilutti S, Patriarca S, et al. Impact of the introduction of organised screening for cervical cancer in Turin, Italy: cancer incidence by screening history 1992-98. *Br J Cancer* 2005;**93**: 376-8.

52. Zappa M, Visioli CB, Ciatto S, et al. Lower protection of cytological screening for adenocarcinomas and shorter protection for younger women: the results of a case-control study in Florence. *Br J Cancer* 2004;**90**: 1784-6.

53. Lynge E, Clausen LB, Guignard R, Poll P. What happens when organization of cervical cancer screening is delayed or stopped? *J Med Screen* 2006;**13**: 41-6.

54. Rebolj M, van Ballegooijen M, Berkers LM, Habbema JDF. Monitoring a national cancer prevention programme: Successful changes in cervical cancer screening in the Netherlands. 2006; 1-29.

55. Nygard JF, Skare GB, Thoresen SO. The cervical cancer screening programme in Norway, 1992-2000: changes in Pap smear coverage and incidence of cervical cancer. *J Med Screen* 2002;**9**: 86-91.

56. Bjorge T, Gunbjorud AB, Haugen OA, et al. Mass screening for cervical cancer in Norway: evaluation of the pilot project. *Cancer Causes Control* 1995;**6**: 477-84.

57. Rodvall Y, Kemetli L, Tishelman C, Tornberg S. Factors related to participation in a cervical cancer screening programme in urban Sweden. *Eur J Cancer Prev* 2005;**14**: 459-66.

58. Habbema D, de Kok IM, Brown ML. Cervical cancer screening in the United States and the Netherlands: a tale of two countries. *Milbank Q* 2012;**90**: 5-37.

6. Question: Can self-sampling (at home) increase screening attendance?

6.1. Background and rationale

Well organized screening programs have reduced the incidence of cervical cancer substantially in industrialized countries[1-3]. Within these screening programs, cervical cytology currently is the recommended test[4-6]. However, screening coverage is not always optimal[7;8]. In high-income countries, a considerable number of women who are diagnosed with cervical cancer have never had a Pap test or were infrequently screened[9;10]. Several reasons have been identified as to why women do not attend cervical cancer screening. Barriers for participation in cytological screening do not only include logistical (i.e. transport to the clinic, inconvenient clinic hours), economical, and educational factors, but also personal-level factors[11]. Among the latter, embarrassment and 'being uncomfortable with a clinician taking a sample' are most often given as the most discouraging factors.

Self-sampling followed by HPV-testing has been presented as a means to improve screening coverage. Indeed, many studies demonstrate an overall positive attitude of women towards self-sampling and a high acceptance[12-14]. The comparability of self-collected samples versus clinician-collected samples for HPV-testing is beyond the scope of this systematic review and will be discussed elsewhere.

In this systematic analysis, the coverage-increasing potential of self-sampling and HPV-testing was evaluated, and compared invitations for conventional screening with cytology.

6.2. Question

Can self-sampling (at home) increase screening attendance?

6.2.1. PICOS

P: Women not screened or under-screened for cervical cancer, in particular women non-responding to invitation letters

I: Invitation to take a self-sample for HPV-testing (by mail, door-to-door offering, etc.)

C: Invitation to undergo a conventional screening test, performed by a clinician (invitation or reminder letter, door-to-door invitation, coupon for free testing in clinic, etc.)

O: Response rate to the invitation, test-positivity rate, sample quality, and compliance to

follow-up among test-positive women in the intervention and comparator groups and ratios intervention vs. comparator groups

S: Randomized intervention studies

Controled non-randomized intervention studies

6.2.2. Importance of outcomes

Detection of cervical precancer (CIN2+, CIN3+) belongs to outcome level 5 (cross-link) whereas participation of the target population and compliance of screen-positive women correspond with outcome level 6 in the list of positive outcomes. The PPV of HPV testing on self-samples and of the conventional test on a clinician sample could also be addressed from retrieved studies. However, PPV can be assessed in more detail in the PICOS addressing the accuracy of self-sampling (cross-link).

6.3. Methods

A recent systematic review related to the PICOS was identified[15]: Racey CS, Withrow DR, Gesink D. Self-collected HPV Testing Improves Participation in Cervical Cancer Screening: A Systematic Review and Meta-analysis. *Can J Public Health* 2013; 104(2):e159-e166. No reviews were previously conducted by the Unit of Cancer Epidemiology. The review by Racey et al. (2013) was evaluated using the AMSTAR tool{Shea, 2009 31674 /id} and updated by including new studies.

6.3.1. Search strategy

The search strategy that was described in the review of Racey et al., was implemented in order to update the list of included studies. Two electronic databases, Medline (Pubmed) and Embase, were searched using the search strings listed in Box 1. The literature search in Racey et al. covered studies with a publication date up to mid-2012. Therefore, in the new search performed by the Unit of Cancer Epidemiology, studies with a publication date in the year 2012 and 2013 were retrieved.

Medline:
[Papillomavirus infections OR cervical intraepithelial neoplasia OR uterine cervical neoplasmas OR vaginal smears OR papillomaviridae]
AND
[Self care OR patient acceptance of health care OR Self-sampl* or Self test*]
AND
[HPV test*]

Embase:
[Vagina smear OR papilloma virus OR papillomavirus infection OR wart virus OR papilloma OR uterine cervix carcinoma OR uterine cervix carcinoma in situ]
AND
[Self evaluation OR patient participation OR patient compliance OR self care OR Self test* OR Self sampling OR Self sampl* OR Self sampling Human papillomavirus test]
AND
[Cancer screening OR HPV test*]

Box 1: Search string set up by Racey et al [15].

6.3.2. Eligibility of studies

In line with the strategy set up by Racey et al. [15], the literature search was restricted to peer-reviewed articles that compared HPV testing on a self-sample to standard Pap testing in women who did not routinely participate in cervical cancer screening programs. Retrieval of relevant studies was restricted to developed countries where Pap testing is the standard for cervical cancer screening..

Studies were included if group allocation was clearly described and attendance rates were reported for both the intervention group and the control group. Accepted control groups were those that offered a standard invitation to undergo Pap testing at a local Health Care Clinic or that offered Pap testing via the normal procedures of the jurisdiction within which the study was conducted. Studies that employed an ecological design were excluded as individual-level crude rates of compliance in testing could not be determined. Conference abstracts, editorials, commentaries and other unpublished manuscripts were excluded, in addition to articles that included duplicate datasets or male participants. There were no language restrictions on publications included.

Eligibility of the studies was appraised by F.V. and was subsequently revised by M.A.

6.3.3. Quality of studies and data extraction

The quality of included studies was evaluated using the Cochrane tool for risk of bias in randomised trials[16]. The quality of the systematic review by Racey et al. was assessed using AMSTAR tool[17;18]. Data extraction from the more recent studies was conducted by F.V. and revised by M.A[19].

6.3.4. Outcome measures

The following outcome measures were assessed:

- Absolute attendance rates in the self-sampling and the control arm.
- Relative attendance rate and attendance rate difference in the self-sampling arm, compared to the control arm.

When data were available, the following outcome measures were assessed as well:

- Sample quality in the self-sampling arm.
- Test-positivity rate in the self-sampling and the control arm.
- Absolute compliance to follow-up instructions in the self-sampling arm and the control arm.
- Relative compliance and difference in compliance in the self-sampling arm, compared to the control arm.

6.3.5. Statistical Analysis

The overall pooled measures, with 95% confidence intervals were calculated using a random effects model[20]. The statistical heterogeneity was assessed using the I^2 statistic, which measures the variation across studies that is due to inter-study heterogeneity rather than chance. Statistical analysis was performed using STATA/SE 10 (Stata Corporation, College Station, TX, USA).

6.4. Results

6.4.1. PRISMA flowchart

A 'Preferred Reporting Items for Systematic Reviews and Meta-Analyses' (PRISMA) flow chart of the exclusion process is depicted in Figure 1. A detailed list of included and excluded studies (+ reason for exclusion) is presented in a recent publication derived from the updated systematic review[19]. Additional to the ten studies included by Racey et al.[21-30], two studies were judged relevant for the PICOS[31;32]. Data of the study of Piana et al. (2011)[26], which was included in the systematic review of Racey, was overruled by a more recent report[32] in the review presented here. Hence a total of 11 studies were selected for the updated systematic review[21-25;27-32].

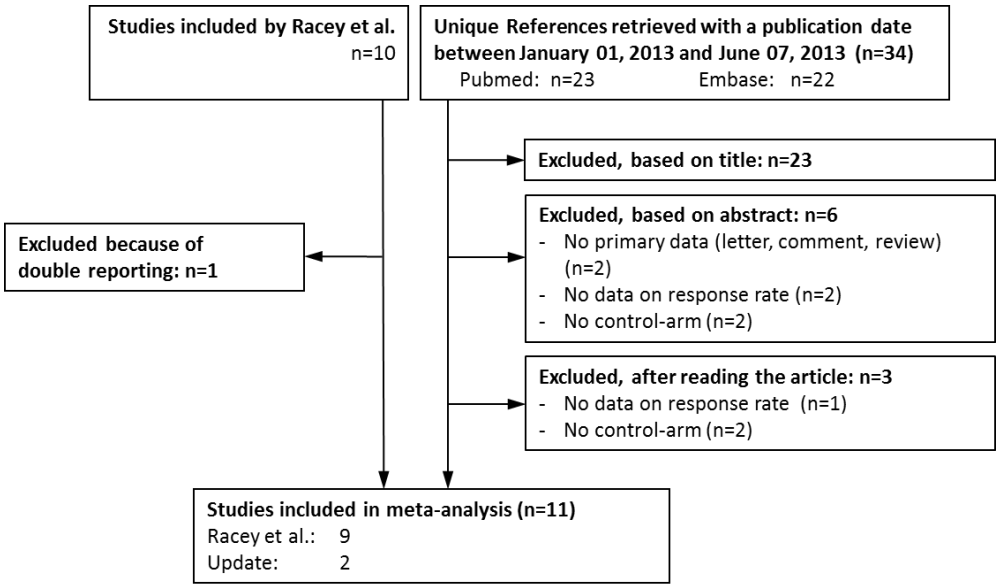


Figure 1: Prisma Flow chart summarizing the selection of relevant studies.

6.4.2. Study characteristics and quality assessment

Of the 11 included studies, the population and study characteristics, and the test, triage and follow-up characteristics are listed in Table 1 and Table 2, respectively.

The quality of these studies was evaluated using the Cochrane tool for risk of bias in randomised trials [16]. A summary table is presented in Table 3.

The study of Castle et al. (2011)[23] did not have a randomised design since allocation was performed by preference of the participant. Hence, this study was given a high risk of bias regarding Random Sequence generation and allocation concealment. Bais et al. (2007) describe that women were allocated on their invitational-procedure number and full concealment of allocation could therefore not be guaranteed and was considered at high risk for bias. Blinding of assessment of participation and test results was not applicable in these studies. Outcome data, at least for the participation rate, was documented in all studies. Timelines were sufficiently reported in all but three studies.

Table 1: Population & study characteristics

Author, year Country	**Study design**	**Population/ setting**	**Method of invitation in self-sampling arm**	**Method of invitation in control arm**	**Self-sampling Arm**	**Control arm**	**Age (range)**
Bais, 2007 The Netherlands	RCT	Women who did not respond to 2 invitations for screening. Urban setting.	Direct mailing of the self- sampling kit.	Extra recall for conventional cytology with an explanatory letter.	2352	272	30-50
Gok, 2010 The Netherlands	RCT	Women who had not had a Pap test in 5 years and did not respond to 1 invitation for screening. Urban setting.	Direct mailing of the self- sampling kit, preceded by a notification.	A second recall for conven-tional cytology.	26886	277	30-60
Castle, 2011 USA	Not randomized cohort study	Women who had not had a Pap test in the last 3 years. Rural setting.	Door-to-door recruitment.	Door-to-door recruitment.	77	42	26-65
Giorgi-Rossi, 2011 Italy	RCT	Women who did not respond to 1 regular invitation for screening. Urban and rural settings.	- Direct mailing of the self-sampling kit, preceded by a notification. - Invitation letter to phone to call center to receive a self-sampling kit.	Recall for conventional cytology.	616	619	35-65
Lazc.-P., 2011 Mexico	RCT	Women in poverty-reduction program, with limited access to health services. Rural setting.	Nurses performed home visits, where a self-sample was taken by the woman herself.	Nurses performed home visits, and made an appointment for conventional cytology in the clinic.	9371	12731	25-65
Szarewski, 2011 United Kingdom	RCT	Women who did not respond to 2 invitations for screening. Urban setting.	Direct mailing of the self-sampling kit.	Recall for conventional cytology.	1500	1500	25-64 (median 48y)
Virtanen,	RCT	Women who did not	Direct mailing of the	Recall for conventional	2397	6302	30-60

Author, year Country	Study design	Population/ setting	Method of invitation in self-sampling arm	Method of invitation in control arm	Self-sampling Arm	Control arm	Age (range)
2011 Finland		respond to 2 invitations for screening. Urban setting.	self-sampling kit, preceded by a notification.	cytology.			
Wikstrom, 2011 Sweden	RCT	Women who had not participated in screening for >6 years. Urban setting.	Direct mailing of the self-sampling kit, preceded by a notification.	Invitation to a midwife reception for conventional cytology, within the framework of the organised screening program.	2000	2060	39-60
Gok, 2012 The Netherlands	RCT	Women who had not attended cervical cancer screening in the last year after a reminder invitation for screening. Urban setting.	Direct mailing of the self-sampling kit, preceded by a notification.	A second recall for conventional cytology.	25561	261	30-60
Darlin, 2013 Sweden	RCT	Women who had not had any cervical smears taken for >9y.	Direct mailing of the self- sampling kit.	Recall for conventional cytology at an outpatient clinic. The invitation included several alternative appointments.	1000	500	32-65
Sancho-Garnier, 2013 France	RCT	Women who did not respond to 1 regular invitation for screening and had not had a Pap test in ≥2y.	Direct mailing of the self-sampling kit, preceded by a notification.	Recall for conventional cytology at an outpatient clinic. The invitation included a list of centers performing the test.	8829	9901	35-69

Table 2: Test, triage & follow-up characteristics

Author, year	Tests	Self-sampling device	Time of response assessment (months after invitation)	Triage of test+	Follow-up
Bias, 2007	PCR genotyping	Cervicovaginal brush	6m	No triage	- Cytology + colposcopy + colpo-directed biopsy
Gok, 2010	HC2	Lavage (Delphi screener)	12m	Self-arm: cytology + repeat HPV	- Colposcopy + colpo-directed biopsy, in case of abnormal cytology (ASC-US+) - repeat testing (Pap + HPV) in 1y, in case of normal cytology
Castle, 2011	HC2	Swab	ND	No triage	Colposcopy + colpo-directed biopsy, in case of screen test+
Giorgi, 2011	HC2	Lavage	3m	Colposcopy	- Cytology in 1y, in case of negative colposcopy - Biopsy in case of positive colposcopy
Lazc.-Ponce, 2011	HC2	Cervicovaginal brush (Digene conical Brush)	ND	No triage	Colposcopy + colpo-directed biopsy, in case of screen test+
Szarewski, 2011	HC2	Swab	6m	Cytology	Colposcopy, in case of triage test+ or screening test+
Virtanen, 2011	HC2	Lavage (Delphi screener)	ND	- <40y: cytology + repeat HPV - ≥40y: no triage	<40y: Colposcopy + colpo-directed biopsy, in case of at least one positive triage test. Repeat testing (cytology + HPV) in 1y, in case of normal triage test. ≥40y: colposcopy + colpo-directed biopsy, in case of a positive screen test
Wikstrom, 2011	HC2	Swab	12m	No triage	- Self-arm: Colposcopy + biopsy; or cytology (with/without repeat HPV) - Control arm: Colposcopy + biopsy, in case of HSIL+ cytology; repeat cytology in case of LSIL cytology
Gok, 2012	HC2	Cervicovagin	12m	Cytology +	- Colposcopy + colpo-directed biopsy, in case of

Author, year	Tests	Self-sampling device	Time of response assessment (months after invitation)	Triage of test+	Follow-up
		al brush		repeat HPV	abnormal cytology (ASC-US+) - Repeat testing (Pap + HPV) in 1y, in case of normal cytology
Darlin, 2013	PCR GP5+/6+	Not documented	ND	No triage	- Colposcopy + colpo-directed biopsy and LBC, in case of hrHPV.
Sancho-Garnier, 2013	PCR genotyping (Abott real time HPV)	Swab (Dacron)	ND	Cytology	- Colposcopy + colpo-directed biopsy, in case of abnormal cytology (LSIL+)

Abbreviations: ND, not documented

Table 3: Summary of the quality of included studies, according to the Cochrane tool for risk of Bias.

Risk of Bias	Selection		Detection		Attrition	Reporting
	Random sequence generation	Allocation concealment	Blinding outcome assessment (participation)	Blinding outcome assessment (testing)	Incomplete outcome data	Reporting of timelines
Bais, 2007	Low	High	N.A.	N.A.	Low	Low
Gok, 2010	Low	?	N.A.	N.A.	Low	Low
Castle, 2011	High	High	N.A.	N.A.	Low	?
Giorgy, 2011	Low	?	N.A.	N.A.	Low	Low
Lazcano-Ponce, 2011	Low	?	N.A.	N.A.	Low	?
Szarewski, 2011	?	?	N.A.	N.A.	Low	Low
Virtanen, 2011	Low	?	N.A.	N.A.	Low	Low
Wikstrom, 2011	?	?	N.A.	N.A.	Low	Low
Gok, 2012	Low	?	N.A.	N.A.	Low	Low
Darlin, 2013	Low	?	N.A.	N.A.	Low	?
Sancho-Garnier, 2013	?	?	N.A.	N.A.	Low	Low

High=high risk of bias, Low=low risk of bias, ?=unclear, N.A.=not applicable.

6.4.3. AMSTAR evaluation of Racey et al. (2013)

Two evaluators (FV and MJ) judged 10 of the 11 AMSTAR key questions equally, which is listed in Table 4. Discordance was noted for item 11 "**Was the conflict of interest included?**", where FV noted "yes" and MJ noted "no". Therefore item 11 was also judged by MA who concluded: "no", since the review authors did not assess the COI of study authors.

The overall AMSTAR score was 8/11 with 3 items scored as negative-
- Item 2: No duplicate study selection and data extraction
- Item 5: No list of excluded studies provided
- Item 11: the review authors did not assess the COI and possible impact on conclusion/ interpretation of the study authors.

It must be remarked that the explanation for AMSTAR item 11 is not clear. The fact that in the study papers COI is declared does not impact on the review. ARMSTAR should reformulate this by splitting into two questions.
a) did the review authors declare their CIO?
b) did the review authors assess the COI of the study authors and the possible impact of evaluated indicators? It could be recommended to the developers of the AMSTAR checklist to split this item 11 in two parts.

Table 4: Amstar evaluation of the systematic review of Racey et al., 2013[15]

AMSTAR key questions	MJ	FV	MA
1. Was an 'a priori' design provided? The research question and inclusion criteria should be established before the conduct of the review. *Note: Need to refer to a protocol, ethics approval, or pre-determined/a priori published research objectives to score a "yes.*	Y	Y	**Y**
2. Was there duplicate study selection and data extraction? There should be at least two independent data extractors and a consensus procedure for disagreements should be in place. *Note: 2 people do study selection, 2 people do data extraction, consensus process or one person checks the other's work.*	N	N	**N**
3. Was a comprehensive literature search performed? At least two electronic sources should be searched. The report must include years and databases used (e.g., Central, EMBASE, and MEDLINE). Key words and/or MESH terms must be stated and where feasible the search strategy should be provided. All searches should be supplemented by consulting current contents, reviews,	Y	Y	**Y**

AMSTAR key questions	MJ	FV	MA
textbooks, specialized registers, or experts in the particular field of study, and by reviewing the references in the studies found. *Note: If at least 2 sources + one supplementary strategy used, select "yes" (Cochrane register/Central counts as 2 sources; a grey literature search counts as supplementary).*			
4. Was the status of publication (i.e. grey literature) used as an inclusion criterion? The authors should state that they searched for reports regardless of their publication type. The authors should state whether or not they excluded any reports (from the systematic review), based on their publication status, language etc. *Note: If review indicates that there was a search for "grey literature" or "unpublished literature," indicate "yes." SINGLE database, dissertations, conference proceedings, and trial registries are all considered grey for this purpose. If searching a source that contains both grey and non-grey, must specify that they were searching for grey/unpublished lit.*	Y	Y	**Y**
5. Was a list of studies (included and excluded) provided? A list of included and excluded studies should be provided. *Note: Acceptable if the excluded studies are referenced. If there is an electronic link to the list but the link is dead, select "no."*	N	N	**N**
6. Were the characteristics of the included studies provided? In an aggregated form such as a table, data from the original studies should be provided on the participants, interventions and outcomes. The ranges of characteristics in all the studies analyzed e.g., age, race, sex, relevant socioeconomic data, disease status, duration, severity, or other diseases should be reported. *Note: Acceptable if not in table format as long as they are described as above.*	Y	Y	**Y**
7. Was the scientific quality of the included studies assessed and documented? 'A priori' methods of assessment should be provided (e.g., for effectiveness studies if the author(s) chose to include only randomized, double-blind, placebo controlled studies, or allocation concealment as inclusion criteria); for other types of studies alternative items will be relevant.	Y	Y	**Y**

AMSTAR key questions	MJ	FV	MA
Note: Can include use of a quality scoring tool or checklist, e.g., Jadad scale, risk of bias, sensitivity analysis, etc., or a description of quality items, with some kind of result for EACH study ("low" or "high" is fine, as long as it is clear which studies scored "low" and which scored "high"; a summary score/range for all studies is not acceptable).			
8. Was the scientific quality of the included studies used appropriately in formulating conclusions? The results of the methodological rigor and scientific quality should be considered in the analysis and the conclusions of the review, and explicitly stated in formulating recommendations. *Note: Might say something such as "the results should be interpreted with caution due to poor quality of included studies." Cannot score "yes" for this question if scored "no" for question 7.*	Y	Y	**Y**
9. Were the methods used to combine the findings of studies appropriate? For the pooled results, a test should be done to ensure the studies were combinable, to assess their homogeneity (i.e., Chi-squared test for homogeneity, I2). If heterogeneity exists a random effects model should be used and/or the clinical appropriateness of combining should be taken into consideration (i.e., is it sensible to combine?). *Note: Indicate "yes" if they mention or describe heterogeneity, i.e., if they explain that they cannot pool because of heterogeneity/variability between interventions.*	Y	Y	**Y**
10. Was the likelihood of publication bias assessed? An assessment of publication bias should include a combination of graphical aids (e.g., funnel plot, other available tests) and/or statistical tests (e.g., Egger regression test, Hedges-Olken). *Note: If no test values or funnel plot included, score "no". Score "yes" if mentions that publication bias could not be assessed because there were fewer than 10 included studies.*	Y	Y	**Y**
11. Was the conflict of interest included? Potential sources of support should be clearly acknowledged in both the systematic review and the included studies. *Note: To get a "yes," must indicate source of funding or support for the systematic review AND for each of the included studies.*	N	Y	**N**

Y=yes, N=No. Evaluation was performed by two evaluators MJ (2nd column), and FV (3rd column). Consensus scores were reached by evaluation of a third evaluator, MA (last column)

6.4.4. Attendance in the self-sampling arm versus the control arm

A total of 80,532 and 34,465 women were allocated to the self-sampling and control arm, respectively. In the self-sampling arm, attendance rates ranged from 10.2% [27] to 98.2% [25], including women who did not take a self-sample but attended regular screening (conventional cytology) instead. In the same way, compliance in the control-arm varied considerably, ranging from 2% [32] to 86.8% [25]. The highest response rates were achieved in two studies that applied a door-to-door approach for inviting women to take a self-sample (80.5% and 98.2% in the self-sampling arm, and 40.5% and 86.8% in the control arm)[23;25]. In the nine other studies, in which invitations were sent by mail, compliance ranged from 10.2% to 39.0% in the self-sampling arm, and 2.0-25.9% in the control arm. As a consequence, subgroup analyses based on invitation method were performed when pooling attendance data .

The pooled response rate of women in the self-sample arm was 25% (95% CI = 20-30%) and 90% (95% CI = 73-100%) for women who were invited by mail and by door-to-door approach, respectively (Figure 2, left). The overall response rate in the control arm was 11% (95% CI = 5-17%) for women invited by mail, and 64% (95% CI = 19-100%) for women recruited at their home (Figure2, right).

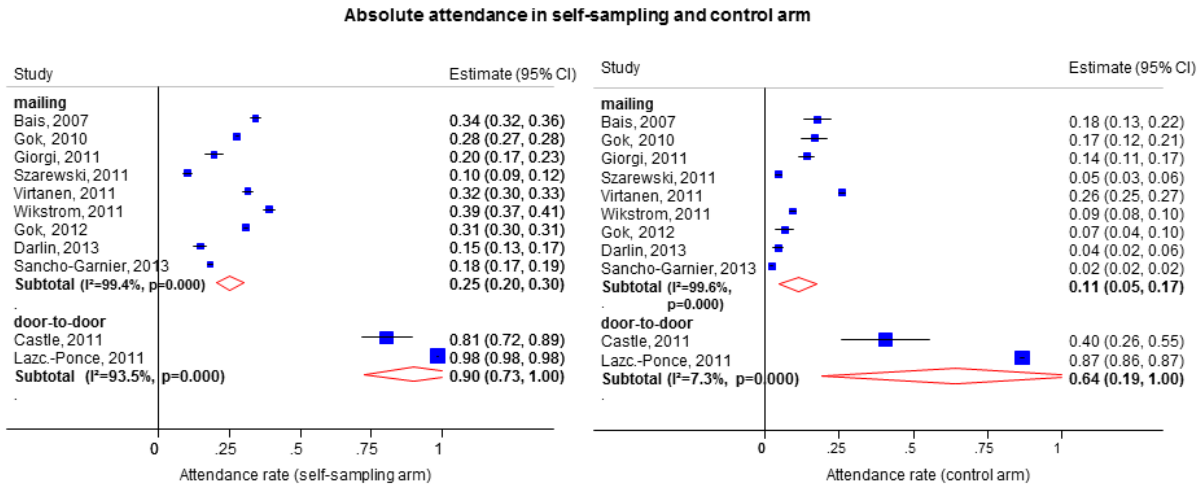


Figure 2: Meta-analysis of the attendance rate of women invited by mail (top) or door-to-door (bottom) in the self-sampling arm (left) and the control arm (right).

In all of the 11 included studies, the relative compliance in the self-sampling arm compared to the control-arm was significantly greater than one. Racey et al. observed a pooled relative attendance of 2.14. In the updated analysis presented here, a subgroup analysis was performed, based on the mode of invitation. The pooled relative attendance in the self-sampling versus the control arm was 2.71 (95% CI = 1.45-5.08) in studies in which self-sample kits were mailed (Figure 3). In studies that applied a door-to-door approach

for both study arms, a lower relative attendance was observed 1.45 (95% CI = 0.83-2.53), not significantly different from unity. The difference in attendance rate for the self-sampling versus the control arm was 14% (95% CI = 9-19%) and 24% (95% CI = -4-52%) in studies where women were invited by mail or door-to-door, respectively. This indicates that studies in which women were invited by mail demonstrated a significantly increased participation in the self-sampling arm, compared to the control arm.

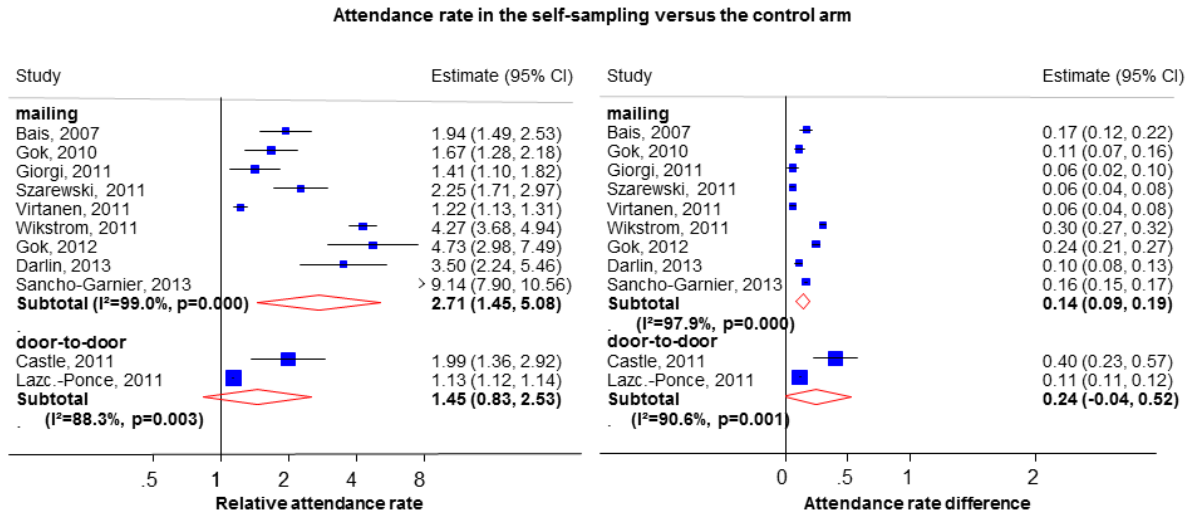


Figure 3: Meta-analysis of the attendance rate in the self-sampling arm, compared to the control arm for women who were invited by mail (top) or door-to-door (bottom).

In the study of Giorgy-Rossi et al. (2011, Br J Cancer)[24], women were invited to participate in screening in three ways: (1) a regular invitation letter for cytology or HPV testing (clinician-obtained sample) was sent, (2) a self-sampling kit was mailed directly, preceded by a notification, and (3) an invitation letter was sent notifying women that if they called a toll free number they would receive a self-sampling kit. Compared to the regular invitation letter (approach 1, 14% attendance rate), direct mailing of a self-sampling kit (20% attendance rate) resulted in a higher participation (relative attendance rate: 1.4 [95% CI=1.1-1.8], p=0.007). However, when women had to call to a phone center in order to receive a self-sampling kit (approach 3) the attendance rate was only 9%, which was significantly lower compared to the regular invitation (relative attendance rate: 0.62 [95% Ci=0.45-0.86], p=0.004) and the direct-mailing of the self-sample kit (relative attendance rate: 0.44 [95% CI= 0.33-0.60], p<0.001).

6.4.5. Sample quality and test-positivity rate

A high specimen quality of self-samples was reported in all studies. When documented, the amount of uninterpretable tests in the self-sampling arm never exceeded 1.4%. Most of the included studies used the Hybrid Capture-2 assay, with exception of three studies in which a PCR-based assay was

used[21;31;32]. Overall, studies using a PCR approach demonstrated the highest proportions of uninterpretable tests (1.2%; 1.4%; and 0.6%).

Ten studies documented the test-positivity in the self-sampling arm, considering HPV-tests only. Women in the self-sampling arm who attended conventional screening by cytology testing were excluded from this analysis. The test-positivity rate (hrHPV) in the self-sampling arm ranged from 6.0%[29] to 20.1%[24], considering only HPV-tests and not cytology results of women who preferred a conventional pap smear. Overall a test-positivity rate of 10.2% (95% CI = 8.6-11.8%) was observed (Figure 4, left). Only four studies provided test results of cytology (at cut-off ASCUS/-US+ or LSIL+) in the control arm. Based on four studies, in the control arm a pooled test-positivity rate of 2.0% (95% CI = 0.0-4.1%) was observed (Figure 4, right).

Figure 4: Meta-analysis of test-positivity rate in the self-sampling arm (left) and the control arm (right).

6.4.6. Compliance to follow-up and detection rate

Ten studies reported the compliance to follow-up for screen-positive women in the self-sampling arm, and 3 of them reported the compliance in the control arm. In the self-sampling arm, compliance to follow-up was on average 85.8% (95% CI = 77.4-94.2%) with an observed minimum and maximum of 41.0%[32] and 100.0%[24;25;28], respectively (Figure 5).

Based on three studies, the compliance to follow-up in the control-arm after a positive screening test ranged from 55.6%[32] to 100.0%[25;31]. Overall, the relative compliance to follow-up of women with a positive test in the self-sampling arm versus the control arm is 0.77 (95% CI = 0.50-1.20) (Figure 6, left). A pooled difference of 10% was observed between the self-sampling arm and the control arm (risk difference: -0.10, 95% CI = -0.38-0.18) (Figure 6, right).

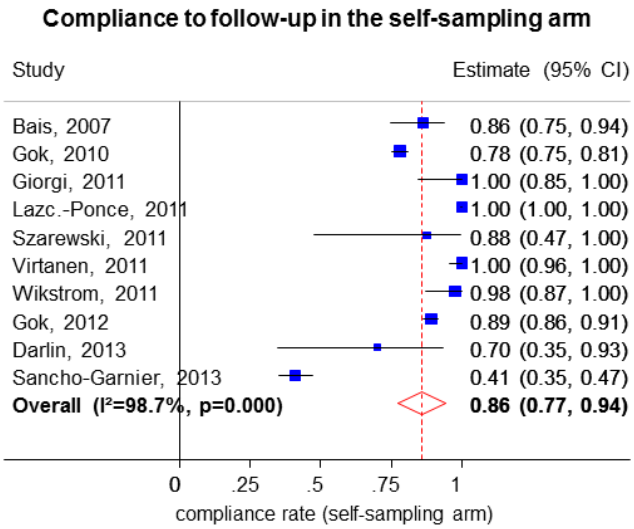


Figure 5: Meta-analysis of the compliance to follow-up in the self-sampling arm.

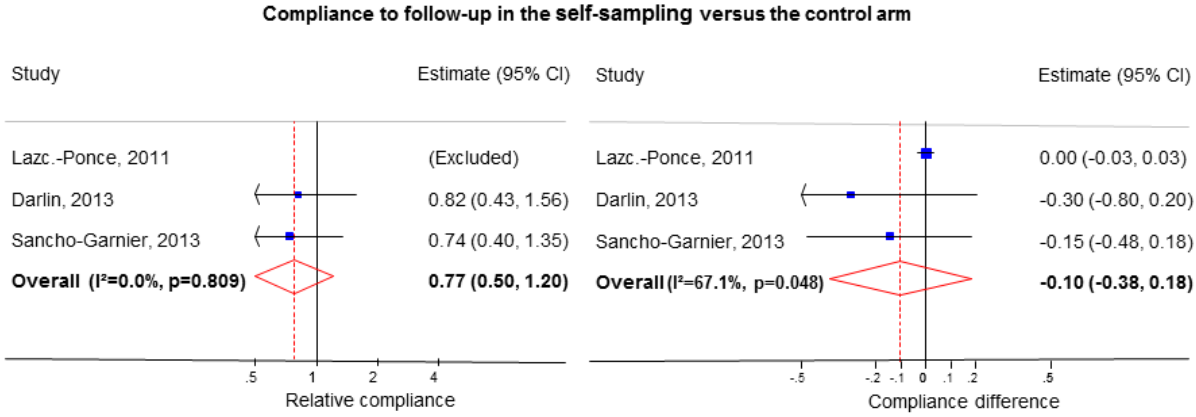


Figure 6: Meta-analysis of the compliance to follow-up instructions after a positive test in the self-sampling arm, compared to the control arm. Compliance to follow-up in Lazc.-Ponce, 2011 was 100% for both arms and is excluded.

Based on ten studies that documented the amount of detected CIN2+ among the total number of women that received an invitation for self-sampling, a pooled detection rate of 0.4% (95% CI=0.3-0.6%) was observed. For the control arm, seven studies reported data on the detection of CIN2+, which resulted in a pooled detection rate for CIN2+ of 0.1% (95% CI=0.0-0.2%). Comparing the detection rate in the self-sampling arm and the control arm, a relative detection rate of 3.05 (95% CI=1.76-5.28) was observed, indicating that in the self-sampling arm three times more CIN2+ was detected compared to the control arm. Still, a detection rate difference of merely 0.2% (95% CI= -0.1-0.5%) was observed for the outcome CIN2+ (Figure 7).

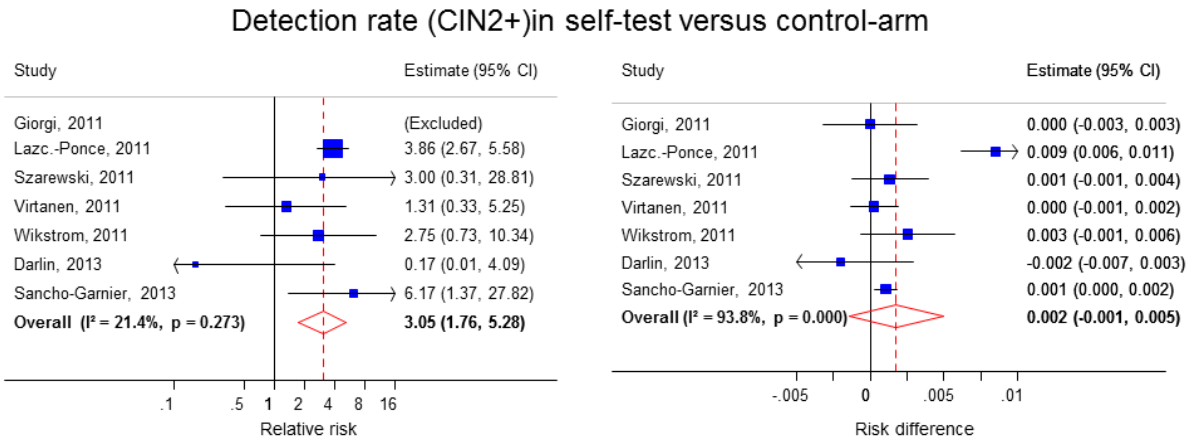


Figure 7: Meta-analysis of the detection rate of CIN2+ in the self-sampling arm compared to the control arm.

6.5. Discussion and interpretation

In this report, a recent systematic review was evaluated which addressed the attendance of women in primary cervical cancer screening when offered a self-sample and HPV-testing, compared to an invitation for a cervical smear performed by a clinician. To improve comparability across studies, middle/low-income study sites were omitted from the meta-analysis.

In the analysis presented here, an update[32] of the data of one included study[26] was made and one new study was added[31], resulting in 11 studies.

6.5.1. Response to invitation for cervical cancer screening

The included studies employed two approaches to invitation for self-sampling. Either women were invited by mail at their home-address, or women were invited door-to-door. Two studies [23;25] employed the latter methodology and documented a considerably higher response rate in both study arms, compared to the nine other studies which made use of invitations by mail. As a consequence, when analyzing the attendance to screening, a subgroup meta-analysis was performed based on the mode of invitation.

Considering invitation by mail, our analysis demonstrated that women who were never screened or are under-screened, are more than twice as likely to undergo screening when offered a self-sample kit, compared to an invitation for regular screening (relative response rate: 2.71, 95% CI = 1.45-5.08). A pooled difference of 14% (95% CI = 9-19%) was observed in favor of self-sampling at home. Interestingly, the two studies that included the most persistent non-responders belonged to the top three of studies with the highest increase in participation in the self-sampling arm compared to the control arm (relative attendance rate = 4). In these studies, participants had not been screened for over six[29] and nine years[31]. The third study included women who failed to participate in screening after they had received an invitation and a recall[30]. This further supports the notion that in persistent

non-responders, offering a self-sample may lead to better participation than sending a conventional screening invitation.
In a large Italian controlled trial where women had to phone to a call center to confirm that they wanted to receive a self-sampling kit, a significantly lower attendance rate was observed compared to the regular recall letter ($p<0.001$)[24].
Based on the two studies that applied a door-to-door approach to invite women in both study arms, a significant difference was not observed when comparing attendance rates in the self-sampling arm and the control arm (relative attendance rate: 1.45, 95% CI = 0.83-2.53; attendance rate difference: 0.24, 95% CI = -0.04-0.52). However, although overall response rates were highest using a door-to-door approach, this method of invitation will most often not be feasible in screening programs. Thus, considering invitations by mail, our analysis clearly indicates that offering self-sampling is the better way to increase participation among not-screened or under-screened women.

One of the most common reasons given for non-attendance to screening is embarrassment and 'being uncomfortable with a vaginal exam performed by a clinician'[27;33;34]. In this review, the studies that quantitatively addressed acceptability of self-sampling, documented that more than 85% of the women describe the sampling procedure as easy and more than 75% prefer this test to a conventional test on material taken by a clinician[21;23;24;35]. The study by Castle et al. did not apply a randomized design, but instead allowed women to choose their preferred sampling method. The self-sampling method at home, rather than going to a clinic for a regular screening test, was chosen by 65% (77/119) of the participants[23]. This information reinforces the results of the meta-analysis presented here, and indicates that offering a self-sample can increase women's willingness to participate in cervical cancer screening. On the other hand, several studies reported a low number of women (on average 1.5%) who preferred to participate in conventional screening, after they had received an invitation for self-sampling[21;22;24;27-29].

6.5.2. Compliance to instructions for follow-up among women with a positive screening test

Compliance to follow-up is an important aspect to assess, since finding just screen-positives without further action will not have an impact on cancer incidence. Our analysis showed a rather good compliance at baseline (86%) to instructions for follow-up in the self-sampling arm. Based on limited data (only three studies), a non-signficantly lower compliance to follow-up was observed in the self-sampling arm versus the control arm (relative compliance 0.77, 95% CI 0.50-1.20; difference in compliance: -10%, 95% CI =

-38-18%). However future studies addressing this topic are needed to enable a stronger conclusion.

6.5.3. Barriers to self-sampling in cervical cancer screening

Our analysis demonstrated a convincingly positive effect on screening participation when offering self-sampling kits to under-screened women. However, previous reports have mentioned weariness among women of doing the test properly[36;37]. Hence, caution is needed when extrapolating levels of acceptance to non-research settings where instructions might be non-uniform and information limited.

Another aspect that needs to be considered is the cost-effectiveness of offering self-samples to non-responders. In the studies presented here, self-sampling kits were sent to all women in the corresponding study arm without them having to confirm beforehand whether they were willing to participate. This approach might be expensive in terms of self-sampling kits used per screened woman. In one of the studies of this meta-analysis, a second self-sampling arm was included in which women were asked to confirm by phone if they wanted to receive a self-sampling kit. However, a considerably lower response rate (9% vs 20%) was observed in the second self-sampling arm compared to the first[24]. These findings demonstrate that the cost-effectiveness of mailing self-sampling kits might be a delicate balance, which should be carefully assessed. Furthermore, prior to introducing self-sampling in a certain demographic setting and geographic region, additional work should be done to ascertain the cost-effectiveness in that particular context.

6.5.4. Conclusions

Self-sampling followed by HPV-testing could significantly improve screening participation, when offered as a complement to conventional screening to reach unscreened or under-screened women. However, the demographic context is an important variable in the effect of offering self-sampling kits on screening participation, and should therefore be carefully evaluated in a pilot study. Furthermore, appropriate follow-up guidelines for women who test positive for high risk HPV are needed.

6.6. GRADE: quality of the evidence and strength of recommendations

From a previous diagnostic test accuracy meta-analysis, it was concluded that HPV testing on a self-sample had a similar sensitivity and specificity as HPV testing on a clinician-taken sample under the condition that a clinically validated PCR-based hrHPV DNA assay was used[38;39]. However, when an assay

based on signal-amplification is used, the sensitivity and specificity of HPV on self-samples is inferior to that on samples taken by a clinician[39]. The quality of evidence of these findings can be judged as moderate, given: consistency and precision of findings and the observational study design (diagnostic test accuracy studies). Some doubt could be raised since the meta-analysis included a mixture of studies conducted in screening and follow-up settings. However since clinical setting did not generate heterogeneity (similarity of the relative sensitivity and relative specificity of HPV testing on self- vs clinician-based samples), no statistical reasons were found to separate data by setting. From both, the generally pooled meta-analyses and setting-stratified meta-analyses the same conclusions could be drawn{Guyatt, 2008 31103 /id}.
The list of hrHPV DNA assays, based on target amplification principle (PCR), including a quality assessment of findings, was published recently[40]. Only a clinically validated HPV PCR test should be used in screening on self-samples. The strength of recommendation can be qualified as a a plain recommendation (not a strong recommendation), given the moderate level of evidence associated with consistent & precise findings from observational studies.

In the current review, the gain in screening attendance among under-screened or non-screened women was assessed. Most included studies were randomized trials comparing response rates in two arms (experimental arm=self-samplers sent to women; control arm, which usually consisted in sending a standard invitation for having a Pap smear taken by clinician). Given the randomized design, a high level of evidence can be anticipated. The quality of the individual trials was good to moderate, the direction of effect (higher participation in self-sampling arm) was consistent but the magnitude of effect was highly heterogeneous and region dependent. To conclusion the level of evidence can be considered as moderate. No universal recommendation can be formulated regarding introduction of strategies offering self-samplers in Germany.
The evidence that opt-in approaches (women receive a self-sampling kit if they request for it) are not more effective than standard invitation letters is weak given the low number of studies and the heterogeneity in findings.
The trials, included the meta-analysis, compared an experimental intervention (sending self-samplers) with the usual invitation which is an inherent part of an organized screening program. However, in Germany no organized invitational system is in place. Therefore, the published trials could be considered as not immediately relevant for the S3 guideline. Therefore, it is recommended to setup pilot studies to assess the potential gain in screening coverage in the German context.

6.7. References

(1) Laara E, Day NE, Hakama M. Trends in mortality from cervical cancer in the Nordic countries: association with organised screening programmes. *Lancet* 1987; 1:1247-1249.

(2) Anttila A, Laara E. Cervix cancer: Geographical correlations. In: Sankila R, Demaret E, Hakama M, Lynge E, Schouten LJ, Parkin DM, editors. Evaluation and Monitoring of Screening Programmes. Brussels, Luxemburg: Europe Against Cancer Programme; 2000. 77-97.

(3) Arbyn M, Raifu AO, Antoine J. Trends of cervical cancer mortality in Europe. *IPH/EPI-REPORTS* 2009; D/2009/2505/41:1-54.

(4) Arbyn M, Anttila A, Jordan J, Ronco G, Schenck U, Segnan N et al. European Guidelines for Quality Assurance in Cervical Cancer Screening. Second Edition - Summary Document. *Ann Oncol* 2010; 21(3):448-458.

(5) Smith RA, Cokkinides V, Brooks D, Saslow D, Shah M, Brawley OW. Cancer screening in the United States, 2011: A review of current American Cancer Society guidelines and issues in cancer screening. *CA Cancer J Clin* 2011; 61(1):8-30.

(6) Saslow D, Solomon D, Lawson HW, Killackey M, Kulasingam SL, Cain J et al. American Cancer Society, American Society for Colposcopy and Cervical Pathology, and American Society for Clinical Pathology screening guidelines for the prevention and early detection of cervical cancer. *CA Cancer J Clin* 2012; 62(3):147-172.

(7) Lancucki L, Fender M, Koukari A, Lynge E, Mai V, Mancini E et al. A fall-off in cervical screening coverage of younger women in developed countries. *J Med Screen* 2010; 17(2):91-96.

(8) Elit L, Saskin R, Raut R, Elliott L, Murphy J, Marrett L. Sociodemographic factors associated with cervical cancer screening coverage and follow-up of high grade abnormal results in a population-based cohort. *Gynecol Oncol* 2012.

(9) Decker K, Demers A, Chateau D, Musto G, Nugent Z, Lotocki R et al. Papanicolaou test utilization and frequency of screening opportunities among women diagnosed with cervical cancer. *Open Med* 2009; 3(3):e140-e147.

(10) Petignat P, Vassilakos P. Is it time to introduce HPV self-sampling for primary cervical cancer screening? *J Natl Cancer Inst* 2012; 104(3):166-167.

(11) Logan L, Mcilfatrick S. Exploring women's knowledge, experiences and perceptions of cervical cancer screening in an area of social deprivation. *Eur J Cancer Care (Engl)* 2011; 20(6):720-727.

(12) Huynh J, Howard M, Lytwyn A. Self-collection for vaginal human papillomavirus testing: systematic review of studies asking women their perceptions. *J Low Genit Tract Dis* 2010; 14(4):356-362.

(13) Wikstrom I, Stenvall H, Wilander E. Attitudes to self-sampling of vaginal smear for human papilloma virus analysis among women not attending organized cytological screening. *Acta Obstet Gynecol Scand* 2007; 86(6):720-725.

(14) Mitchell S, Ogilvie G, Steinberg M, Sekikubo M, Biryabarema C, Money D. Assessing women's willingness to collect their own cervical samples for HPV testing as part of the ASPIRE cervical cancer screening project in Uganda. *Int J Gynaecol Obstet* 2011; 114(2):111-115.

(15) Racey CS, Withrow DR, Gesink D. Self-collected HPV Testing Improves Participation in Cervical Cancer Screening: A Systematic Review and Meta-analysis. *Can J Public Health* 2013; 104(2):e159-e166.

(16) Higgins JP, Altman DG, Gotzsche PC, Juni P, Moher D, Oxman AD et al. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. *BMJ* 2011; 343:d5928.

(17) Shea BJ, Grimshaw JM, Wells GA, Boers M, Andersson N, Hamel C et al. Development of AMSTAR: a measurement tool to assess the methodological quality of systematic reviews. *BMC Med Res Methodol* 2007; 7:10.

(18) Shea BJ, Hamel C, Wells GA, Bouter LM, Kristjansson E, Grimshaw J et al. AMSTAR is a reliable and valid measurement tool to assess the methodological quality of systematic reviews. *J Clin Epidemiol* 2009; 62(10):1013-1020.

(19) Verdoodt F, Jentschke M, Hillemanns P, Racey CS, Snijders PJF, Arbyn M. Reaching women who do not participate in the regular cervical cancer screening program by offering self-sampling kits: A systematic review and meta-analysis of randomised trials. *Eur J Cancer* 2015; 51(16):2375-2385.

(20) Dersimonian R, Laird NM. Meta-analysis in clinical trials. *Controlled Clin Trials* 1986; 7:177-188.

(21) Bais AG, van Kemenade FJ, Berkhof J, Verheijen RH, Snijders PJ, Voorhorst F et al. Human papillomavirus testing on self-sampled cervicovaginal brushes: An effective alternative to protect nonresponders in cervical screening programs. *Int J Cancer* 2007; 120(7):1505-1510.

(22) Gok M, Heideman DA, van Kemenade FJ, Berkhof J, Rozendaal L, Spruyt JW et al. HPV testing on self collected cervicovaginal lavage specimens as screening method for women who do not attend cervical screening: cohort study. *BMJ* 2010; 340:c1040-*.

(23) Castle PE, Rausa A, Walls T, Gravitt PE, Partridge EE, Olivo V et al. Comparative community outreach to increase cervical cancer screening in the Mississippi Delta. *Prev Med* 2011; 52(6):452-455.

(24) Giorgi-Rossi P, Marsili LM, Camilloni L, Iossa A, Lattanzi A, Sani C et al. The effect of self-sampled HPV testing on participation to cervical cancer screening in Italy: a randomised controlled trial (ISRCTN96071600). *Br J Cancer* 2011; 104(2):248-254.

(25) Lazcano-Ponce E, Lorincz AT, Cruz-Valdez A, Salomon J, Uribe P, Velasco-Mondragon E et al. Self-collection of vaginal specimens for human papillomavirus testing in cervical cancer prevention (MARCH): a community-based randomised controlled trial. *The Lancet* 2011;(0).

(26) Piana L, Leandri FX, Le RL, Heid P, Tamalet C, Sancho-Garnier H. [HPV-Hr detection by home self sampling in women not compliant with pap test for cervical cancer screening. Results of a pilot programme in Bouches-du-Rhone]. *Bull Cancer* 2011; 98(7):723-731.

(27) Szarewski A, Cadman L, Mesher D, Austin J, Ashdown-Barr L, Edwards R et al. HPV self-sampling as an alternative strategy in non-attenders for cervical screening - a randomised controlled trial. *Br J Cancer* 2011; 104(6):915-920.

(28) Virtanen A, Nieminen P, Luostarinen T, Anttila A. Self-sample HPV tests as an intervention for nonattendees of cervical cancer screening in Finland: a randomized trial. *Cancer Epidemiol Biomarkers Prev* 2011; 20(9):1960-1969.

(29) Wikstrom I, Lindell M, Sanner K, Wilander E. Self-sampling and HPV testing or ordinary Pap-smear in women not regularly attending screening: a randomised study. *Br J Cancer* 2011; 105(3):337-339.

(30) Gok M, van Kemenade FJ, Heideman DA, Berkhof J, Rozendaal L, Spruyt JW et al. Experience with high-risk human papillomavirus testing on vaginal brush-based self-samples of non-attendees of the cervical screening program. *Int J Cancer* 2012; 130(6):1228-1235.

(31) Darlin L, Borgfeldt C, Forslund O, Henic E, Hortlund M, Dillner J et al. Comparison of use of vaginal HPV self-sampling and offering flexible appointments as strategies to reach long-term non-attending women in organized cervical screening. *J Clin Virol* 2013; 58(1):155-160.

(32) Sancho-Garnier H, Tamalet C, Halfon P, Leandri F, Le RL, Djoufelkit K et al. HPV self-sampling or the pap-smear: A randomized study among

cervical screening non-attenders from lower socio-economic groups in France. *Int J Cancer* 2013; 133(11):2681-2687.

(33) Scarinci IC, Litton AG, Garces-Palacio IC, Partridge EE, Castle PE. Acceptability and Usability of Self-Collected Sampling for HPV Testing among African-American Women Living in the Mississippi Delta. *Womens Health Issues* 2013;in-press.

(34) Waller J, McCaffery K, Forrest S, Szarewski A, Cadman L, Austin J et al. Acceptability of unsupervised HPV self-sampling using written instructions. *J Med Screen* 2006; 13(4):208-213.

(35) Leniz J, Barriga MI, Lagos M, Ibanez C, Puschel K, Ferreccio C. HPV vaginal self-sampling among women non-adherent to Papanicolaou screening in Chile. *Salud Publica Mex* 2013; 55(2):162-169.

(36) Howard M, Lytwyn A, Lohfeld L, Redwood-Campbell L, Fowler N, Karwalajtys T. Barriers to acceptance of self-sampling for human papillomavirus across ethnolinguistic groups of women. *Can J Public Health* 2009; 100(5):365-369.

(37) Anhang R, Nelson JA, Telerant R, Chiasson MA, Wright TC, Jr. Acceptability of self-collection of specimens for HPV DNA testing in an urban population. *J Womens Health (Larchmt)* 2005; 14(8):721-728.

(38) Arbyn M, Verdoodt F, Snijders PJF, Verhoef VM, Suonio E, Dillner J et al. Accuracy of human papillomavirus testing on self-collected versus clinician-collected samples: a meta-analysis. *Lancet Oncol* 2014; 15(2):172-183.

(39) Arbyn M, Castle P. Offering self-sampling kits for HPV testing to reach women who do not attend in the regular cervical cancer screening program. *Cancer Epidemiol Biomarkers Prev* 2015; 24(5):769-772.

(40) Arbyn M, Snijders PJ, Meijer CJ, Berkhof H, Cuschieri K, Kocjan BJ et al. Which high-risk HPV assays fulfil criteria for use in primary cervical cancer screening? *Clin Microbiol Infect* 2015; 21(9):817-826.

7. Question: Is testing for presence of high-risk types of the human papillomavirus (hrHPV) on vaginal samples taken by the woman herself as accurate to detect cervical pre-cancer than hrHPV testing or cytological processing on a sample taken by a health professional?

7.1. Clinical questions and PICOS components

Study question:
Is testing for presence of high-risk types of the human papillomavirus (hrHPV) on vaginal samples taken by the woman herself as accurate to detect cervical pre-cancer than hrHPV testing or cytological processing on a sample taken by a health professional?

PICOS components

***P**opulation*: women participating in cervical cancer screening, or women with cervical abnormalities detected previously and under follow-up, presenting at a colposcopy clinic.

Index test: hrHPV testing on a self-sample.

Comparator tests:
C1: hrHPV testing on a clinician sample
C2: cytology on a clinician sample.

Outcomes:
O1: absolute sensitivity and specificity for detection of CIN2+ or CIN3+ of the index test and of the two comparator tests.
O2: relative sensitivity and specificity for CIN2+ and CIN3+ of:
- hrHPV testing on a self-sample versus hrHPV testing on a clinician-sample
- hrHPV testing on a self-sample versus cytology on a clinician-sample

Studies:
Cross-sectional diagnostic test accuracy studies where a self-sample and a clinician-sample are taken and a hrHPV test is performed on the self-sample and one or two comparator tests are performed on the clinician sample.
Randomised trials with hrHPV testing on a self-sample in one arm and one or two comparator tests on a clinician-sample in the other arm.

7.2. Literature retrieval strings

A. In Pubmed-Medline

#1: Cervix OR cervico* OR cervica*
#2: Cancer OR carcinoma OR neoplas* OR dysplas* OR CIN[tw] OR CINII*[tw] OR CIN2*[tw] OR CINIII*[tw] OR CIN3[tw] OR SIL[tw] OR SIL OR HSIL[tw] OR H-SIL OR LSIL[tw] OR L-SIL OR OR ‘‘low grade’’ OR low-grade OR mild OR equivocal OR borderline.

#3: #1 AND #2.
#4: HPV OR "Human Papillomavirus DNA Tests"[Mesh] OR ‘‘human papillomavirus’’ OR papillomavir* OR viral OR virus
#5: self-collection OR “self collection” OR self-sampling OR self-collect* OR self-sampl* OR self OR "Self-Examination"[Mesh]
#6: #4 AND #5
#7: #3 AND #6
#8: Publication Date to May 25, 2013
#9: #7 AND #8

B. In Embase

#1: 'cervix'/exp OR cervix OR cervico* OR cervica*
#2: 'cancer'/exp OR cancer OR 'carcinoma'/exp OR carcinoma OR neoplas* OR dysplas* OR cin OR 'cin2' OR 'cin3' OR sil OR hsil OR h+sil OR lsil OR l+sil OR 'low grade' OR low+grade OR mild OR equivocal OR 'borderline'/exp OR borderline
#3: 'hpv'/exp OR hpv OR 'human papillomavirus'/exp OR 'human papillomavirus' OR papillomavir* OR viral OR 'virus'/exp OR virus
#4: self+collection OR 'self collection' OR self+sampling OR 'self-sampling' OR self+collect* OR self+sampl* OR 'self'/exp OR self
#5: #1 AND #2 AND #3 AND #4

With the following limits:

- Map to preferred terminology (with spell check)
- Also search as free text
- Include sub-terms/derivatives (explosion search)
- Search publications: all years

C. In Cochrane Library

#1: Cervix or cervico* or cervica*
#2: Cancer or carcinoma or neoplas* or dysplas* or CIN or CIN2 or CIN3 or SIL or SIL or HSIL or H-SIL or LSIL or L-SIL or "low grade" or low-grade or mild or equivocal or borderline.
#3: HPV or ‘‘human papillomavirus’’ or papillomavir* or viral or virus
#4: self-collection or "self collection" or self-sampling or ‘‘self-sampling’’ or self-collect* or

self-sampl* or self

With the following limits:

- Cochrane reviews (reviews + protocols)
- Other reviews

Search for word variations

7.3. List of excluded studies, after reading of abstract or material & methods of full papers

Author	Year	Journal	Vol	Start page	End page	Code exclusion*
Fairley	1992	J Infect Dis	165	1103	1106	D
Fairly	1994	Genitourin Med	70	171	174	D
Coutlee	1997	J Med Virol	51	42	47	D
Bowden	1998	AIDS Patient Care STDS	12	29	32	E
Bowden	1998	Int J Gynecol Cancer	8	471	476	E
Belinson	1999	Int J Gynecol Cancer	9	411	417	A
Bowden	1999	Sex Transm.Infect.	75	431	434	E
Harper	1999	J FamPract	48	531	535	D
Serwadda	1999	J Infect Dis	180	1316	1319	E
Gravitt	2001	Cancer Epidemiol Biomarkers Prev	10	95	100	D
Rompalo	2001	Clin Infect Dis	33	1455	1461	D
Smith	2001	Arch PediatrAdolesc Med	155	676	679	E
Smith	2001	Arch PediatrAdolesc Med	155	676	679	E
Chang	2002	Chang Gung Med J	25	664	671	D
Dzuba	2002	J Womens Health Gend Based	11	265	275	B
Flores	2002	SaludPublicaMex	44	335	344	A
Flores	2002	SaludPublicaMex	44	335	344	A
Harper	2002	Arch PedAdolesc Med	156	1154	1155	A
Harper	2002	Am J ObstetGynecol	186	365	373	D
Harper	2002	Sex Transm Dis	29	628	636	E

Author	Year	Journal	Vol	Start page	End page	Code exclusion*
Harper	2002	Sex Transm Dis	29	628	636	E
Holland-Hall	2002	J PediatrAdolescGynecol	15	307	313	B
Knox	2002	Sex Transm Dis	29	647	654	B
Belinson	2003	Int J Gynecol Cancer	13	819	826	F
Belinson	2003	Int J Gynecol Cancer	13	819	826	G
Harper	2003	Ann Fam Med	1	221	227	E
Palmisano	2003	Int J STD AIDS	14	560	567	D
Tisci	2003	J Lower Gen Tract Dis	7	107	116	B
Baay	2004	Scand J Infect Dis	36	456	459	C
Dannecker	2004	Ann Oncol	15	863	869	E
Forrest	2004	J Med Screen	11	85	88	B
Howard	2004	J Lower Gen Tract Dis	8	33	37	E
Kahn	2004	ObstetGynecol	103	952	959	D
Agorastos	2005	Int J STD AIDS	16	727	729	D
Bidus	2005	ClinObstetGynecol	48	127	132	A
Budge	2005	Aust NZ J ObsetGynecol	45	215	219	B
Lack	2005	Sex Transm Infect	81	239	241	D
Prusty	2005	Int J GynecolObst	90	223	227	D
Castle	2006	J ClinMicrobiol	44	2158	2159	D
Karwalajtys	2006	Sex Transm Infect	82	337	339	D
Stenvall	2006	ActaDermVenereol	86	465	467	D
Van De Wijgert	2006	Sex Transm Dis	33	516	523	D
Bais	2007	Int J Cancer	120	1505	1510	B
Bais	2007	Int J Cancer	120	1505	1510	D
Jones	2007	J ClinMicrobiol	45	1679	1683	D
Khanna	2007	Int J Gynecol Cancer	17	615	622	D

Author	Year	Journal	Vol	Start page	End page	Code exclusion*
Kim	2007	Int J STD Aids	16	163	166	D
Lippman	2007	Sex Transm Dis	34	421	428	B
Morris	2007	ClinChem Lab Med	45	577	591	A
Ogilvie	2007	Can Med Assoc J	177	480	483	B
Safaeian	2007	Sex Transm Dis	34	429	436	D
Stewart	2007	J ObstetGynaecol Can	29	817	828	A
Winer	2007	Sex Transm Dis	34	371	377	D
De Alba	2008	Cancer Epidemiol Biomarkers Prev	17	2163	2168	D
Longatto-Filho	2008	Eur J GynaecolOncol	29	327	332	G
Lenselink	2009	J ClinMicrobiol	47	2564	2570	D
Sanner	2009	Br J Cancer	101	871	874	E
Sanner	2009	Br J Cancer	101	871	874	G
Belinson	2010	Int J Cancer	127	1151	1157	G
Gök	2010	BMJ	340	c1040		E
Giorgi Rossi	2011	Br J Cancer	104	248	254	B
Gravitt	2011	Int J Cancer	129	517	527	A
Silva	2011	J Oncol	Epub 953 469			E
Szarewski	2011	Br J Cancer	104	915	920	D
Virtanen	2011	Int J Cancer	128	2681	2687	B
Virtanen	2011	Cancer Epidemiol Biomarkers Prev	20	1960	1969	D
Cerigo	2012	J Med Sceen	19	42	48	D
Cuzick	2012	Vaccine	30	Suppl 05		A
daSilvaRocha	2012	ActaCytol	56	520	526	D

Author	Year	Journal	Vol	Start page	End page	Code exclusion*
Gök	2012	Int J Cancer	130	1228	1235	D
Guan	2012	Sex Transm Infect	88	490	494	B
Jones	2012	J Womens Health	21	1275	1281	B
Levinson	2012	Int J Gynecol Cancer	23	141	147	D
Lindell	2012	BJOG	119	245	248	D
Lindell	2012	BJOG	119	245	248	G
Nabandith	2012	Asian Pac J Cancer Prev	13	4665	4667	D
Nilaweera	2012	Asian Pac J Cancer Prev	13	1193	1196	B
Okayama	2012	Asian Pac J Cancer Prev	13	4521	4524	D
Quincy	2012	J ObstetGynaecol	32	87	91	B
Rositch	2012	Plos One	7	e40766		B
Suba	2012	Lancet	379	1587	1588	A
Tamalet	2012	ClinMicrobiol Infect	19	E44	E50	D
Wolfrum	2012	J Med Microbiol	61	1538	1545	E
Wu	2012	J Low Genit Tract Dis	16	416	420	G
Berner	2013	J Low Genit Tract Dis	17	235	241	B
Camilloni	2013	BMC Public Health	13	464	499	B
Hoque	2013	J Cancer Res Ther	9	25	28	C
Levinson	2013	gynecoloncol	129	318	323	D
ortiz	2013	J Low Genit Tract Dis	17	210	217	D
Sancho-Garnier	2013	Int J Cancer	in press			D
Scarinci	2013	Women's Health Issues	23	e123	e130	B
Schmeink	2013	Int J Cancer	in press			D
Silva	2013	Arch GynecolObstet	in press			D
Snijders	2013	Int J Cancer	132	2223	2236	A

Author	Year	Journal	Vol	Start page	End page	Code exclusion*
Verhoef	2013	BMC Women's health	in press			D

Number of studies excluded because:

Code	Reason for exclusion	Number
A	no primary data (letter, comment, review)	11
B	no HPV-testing data	19
C	no self-sampling	2
D	no accuracy data for CIN2+ or CIN3+ (or follow-up cytology)	41
E	no comparison with clinician-obtained sample	15
F	long time interval between self &clin sample	1
G	double reporting	6

7.4. Population and study characteristics.

Supplementary Table 1. Population and study characteristics.

Author, year Country	**Study design**	**Population/ setting**	**Inclusion & exclusion criteria**	**Study size**	**Age**	**Involvement of company**
Morrison, 1992 USA	Cross-sectional; All had self- & clinician-samples	Follow up (colposcopy clinic)	Incl: - equivocal cervical cytology	25	Not specified	1
Hillemanns, 1999 Germany	Cross-sectional; All had self- & clinician-samples	Follow up (colposcopy clinic)	Not specified	247	Not specified	0
Sellors, 2000 Canada	Cross-sectional; All had self- & clinician-samples	Follow up (colposcopy clinic)	Incl: - equivocal cervical cytology	200	Mean: 31.5y (SD=9.4y) Range: ≥ 18y	5
Wright, 2000 South-Africa	Cross-sectional; All had self- & clinician-samples	Screening	Incl: - unscreened for ≥ 3y	1415	Median: 39y Range: 35-65y	5
Belinson, 2001 China	Cross-sectional; All had self- & clinician-samples	Screening	Excl: - pregnancy, history of cervical screening, pelvic radiation, or hysterectomy	1997	Mean: 39.1y (SE=3.16y)	5
Lorenzato, 2002 Brazil	Cross-sectional; All had self- & clinician-samples	Screening (high-risk population)	Excl: - cervix removed - illness impeding participation	253	Mean: 38.1y (SD=13.7y) Median: 38y	1
Nobbenhuis, 2002 The Netherlands	Cross-sectional; All had self- & clinician-samples	- Follow up + healthy participants (colposcopy clinic)	Incl: - equivocal cervical cytology (very mild dyskaryosis or worse) - normal cervical cytology	71	Mean: 35y	1

Author, year Country	**Study design**	**Population/ setting**	**Inclusion & exclusion criteria**	**Study size**	**Age**	**Involvement of company**
Garcia, 2003 Mexico, Peru, USA	Cross-sectional; All had self- & clinician-samples	Follow up (colposcopy clinic)	Excl: - history of hysterectomy, vaginal trauma or laceration - pregnancy	334	Mean: 36.9y	2
Salmeron, 2003 Mexico	Cross sectional; All had self- & clinician-samples	Screening	Excl: - history of CIN 2+ - hysterectomy - pregnancy	7868	Mean: 42.5y	3
Brink, 2006 The Netherlands	Cross-sectional; All had self- & clinician-samples	- Follow up + healthy participants (colposcopy clinic)	Incl: - repeat equivocal cervical cytology	96	Median: 35y Range: 18-59y	0
Daponte, 2006 Greece	Clinical prospective evaluation study; All had self- & clinician-samples	Follow up (colposcopy clinic)	Excl: - HIV-positivity	98	Not documented	1
Girianelli, 2006 Brazil	Cross-sectional; All had self- & clinician-samples	Screening (high-risk population)	Incl: - unscreened for >3year Excl: - pregnancy - delivery <6month ago - never having had sexual intercourse - hysterectomy	1777	Median: 39y Mean: 39y Range: 25-59y	2
Holanda, 2006 Brazil	Cross-sectional; All had self- & clinician-samples	Screening	Incl: - sexually active Excl: - pregnancy - hysterectomy	878	Range: 15-69y	5

Author, year Country	**Study design**	**Population/ setting**	**Inclusion & exclusion criteria**	**Study size**	**Age**	**Involvement of company**
Seo, 2006 South-Korea	Cross-sectional; All had self- & clinician-samples	Follow up (colposcopy clinic)	Incl: - equivocal cervical cytology (ASC-US+)	118	Mean: 46.2y	1
Szarewski, 2007 United Kingdom	Cross-sectional; All had self- & clinician-samples	Screening	Excl: - history of ablative or excisional treatment of the cervix	920	Median: 29y (pop 1) Median: 41y (pop 2)	2
Qiao, 2008 China	Cross sectional; All had self- & clinician-samples	Screening	Excl: - pregnancy - history of CIN, pelvic radiation, or hysterectomy	2530	Mean: 43.4y (SD=6.2y) Range: 30-55y	6
Bhatla, 2009 India	Cross sectional; All had self- & clinician-samples	Screening (high-risk population)	Incl: - sexually active women - persistent vaginal discharge, intermenstrual or postcoital bleeding or an unhealthy cervix Excl: - age <30 years - unmarried - hysterectomy or prior surgical procedures on the cervix - gross tumour on the cervix - pregnancy	546	Median: 36y	2
Balasubramanian, 2010 USA	Cross-sectional; All had self- & clinician-samples	Screening (high-risk population)	Excl: - pregnancy - chronically immunecompromised - prior treatments for cervical neoplasia	1665	Median: 23y Range: 18-50y	1

Author, year Country	**Study design**	**Population/ setting**	**Inclusion & exclusion criteria**	**Study size**	**Age**	**Involvement of company**
Gustavsson, 2011 Sweden	Cross-sectional; All had self- & clinician-samples	Follow up (colposcopy clinic)	Incl: - unscreened for ≥ 6years - HPV-positivity in a previous self-obtained sample §	50	Range: 39-60y	1
Lazcano-Ponce, 2011 Mexico	Randomized trial; - Study group (self-obtained sample) - Control group (clinician-obtained sample-	Screening	Incl: - participating in a poverty-reduction programme Excl: - hysterectomy - pregnancy	HPV: 9202 Cyto: 11054	Range: 25-65y	3
Taylor, 2011 South-Africa	Cross-sectional; All had self- & clinician-samples	Post-treatment follow up + healthy participants	Incl: - subjects derived from randomized clinical trial evaluating the safety and efficacy of two screen-and-treat approaches for cervical cancer prevention - women who had undergone cryotherapy in the two screen and-treat groups + all women in the control group who did not undergo cryotherapy.	2670	Mean: 43y Range: 35-65y	3
Twu, 2011 Taiwan	Cross-sectional; All had self- & clinician-samples	Follow up (colposcopy clinic)	Incl: - unscreened for ≥ 3years Excl: - acute cervicitis or vaginitis - pregnancy - menstruating period - sexual intercourse <2d before the study	252	Median: 42y Range: 26-79y	1

Author, year Country	**Study design**	**Population/ setting**	**Inclusion & exclusion criteria**	**Study size**	**Age**	**Involvement of company**
Wikstrom, 2011 Sweden	Randomized trial; - Study group (self- and/or clinician-obtained sample) - Control group (clinician-obtained sample)	Screening	Incl: - unscreened for ≥ 6year	4060	Range: 39-60y	4
Belinson, 2012 China	Cross-sectional; All had self- & clinician-samples	Screening	Incl: - unscreened for ≥ 3years Excl: - pregnancy - hysterectomy - history of pelvic radiation	8556	Mean: 38.9y Range: 25-59y	0
Dijkstra, 2012 The Netherlands	Cross-sectional; All had self- & clinician-samples	Follow up (colposcopy clinic)	Incl: - equivocal cervical cytology (moderate dyskaryosis or worse) - post-coital bleeding (normal cytology)	135	Median: 34y	4
Longatto-Fillho, 2012 Argentina, Brazil	Cross-sectional and prospective cohort; All had self- & clinician-samples	Screening	Incl: - consecutive series of women at their first visit to the clinic	12114	Mean: 37y* Range: 14-67y*	2
van Baars, 2012 The Netherlands	Cross-sectional; All had self- & clinician-samples	Follow up (colposcopy clinic)	Incl: - equivocal cervical cytology	134	Mean: 40y (SD=9.5y) Range: 21-66y	2

Author, year Country	**Study design**	**Population/ setting**	**Inclusion & exclusion criteria**	**Study size**	**Age**	**Involvement of company**
Zhao, 2012 China	Cross-sectional; All had self- & clinician-samples	Screening	Incl: - sexually active - having an intact uterus - unscreened for ≥ 5years Excl: - pregnancy - history of CIN2+, or pelvic radiation	13004	Mean: 37.9y (SD=11.2y)	3
Darlin, 2013 Sweden	Cross-sectional; All had self- & clinician-samples	Follow up (colposcopy clinic)	Incl: - equivocal cervical cytology	108	Mean: 34y Range: 18-65y	1
Geraets, 2013 Spain	Cross-sectional; All had self- & clinician-samples	Follow up (colposcopy clinic)	Incl: - equivocal cervical cytology	182	Median: 34y Range: 16-76y	4
Guan, 2013 China	Cross-sectional; All had self- & clinician-samples	Screening	Incl: - VIA or VILI positive - VIA or VILI negative (random sample)	174	Not documented	1
Jentschke, 2013 Germany	Retrospective; All had self- & clinician-samples	Follow up (colposcopy clinic)	Not documented	72	Mean: 37.1y Range: 16-68y	2
Nieves, 2013 Mexico	Cross-sectional; All had self- & clinician-samples	Screening	Excl: - pregnancy - history of hysterectomy or pelvic irradiation	2049	Median: 39.2y Range: 30-50y	1

Author, year Country	**Study design**	**Population/ setting**	**Inclusion & exclusion criteria**	**Study size**	**Age**	**Involvement of company**
Involvement of company: 0, not documented; 1, no connection with company; 2, material or financial support from company; 3, author linked (consultant, shareholder, speakers honorarium, advisory board, etc.) to manufacturer of tests/clinician-sample device; 4, author linked to manufacturer of self-sampling device; 5, author employed by manufacturer of tests/clinician-sample devices; 6, author employed by manufacturer of self-sampling device. * retrieved from Syrjänen, 2005 (Anticancer Research) § Sanner K, Wikstrom I, Strand A, Lindell M, Wilander E. Self-sampling of the vaginal fluid at home combined with high-risk HPV testing. Br J Cancer 2009; 101(5):871-874.						

Supplementary Table 2. Characteristics of tests and disease verification.

Author, year	Tests	Test	Sampling device	Storage medium	Golden standard	Test cut-off	Outcome
Morrison, 1992	Self: HPV Clin: HPV	PCR (L1 consensus)	Self: lavage (My-PAP)	ethanol carbowax	Colposcopy + colpo-directed biopsy - All participants	Not Specified	CIN2+
Hillemanns, 1999	Self: HPV Clin: HPV	HC2	Self: cytobrush Clin: cytobrush	STM (Digene)	Colposcopy + colpo-directed biopsy and/or endocervical curettage - All participants	HC2: 1pg/ml	CIN2+
Sellors, 2000	Self: HPV Clin: HPV, cytology	HPV: HC2, PCR (L1 consensus) Cyto: cPap	Self: Dacron Polyester Swab Clin: soft cone-shaped cervical brush, Ayre spatula, and Dacron Polyester Swab	Self: STM Clin brush: STM Clin swab: sterile phosphate-buffered saline	Colposcopy + colpo-directed biopsy and/or endocervical curettage - All participants	HC2: 1pg/ml	CIN2+
Wright, 2000	Self: HPV Clin: HPV, cytology	HPV: HC2 Cyto: cPap	Self: Dacron Polyester Swab Clin: AccelonCombi Cervical Biosampler (cyto), conical brush (HPV)	STM	Colposcopy + colpo-directed biopsy/loop excision or endocervical curettage - Participants with at least one positive test result	HC2: 1pg/ml Cyto: ASCUS+, LSIL+	CIN2+

Author, year	Tests	Test	Sampling device	Storage medium	Golden standard	Test cut-off	Outcome
Belinson, 2001	Self: HPV Clin: HPV, cytology	HPV: HC2 Cyto: LBC (ThinPrep)	Self: Dacron Swab Clin: plastic spatula, endocervical brush	Self: STM Clin: PreservCyt	Colposcopy + colpo-directed biopsy + multiple random biopsies - All participants	HPV: 1 pg/ml	CIN2+
Lorenzato, 2002	Self: HPV Clin: HPV	PCR (MY 9/11)	Self: cotton swab Clin: Ayre spatula, cytobrush	Self + Clin: PBS	Colposcopy + colpo-directed biopsy - All participants	Not documented	CIN2+ CIN3+
Nobbenhuis, 2002	Self: HPV Clin: HPV, cytology	HPV: PCR (GP 5+/6+) Cyto: cPap	Self: lavage Clin: lavage, Cervex-Brush	Self + Clin: PBS	Colposcopy + colpo-directed biopsy - All participants	Cyto: ASC-US+	CIN2+
Garcia, 2003	Self: HPV Clin: HPV, cytology	HPV: PCR (PGMY09/11) Cyto: LBC (ThinPrep)	Self: cytobrush Clin: Ayre spatula and endocervical brush	Self: PreservCyt Clin: methanol buffer solution	Colposcopy + colpo-directed biopsy (+ endocervical curettage) - All participants	Cyto: ASCUS+	CIN2+
Salmeron, 2003	Self: HPV Clin: HPV, cytology	HPV: HC2 Cyto: cPap	Self: Dacron Swab Clin: conical cytobrush	Self + Clin: STM	Colposcopy + colpo-directed biopsy (+ endocervical curettage) - Participants with at least one positive test result	HPV: 1 pg/ml Cyto: ASC-US+	CIN2+
Brink, 2006	Self: HPV Clin: HPV,	HPV: PCR (GP5+/6+) Cyto: LBC (SurePath)	Self: cervicovaginal	Self + Clin: SurePath	Colposcopy + colpo-directed	Cyto: ASC-US+, HSIL+	CIN2+

Author, year	Tests	Test	Sampling device	Storage medium	Golden standard	Test cut-off	Outcome
	cytology		lavage (Mermaid) Clin: endocervical brush		biopsy - Participants with equivocal cytology		
Daponte, 2006	Self: HPV Clin: HPV	PCR (L1 and E6 type-specific primers for HPV16)	Self +Clin: cytobrush	Self + Clin: PBS	Colposcopy + colpo-directed biopsy - All participants	Not documented	CIN2+
Girianelli, 2006	Self: HPV Clin: HPV, cytology	HPV: HC2 Cyto: cPap, LBC (Citoliq)	Self: conical brush Clin: conical brush (HPV, LBC), Ayre spatula and endocervical brush (cPap)	HPV: Citoliq	Colposcopy + colpo-directed biopsy - Participants with at least one positive test result - Systematic sample of 70 women with negative tests	HPV: 1pg/ml Cyto: ASC-US+	CIN2+
Holanda, 2006	Self: HPV Clin: HPV	HC2	Self: collection brush Clin: small conical brush	Self + Clin: UCM	Colposcopy + colpo-directed biopsy - All participants	1pg/ml	CIN2+
Seo, 2006	Self: HPV Clin: HPV	HPV DNA Chip	Self: Dacron Polyester Swab Clin: Dacron Polyester Swab	Not documented	Colposcopy + colpo-directed biopsy (+ endocervical curettage) or random biopsies - All participants		CIN2+ CIN3+
Szarewski, 2007	Self: HPV	HPV: HC2	Self: cotton swab	Not	Colposcopy +	HPV: 1 pg/ml	CIN2+

Author, year	Tests	Test	Sampling device	Storage medium	Golden standard	Test cut-off	Outcome
	Clin: HPV, cytology	Cyto: cPap	ClinCyto: pointed spatula, endocervical brush Clin HPV: Cervical Sampler Brush	documented	colpo-directed biopsy - Participants with at least one positive test result - Random sample (5%) of women with negative tests	Cyto: ASC-US+	
Qiao, 2008	Self: HPV Clin: HPV, cytology	HPV: CareHPV Cyto: LBC (SurePath)	Self: vaginal brush (careHPV) Clin: nylon swab (LBC), cervical brush (care HPV)	- HPV: collection medium (QIAGEN) - LBC: SurePath	Colposcopy + colpo-directed biopsy and endocervical curettage - All participants $	careHPV: 0.5 pg/ml, 1 pg/ml Cyto: ASC-US+, LSIL+	CIN2+ CIN3+
Bhatla, 2009	Self: HPV Clin: HPV, cytology	HPV: HC2, PCR (PGMY09/11) Cyto: cPap	Self: cervical sampling brush Clin: Ayre spatula and endocervical brush (cyto), cervical sampling brush (HPV)	HPV: STM	Colposcopy + colpo-directed biopsy - All participants	- HPV: 1pg/ml - Cyto: ASC-US+, LSIL+	CIN2+
Balasubramanian, 2010	Self: HPV Clin: HPV	HC2	Self: Dacron Swab Clin: Dacron Swab	HPV: STM	Colposcopy + colpo-directed biopsy or random biopsy - Participants with	HPV: 1 pg/ml	CIN2+

Author, year	Tests	Test	Sampling device	Storage medium	Golden standard	Test cut-off	Outcome
					at least one positive test result - Random sample of women with negative tests		
Gustavson, 2011	Self: HPV Clin: HPV	PCR (primers for E6/E7/L1) *	Self: Viba-Brush Clin: cytobrush	Self + Clin: FTA cartridge	Colposcopy + colpo-directed biopsy - All participants	10 Geq/PCR	CIN2+
Lazcano-Ponce, 2011	Self: HPV (study-group) Clin: cytology (control-group)	HPV: HC2 Cyto: cPap	HPV: conical brush Cyto: Ayre wooden spatula, cytobrush	Self: STM	Colposcopy + colpo-directed biopsy - Participants with at least one positive test result	Cyto: LSIL+ HPV: 1 pg/mL, 2 pg/mL, 5pg/ml	CIN2+
Taylor, 2011	Self: HPV Clin: HPV, cytology	HPV: HC2 Cyto: cPap, LBC (ThinPrep)	Self: Dacron Swab Clin: plastic spatula, cytobrush	Self: STM Clin: PreservCyt	Colposcopy + endocervical curettage and/or colpo-directed biopsy - All participants	HPV: 1 pg/mL Cyto: ASC-US+, LSIL+	CIN2+
Twu, 2011	Self: HPV Clin: HPV	PCR (MY9/11 nested GP5+/6+)	Self: cytobrush Clin: Ayre spatula, cytobrush	Self + Clin: STM	Colposcopy + colpo-directed biopsy - participants with acetowhite lesions (VIA), or a positive cPap	- HPV Blot : 1-50 copies of HPV geq per PCR.	CIN2+ CIN3+
Wikstrom, 2011	Self: HPV	HPV: HC2	Self: plastic wand	Not	colposcopy +	HPV: 1	CIN2+

Author, year	Tests	Test	Sampling device	Storage medium	Golden standard	Test cut-off	Outcome
	(Study group) Clin: cytology (Study and Control group)	Cyto: cPap	(Qvintip)	documented	colpo-directed biopsy or random biopsy; or repeat cytology - Participants with a positive test result	pg/mL Cyto: ASC-US	
Belinson, 2012	Self: HPV Clin: HPV, cytology	Self HPV: Cervista, MALDI-TOF Clin HPV: HC2, Cervista, MALDI-TOF Clincyto: cPap	Self: POI/NIH self-sampler, conical brush Clin: broom sampler	Self + clin: PreservCyt	Colposcopy + colpo-directed biopsy, or random biopsy and endocervical curettage - Participants with at least one positive test result	ASC-US	CIN3+
Dijkstra, 2012	Self: HPV Clin: HPV	PCR (GP5+/6+)	Self: Viba-Brush Clin: Viba-Brush, Cervex-Brush	Self + Clin: PreservCyt	Colposcopy + colpo-directed biopsy, or random biopsy (≥1) - All participants	Not documented	CIN2+
Longatto-Fillho, 2012	Self: HPV Clin: HPV, cytology	HPV: HC2 Cyto:cPap, LBC (SurePath), LBC (Citoliq)	Self HPV: tampon Clin HPV: cervical swab Clincyto: cervix brush (Surepath), DNA-Citoliq Brush	Self + Clin HPV: STM ClinCyto: SurePath, Citoliq	Colposcopy + colpo-directed biopsy - Participants with at least one positive test result	HPV: 1pg/ml hrHPV Cyto: ASC-US+, LSIL+, HSIL+	CIN2+ cancer

Author, year	Tests	Test	Sampling device	Storage medium	Golden standard	Test cut-off	Outcome
					- Random sample (>5%) of women with negative tests **		
van Baars 2012	Self: HPV Clin: HPV	SPF10-PCR, PCR (GP5+/6+)	Self: Evalyn-Brush Clin: Cervex-Brush	Self: FTA cartridge Clin: ThinPrep, SurePath	Colposcopy + colpo-directed biopsy - 44 out of 134 women (if histological result was available), for others follow-up cytology - All participants	Not documented	CIN2+ CIN3+
Zhao, 2012	Self: HPV Clin: HPV, cytology	HPV: HC2 Cyto: LBC (ThinPrep)	Self***: Dacron Swab Clin***: plastic spatula, endocervical brush	Self***: STM Clin***: PreservCyt	Colposcopy + colpo-directed or random biopsy (4) - Participants with at least one positive test result	HPV: 1 pg/ml Cyto: ASC-US	CIN2+ CIN3+
Darlin, 2013	Self: HPV Clin: HPV	PCR (GP5+/6+)	Self: coton swab Clin: Cervex-Brush Combi	Clin: PreservCyt	Colposcopy + colpo-directed biopsy - All participants	Not documented	CIN2+
Geraets, 2013	Self: HPV Clin: HPV	SPF10-PCR, PCR (GP5+/6+)	Self: Viba-Brush Clin: Cervex-Brush	Self: FTA cartridge Clin: PreservCyt	Colposcopy + colpo-directed biopsy - All participants	Not documented	CIN2+ CIN3+
Guan, 2013	Self: HPV	Linear Array	cervical sampler	Self + Clin: FTA	Colposcopy +	Not	CIN2+

Author, year	Tests	Test	Sampling device	Storage medium	Golden standard	Test cut-off	Outcome
	Clin: HPV		brush	cartridge	colpo-directed biopsy - All participants	documented	
Jentschke, 2013	Self: HPV Clin: HPV, p16	HPV: HC2 P16: $p16^{INK4a}$ ELISA	Self: lavage (Delphi screener) Clin: Not documented	Self: buffered saline Clin: PreservCyt, Cervatec	Colposcopy + colpo-directed biopsy - All participants	Not documented	CIN2+ CIN3+
Nieves, 2013	Self: HPV Clin: HPV, cytology	HPV: HC2, APTIMA Cyto: LBC (Thinprep)	Self: POI/NIH self-sampler Clin : broom sampler	Self + Clin: PreservCyt	Colposcopy + colpo-directed cryoterapy or colpo-directed biopsy and/or multiple random biopsies - Participants with at least one positive test result	Cyto: ASCUS HPV: Not documented	CIN2+ CIN3+

⌗ (30 participants did not undergo colposcopy (likely part of the group of 32 healthy volunteers)

$ participants with a negative colposcopy, but abnormal cytology or a positive HPV-test, had a second colposcopy + four-quadrant biopsies + endocervical curettage

¥ women who had a positive HC2-test, and a random sample of women with a negative HC2-test or who were CIN2+

* Retrieved from: Moberg M, Gustavsson I, Gyllensten U. Real-time PCR-based system for simultaneous quantification of human papillomavirus types associated with high risk of cervical cancer. *J Clin Microbiol* 2003; 41(7):3221-3228.

** Retrieved from: Syrjanen K, Naud P, Derchain S, Roteli-Martins C, Longatto-Filho A, Tatti S et al. Comparing PAP smear cytology, aided visual inspection, screening colposcopy, cervicography and HPV testing as optional screening tools in Latin America. Study design and baseline data of the LAMS study. *Anticancer Res* 2005; 25(5):3469-3480.

*** Retrieved from: Belinson JL, Qiao YL, Pretorius R, Zhang WH, Elson P, Li L et al. Shanxi province cervical cancer screening study: a cross-sectional comparative trial multiple techniques to detect cervical neoplasia. *Gynecol Oncol* 2001; 83:439-444.

Abbreviations: ASC-US, Atypical Squamous Cells of Undetermined Significance ; CIN, Cervical Intraepithelial Neoplasia; colpo, colposcopy; cPap,

Author, year	Tests	Test	Sampling device	Storage medium	Golden standard	Test cut-off	Outcome
conventional Pap smear; cyto, cytology; HPV, Human Papillomavirus; LBC, Liquid-Based Cytology; PBS, Phosphate Buffered Saline; STM, Specimen Transport Medium; UCM, Universal Collection Medium.							

7.5. Assays used for HPV and cytology testing.

Supplementary table 3. Used tests (abbreviations, manufacturer and study where the tests were applied).

Abbreviation	Assay	Manufacturer	Studies
AB	Abbott RT PCR hrHPV (Multiplex real-time PCR test that targets the (GP5+/6+) **L1 region** of 14 hrHPV types PCR)	Abbott Molecular, Inc., Des Plaines, IL, USA	Jen13a, Jen13b
APTI	APTIMA (A multiplex *in vitro* nucleic acid amplification test targeting **E6/E7** mRNA from 14 hrHPV types)	Gen-Probe Inc., San Diego, CA, USA	Nie13
Cerv	Cervista (A signal amplification method by Invader chemistry using 3 ologinucleotide mixtures, together targeting 14 hrHPV types)	Hologic, Bedford, MA, USA	Bel12
cHPV	careHPV (A signal amplification method (simplified HC2) targeting 14 hrHPV types) [0.5]: cutoff at RLU>0.5 [1]: cutoff at RLU>1	QIAGEN Corporation, Gaithersburg, MD, USA	Qia08
DNAch	DNAchip (A broad spectrum PCR based on GP5+/6+PCR targeting the **L1 region** to detect and genotype 15 hrHPVs and 9 lrHPVs)	Biomedlab Co., Seoul, Korea	Seo06
GP5+/6+-Lum	modified GP5+/6+ with Luminex read-out targeting the **L1 region** of hr- and lr HPVs)	-	Dar13

HC2	Hybrid Capture-2 (A signal amplification method targeting 13 hrHPV types)	Qiagen Corporation, Gaithersburg, MD, USA	Hil99, Sel00, Wri00, Bel01, Sal03, Gir06, Hol06, Sza07, Bha09, Bal10, Laz11, Tay11, Wik11, Lon12, Zha12a, Zha12b, Zah12c, Jen13a, Jen13b, Nie13
LBC-TP	Liquid-Based Cytology (ThinPrep)	Cytyc Corporation, Boxborough, MA, USA	Gar03, Bri06
M-TOF	MALDI-TOF (GP5+/6+ based PCR with MALDI-TOF read out to detect **L1 region**)	AB SCIEX, Foster City, CA, USA	Bel12
Pap	Conventional cytology	-	Nob02
PCR other	PCR using primers, other than GP5+/6+:	-	
	MY9/11 (**L1 region**)		Mor92, Lor02,
LA	PGMY9/11 (**L1 region**)		Gar03, Bha09
	Linear Array (PGMY09/11 **L1** consensus primer PCR test that detects 37 HPV types by reverse line blot hybridization)	Roche MolecularSystems, Alameda, CA, USA	Gua13

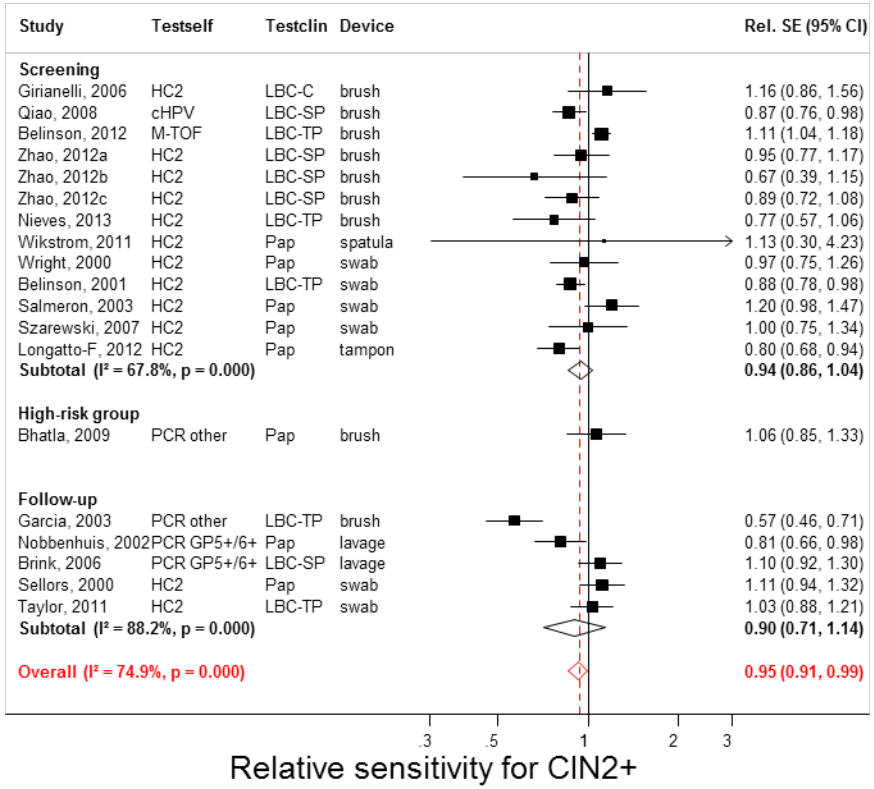
7.6. Devices used for self-sampling

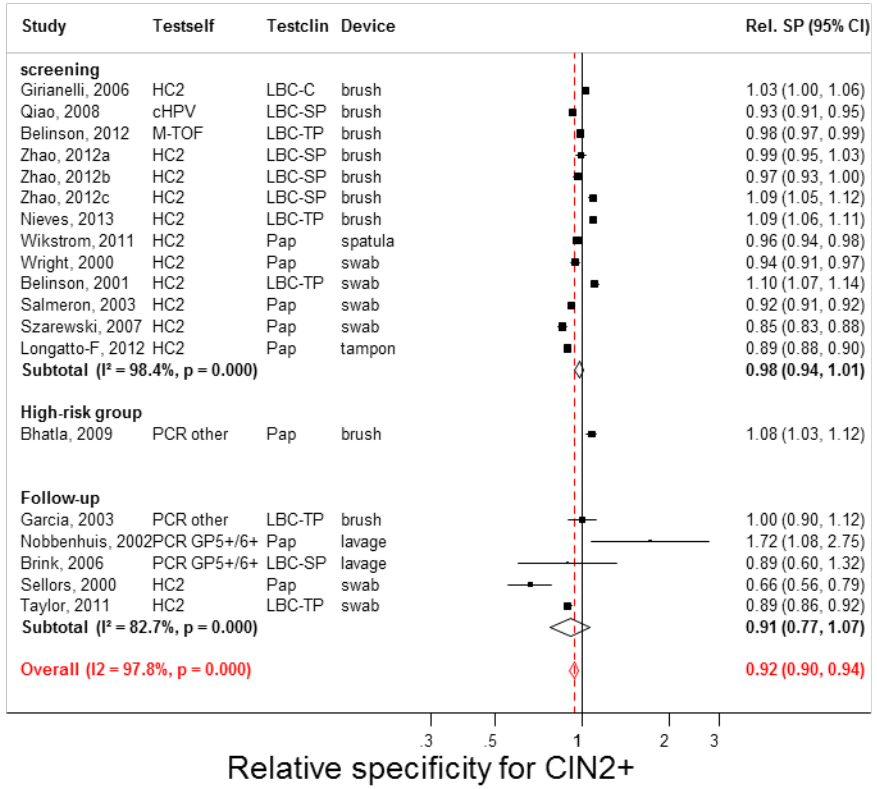
Supplementary Table 4. Devices used for self-sampling, grouped in five categories.

Device group	Device	Manufacturer	Studies
Brush	Cervical brush	-	Hol06
	Conical brush	-	Bha09, Laz11, Gir06, Bel12
	Cyto-Brush	-	Hil99, Dap06, Twu11
	Cyto-Brush PLUS	Cooper Surgical, Trumbull, CT, USA	Gar03
	Evalyn Brush	Rovers Medical Devices B.V., Oss, Netherlands	Van12
	POI self-sampler	-	Nie13
	POI/NIH self-sampler	-	Bel12
	Qiagen/Digene cervical sampler	Qiagen (previously Digene Corporation), Gaithesburg, MD, USA	Zha12a, Zha12b, Zha12c,Gua13
	Vaginal brush	-	Qia08, Dij12
	Viba-Brush + FTA Elute cartridge	Vibabrush: Rovers Medical Devices B.V., Oss, Netherlands FTA cartridge: GE Healthcare, Buckinghamshire, United Kingdom	Gus11, Ger13
Lavage	*Delphi Screener*	Delphi Bioscience, Scherpenzeel, the Netherlands	Jen13a, Jen13b
	Lavage (15 ml)	-	Nob02
	Mermaid (5 ml)	(previous Delphi Screener)	Bri06
	MY-PAP (21ml)	Medtech, Bohemia, NY, USA	Mor92
Spatula	*Qvintip*	AprovixAB, Uppsala, Sweden	Wik11
Swab	Cotton swab	-	Lor02, Sza07, Dar13
	Dacron swab	-	Sel00, Wri00, Bel01, Sal03, Seo06, Bal10, Tay11,
Tampon	Tampon	-	Lon12

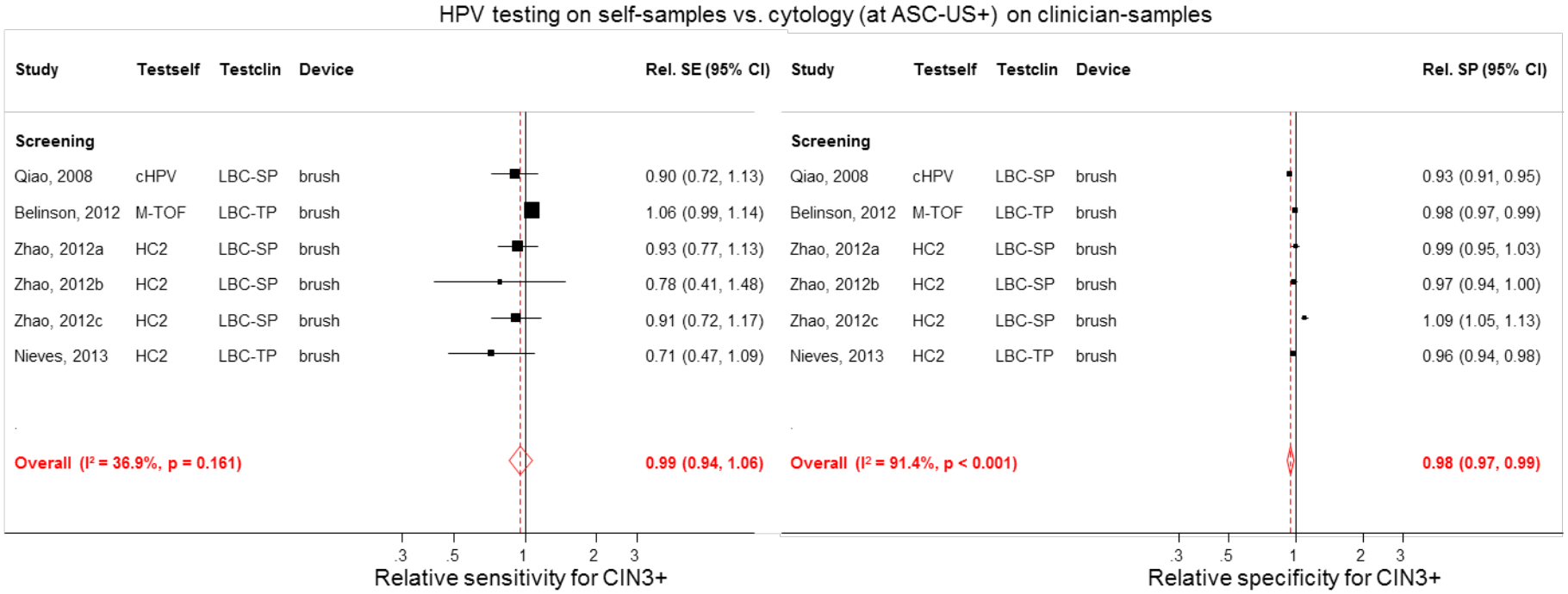
7.7. Forrest plots of the relative accuracy of hrHPV testing on self-samples versus hrHPV testing or cytology on clinician samples

Supplementary Figure 1a. Relative sensitivity (left) and specificity (right) of hrHPV testing on self-samples compared to cytology at cut-off ASC-US+ on clinician samples to detect CIN2+, by clinical setting.

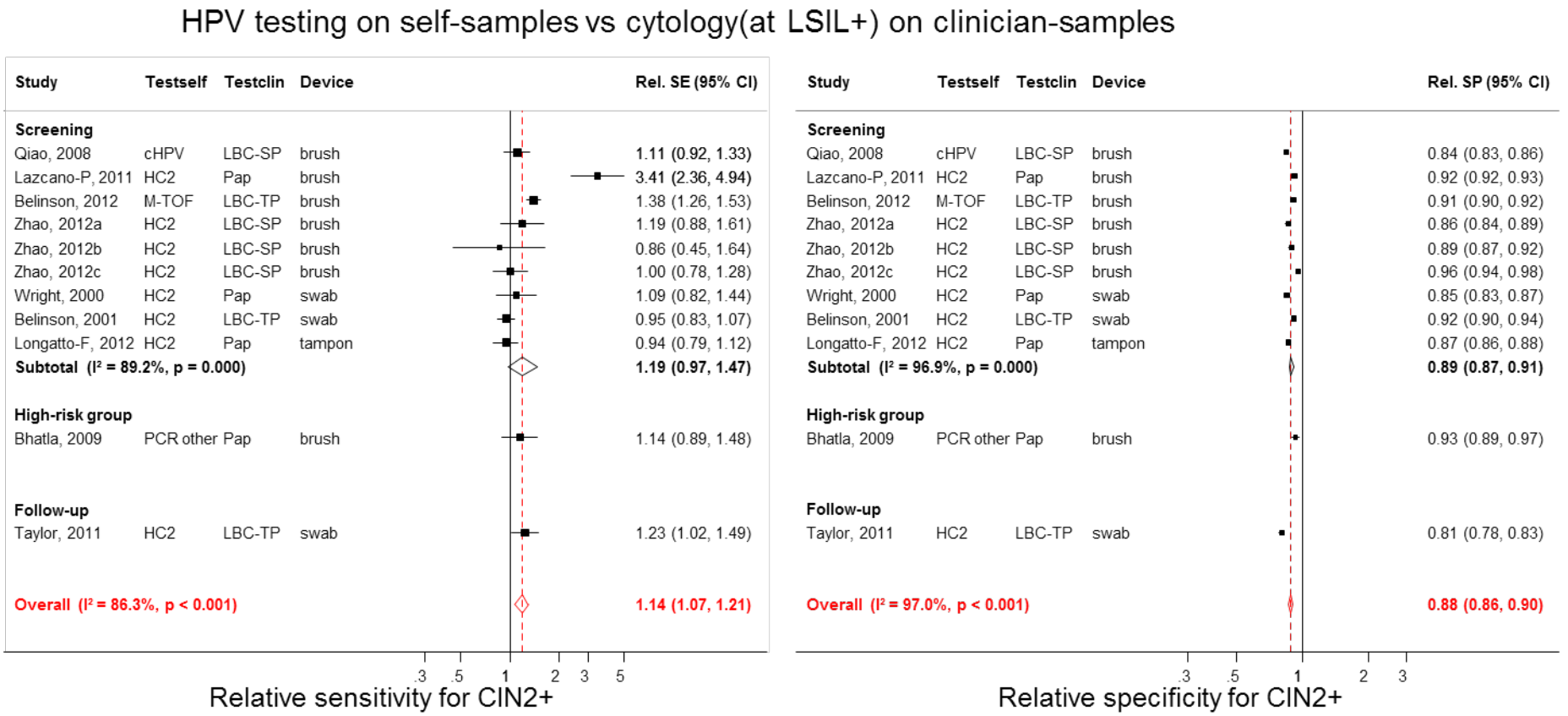




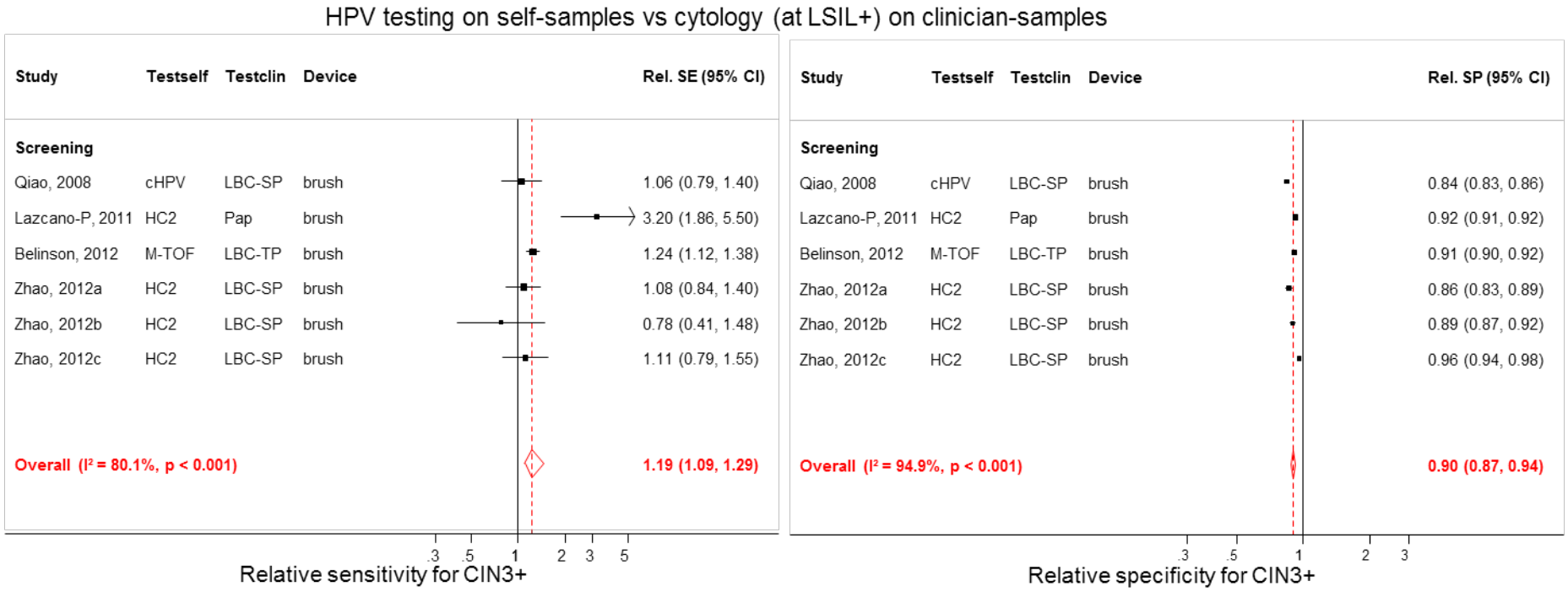
Supplementary Figure 1b. Relative sensitivity (left) and specificity (right) of hrHPV testing on self-samples compared to cytology at cut-off ASC-US+ on clinician samples to detect CIN3+, by clinical setting.



Supplementary Figure 1c. Relative sensitivity (left) and specificity (right) of hrHPV testing on self-samples compared to cytology at cut-off LSIL+ on clinician samples to detect CIN2+, by clinical setting.

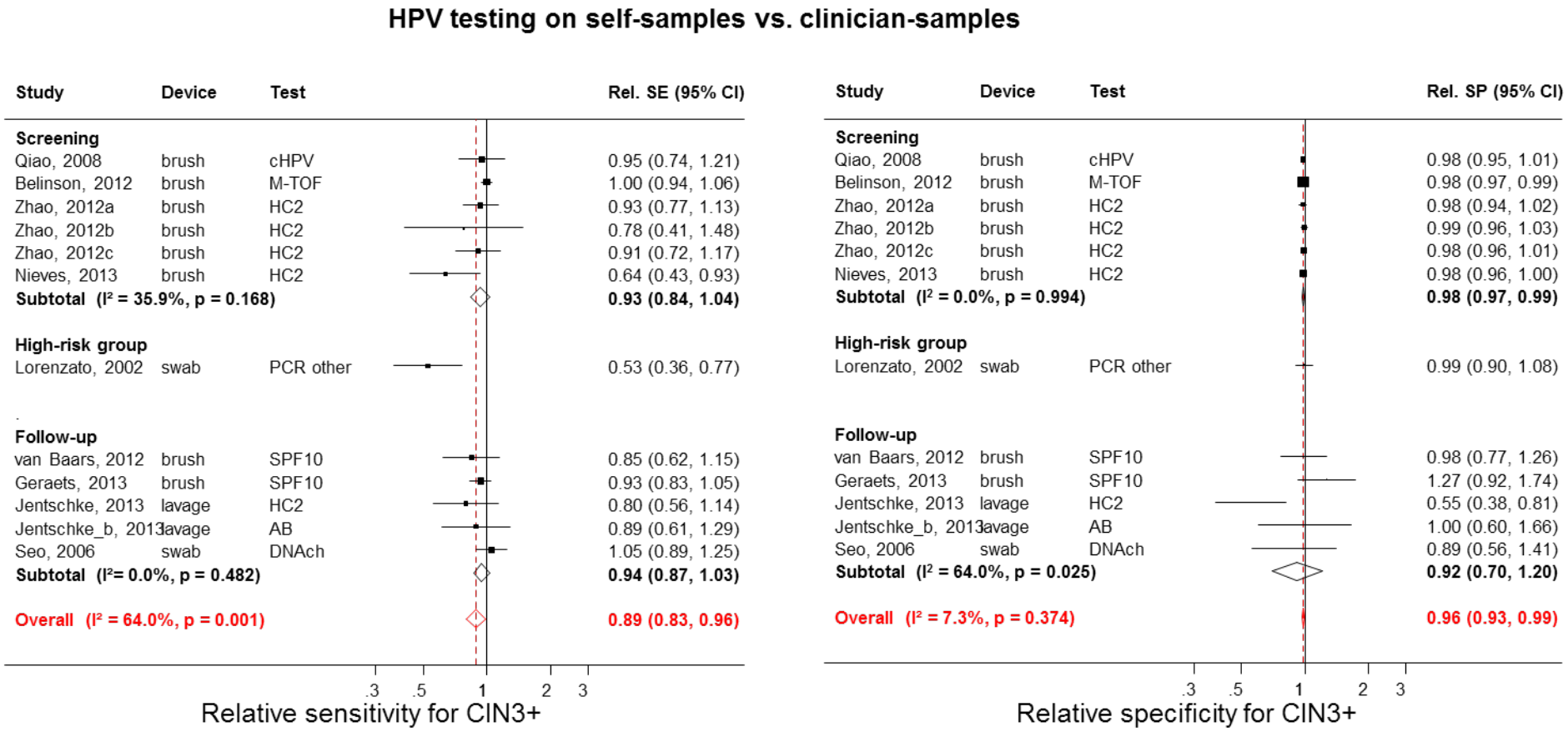


Supplementary Figure 1d. Relative sensitivity (left) and specificity (right) of hrHPV testing on self-samples compared to cytology at cut-off LSIL+ on clinician samples to detect CIN3+, by clinical setting.



Supplementary Figure 1f. Relative sensitivity (left) and specificity (right) of hrHPV testing on self-samples compared to hrHPV testing on clinician samples to detect CIN3+, by clinical setting

for relative sensitivity and specificity of hrHPV testing on self-samples compared to hrHPV testing on clinician samples to detect CIN2+, by clinical setting see Figure 3 in the publication of M. Arbyn et al. 2014:)
(http://www.ncbi.nlm.nih.gov/pubmed/?term=Accuracy+of+human+papillomavirus+testing+on+self-collected+versus+clinician-collected+samples



7.8. Variation in the absolute accuracy of HPV testing on self-samples to detect CIN2+ in women attending cervical cancer screening by covariates

Supplementary Table 5. Variation in absolute sensitivity and specificity (and 95% CI) of HPV testing on self-samples by covariates. Relative values were computed using a bivariate normal model(relative values are computed using a bivariate normal model).

Co-variate	Number of studies (test/device combinations)	Sensitivity	Relative to reference	Specificity	Relative to reference
Test					
HC2 (ref)	11	0.72 (0.66-0.79)	1	0.86 (0.83-0.89)	1
APTIMA	1	0.61 (0.40-0.79)	0.83 (0.61-1.15)	0.90 (0.64-0.92)	1.04 (1.02-1.07)*
Linear Array	1	0.78 (0.39-0.95)	1.06 (0.72-1.55)	0.79 (0.61-0.90)	0.92 (0.76-1.12)
Cervista	1	0.71 (0.54-0.83)	0.96 (0.76-1.20)	0.87 (0.75-0.94)	1.02 (0.91-1.14)
MALDI-TOF	1	0.93 (0.85-0.97)	1.25 (1.12-1.39)*	0.88 (0.77-0.94)	1.04 (0.94-1.14)
CareHPV (≥0.5)†	1	0.82 (0.64-0.92)	1.12 (0.93-1.35)	0.82 (0.67-0.91)	0.96 (0.82-1.11)
CareHPV (≥1)†	1	0.73 (0.53-0.87)	1.00 (0.78-1.29)	0.88 (0.76-0.94)	1.02 (0.91-1.14)
Collection device for self-sampling					
Brush (ref)	9(12)	0.78 (0.71-0.84)	1	0.86 (0.82-0.89)	1
Swab	4	0.76 (0.66-0.84)	0.97 (0.84-1.12)	0.85 (0.79-0.90)	0.99 (0.92-1.07)
Tampon	1	0.57 (0.37-0.75)	0.73 (0.51-1.06)	0.86 (0.71-0.94)	1.00 (0.87-1.14)
Risk of bias with respect to enrolment of study subjects					
Low concern (ref)	8(10)	0.73 (0.64-0.80)	1	0.87 (0.84-0.90)	1
Moderate concern	5(6)	0.77 (0.67-0.85)	1.06 (0.90-1.25)	0.84 (0.79-0.88)	0.96 (0.90-1.03)
High concern	1	0.78 (0.37-0.96)	1.08 (0.71-1.62)	0.79 (0.60-0.90)	0.90 (0.74-1.090)
Risk of bias with respect to index & comparator tests					
Low concern (ref)	13(16)	0.75 (0.68-0.81)	1	0.86 (0.83-0.89)	1
Moderate concern	1	0.79 (0.36-0.96)	1.05 (0.70-1.57)	0.79 (0.60-0.91)	0.92 (0.75-1.12)

Co-variate	Number of studies (test/device combinations)	Sensitivity	Relative to reference	Specificity	Relative to reference
High concern	0	-	-	-	-
Risk of bias with respect of use and interpretation of the reference standard					
Low concern (ref)	11(13)	0.75 (0.68-0.81)	1	0.86 (0.83-0.89)	1
Moderate concern	3(4)	0.79 (0.36-0.96)	1.05 (0.70-1.57)	0.79 (0.60-0.91)	0.92 (0.75-1.12)
High concern	0	-	-	-	-
Time delay between index, comparator and reference tests					
Acceptable (ref)	10(12)	0.81 (0.78-0.83)	1	0.85 (0.81-0.88)	1
Unclear	2(3)	0.61 (0.55-0.66)	0.75 (0.69-0.82)*	0.89 (0.86-0.92)	1.06 (1.00-1.12)*
Inacceptable	2	0.66 (0.52-0.78)	0.82 (0.67-1.00)	0.81 (0.79-0.84)	0.96 (0.91-1.01)
Partial verification					
Avoided (ref)	11(13)	0.75 (0.67-0.82)	1	0.86 (0.82-0.89)	1
Unclear	3(4)	0.75 (0.62-0.84)	0.99 (0.83-1.19)	0.87 (0.80-0.92)	1.02 (0.94-1.10)
Not avoided	0	-	-	-	-
Differential verification					
Avoided (ref)	14(17)	0.75 (0.68-0.81)	1	0.86 (0.83-0.89)	1
Unclear	0	-	-	-	-
Not avoided	0	-	-	-	-
Withdrawals					
Explained (ref)	12(15)	0.76 (0.69-0.82)	1	0.87 (0.83-0.89)	1
Unclear	0	-	-	-	-
Not explained	2	0.70 (0.48-0.86)	0.92 (0.69-1.24)	0.80 (0.69-0.88)	0.93 (0.82-1.05)
Uninterpretable results of index & comparator tests					
Reported (ref)	6(8)	0.74 (0.67-0.80)	1	0.88 (0.83-0.91)	1
Unclear	1	0.57 (0.39-0.74)	0.77 (0.55-1.08)	0.86 (0.72-0.93)	0.98 (0.86-1.11)
Not Reported	7(8)	0.80 (0.72-86)	1.08 (0.95-1.22)	0.84 (0.79-0.88)	0.96 (0.90-1.03)
Uninterpretable results of the reference standard					
Reported (ref)	7(10)	0.76 (0.69-0.82)	1	0.88 (0.84-0.91)	1
Unclear	1	0.57 (0.37-0.75)	0.75 (0.52-1.07)	0.86 (0.73-0.93)	0.98 (0.86-1.11)
Not reported	6	0.77 (0.66-0.85)	1.01 (0.87-1.18)	0.88 (0.84-0.91)	0.95 (0.89-1.02)
Involvment of company					

Co-variate	Number of studies (test/device combinations)	Sensitivity	Relative to reference	Specificity	Relative to reference
No involvement	2(3)	0.61 (0.45-0.76)	1	0.88 (0.81-0.92)	1
Not documented	1(2)	0.82 (0.74-0.88)	1.34 (1.01-1.76)*	0.88 (0.78-0.93)	1.00 (0.90-1.11)
Material/financial supp.	3	0.66 (0.52-0.77)	1.07 (0.80-1.42)	0.85 (0.79-0.90)	0.97 (0.89-1.06)
Author linked to manufacturer of					
Tests	4	0.76 (0.66-0.84)	1.24 (0.95-1.62)	0.89 (0.85-0.92)	1.02 (0.94-1.10)
self-sampling devices	0	-	-	-	-
Author employed by manufacturer of					
Tests	3	0.78 (0.68-0.86)	1.28 (0.99-1.65)	0.79 (0.72-0.85)	0.90 (0.81-1.00)
self-sampling devices	1(2)	0.77 (0.66-0.86)	1.26 (0.94-1.69)	0.85 (0.74-0.92)	0.97 (0.86-1.09)
Developing/Developed country					
Developing (ref)	13(16)	0.75 (0.68-0.80)	1	0.86 (0.83-0.89)	1
Developed	1	0.81 (0.51-0.95)	1.09 (0.83-1.44)	0.82 (0.66-0.92)	0.96 (0.81-1.12)
Continent					
Europe (ref)	1	0.81 (0.57-0.93)	1	0.82 (0.66-0.92)	1
Africa	1	0.66 (0.50-0.79)	0.82 (0.59-1.12)	0.81 (0.65-0.91)	0.99 (0.79-1.24)
America (Centr. + South)	5(6)	0.65 (0.57-0.76)	0.80 (0.62-1.03)	0.86 (0.81-0.90)	1.05 (0.89-1.24)
Asia	7(9)	0.81 (0.76-0.85)	1.00 (0.80-1.25)	0.87 (0.82-0.90)	1.05 (0.90-1.24)

*significantly different from reference category; † cutoff expressed in relative light units;
Ref: reference category with default relative value =1.

7.9. Subgroup- meta-analysis: variation in the relative accuracy of hrHPV testing on self- versus clinician samples.

Supplementary Table 6. Sub-group meta-analysis of the relative sensitivity and specificity (and 95% CI) of HPV testing in self- compared to clinician-samples to detect CIN2+, by covariate.

Co-variate	Nb. of studies (test/device combinations)	Relative sensitivity	Relative Specificity
Test			
HC2	18	0.85 (0.81-0.90)*	0.96 (0.93-0.98)*
PCR GP5+/6+	5	0.95 (0.89-1.01)	1.11 (0.95-1.29)
CareHPV (at RLU≥0.5)	1	0.90 (0.79-1.04)	0.98 (0.95-1.00)
CareHPV (at RLU≥1)	1	0.86 (0.73-1.03)	1.00 (0.98-1.02)
PCR-SPF10	2	0.96 (0.89-1.02)	1.10 (0.85-1.41)
Abbott Real Time hrHPV Test	1	1.00 (0.75-1.34)	1.07 (0.65-1.78)
Cervista	1	0.76 (0.70-0.83)*	0.95 (0.94-0.96)*
APTIMA	1	0.64 (0.46-0.90)*	0.99 (0.98-1.01)
DNAchip	1	1.03 (0.89-1.19)	0.88 (0.55-1.42)
modified GP5+/6+ PCR with Luminex reading	1	0.96 (0.75-1.24)	0.94 (0.67-1.33)
Linear Array	1	0.79 (0.54-1.16)	1.00 (0.89-1.12)
MALDI-TOF	1	1.00 (0.95-1.05)	0.98 (0.97-0.99)*
Other nonGP5+/6+ PCR	7	0.82 (0.66-1.01)	1.02 (0.97-1.07)
Collection device for self-sampling			
Brush	18 (24)	0.89 (0.83-0.94)*	0.98 (0.97-0.99)*
Lavage	5 (6)	0.94 (0.85-1.03)	0.95 (0.68-1.34)
Swab	10	0.86 (0.80-0.92)*	0.95 (0.90-1.01)
Tampon	1	0.71 (0.62-0.83)*	1.01 (1.00-1.02)*
Developing and developed countries			
Developing	19(23)	0.85 (0.79-0.91)*	.97 (0.95-0.99)*
Developed	14(17)	0.94 (0.90-0.97)*	.99 (0.93-1.06)
Type of reference test			
One colpo-targeted biopsy accepting - colpo as –for CIN2+	17 (20)	0.88 (0.82-0.95)*	0.98 (0.93-1.03)
Multiple colpo-targeted biopsies accepting - colpo as – for CIN2+	6	0.84 (0.74-0.95)*	0.99 (0.95-1.03)
Colpo-targeted + one random biopsy	4	0.89 (0.83-0.97)*	0.97 (0.95-1.00)*
Colpo-targeted + random biopsies/EC	5 (8)	0.86 (0.77-0.97)*	0.98 (0.97-1.00)*
Biopsy + follow-up	2 (3)	0.93 (0.80-1.07)	0.99 (0.86-1.13)

Co-variate	Nb. of studies (test/device combinations)	Relative sensitivity	Relative Specificity
cytology if no biopsy			
Partial or complete verification			
All verified	22(25)	0.90 (0.86-0.95)*	0.97 (0.93-1.02)
Test+ & random fraction test- verified	4(5)	0.92 (0.87-0.97)*	0.97 (0.93-1.01)
Only test+ for or more tests verified, assumption absence CIN2+ if -all tests are negative	8 (10)	0.79 (0.69-0.91)*	0.98 (0.96-0.99)*
Involvement of manufacturers of HPV assays or sampling devices			
No involvement	10 (11)	0.85 (0.75-0.97)	0.99 (0.98-1.00)
Material support	6 (8)	0.81 (0.71-0.91)	0.99 (0.96-1.03)
Financial support, (co-)authorship	14 (19)	0.88 (0.84-0.92)	0.97 (0.96-0.99)
Support not reported	3 (4)	0.96 (0.93-0.99)	0.96 (0.93-0.99)

*statistically significantly different from unity.

EC: endo-cervical curettage; test-: test negative; test+: test positive; - colposcopy: negative colposcopy; RLU: relative light units.

7.10. Small study effects

7.10.1. Absolute accuracy

Supplementary Table 7. Small study effects in the absolute accuracy of hrHPV testing in self-samples and in the accuracy of hrHPV testing and cytology testing on clinician samples. p values are assessed from a regression of the effective study size against the logarithm of the diagnostic odds ratio as proposed by Deeks*et al.* (J Clin Epidemiol 2005;58: 882-93)

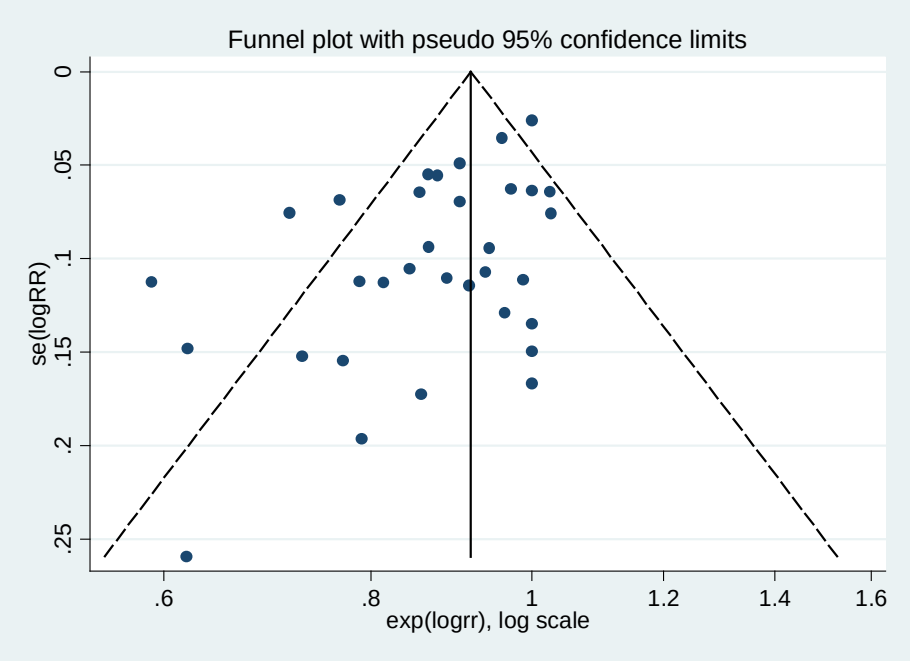
Collection type	Test	Outcome	p value
Self-samples	hrHPV testing	CIN2+	0.990
		CIN3+	0.759
Clinician-samples	hrHPV testing	CIN2+	0.621
		CIN3+	0.560
	Cytology at ASC-US+	CIN2+	0.127
		CIN3+	0.120
	Cytology at LSIL+	CIN2+	0.109
		CIN3+	0.141

The effective size funnel plots can be obtained from the first author upon request.

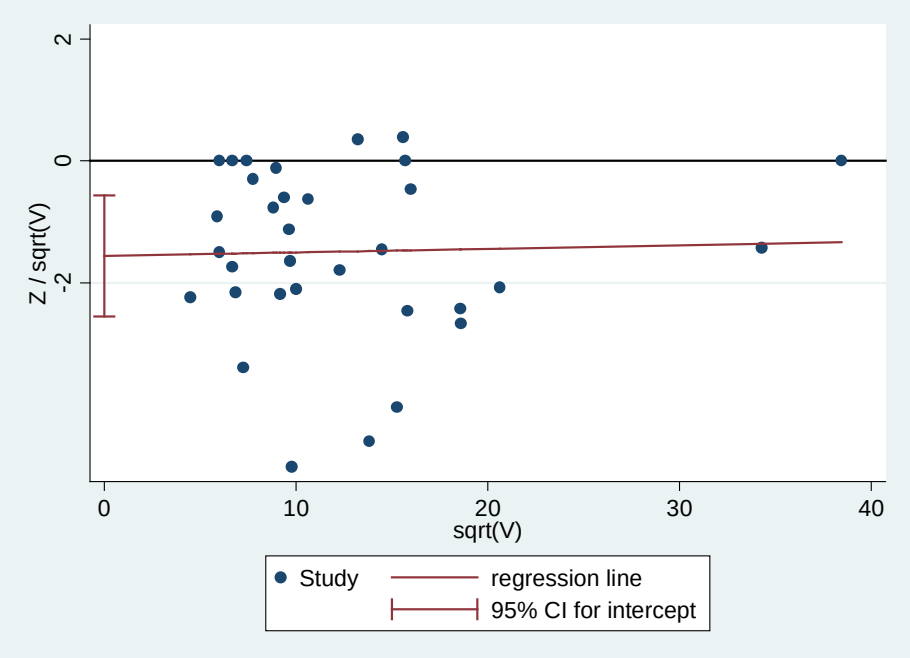
7.10.2. Relative accuracy

Supplementary Table 7. Small study effects in the relative accuracy of hrHPV testing on self-samples vs. comparator tests in clinician samples. P values in the relative sensitivity and specificity are assessed as proposed by Harbord*et al* (Stat Med 2006; 20: 641–54). The intercept and the slope of the regression of $Z/\sqrt{V}$ against $\sqrt{V}$ are shown when the p-values are significant (p<0.05).

Comparator test	Outcome	p (relative Sensitivity)	Intercept (bias)	Slope	p (relative Specificity)
hrHPV testing	CIN2+	0.003	-1.56 (-2.56 to -0.56)	0.01 (-0.06 to 0.07)	0.541
	CIN3+	0.006	-1.61 (-2.67 to -0.55)	0.05 (-0.03 to 0.13)	0.741
Cytology at ASC-US+	CIN2+	0.232	-	-	0.107
	CIN3+	0.004	-2.12 (-3.11 to -1.13)	0.13 (0.06 to 0.20)	0.733
Cytology at LSIL+	CIN2+	0.930	-	-	0.491
	CIN3+	0.975	-	-	0.327



Supplementary Figure 2. Funnel plot of the relative sensitivity for CIN2+ of hrHPV testing on self-samples versus clinician samples. Effect size (X axis: log of the relative sensitivity, on an exponentiated scale); study size measure: standard error of the relative sensitivity. Some asymmetry can be discerned, with more small studies (at the bottom) showing a low relative sensitivity.



Supplementary Figure 3. Small study effect inthe relative sensitivity for CIN2+ of hrHPV testing on self- versus clinician-samples. Harbord's plot based on the regression of $Z/\sqrt{V}$ against $\sqrt{V}$. *Z* is the efficient score and *V* is the variance of *Z* under the null hypothesis (Harbord*et al.*, Stat Med 2006; 20: 641-54). The intercept is statistically significant from zero, whereas the slope is not significantly different from a horizontal line.

Funnel plots and Harbord's plots for other test comparisons can be obtained from the first author upon request.

7.11. Risk of cervical pre-cancer after a positive or negative screening test

Sensitivity and specificity are test characteristics reflecting the capacity to identify diseased subjects by a positive test result and non-diseased subjects by a negative test result. These are test characteristics which are typically not influenced by disease prevalence. Therefore, in systematic reviews and meta-analyses, sensitivity and specificity are the test measures that are pooled to synthesize knowledge on test performance.

However, patients clinicians, and decision makers defining policies for good clinical practice, are in the first place interested in the probability of disease when a test is positive (positive predictive value: PPV) and the risk of disease when a test is negative (complement of the negative predictive value: 1-NPV=cNPN). The PPV provides information on the risk of underlying pre-cancer and consequently on the efficiency of referral for further management. The inverse of the PPV (1/PPV) corresponds with the number needed to refer [colposcopy/biopsy] to find 1 case of cervical pre-cancer. The NPV provides assurance on the safety that a women does not have (pre)-cancer and will have a very low risk to develop (pre-) by the next screening round.

Below, we computed for a plausible series of background prevalence of CIN3+ (possible pretest probabilities) which are relevant for the settings where the evaluated tests will possibly be used. The predictive values, computed for a given setting/area, allow decision making regarding the use of a test in this setting/area. The risk of underlying pre-cancer or cancer (CIN3+) should be sufficiently low to reassure women and to refer them back to the normal screening schedule. Whereas the risk of CIN3+ should be sufficiently high if the screening test is positive (=PPV). If the PPV is not high enough a triage test is needed.
We considered the following range of background risk of cervical pre-cancer or cancer (prevalence of CIN3+):

- Low: 0.25%
- Intermediate: 0.50%
- High: 2%.

We accepted the following cutoffs for the measures of efficiency (PPV) and safety (cNPV), considering prevalent CIN3+ as targeted prevalent disease:

- PPV: >10%
- cNPV: <1%.

In addition, the following cut-offs for longitudinal PPV and cNPV over a period of five years after the screening test were accepted.

- PPV_{long}: >20%
- $cNPV_{long}$: <1%.

Longitudinal predictive values were estimated using ratios PV_{long} / PV_{prev} derived from a cancer screening registry involving combined cytology and HPV screening in the USA Katki et al 2011, 2013 (1;2).

Supplementary Table 8. Number of true- and false positives and negatives among 10,000 women attending cervical cancer screening, post-test probabilities of CIN3+ in case of a positive (PPV) or negative (1-NPV) at screening or 5 years after screening using 5 screening tests: HC2 and MALDITOF on self-samples and hrHPV DNA or cytology (at ASC-US+ or LSIL+) on clinician samples, applied in 3 situations of low, medium and high background risk (prevalence=pretest probabilities of CIN3+ = 0.25%, 0.50% or 2%).

				At enrollment							5 years after enrollment			
Test		**Sensitivity/ Specificity***	**Pre-valence**		**Useful referrals**	**Missed cases**	**Un-necessary referrals**	**True reassurance**	**Cross-sectional**		**Longitudinal**			
(sample)				**%test +ve**	**TP**	**FN**	**FP**	**TN**	**PPV**	**1-NPV**	**PPV**	**1-NPV**	**factor if test+ve†**	**factor if test-ve†**
HC2 (self)	SE	72%	0.25%	14.1%	18	7	1397	8579	1.3%	0.08%	4.6%	0.24%	3.62	3.00
	SP	86%	0.50%	14.3%	36	14	1393	8557	2.5%	0.16%	9.1%	0.49%	3.62	3.00
			2.00%	15.2%	144	56	1372	8428	9.5%	0.66%	34.4%	1.98%	3.62	3.00
MALDI-TOF (self)	SE	93%	0.25%	12.2%	23	2	1197	8778	1.9%	0.02%	6.8%	0.07%	3.62	3.00
	SP	88%	0.50%	12.4%	47	3	1194	8756	3.8%	0.03%	13.7%	0.10%	3.62	3.00
			2.00%	13.6%	186	14	1176	8624	13.7%	0.16%	49.4%	0.49%	3.62	3.00
hrHPV (clin)	SE	95%	0.25%	11.2%	24	1	1097	8878	2.1%	0.01%	7.7%	0.03%	3.62	3.00
	SP	89%	0.50%	11.4%	48	2	1095	8856	4.2%	0.02%	15.2%	0.07%	3.62	3.00
			2.00%	12.7	190	10	1078	8722	15.0	0.1	54.	0.3	3.62	3.00

				At enrollment							5 years after enrollment			
Test		**Sensitivity/ Specificity***	**Pre-valence**		**Useful referrals**	**Missed cases**	**Un-necessary referrals**	**True reassurance**	**Cross-sectional**		**Longitudinal**			
				%					%	1%	2%	4%		
Cyto ASC-US+ (clin)	SE	91%	0.25%	11.2 %	23	2	1097	8878	2.1%	0.02%	3.4 %	0.11%	1.65	5.00
	SP	89%	0.50%	11.4 %	46	4	1095	8856	4.0%	0.05%	6.7 %	0.23%	1.65	5.00
			2.00%	12.6 %	182	18	1078	8722	14.4 %	0.21%	23.8%	1.03%	1.65	5.00
Cyto clin LSIL+ (clin)	SE	78%	0.25%	3.2%	20	5	299	9676	6.3%	0.05%	9.7 %	0.15%	1.55	3.00
	SP	97%	0.50%	3.4%	39	11	299	9652	11.6 %	0.11%	17.9%	0.34%	1.55	3.00
			2.00%	4.5%	156	44	294	9506	34.7 %	0.46%	53.7%	1.38%	1.55	3.00

* derived from the meta-analysis; † derived from Katki et al 2011, 2013 (1;2)

TP: number true positives, FN: number of false-negatives; FP: number of false-positives; TN: number of true negatives; PPV: positive predictive value; NPV: negative predictive value; HC2: Hybrid Capture-2 assay; hrHPV: assay detecting high-risk types of human papillomavirus; ASC-US: atypical squamous cells of unspecified significance; LSIL: low-grade squamous intraepithelial lesion.

Red: PPV<10% or PPV_{long}<20% or 1-NPV≥1%. Green: PPV≥10% or PPV_{long}≥20% or 1-NPV>1%.

7.12. Grade Profil

Grading of Recommendations Assessment, Development and Evaluation (GRADE)

GRADE has built on previous systems to create a highly structured, transparent, and informative system for rating quality of evidence(Guyatt *et al.,* 2008b).

Steps in evidence assessment for making guidelines
1) Formulate a question
2) Identify the PICO(S) components
3) Qualify outcomes as critical, important, not important

1) Questions
Is HPV testing on a self-sample as accurate as HPV testing or cytology testing on a clinician sample to detect high-grade cervical intraepithelial neoplasia.

Other question
Does offering a self-sample for HPV testing result in higher participation rates than sending a reminder for conventional screening in women a) non-participating in the regular screening programme b) the general population. See other systematic review: Racey Can J Pub Health 2013.

See further below.

2) PICOS
P: Women attending cervical cancer screening. Women being tested for cervical cancer precursors (high-risk group) or under follow-up because of previously found cervical lesions will be included also, but will be considered as less relevant for answering the study question.
I: hrHPV DNA or RNA testing on a self-sample (index test).
C1: hrHPV DNA or RNA testing on a clinician-sample (comparator test 1).
C2: cytological interpretation on a clinician sample (comparator test 2).
O: absolute sensitivity and specificity of index- and comparator tests; relative sensitivity and specificity of index versus comparator tests to detect CIN2+ and CIN3+.
S: diagnostic test accuracy studies (with complete verification with a reference standard); screening studies with different screening tests involving at least complete verification of women with one or more positive

screening test results; randomised trials with different screening tests in separate study arms (these studies will only include relativity sensitivity) .

3) Importance of outcomes

Outcome:

25. Reduction of mortality from cervical cancer, (quality-adjusted) life-years gained.
26. Reduction of morbidity due to cervical cancer: incidence of cancer (Ib+).
27. Reduction of incidence of cancer (including micro-invasive cancer).
28. Reduction of incidence of CIN3 or worse disease (CIN3+).
29. Increased detection rate of CIN3+ or CIN2+.
30. Increased test positivity with increased, similar or hardly reduced positive predictive value.

Quality of evidence for each outcome in four categories;

- High: ++++
- Moderate:+++
- Low:++
- Very low: +

The included studies address diagnostic accuracy derived from observational studies. Only 2 randomised studies were retrieved, but these trials only described relative detection rates of CIN2/3+ (equivalent to relative sensitivity) and PPVs. None described reduction of disease. Given the observational character, we must attribute a priori the category low evidence.

Given the direct link to accuracy of cervical cytology and HPV testing on clinician-samples, we may assume that treating cervical cancer precursors detected through screening by HPV testing on self samples will result in reduced incidence of invasive cervical cancer, which has been thoroughly documented through cohort studies, registry linkage studies, case control studies and randomised trials.

5 factors that lower the quality of the evidence (Guyatt *et al.*, 2008a)

21. Risk of bias due to limitations in design or execution: see CONSORT, STROBE, QUADAS
22. Inconsistency or heterogeneity: if consistency unexplained, lower quality
23. Indirectness, applicability (relevance of studies for answering the PICPO question)
24. Imprecision: number of studies, width of CI
25. Reporting bias, publication bias.

3 factors that increase the quality

10. Large effect
11. Dose effect gradient
12. Confounders (presence of confounders that have lowered the observed effect, absence of these would have increased the effect) (Guyatt *et al.,* 2008a)

Items downgrading quality of evidence		Downgrading
Bias, design	The QUADAS assessment generally provided a moderately good scoring of the majority of studies. The QUADAS issues did not influence study outcomes significantly.	No (-0)
Inconsistency	No major inconsistency was observed in the comparisons of HPV testing on self- versus on clinician samples.. Influential studies with outlying results were identified, in particular in the comparison of HPV testing on self-samples with cytology on clinician-samples were observed resulting in lower relative sensitivity. Omission yielded relative sensitivity not significantly different from unity. Important test effects were noted which nuanced conclusions.	No (-0)
Indirectness	The comparison of index with comparator tests among women under follow-up for previous cervical abnormalities could be considered as not relevant for screening. However, given the similaritity in the pooled relative accuracy measures across settings, we can accept the studies conducted in colposcopy clinics.	No (-0)
Imprecision	Confidence intervals were quite narrow given the number of studies (N=36) and cumulative number of enrolled women (~155,000).	No (-0)
Publication bias, other	Some evidence of publication bias was noted. However, the direction was surprising. In the sense that more small studies with lower sensitivity of HPV testing were retrieved. So, there is no evidence that small study effects may have improved pooled relative accuracy estimates spuriously.	No (-0)
Items downgrading quality of evidence		
Large effect	The clear lower sensitivity and specificity of HPV testing with HC2 on self- versus clinician samples is notable.	Yes (+1)
Dose-effect correlation	The change in accuracy by test threshold was rarely documented though-out retrieved studies.	No (+0)

Confounding factors neutralising effects	The possible impact of small study effects may have had an unfavourable effect on the sensitivity of HPV testing on self-samples. However, probably the impact of publication bias is small. Age could not be assessed systematically throughout studies by lack of age-specific data or lack of commonly presented age-categories impeding inclusion of age as a covariate.	No (+0)˚

Conclusion: evidence of moderate quality.

Overall Grade of the quality of evidence is assigned and this is based on the outcome with the lowest quality of evidence given that it is a critical outcome.

For the accuracy we consider the relative sensitivity (outcome CIN3+) and specificity (outcome CIN2+) as critical. The other outcomes: absolute accuracy, relative sensitivity for CIN2+ and relative specificity for CIN3+ are considered as important.

GRADE evidence profile f

	Quality of evidence								
# studies (N)	Absence of study limitations	Consistency	Directness (outcome, representativity Germany)	Precision	Absence publication bias	Large effect	Dose-effect relation	Bias lowering effect	Comment
Outcome 1: Relative sensitivity for CIN3+ of HPV testing on self-samples versus clinician samples [CRITICAL]									
12	Yes	Yes	No, Yes	Yes	Yes	No	No	No	Low-moderate

Outcome 2: Relative specificity for CIN2+ of HPV testing on self-samples versus clinician samples [CRITICAL]									
34	Yes	Yes	No, Yes	Yes	Yes	No	No	No	Low-moderate

Summary of findings

Outcome	Absolute accuracy in control group	Absolute accuracy in intervention group	Relative accuracy (intervention/control)	Absolute difference (RD: intervention-control)
1. sensitivity for CIN3+: HPV on self vs clin-samples	95% (91-97%)	84% (72-92%)	0.89˚(0.83-0.96)	11% (-%)
2. specificity for CIN2+: HPV on self vs clin-samples	88% (87-94%)	86% (83-89%)	0.96 (0.95-0.97)	2% (-)

To add:

Risk among screen test negatives, nb of TP,FN,FP and TN among 100,000 screened women screened with HPV on self- versus a clinician sample

PARTICIPATION IN SCREENING
GRADE evidence profile for interventions

	Quality of evidence								
# studies (N)	Absence of study limitations	Consistency	Directness (outcome, representativity Germany)	Precision	Absence publication bias	Large effect	Dose-effect relation	Bias lowering effect	Comment
Outcome 1: Difference participation in screening among women receiving self-sampling kit vs women receiving a conventional reminder [CRITICAL]									
9	Yes	Yes	No, No	Yes	Possible bias cannot be excluded	Yes	No	No	-

Outcome 2: Difference in compliance to follow-up among women who received a self-sampling kit vs. women who received a conventional reminder [CRITICAL]									
3	Yes	No	No, No	No	Possible bias cannot be excluded	Yes	No	No	-

Summary of findings

Outcome	Absolute effect in control group	Absolute effect in intervention group	Relative effect (RR: intervention/control)	Absolute effect (RD: intervention/control)	Quality of evidence
Outcome 1: Participation	11% (5-17%)	25% (20-30%)	2.71 (1.45-5.08)	14% (9-19%)	Moderate-high
Outcome 2 Compliance follow-up	96%	86% (77%-94%)	0.77 (0.50-1.20)	-10% (-38-18%)	Low

Formulation of recommendation: Pro/against, strong (we recommend)/weak (we suggest = conditional on). In concertation between guideline and systematic review group.
(Guyatt *et al.*, 2008a)

Factors that influence the strength of recommendation:

- Quality of evidence: by outcome and across outcomes

- Balance benefits/harms
- Values and preferences
- Resource use, costs

References

Guyatt G.H., Oxman A.D., Kunz R., Falck-Ytter Y., Vist G.E., Liberati A., & Schunemann H.J. (2008a) Going from evidence to recommendations. *BMJ* **336:** 1049-1051.

Guyatt G.H., Oxman A.D., Kunz R., Vist G.E., Falck-Ytter Y., & Schunemann H.J. (2008b) What is "quality of evidence" and why is it important to clinicians? *BMJ* **336:** 995-998.

Guyatt G.H., Oxman A.D., Vist G.E., Kunz R., Falck-Ytter Y., Alonso-Coello P., & Schunemann H.J. (2008c) GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *BMJ* **336:** 924-926.

Schunemann H.J., Oxman A.D., Brozek J., Glasziou P., Jaeschke R., Vist G.E., Williams J.W., Jr., Kunz R., Craig J., Montori V.M., Bossuyt P., & Guyatt G.H. (2008) GRADE: Grading quality of evidence and strength of recommendations for diagnostic tests and strategies. *BMJ* **336:** 1106-1110.

8. Question: Follow-up after treatment of CIN: Is testing with a biomarker more accurate than follow-up using hrHPV-DNA testing or cytology to predict failure or success after treatment of cervical precancer?

8.1. Background and rationale

The presence of cervical intraepithelial neoplasia of grade two and three (CIN2-3) is linked with a risk of developing invasive carcinoma if not treated[1]. Conservative treatment methods for high-grade CIN, which can be performed under local anaesthesia in an outpatient setting, are performed to prevent progression to cervical cancer[2]. Such methods can be excisional or ablative[3]. The former type includes laser electrosurgical excision (LLETZ), and laser conisation. The latter type includes cryotherapy, laser ablation, and electrocoagulation.

However, success of treatment is suboptimal and residual or recurrent high grade disease (CIN2+) can be detected in on average 8% (ranging from 4% to 18%) of treated women[4], with the majority of treatment failure occurring in the first two post-operative years[5-7]. Moreover, women with a history of treatment for cervical precancer are still at higher risk of developing invasive cervical cancer in the future compared to the general population during at least 10 years and maybe up to 20 years after treatment[8;9]. A recent Swedish trend study also confirmed that women with prior treatment for CIN3+ show an increased incidence of and even mortality from cervical or vaginal cancer in particular in older age groups (aged 60 or older) and among those treated in more recent periods[10;11]. Therefore, finding an indicator to predict this risk with great accuracy would be particularly helpful.

Guidelines for surveillance strategies after treatment of CIN vary greatly among countries with respect to timelines, type of tests performed, and length of follow-up. In this report, a meta-analysis is updated on the occurrence of treatment failure and on the accuracy of testing with an hrHPV-DNA test versus cytology to detect residual/recurrent CIN2+[4;12].

8.2. Question

8.2.1. PICOS

P: Women treated for histologically confirmed CIN2+ or CIN3+ (or CIN1+) by an excisional (LEEP, LLETZ, laser conisation) or ablative (cryotherapy, laser ablation, electrocoagulation) procedure.

I1: Assessment of margin status at time of treatment.

I2: Testing with a biomarker (mRNA,p16, HPV type-specific persistence, or other) three to nine months post-treatment.

C: HPV-DNA testing or Cytology, three to nine months post-treatment.

O: Accuracy to detect residual/recurrent CIN2+ or CIN3+ (or CIN1+), confirmed by histology.

Prospective studies: - accuracy (sensitivity, specificity, positivie predictive value [PPV], negative predictive value [NPV], and complement to NPV [cNPV=1-NPV]),
- prediction values

Case control studies: - accuracy (sensitivity, specificity, positivie predictive value [PPV], negative predictive value [NPV], and complement to NPV [cNPV=1-NPV])

S: Prospective studies: - all participants tested with a biomarker and a HPV test and/or cytology
- 18 months follow-up
- verification with a golden reference standard: colposcopy completed with targeted biopsy for all participants, or in case of a positive test result (biomarker, HPV-test or cytology)

Case control studies: - all participants tested with a biomarker and a HPV test and/or cytology
- 18 months follow-up

8.3. Methods

Is testing with a biomarker more accurate than follow-up using hrHPV DNAtesting or cytology to predict failure or success after treatment of cervical precancer?

8.3.1. Search strategy

The search-string, shown in Box1 was used to identify relevant studies.

((cancer OR carcinoma OR dysplas* OR neoplas* OR CIN OR SIL) AND (cervix OR cervical) OR vaginal smears[MeSH] OR Cervix Neoplasm [Mesh]))
AND
(treatment OR conisation OR conization OR leep OR LEEP OR LETZ OR cryotherapy)
AND
(HPV OR human papillomavirus OR papillomavirus OR mRNA OR p16)

8.3.2. Eligibility of studies

Studies were eligible if (1) women were treated for histologically-confirmed CIN, (2) an assessment of margin status was performed at time of treatment or women were tested with another biomarker (hrHPV genotyping, HPV type-specific persistence, mRNA, p16) between 3-9 months post-treatment, (3) women had a hrHPV-DNA test and/or cytology between at 3-9 months post-treatment, (4) women were followed-up for at least 18 months, and (5) colposcopy and targeted biopsy was performed on all women or women with at least one positive test result. Treatment failure was defined as follow-up histological diagnosis of CIN2+ or CIN3+ (or CIN1+).

8.3.3. Outcome measures

The following outcome measures were assessed:
- treatment failure rate (prevalence of residual/recurrent CIN2+)
- absolute accuracy of margin status to detect CIN2+ or CIN3+
- relative accuracy of margin status versus HPV-DNA testing or cytology to detect CIN2+ or CIN3+
- absolute accuracy of testing with a biomarker to detect CIN2+ or CIN3+
- relative accuracy of testing with a biomarker versus HPV-DNA testing or cytology to detect CIN2+ or CIN3+

8.3.4. Statistical Analysis

The pooled absolute sensitivity and specificity of the tests were estimated jointly using *metandi*, a procedure in STATA, based on the bivariate normal model for the logit transforms of sensitivity and specificity taking the intrinsic correlation between true and false-positivity rates and the variability between studies into account [13;14]. The relative sensitivity and specificity of hrHPV-DNA testing compared with cytology were computed using *metadas*, a SAS macro for meta-analysis of diagnostic accuracy studies which allows the inclusion of test as a covariate making comparison of tests possible[15;16].
Forrest plots were also produced showing study-specific and pooled absolute and relative accuracy estimates using a random effects model[17]. In these forest plots, the statistical heterogeneity was assessed using Cochran's Q

test and the I^2 statistic, which measures the proportion of variation that is due to inter-study heterogeneity. Statistical analysis was performed using STATA/SE 10 (Stata Corporation, College Station, TX, USA) and SAS 9.3(SAS Institute, Campus Drive Cary, NC, USA).

8.4. Results and interpretation

8.4.1. Margin status to predict residual/recurrent disease

8.4.1.1. Literature retrieval

Our systematic literature search resulted in 30 studies containing accuracy data on the status of the resection margins. From this list, nine studies[18-26] could be pooled in a meta-analysis because they had similar characteristics. The PRISMA flow chart for literature retrieval is shown in Figure 8. Firstly, the disease treated and the disease outcome was histologically identified CIN2+. Secondly, a comparator HPV-DNA test and/or cytology was performed three to nine months post-treatment. Thirdly, the follow-up time was 18 months or more. One study had a case-control design[18], while the other eight were prospective studies[19-26]. A summary of study characteristics is listed in Table 55 and Table 56.

8.4.1.2. Absolute accuracy of margin status

The pooled positivity rate of margin involvement was 24.7% (95% CI: 18.3 -31.1%) among a total of 1912 women treated for CIN2+ (Figure 9). The pooled estimate of occurrence of residual/recurrent CIN2+ was 8.2% (95% CI: 6.2-10.4%) (Figure 9).

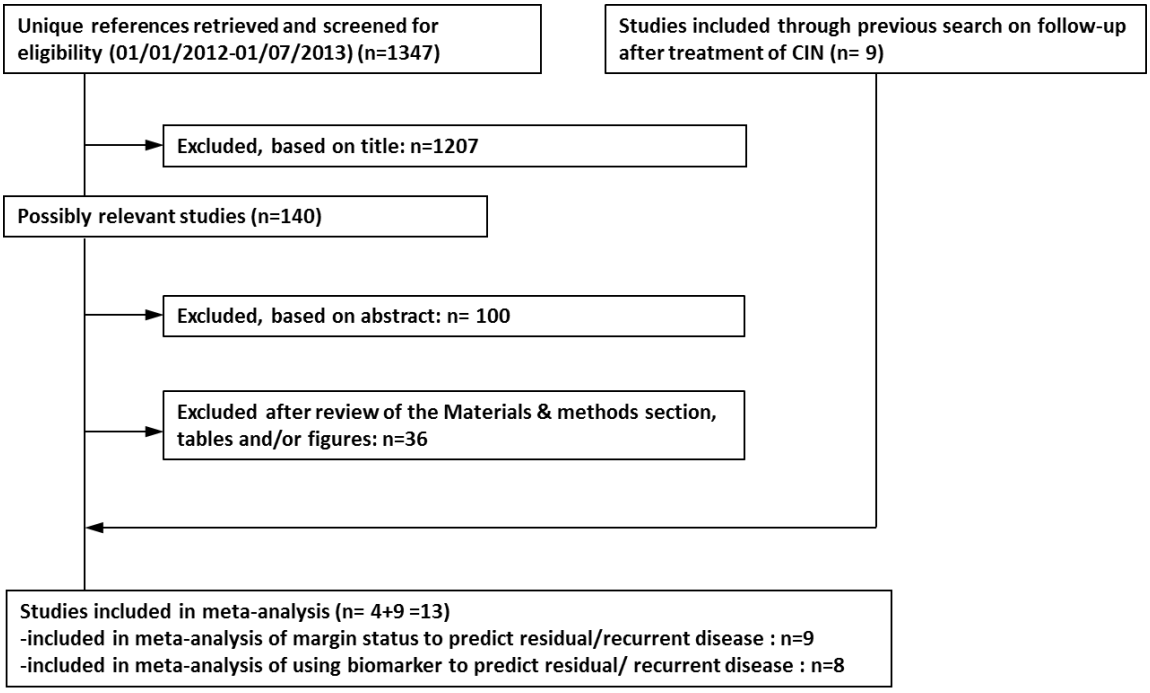


Figure 8: PRISMA flow chart for the retrieval of studies

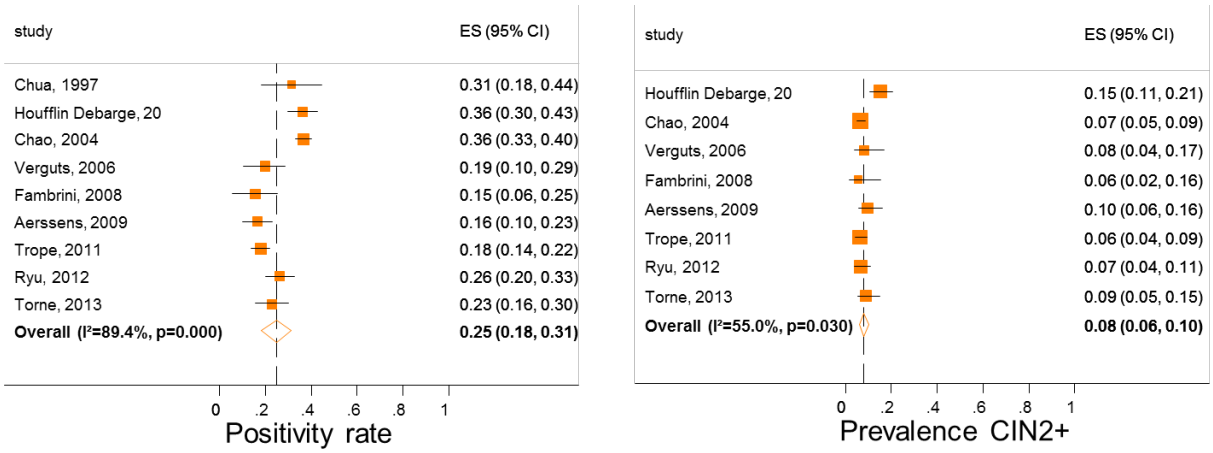


Figure 9: Positivity rate of margin status and occurence of residual/recurrent CIN2+.

The pooled absolute sensitivity and specificity of margin status to predict residual/recurrent CIN2+ was 58.6% (95% CI: 48.4-68.1%) and 80.4% (95% CI: 74.7-85.1%), respectively (Figure 10).

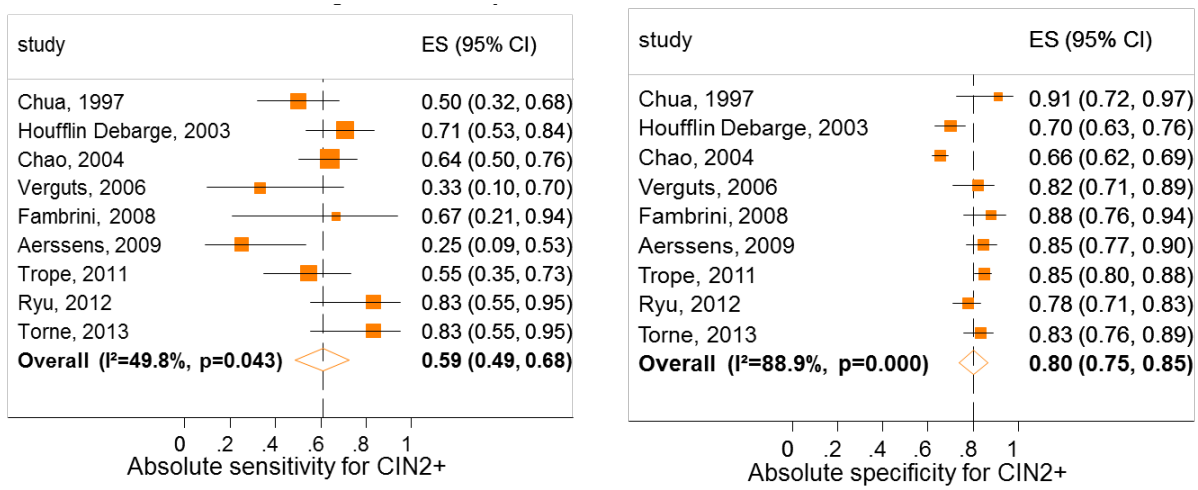


Figure 10: Absolute accuracy of margin status to detect residual/recurrent CIN2+.

8.4.1.3. Relative accuracy of margin status versus HPV-DNA testing

All studies allowed comparison of the accuracy of margin status at time of treatment with HPV-DNA testing 3-9 months post-treatment. The Hybrid Capture 2 assay was used in five studies[19-21;25;26], while four other studies used PCR-based techniques[18;22-24]. The sensitivity of margin status was 27% lower than that of HPV-DNA testing, resulting in a relative sensitivity of 0.73 (95% CI: 0.63-0.86) (Figure 11, Figure 12 left pane). The specificities of margin status and HPV-DNA testing were similar, with a specificity ratio of 1.04 (95% CI: 0.97-1.11) (Figure 11, Figure 12 right pane).

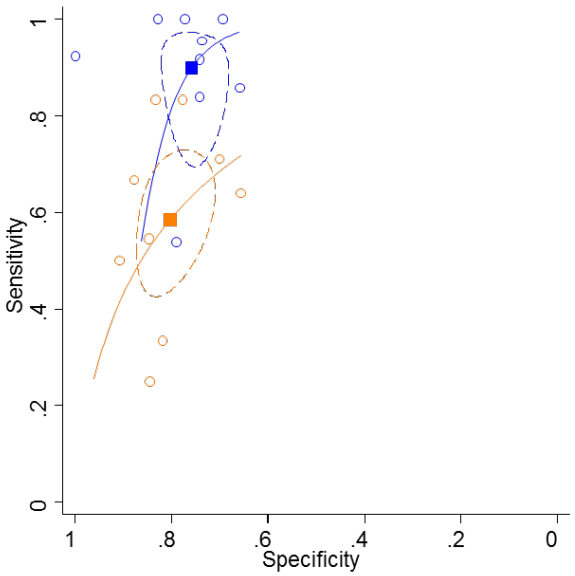


Figure 11: Pooled sensitivity and specificity of margin status (orange) versus HPV-DNA testing (blue) to predict residual/recurrent CIN2+. Pooled estimate (square), study estimates (circles), ROC-curve (full line), 95% confidence ellipse (dashed line).

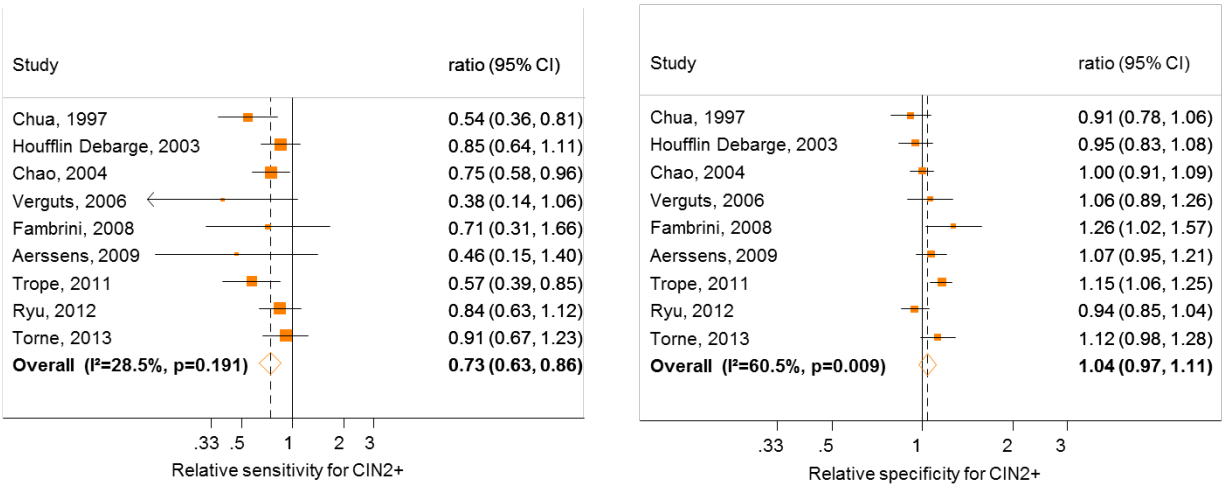


Figure 12: Sensitivity ratio(left pane) and specificity ratio (right pane) of margin status versus HPV-DNA testing to detect residual/recurrent CIN2+.

8.4.1.4. Relative accuracy of margin status versus cytology

Eight studies allowed comparison of the accuracy of margin status versus that of cytology[18;20-26].

Compared to cytology 3-9 months post-treatment, evaluation of the margin status had a lower sensitivity and specificity (Figure 13), albeit not significantly. The pooled sensitivity and specificity ratio were 0.90 (95% CI: 0.75-1.10; Figure 14) and 0.95 (95% CI: 0.85-1.07; Figure 14), respectively.

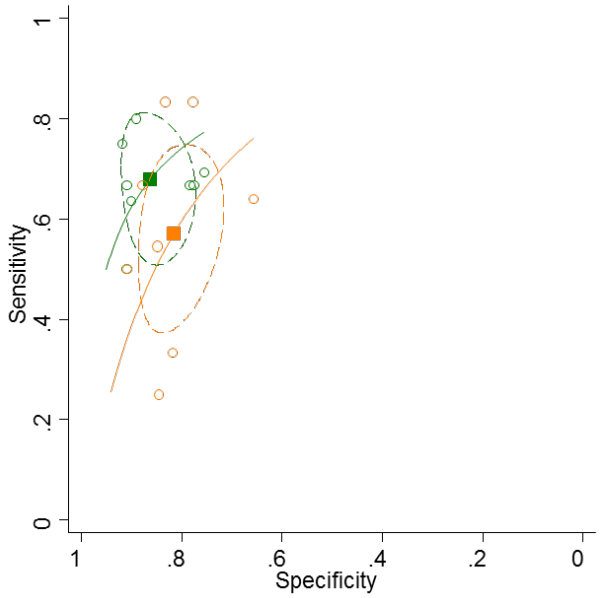


Figure 13: Pooled sensitivity and specificity of margin status (orange) versus cytology (green) to detect residual/recurrent CIN2+. Pooled estimates (square), study estimates (circles), ROC-curve (full line), 95% confidence ellipse (dashed line).

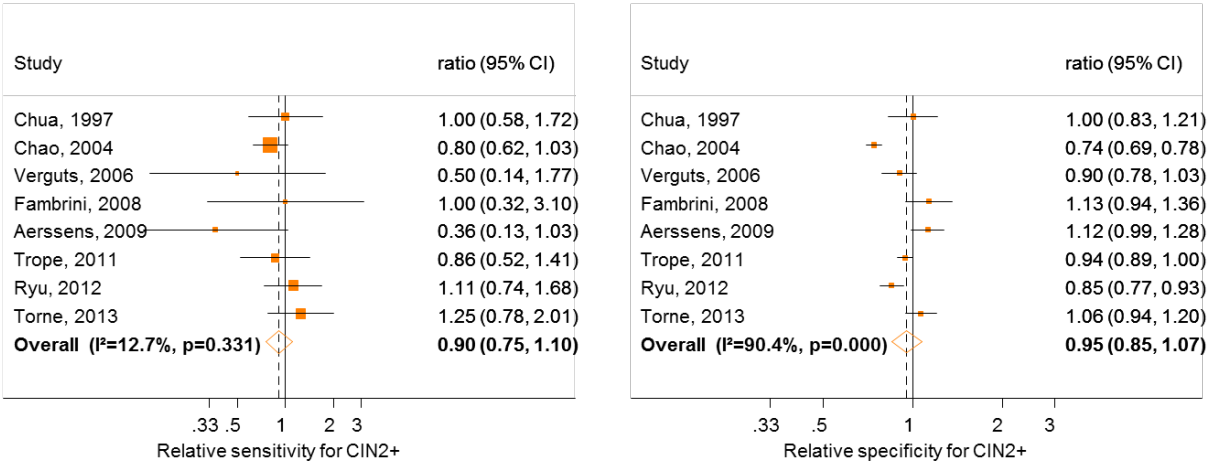


Figure 14: Sensitivity ratio (left pane) and specificity ratio (right pane) of margin status versus cytology to detect residual/recurrent CIN2+.

8.4.1.5. Interpretation

The importance of margin involvement to assess the risk of post-treatment disease is controversial, especially since a direct link between the size of the excisional specimen and obstetrical outcomes has been shown[27;28]. In order to minimize immediate and future complications for treated women, complete excision is not always regarded as a treatment objective[29]. Furthermore, the destruction of the excision site by the treatment procedure, has been documented to hamper interpretation of the margin status[23]. Consequently, local guidelines and procedures are likely to influence the margin status and the interpretation thereof. This is demonstrated by the

significant heterogeneity of the margin positivity rate in our meta-analysis (I^2=89%, p=0.000, Figure 9), which could explain the poor correlation between positive section margins and residual/recurrent disease.
Of the nine included studies, merely two documented the involvement of endo- and exocervical margins separately [19;25], however accuracy data for the detection of residual/recurrent disease were not available. One study providing such data with at least 18 months follow-up was identified, but was not included in our analysis because the grade of CIN was not specified in the disease outcome[30]. In this study by Leguevaque et al., the sensitivity to detect residual/recurrent CIN1+ for endo- and exocervical margins was 34.9% and 7.0%, respectively. Specificities of 89.3% and 93.5% were observed, respectively.
Overall, pooling of the accuracy data of nine, demonstrated a poor sensitivity (60%) for the involvement of section margins and CIN2+ disease. This implied a sensitivity loss of 27% compared to follow-up hrHPV-DNA testing (relative sensitivity: 0.73). The specificities for margin status and HPV-DNA testing were similar (relative specificity: 1.04). Similar accuracy estimates were observed for margin status and follow-up cytology (relative sensitivity: 0.90; relative specificity: 0.95).

8.4.2. Using other biomarkers to predict residual/recurrent disease

8.4.2.1. Literature retrieval

A systematic literature search, using the predetermined inclusion criteria, yielded only a limited number of studies that documented accuracy data for the use of a biomarker to predict post-treatment disease and had a sufficient follow-up interval (at least 18 months). Therefore inclusion criteria were extended. As a result, three studies[24;31;32] were found assessing the accuracy of mRNA testing, and five studies[23;33-36] using HPV type-specific persistence as a biomarker for residual/recurrent disease (See Figure 8 for the process of literature retrieval). All eight studies allowed comparison with hrHPV-DNA testing and/or cytology. However, the study characteristics of these studies were too diverse to allow meta-analytical pooling. Consequently, in this report, a literature review is performed enlisting and discussing the available evidence. A summary of study characteristics is listed in Table 55 and Table 56

8.4.2.2. Accuracy of mRNA testing

Table 50 and Table 51 respectively list the absolute and relative accuracy for testing with a hrHPV-mRNA assay, compared to testing with an hrHPV-DNA test or cytology.

Merely two studies listed accuracy values for mRNA testing when the treated disease and the disease outcome both were CIN2+[24;32]. Both studies used assays that are based on the detection of E6/E7 mRNA of 5 high-risk (hr) HPV types (HPV 16/18/31/33/45)[37]. Although no firm conclusions can be drawn, one study demonstrated a 50% sensitivity loss (ratio: 0.48 [95% CI=0.30-0.76]) but a 30% gain in specificity (ratio: 1.29 [95% CI= 1.20-1.38]) when using 5-type mRNA testing compared to hrHPV-DNA testing[19]. In the same study, sensitivity was similar (ratio: 0.71 [95% CI=0.41-1.25]) but specificity was significantly increased (ratio: 1.06 [95% CI=1.01-1.10]), compared to cytology. In the study of Tinelli and colleagues, sensitivity of the Nuclisens assay was three times higher than that of cytology (ratio: 3.00 [95% CI= 0.50-17.95]) , due to the remarkably low sensitivity of the latter (25.0% [6.3-80.6] for CIN2+)[37]. One study used the APTIMA assay and demonstrated a significant gain in specificity (ratio: 1.15 [95% CI=1.05-1.27] for CIN2+) compared to hrHPV-DNA testing[31].

8.4.2.3. Accuracy of HPV type-specific persistence

Table 50 and Table 51 respectively list the absolute and relative accuracy for testing for type-specific HPV-persistence, compared to testing with an hrHPV-DNA test or cytology.

Four studies provided accuracy data for HPV genotype testing to detect residual/recurrent CIN2+ after treatment for CIN2+[23;33;35;36]. Compared to testing with a hrHPV-DNA assay, the sensitivity of type-specific persistence was not significantly different (95% CI's containing unity). Specificity of type-specific persistence was significantly higher in two studies[33;35]. Compared to cytology, testing for type-specific persistence was significantly more sensitive and specific in two[35;36] and three studies[23;33;35], respectively. One study included women that were treated for CIN1 or worse, and demonstrate equal accuracy for type-specific persistence and hrHPV-detection (sensitivity ratio: 0.60 [95% CI=0.29-1.23]; specificity ratio: 1.10 [95% CI=1.00-1.22] for CIN2+)[34].

8.4.2.4. Interpretation

Study characteristics of the studies included are heterogeneous. As a result, the data listed here demonstrate substantial variability in accuracy values of biomarkers to predict residual/recurrent disease. This aspect, together with the scarceness of evidence and the diverse characteristics of the current studies demonstrate the pressing need for additional well-designed studies.

Table 50: Absolute sensitivity and specificity for testing with a biomarker, an hrHPV-DNA test, and cytology.

Study	Disease treated	Disease outcome	Test	n	Test time	Follow-up time	Biomarker		HPV		Cytology	
							Sensitivity	Specificity	Sensitivity	Specificity	Sensitivity	Specificity
mRNA												
Trope 2011	CIN2+	CIN2+	Pretect	344	6m	18m	**45.5%** (24.4-67.8)	**95.0%** (92.1-97.1)	**95.5%** (77.2-99.9)	**73.6%** (68.4-78.3)	**63.6%** (40.7-82.8)	**90.1%** (86.3-93.1)
Tinelli 2013	CIN2+	CIN2+	Nuclisens	107	6m	28m*	**75.0%** (19.4-99.4)	**83.5%** (74.9-90.1)	□	□	**25.0%** (6.3-80.6)	**75.7%** (66.3-83.6)
Tinelli 2013†	CIN2+	CIN1+	Nuclisens	107	6m	30m*	**76.5%** (50.1-93.2)	**92.2%** (84.6-96.8)	□	□	**23.5%** (6.8-49.9)	**75.6%** (65.4-84.0)
Persson 2012†	CIN1+	CIN2+	Aptima	143	0m	184m	**57.1%** (18.4-90.1)	**93.4%** (87.8-96.9)	**100%** (59.0-100)	**80.9%** (73.3-87.1)	**85.7%** (42.1-99.6)	**87.5%** (80.7-92.5)
Persson 2012†	CIN1+	CIN1+	Aptima	143	0m	184m	**24.0%** (9.4-45.1)	**94.9%** (89.3-98.1)	**52.0%** (31.3-72.2)	**84.7%** (77.0-90.7)	**72.0%** (50.6-87.9)	**99.1%** (95.4-100)
Type specific persistence												
Kreimer 2006	CIN2+	CIN2+	PCR	485	4.5m	24m	**78.1%** (60.0-90.7)	**79.9%** (75.9-83.5)	**90.6%** (75.0-98.0)	**63.8%** (59.2-68.2)	**78.1%** (60.0-90.7)	**69.1%** (64.6-73.3)
Aerssens 2009	CIN2+	CIN2+	PCR-SPF10	131	6m	32m	**38.5%** (13.9-68.4)	**85.6%** (77.9-91.4)	**53.8%** (25.1-80.8)	**79.0%** (70.8-85.8)	**69.2%** (38.6-90.9)	**75.4%** (66.8-82.8)
Kang 2010†	CIN2+	CIN2+	PCR	672	24m	24m	**100%** (90.5-100)	**97.0%** (95.4-98.2)	**97.3%** (85.8-99.9)	**93.1%** (90.8-94.9)	**70.3%** (53.0-84.1)	**90.7%** (88.2-92.9)
Heymans 2011§	CIN2+	CIN2+	PCR	63	6m	24m	**100%** (83.8-100)	**71.4%** (55.4-84.3)	**100%** (83.9-100)	**57.1%** (41.0-72.3)	**76.2%** (52.8-91.8)	**64.3%** (48.0-78.4)

Brismar 2009†	CIN1+	CIN2+	Linear Array	84	12m	39m*	**60.0%** (14.7-94.7)	**94.9%** (87.5-98.6)	**100%** (47.8-100)	**86.1%** (76.5-92.8)	□	□
Brismar 2009†	CIN1+	CIN1+	Linear Array	84	12m	39m*	**23.8%** (8.2-47.2)	**96.8%** (89.0-99.6)	**52.4%** (29.8-74.3)	**92.1%** (82.4-97.4)	□	□

*mean FUtime (rest: max futime); §case-control study; □ no data available; † studies that do not fulfill the original inclusion criteria. Abbreviations: CIN1+, cervical intraepithelial neoplasia grade one or worse; CIN2+, cervical intraepithelial neoplasia grade two or worse.

Table 51: Relative sensitivity and specificity of testing with a biomarker, compared to hrHPV-DNA testing or cytology (cut-off ASC-US+).

							Comparator			
							hrHPV-DNA testing		**Cytology**	
Study	**Disease treated**	**Disease outcome**	**Test**	**n**	**Test time**	**Follow-up time**	**Sensitivity**	**Specificity**	**Sensitivity**	**Specificity**
mRNA										
Trope 2011	CIN2+	CIN2+	Pretect	344	6m	18m	0.48 (0.30-0.76)	1.29 (1.20-1.38)	0.71 (0.41-1.25)	1.06 (1.01-1.10)
Tinelli 2013	CIN2+	CIN2+	Nuclisens	107	6m	28m*	□	□	3.00 (0.50-17.95)	1.10 (0.96-1.27)
Tinelli 2013†	CIN2+	CIN1+	Nuclisens	107	6m	30m*	□	□	3.25 (1.33-7.97)	1.22 (1.07-1.39)
Persson 2012†	CIN1+	CIN2+	Aptima	143	0m	184m	0.57 (0.30-1.09)	1.15 (1.05-1.27)	0.67 (0.33-1.35)	1.07 (0.99-1.15)
Persson 2012†	CIN1+	CIN1+	Aptima	143	0m	184m	0.46 (0.21-1.02)	1.03 (0.96-1.10)	0.33 (0.16-0.70)	0.96 (0.92-1.00)
Type specific persistence										

Kreimer 2006	CIN2+	CIN2+	PCR	485	4.5 m	24m	0.86	(0.70-1.07)	1.25	(1.15-1.36)	1.00	(0.77-1.30)	1.16	(1.07-1.25)
Aerssens 2009	CIN2+	CIN2+	PCR-SPF10	131	6m	32m	0.71	(0.30-1.67)	1.08	(0.96-1.22)	0.56	(0.26-1.21)	1.14	(1.00-1.29)
Kang 2010†	CIN2+	CIN2+	PCR	672	24 m	24m	1.03	(0.97-1.08)	1.04	(1.02-1.07)	1.42	(1.15-1.75)	1.07	(1.04-1.10)
Heymans 2011§	CIN2+	CIN2+	PCR	63	6m	24m	1.00	(1.00-1.00)	1.25	(0.90-1.73)	1.31	(1.03-1.67)	1.11	(0.83-1.49)
Brismar 2009†	CIN1+	CIN2+	Linear Array	84	12 m	39m*	0.60	(0.29-1.23)	1.10	(1.00-1.22)	□		□	
Brismar 2009†	CIN1+	CIN1+	Linear Array	84	12 m	39m*	0.45	(0.19-1.08)	1.05	(0.97-1.15)	□		□	

*mean FUtime (rest: max futime); §case-control study; □ no data available; † studies that do not fulfill the original inclusion criteria. Abbreviations: CIN1+, cervical intraepithelial neoplasia grade one or worse; CIN2+, cervical intraepithelial neoplasia grade two or worse.

8.5. References

(1) Moscicki AB, Schiffman M, Burchell A, Albero G, Giuliano AR, Goodman MT et al. Updating the natural history of human papillomavirus and anogenital cancers. *Vaccine* 2012; 30 Suppl 5:F24-F33.

(2) Paraskevaidis E, Arbyn M, Diakomanolis E, Martin-Hirsch P, Koliopoulos G, Makrydimas G et al. The role of HPV DNA testing in the follow-up period after treatment for CIN: a systematic review of the literature. *Cancer Treat Rev* 2004; 30(2):205-211.

(3) Lacey CJ. Therapy for genital human papillomavirus-related disease. *J Clin Virol* 2005; 32 Suppl 1:S82-S90.

(4) Arbyn M, Ronco G, Anttila A, Meijer CJLM, Poljak M, Ogilvie G et al. Evidence regarding HPV testing in secondary prevention of cervical cancer. *Vaccine* 2012; 30 Suppl 5:F88-F99.

(5) Alonso I, Torne A, Puig-Tintore LM, Esteve R, Quinto L, Campo E et al. Pre- and post-conization high-risk HPV testing predicts residual/recurrent disease in patients treated for CIN 2-3. *Gynecol Oncol* 2006; 103(2):631-636.

(6) Stern PL, van der Burg SH, Hampson IN, Broker TR, Fiander A, Lacey CJ et al. Therapy of human papillomavirus-related disease. *Vaccine* 2012; 30 Suppl 5:F71-F82.

(7) Melnikow J, McGahan C, Sawaya GF, Ehlen T, Coldman A. Cervical Intraepithelial Neoplasia Outcomes After Treatment: Long-term Follow-up From the British Columbia Cohort Study. *J Natl Cancer Inst* 2009; 110(10):721-728.

(8) Soutter WP, Sasieni P, Panoskaltsis T. Long-term risk of invasive cervical cancer after treatment of squamous cervical intraepithelial neoplasia. *Int J Cancer* 2005; 118(8):2048-2055.

(9) Kalliala I, Anttila A, Pukkala E, Nieminen P. Risk of cervical and other cancers after treatment of cervical intraepithelial neoplasia: retrospective cohort study. *BMJ* 2005; 331(7526):1183-1185.

(10) Strander B, Hallgren J, Sparen P. Effect of ageing on cervical or vaginal cancer in Swedish women previously treated for cervical intraepithelial neoplasia grade 3: population based cohort study of long term incidence and mortality. *BMJ* 2014; 348:f7361.

(11) Arbyn M, Kyrgiou M, Gondry J, Petry KU, Paraskevaidis E. Long term outcomes for women treated for cervical precancer. *BMJ* 2014; 348:f7700.

(12) Clinical applications of HPV testing: a summary of meta-analyses. 06 Apr 8; Barcelona: 2006.

(13) Harbord RM, Deeks JJ, Egger M, Whiting P, Sterne JA. A unification of models for meta-analysis of diagnostic accuracy studies. *Biostatistics* 2007; 8:239-251.

(14) Harbord RM, Whiting P. metandi: Meta-analysis of diagnostic accuracy using hierarchical logistic regression. *The Stata Journal* 2009; 9(2):211-229.

(15) Takwoingi Y, Diagnostic Test AccuracyWorking Group. METADAS: A SAS macro for meta-analysis of diagnostic accuracy studies. http://srdta cochrane org/sites/srdta cochrane org/files/uploads/MetaDAS%20Quick%20Reference%20v1 3%20May%202012 pdf [2009 :[-*]

(16) Diagnostic Test Accuracy Working Group. Handbook for Diagnostic Test Accuracy Reviews. Available from: http://srdta.cochrane.org/handbook-dta-reviews. 2011.

(17) Dersimonian R, Laird NM. Meta-analysis in clinical trials. *Controlled Clin Trials* 1986; 7:177-188.

(18) Chua KL, Hjerpe A. Human papillomavirus analysis as a prognostic marker following conization of the cervix uteri. *Gynecol Oncol* 1997; 66(1):108-113.

(19) Houfflin Debarge V, Collinet P, Vinatier D, Ego A, Dewilde A, Boman F et al. Value of human papillomavirus testing after conization by loop electrosurgical excision for high-grade squamous intraepithelial lesions. *Gynecol Oncol* 2003; 90(3):587-592.

(20) Chao A, Lin CT, Hsueh S, Chou HH, Chang TC, Chen MY et al. Usefulness of human papillomavirus testing in the follow-up of patients with high-grade cervical intraepithelial neoplasia after conization. *Am J Obstet Gynecol* 2004; 190(4):1046-1051.

(21) Verguts J, Bronselaer B, Donders G, Arbyn M, Van Eldere J, Drijkoningen M et al. Prediction of recurrence after treatment for high-grade cervical intraepithelial neoplasia: the role of human papillomavirus testing and age at conisation. *BJOG* 2006; 113(11):1303-1307.

(22) Fambrini M, Penna C, Pieralli A, Bussani C, Fallani MG, Andersson KL et al. PCR detection rates of high risk human papillomavirus DNA in paired self-collected urine and cervical scrapes after laser CO2 conization for high-grade cervical intraepithelial neoplasia. *Gynecol Oncol* 2008.

(23) Aerssens A, Claeys P, Beerens E, Garcia A, Weyers S, Van Renterghem L et al. Prediction of recurrent disease by cytology and HPV testing after treatment of cervical intraepithelial neoplasia. *Cytopathology* 2009; 20(1):27-35.

(24) Trope A, Jonassen CM, Sjoborg KD, Nygard M, Dahl FA, Alfsen GC et al. Role of high-risk human papillomavirus (HPV) mRNA testing in the prediction of residual disease after conisation for high-grade cervical intraepithelial neoplasia. *Gynecol Oncol* 2011; 123(2):157-162.

(25) Ryu A, Nam K, Kwak J, Kim J, Jeon S. Early human papillomavirus testing predicts residual/ recurrent disease after LEEP. *J Gynecol Oncol* 2012; 23(4):217-225.

(26) Torne A, Fuste P, Rodriguez-Carunchio L, Alonso I, del PM, Nonell R et al. Intraoperative post-conisation human papillomavirus testing for early detection of treatment failure in patients with cervical intraepithelial neoplasia: a pilot study. *BJOG* 2013; 120(4):392-399.

(27) Simoens C, Goffin F, Simon P, Barlow P, Antoine J, Foidart J-M et al. Adverse obstetrical outcomes after treatment of precancerous cervical lesions: a Belgian multicentre study. *BJOG* 2012; 119(10):1247-1255.

(28) Arbyn M, Kyrgiou M, Simoens C, Raifu AO, Koliopoulos G, Martin-Hirsch P et al. Peri-natal mortality and other severe adverse pregnancy outcomes associated with treatment of cervical intraepithelial neoplasia: a meta-analysis. *BMJ* 2008; 337:a1284.

(29) Lubrano A, Medina N, Benito V, Arencibia O, Falcon JM, Leon L et al. Follow-up after LLETZ: a study of 682 cases of CIN 2-CIN 3 in a single institution. *Eur J Obstet Gynecol Reprod Biol* 2012; 161(1):71-74.

(30) Leguevaque P, Motton S, Decharme A, Soulé-Tholy M, Escourrou G, Hoff J. Predictors of recurrence in high-grade cervical lesions and a plan of management. *Eur J Surg Oncol* 2011; 40(2):174-177.

(31) Persson M, Brismar WS, Ljungblad L, Johansson B, Weiderpass E, Andersson S. High-risk human papillomavirus E6/E7 mRNA and L1 DNA as markers of residual/recurrent cervical intraepithelial neoplasia. *Oncol Rep* 2012; 28(1):346-352.

(32) Tinelli A, Guido M, Zizza A, Pellegrino M, Greco M, Vergara D et al. The mRNA-HPV Test Utilization in the Follow up of HPV Related Cervical Lesions. *Curr Pharm Des* 2013; 19(8):1458-1465.

(33) Kreimer AR, Guido RS, Solomon D, Schiffman M, Wacholder S, Jeronimo J et al. Human papillomavirus testing following loop electrosurgical excision procedure identifies women at risk for posttreatment cervical intraepithelial neoplasia grade 2 or 3 disease. *Cancer Epidemiol Biomarkers Prev* 2006; 15(5):908-914.

(34) Brismar S, Johansson B, Borjesson M, Arbyn M, Andersson S. Follow-up after treatment of cervical intraepithelial neoplasia by HPV-genotyping. *Am J Obstet Gynecol* 2009; 201(1):17.e1-17.e8.

(35) Kang WD, Jeong OM, Kim SM, Nam JH, Soo PC, Choi HS. Significance of human papillomavirus genotyping with high-grade cervical intraepithelial neoplasia treated by a loop electrosurgical excision procedure. *Am J Obstet Gynecol* 2010; 203(1):72-76.

(36) Heymans J, Benoy IH, Poppe W, Depuydt CE. Type-specific HPV geno-typing improves detection of recurrent high-grade cervical neoplasia after conisation. *Int J Cancer* 2011; 129(4):903-909.

(37) Boulet GA, Micalessi IM, Horvath CA, Benoy IH, Depuydt CE, Bogers JJ. Nucleic acid sequence-based amplification assay for human papillomavirus mRNA detection and typing: evidence for DNA amplification. *J Clin Microbiol* 2010; 48(7):2524-2529.

8.6. GRADE-Profil

Follow-up – Biomarker vs. HPV oder Zytologie - GRADE process

Authors: M. Arbyn, M. Jentschke

GRADE has built on previous systems to create a highly structured, transparent, and informative system for rating quality of evidence(Guyatt *et al.,* 2008b).

Steps in evidence assessment for making guidelines
1) Formulate a question
2) Identify the PICO(S) components
3) Qualify outcomes as critical, important, not important

1) Questions
Is testing with a biomarker more accurate than follow-up using hrHPV-DNA testing or cytology to predict failure or success after treatment of cervical precancer?

2) PICOS

- P: Women treated for histologically confirmed CIN2+ or CIN3+ (or CIN1+) by an excisional (LEEP, LLETZ, laser conisation) or ablative (cryotherapy, laser ablation, electrocoagulation) procedure.
- I1: Assessment of margin status at time of treatment.
- I2: Testing with a biomarker (mRNA,p16, HPV type-specific persistence, or other) three to nine months post-treatment.
- C: HPV-DNA testing or Cytology, three to nine months post-treatment.
- O: Accuracy to detect residual/recurrent CIN2+ or CIN3+ (or CIN1+), confirmed by histology.
    - Prospective studies:
        - accuracy (sensitivity, specificity, positivie predictive value [PPV], negative predictive value [NPV], and complement to NPV [cNPV=1-NPV]),
        - prediction values
    - Case control studies:
        - accuracy (sensitivity, specificity, positivie predictive value [PPV], negative predictive value [NPV], and complement to NPV [cNPV=1-NPV])
- S: Prospective studies:
    - all participants tested with a biomarker and a HPV test and/or cytology
    - 18 months follow-up

- verification with a golden reference standard: colposcopy completed with targeted biopsy for all participants, or in case of a positive test result (biomarker, HPV-test or cytology)

- Case control studies:
  - all participants tested with a biomarker and a HPV test and/or cytology
  - 18 months follow-up

3) Importance of outcomes

Outcome:

31. Reduction of mortality from cervical cancer, (quality-adjusted) life-years gained.
32. Reduction of morbidity due to cervical cancer: incidence of cancer (Ib+).
33. Reduction of incidence of cancer (including micro-invasive cancer).
34. Reduction of incidence of CIN3 or worse disease (CIN3+).
35. Increased detection rate of CIN3+ or CIN2+.
36. Increased test positivity with increased, similar or hardly reduced positive predictive value.

Increased detection rate of cervical disease after treatment for CIN.

The following outcome measures were assessed:

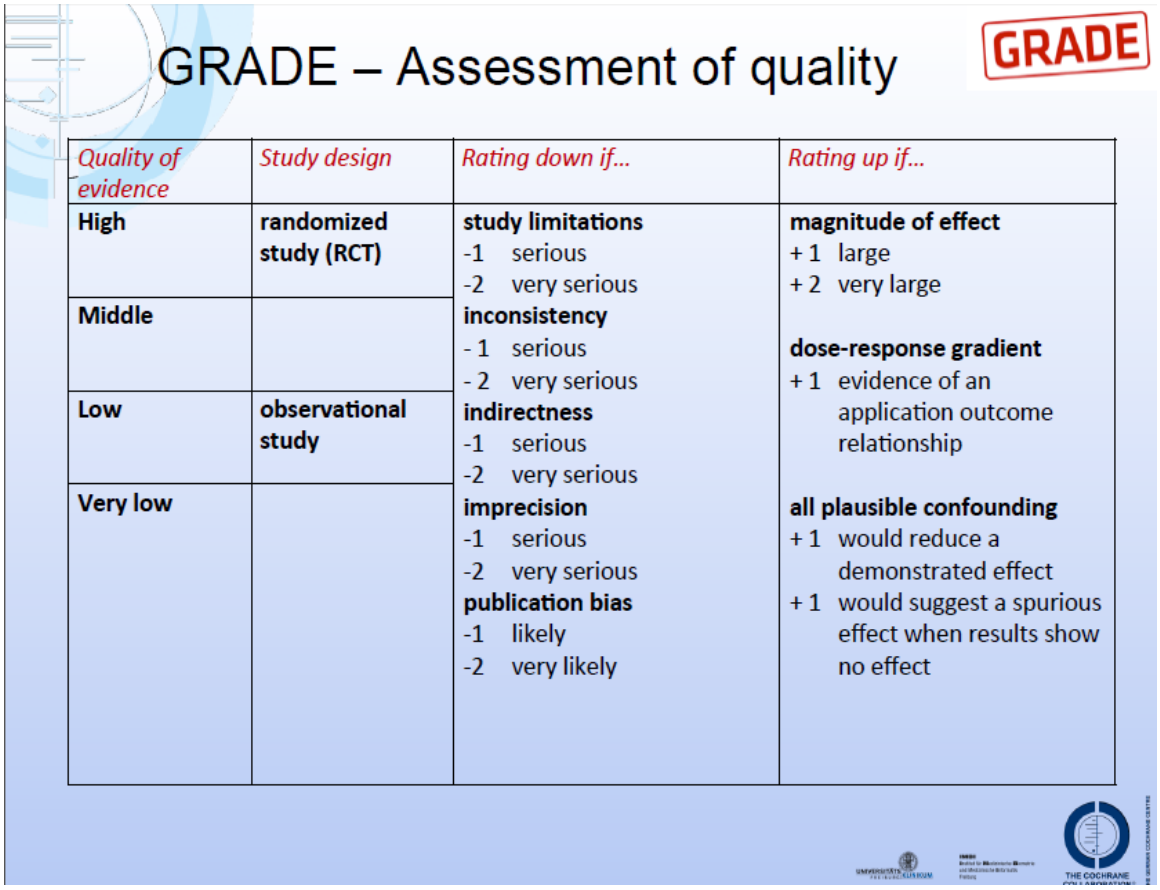
- treatment failure rate (prevalence of residual/recurrent CIN2+)
- absolute accuracy of margin status to detect CIN2+ or CIN3+
- relative accuracy of margin status versus HPV-DNA testing or cytology to detect CIN2+ or CIN3+
- absolute accuracy of testing with a biomarker to detect CIN2+ or CIN3+
- relative accuracy of testing with a biomarker versus HPV-DNA testing or cytology to detect CIN2+ or CIN3+

4) Quality of evidence for each outcome in four categories;

- High: ++++
- Moderate: +++
- Low: ++
- Very low: +

4.1 Margin status to predict residual/recurrent disease

Nine studies could be pooled in a meta-analysis because they had similar characteristics.



5 factors that lower the quality of the evidence (Guyatt *et al.*, 2008a)

26. Risk of bias due to limitations in design or execution: see CONSORT, STROBE, QUADAS
27. Inconsistency or heterogeneity: if consistency unexplained, lower quality
28. Indirectness, applicability (relevance of studies for answering the PICPO question)
29. Imprecision: number of studies, width of CI
30. Reporting bias, publication bias.

3 factors that increase the quality

1. Large effect
2. Dose effect gradient
3. Confounders (presence of confounders that have lowered the observed effect, absence of these would have increased the effect) (Guyatt *et al.*, 2008a)

Items downgrading quality of evidence		**Downgrading**
Bias, design	No	No (-0)
Inconsistency	Significant heterogeneity of the margin positivity rate in the meta-analysis (I^2=89%, p=0.000), which could explain the poor correlation between positive section margins	Yes (-1)

	and residual/recurrent disease.	
Indirectness	No	No (-0)
Imprecision	The sensitivity of margin status was 27% lower than that of HPV-DNA testing, resulting in a relative sensitivity of 0.73 (95% CI: 0.63-0.86). The specificities of margin status and HPV-DNA testing were similar, with a specificity ratio of 1.04 (95% CI: 0.97-1.11) The pooled sensitivity and specificity ratio were 0.90 (95% CI: 0.75-1.10) and 0.95 (95% CI: 0.85-1.07)	No (-0)
Publication bias, other	No information	No (-0)
Items upgrading quality of evidence		
Large effect	No	No (+0)
Dose-effect correlation	No	No (+0)
Confounding factors neutralising effects	No	No (+0)°

Conclusion: evidence of low quality.

4.2 Using other biomarkers to predict residual/recurrent disease

Three studies were found assessing the accuracy of mRNA testing, and five studies using HPV type-specific persistence as a biomarker for residual/recurrent disease. All eight studies allowed comparison with hrHPV-DNA testing and/or cytology. However, the study characteristics of these studies were too diverse to allow meta-analytical pooling.

Items downgrading quality of evidence		**Downgrading**
Bias, design	No	No (-0)
Inconsistency	The study characteristics of these studies were too diverse to allow meta-analytical pooling.	Yes (-1)
Indirectness	No	No (-0)
Imprecision	No	No (-0)
Publication	No information	No (-0)

bias, other		
Items upgrading quality of evidence		
Large effect	No	No (+0)
Dose-effect correlation	No	No (+0)
Confounding factors neutralising effects	No	No (+0)˚

Conclusion: evidence of very low quality.

Table 52 GRADE evidence profile

	Quality of evidence								
# studies (N)	Absence of study limitations	Consistency	Directness (outcome, representativity Germany)	Precision	Absence publication bias	Large effect	Dose-effect relation	Bias lowering effect	Comment
Outcome 1: Margin status to predict residual/recurrent disease									
9	Yes	No	Yes	Yes	Yes	No	No	No	Low
Outcome 2: Using other biomarkers to predict residual/recurrent disease									
8	Yes	No	Yes	Yes	Yes	No	No	No	Very Low

Table 53 Pooled relative sensitivity and specificity of mRNA testing compared to HPV-DNA testing to detect CIN2+ and CIN3+.

	Number of studies (test combination)		Sensitivity ratio (95% CI)		Specificity ratio (95%CI)	
	CIN2+	CIN3+	CIN2+	CIN3+	CIN2+	CIN3+
mRNA versus validated HPV-DNA testing						
mRNA >5types	6	4	0.99 (0.95-1.04)	1.01 (0.98-1.04)	1.05 (1.03-1.07)	1.05 (1.02-1.07)
mRNA 5types	2	1	0.77 (0.65-0.90)	0.69 (0.50-0.97)	1.12 (1.10-1.13)	1.12 (1.10-1.13)

Table 54 Pooled relative sensitivity and specificity (95% CI) of mRNA testing compared to LBC testing at cut-off ASC-US and LSIL, to detect CIN2+ and CIN3+.

	Number of studies (test combination)		Sensitivity ratio (95% CI)		Specificity ratio (95%CI)	
	CIN2+	CIN3+	CIN2+	CIN3+	CIN2+	CIN3+
mRNA versus LBC (ASC-US)						
mRNA >5types	4	3	1.14 (0.88-1.49)	1.21 (1.03-1.42)	1.01 (0.94-1.08)	0.98 (0.94-1.01)
mRNA 5types	2	1	0.87 (0.58-1.31)	0.69 (0.50-0.97)	1.00 (0.99-1.01)	1.00 (0.99-1.01)
mRNA versus LBC (LSIL)						
mRNA >5types	3	2	1.32 (0.97-1.81)	1.25 (0.77-2.03)	0.95 (0.91-0.98)	0.94 (0.90-0.98)
mRNA 5types	2	1	1.00 (0.60-1.67)	0.73 (0.51-1.03)	0.96 (0.96-0.99)	0.97 (0.96-0.98)

References

Guyatt G.H., Oxman A.D., Kunz R., Falck-Ytter Y., Vist G.E., Liberati A., & Schunemann H.J. (2008a) Going from evidence to recommendations. *BMJ* **336:** 1049-1051.

Guyatt G.H., Oxman A.D., Kunz R., Vist G.E., Falck-Ytter Y., & Schunemann H.J. (2008b) What is "quality of evidence" and why is it important to clinicians? *BMJ* **336:** 995-998.

Guyatt G.H., Oxman A.D., Vist G.E., Kunz R., Falck-Ytter Y., Alonso-Coello P., & Schunemann H.J. (2008c) GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *BMJ* **336:** 924-926.

Schunemann H.J., Oxman A.D., Brozek J., Glasziou P., Jaeschke R., Vist G.E., Williams J.W., Jr., Kunz R., Craig J., Montori V.M., Bossuyt P., & Guyatt G.H. (2008) GRADE: Grading quality of evidence and strength of recommendations for diagnostic tests and strategies. *BMJ* **336:** 1106-1110.

9. Question: Follow-up after treatment of CIN: Is hrHPV-DNA testing more accurate than follow-up cytology to predict failure or success after treatment of CIN?

9.1. Background and rationale

The presence of cervical intraepithelial neoplasia of grade two and three (CIN2-3) is linked with a risk of developing invasive carcinoma if not treated[1]. Conservative treatment methods for high-grade CIN, which can be performed under local anaesthesia in an outpatient setting, are performed to prevent progression to cervical cancer[2]. Such methods can be excisional or ablative[3]. The former type includes laser electrosurgical excision (LLETZ), and laser conisation. The latter type includes cryotherapy, laser ablation, and electrocoagulation.

However, success of treatment is suboptimal and residual or recurrent high grade disease (CIN2+) can be detected in on average 8% (ranging from 4% to 18%) of treated women[4], with the majority of treatment failure occurring in the first two post-operative years[5-7]. Moreover, women with a history of treatment for cervical precancer are still at higher risk of developing invasive cervical cancer in the future compared to the general population during at least 10 years and maybe up to 20 years after treatment[8;9]. A recent Swedish trend study also confirmed that women with prior treatment for CIN3+ show an increased incidence of and even mortality from cervical or vaginal cancer in particular in older age groups (aged 60 or older) and among those treated in more recent periods[10;11]. Therefore, finding an indicator to predict this risk with great accuracy would be particularly helpful.

Guidelines for surveillance strategies after treatment of CIN vary greatly among countries with respect to timelines, type of tests performed, and length of follow-up. In this report, a meta-analysis is updated on the occurrence of treatment failure and on the accuracy of testing with an hrHPV-DNA test versus cytology to detect residual/recurrent CIN2+[4;12].

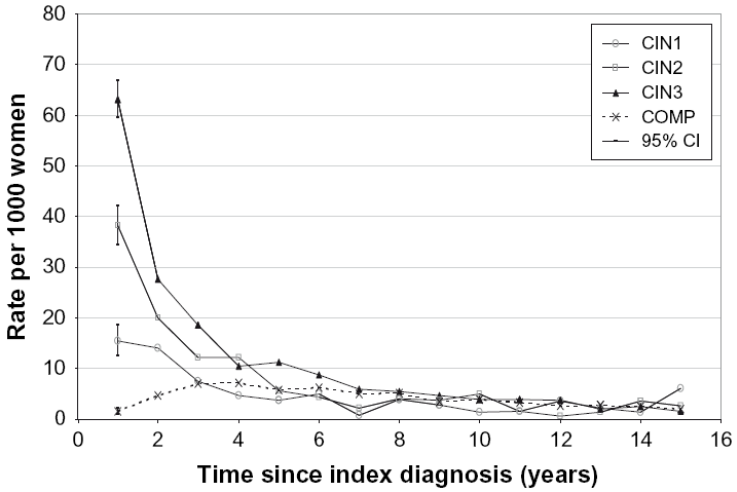


Figure 15: Adapted from Melnikow et al., 2009[7]. Incidence rates of CIN2/3 per 1000 women by index diagnosis among women treated for CIN and women with normal cytology (COMP). Error bars: 95% confidence intervals.

9.2. Question

Is hrHPV-DNA testing more accurate than follow-up cytology to predict failure or success after treatment of CIN?

9.2.1. PICOS components

P: Women treated for histologically confirmed CIN2/3 by an excisional (LLETZ, laser conisation) or ablative (cryotherapy, laser ablation, electrocoagulation) procedure.

I: hrHPV-DNA testing, three to nine months post-treatment.

C: Cytology, three to nine months post-treatment.

O: Occurrence of residual or recurrent CIN2/3+ after treatment.

Absolute sensitivity and specificity of hrHPV-DNA testing, cytology testing and co-testing (hrHPV-DNA and cytology) to detect residual/recurrent CIN2+ or CIN3, confirmed by histology[§§§§]. Relative sensitivity and specificity of hrHPV-DNA testing versus cytology and of co-testing versus hrHPV-DNA testing alone.

S: Prospective studies and case-control studies:

9.3. Methods

Previous meta-analyses have been performed and published by the Unit of Cancer Epidemiology on the accuracy of hrHPV-DNA testing and cytology to predict residual/recurrent disease after treatment of CIN2/3[4;12].

[§§§§] From cohort studies predictive values (PPV, NPV and cNPV) can be derived as well but these accuracy indicators will not be pooled for the current systematic review.

9.3.1. Search strategy

The search-string used to identify relevant studies in the previously published meta-analyses [4;12] was used to update the list of included studies (Box1).

((cancer OR carcinoma OR dysplas* OR neoplas* OR CIN OR SIL) AND (cervix OR cervical) OR vaginal smears[MeSH] OR Cervix Neoplasm [Mesh]))
AND
(treatment OR conisation OR conization OR LLETZ OR leep OR cryotherapy)
AND
(HPV OR human papillomavirus OR papillomavirus)

Box 1: search string

9.3.2. Eligibility of studies

Studies were eligible if (1) women were treated for histologically-confirmed CIN2 or CIN3, (2) women had a cytology and hrHPV-DNA test between three and nine months post-treatment, (3) women were followed-up for at least 18 months, and (4) colposcopy and targeted biopsy was performed on all women or women with a positive hrHPV-DNA test or abnormal cytology. Treatment failure was defined as follow-up histological diagnosis of CIN2+.

9.3.3. Outcome measures

The following outcome measures were assessed:
- treatment failure rate (occurence of residual/recurrent CIN2+)
- absolute accuracy of hrHPV-DNA testing and cytology to detect CIN2+
- absolute accuracy of combined testing with hrHPV-DNA and cytology to detect CIN2+
- relative accuracy hrHPV-DNA testing versus cytology to detect CIN2+
- relative accuracy of combined testing with hrHPV-DNA and cytology to detect CIN2+

9.3.4. Statistical analysis

The pooled absolute sensitivity and specificity of the tests were estimated jointly using *metandi*, a procedure in STATA, based on the bivariate normal model for the logit transforms of sensitivity and specificity taking the intrinsic correlation between true and false-positivity rates and the variability between studies into account [13;14]. The relative sensitivity and specificity of hrHPV-DNA testing compared with cytology were computed using *metadas*, a SAS macro for meta-analysis of diagnostic accuracy studies which allows the inclusion of test as a covariate making comparison of tests possible[15;16].

Forrest plots were also produced showing study-specific and pooled absolute and relative accuracy estimates using a random effects model[17]. In these forest plots, the statistical heterogeneity was assessed using Cochran's Q test and the I^2 statistic, which measures the proportion of variation that is due to inter-study heterogeneity. Statistical analysis was

performed using STATA/SE 10 (Stata Corporation, College Station, TX, USA) and SAS 9.3(SAS Institute, Campus Drive Cary, NC, USA).

9.4. Results

9.4.1. Relevant studies and study characteristics

Three studies[18-20] were found eligible and were added to the list of studies that had been included in the previously published meta-analysis[4], resulting in a total of 17 studies. A PRISMA flow chart for the retrieval of studies is shown in Figure 16. Overall, included studies were heterogeneous with regard to their study characteristics. A comprehensive summary of study and test-characteristics is listed in Table 55 and Table 56.

All included studies were prospective, except for two that had a case-control design[21;22]. HrHPV-DNA testing was performed by the Hybrid Capture 2 assay (HC2) in ten studies[5;19;20;23-28], by a PCR method in eight studies[18;21;22;29-33]. Overall, follow-up data of 3041 women treated for CIN2+ or CIN3+ were included in the analysis.

The quality of included studies was evaluated using the QUADAS_2 tool and is summarized in Table 57. Overall, studies scored well for the majority of QUADAS-items, except that the blinding of tests (hrHPV-DNA testing and follow-up cytology) and gold standard results was in some cases not assured or not sufficiently documented.

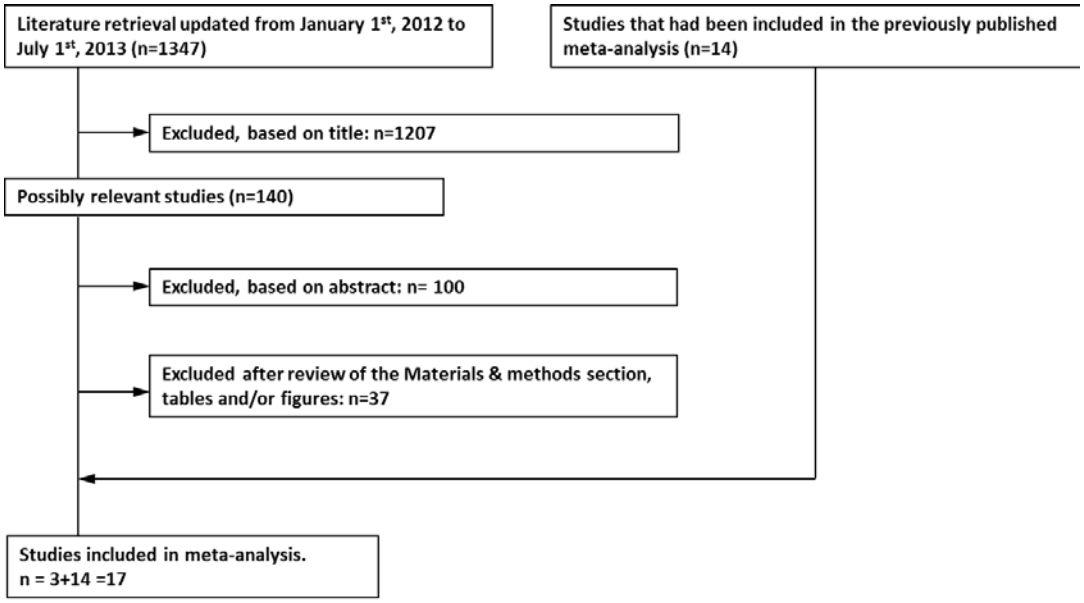


Figure 2: PRISMA flow chart for the retrieval of studies.

Table 55: Study characteristics of included reports

Author	Year	Country	Study type	Study size final/initial	Age	Inclusion criteria	Exclusion criteria	Treatment procedure	Treated disease
Chua	1997	Sweden	Case-control	26 cases 22 controls	Cases: 38.2 Controls: 33.5	Archived biopsies and pap smears. Cases: 26 sections who have had at least one histologically confirmed recurrence of CIN 2 or CIN 3 after the first cone biopsy. Controls: 22 consecutive patients who have had conization due to CIN 3 and have remained disease-free for more than 46 months.	Not documented	Conisation (type not specified)	Cases: CIN2/3 (n=26) Controls: CIN3 (n=22)
Nobbenhuis	2001	The Netherlands	Prospective	184/184	34 (21-70)	Women diagnosed with CIN 2 or 3 at the colposcopy outpatient clinic and consecutively treated for CIN. All fulfilled the following inclusion criteria: adequate HPV sample at initial treatment; ≥1 adequate HPV samples after treatment.	Previous history of cervical pathology, prenatal DES (diethylstilbestrol) exposure, concomitant cancer	LLETZ (n=152) or cone biopsy (n=32)	CIN2/3 (n=184)
Zielinski	2003	The Netherlands	Prospective	108/111	35 (23-56)	Women treated for histologically confirmed CIN3.	No valid post-treatment HPV test.	Cone biopsy (n=23) or LLETZ (n=85)	CIN3 (n=108)
Cecchini	2004	Italy	Prospective	84/84	34	Women, consecutively treated for histologically confirmed high-grade CIN in the Florence screening program.	Not documented	LLETZ	CIN2 (n=27) CIN3 (n=57)
Sarian	2004	Brasil	Prospective	88/107	34 (20-60)	Women treated for CIN, with histologically confirmed CIN 2/3 on the conisation specimen, and ≥1 follow-up visit.	pregnancy, clinical signs of immunosuppression, HIV positivity	LLETZ or cold knife conisation	CIN2/3 (n=88)
Alonso	2006	Spain	Prospective	203/224	39 (22-83)	Women treated for CIN, with histologically confirmed CIN 2/3 on the conisation specimen, and ≥1 follow-up visit.	Not documented	LLETZ	CIN2/3 (n=203)
Kreimer	2006	USA	Prospective	485/610	24 (21-28)	Not documented	<CIN2 on the baseline biopsy or treatment specimen (n=20). >6m between baseline biopsy and treatment (n=56).	LLETZ	CIN2 (n=312) CIN3 (n=298)
Verguts	2006	Belgium	Prospective	72/72	40 (22-78)	Not documented	Not documented	LLETZ	CIN2 (n=12) CIN3 (n=60)
Fambrini	2008	Italy	Prospective	52/103	38 (18-59)	Women with histologically confirmed high grade CIN on conisation specimen, providing informed consent and with 12 months of follow-up including all the scheduled examinations.	Diabetes, HIV positivity and chronic steroidal therapy.	Laser CO2 conisation	CIN2/3 (n=52)
Aerssens	2009	Belgium, Nicaragua	Prospective	137/138	34.8 (20-60)	Women with histologically confirmed CIN2/3.	Not documented	LLETZ	CIN2 (n=73) CIN3 (n=65)

Author	Year	Country	Study type	Study size final/initial	Age	Inclusion criteria	Exclusion criteria	Treatment procedure	Treated disease
Bais	2009	The Netherlands	Prospective	89/102	not documented	Women who were to be treated for high-grade CIN lesions and agreed to participate.	Previous treatment for high-grade CIN; immune compromising conditions; previous or current cancer.	LETZ, cold-knife conisation, or laser conisation.	CIN2/3 (n=89)
Kang	2010	South-Korea	Prospective	672/672	39.7 (21–62)	Women with histologically confirmed CIN2/3 on conisation specimen, valid pre- and post-LLETZ HPV-tests, and ≥ 24m follow-up	Hysterectomy	LLETZ	CIN2/3 (n=672)
Smart	2010	Australia	Prospective	100/100	32 (19-66)	Women treated for histologically confirmed CIN2/3, attending the first follow-up visit.	Pregnancy, history of cervical cancer	LLETZ (n=85), cold-knife conisation (n=14), or laser ablation (n=1)	CIN2 (n=30) CIN3 (n=70)
Heymans	2011	Belgium	Case-control	21 cases 42 controls	Cases: 40.9 Controls: 35.5	Women with histologically confirmed CIN2/3 after conisation (surgical or LLETZ); test-results for LBC and HPV genotyping pre- and 6m post-treatment. Cases: histological recurrence of CIN2/3. Controls: no recurrence of CIN2/3, and ≥2 consecutive normal cervical smears in ≥24m follow-up	Invasive cervical cancer	LLETZ, or surgical conisation	CIN2/3 (n=63)
Trope	2011	Norway	Prospective	344/604	36 (20-75)	Woman with histologically confirmed CIN2+, and treated by conisation (LLETZ or carbon dioxide laser conisation), valid hrHPV mRNA and DNA test results ≤2m pre-treatment.	Invalid cytology, mRNA or DNA test results at 6m follow-up. No biopsy ≤18m post-treatment and invalid cytology or hrHPV test at 12m. Invasive carcinoma, treated with radical hysterectomy (n=9).	LLETZ Laser conisation	CIN2+ (n=604)
Ryu	2012	South-Korea	Prospective	180/371	39.3 (22-73)	Inclusion criteria were CIN 2/3 on a colposcopic punch biopsy and/or excised specimen, adequate 3- and 6month follow-up after LLETZ, and an HPV HC2 test and/or HPV DNA chip test before and after LLETZ.	Hysterectomy (n=59) CIN1 at treatment (n=29)	LLETZ	CIN2/3 (n=180)
Torne	2012	Spain	Prospective	132/132	36.2	patients consecutively diagnosed with CIN2/3 by colposcopically directed biopsy or ECC ≤90d pre-treatment, who underwent conisation by LLETZ	Not documented.	LLETZ	CIN2/3 (n=132)

Abbreviations: AIS, adenocarcinoma in situ; CIN2, cervical intra-epithelial neoplasia grade two; CIN3, cervical intra-epithelial neoplasia grade three; DNA, deoxyribonuclease; HC2, hybrid capture-2 assay; HIV, hunan immunodeficiency virus; HPV, human papillomavirus; LBC, liquid-based cytology; LLETZ, large loop excision of the transformation zone; mRNA, messenger ribonucleic acid; Pap, papanikolau smear.

Table 56: Test characteristics, and duration of follow-up of included studies

Study	**hrHPV-DNA test** (primer)	**Timing testing**	**Gold standard verification**	**Timing Follow-up** Mean (range)
Chua, 1997	Nested PCR (my09/11-gp5+/6+)	3	Not documented	46
Nobbenhuis, 2001	PCR (GP5+/6+)	6	Colposcopy + targeted biopsy in case of abnormal cytology.	24
Zielinski, 2003	HC2	3	Colposcopy + targeted biopsy.	29 (2-65)
Cecchini, 2004	PCR (non-consensus)	6	Colposcopy + targeted biopsy for all women at 6m and 12m post-treatment. Subsequently, annually in case of abnormal cytology (ASCUS+)	23 (11-40)
Sarian, 2004	HC2	5	Colposcopy + targeted biopsy for all women during 2 follow-up visits (5m,12m post-treatment) .	18
Alonso, 2006	HC2	6	Colposcopy for all women at 6m, 12m, 18m, 24m post-treatment. Biopsies were performed in case of an abnormal transformation zone, or in case of ≥1 positive test result (ASCUS+, hrHPV+). When the transformation zone was not/partially visible or no colposcopy abnormality was identified, an endocervical curettage was performed.	20 (6-66)
Kreimer, 2006	HC2	4.5	Colposcopy + targeted biopsy every 6m for 2y (after onset of the study).	24
Verguts, 2006	HC2	4.5	Colposcopy + targeted biopsy for all women at 6m intervals for 2y. An additional colposcopy was performed in case of abnormal cytology (ASCUS+)	24
Fambrini, 2008	PCR (non-consensus, E6/E7)	6	Colposcopy + targeted biopsy for all women at 3m, 6m, 12m, 18m, and 24m post-treatment, and annually afterwards.	25 (19-30)
Aerssens, 2009	PCR (SPF10)	6	Colposopy for all women at 6w, 6m, 12m, 24m post-treatment . Biopsy was taken in case of abnormal colposcopy (LSIL+) or cytology (ASCUS+). ECC was performed if cytology was positive and colposcopy was negative.	22 (4-32)
Bais, 2009	PCR (GP5+/6+)	6	Colposcopy + targeted biopsy for all women at the end of the study (24m post-treatment, or at time of re-treatment). During the study, colposcopy was performed in case of ASCUS+ and hrHPV-positivity.	24
Kang, 2010	HC2	24	Colposcopy + targeted biopsy in case of ASCUS+ or hrHPV-positivity at 6, 12, 18, 24m follow-up.	24
Smart, 2010	HC2	6	Colposcopy + targeted biopsy in case of ASCUS+.	9 (3-18)
Heymans, 2011	PCR (non-consensus, E6/E7)	6	Colposcopy + targeted biopsy for all women post-treatment Controls: ≥ 2 subsequent negative cytology smears during ≥24m follow-up	24
Trope, 2011	Amplicor (L1)	6	Colposcopy (+ biopsies) for all women. At 6m follow-up: biopsy was taken in case of LSIL+/hrHPV+, HSiL+, or abnormal colposcopy. At 12m follow-up: a biopsy was taken if indicated by mRNA+, or DNA+, or ASCUS+, or abnormal colposcopy. If cytology was abnormal and/or hrHPV test was positive and colposcopy was normal, random biopsies and ECC were taken. All test results during the 18-month follow-upwere reconciled with the Norwegian Cancer Registry.	(6-18)
Ryu, 2011	HC2	6	Colposcopy + targeted biopsy for all women at 3, 6, 12, 18, 24m post-treatment.	25 (4-60)
Torne, 2012	HC2	6	Patients with ASC-H, LSIL, or HSIL: Colposcopy + targeted biopsy or ECC (TZ not/partially visible, or normal colposcopy) . Patients with ASC-US and/or an isolated HPV+ test result: colposcopy. Targeted biopsy or ECC only in case of abnormal colposcopy or if TZ not/partially visible. Colposcopies were performed at 6, 12, 18, and 24m follow-up.	24

Table 57. Evaluation of the quality of each included study according to the QUADAS-2 check list[34].

	Risk of Bias													
	Patient Selection			Screening Test		Reference Test			Flow & Timing					
Author, Year	P1	P2	P3	T1	T2	R1	R2	R3	F1	F2	F3	F4	F5	F6
Chua,1997	Y	Y	Y	Y	?	Y	Y	Y	Y	Y	?	Y	Y	Y
Nobbenhuis,2001	Y	Y	Y	Y	N	Y	Y	Y	Y	Y	?	Y	Y	Y
Zielinski,2003	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Cecchini,2004	Y	Y	Y	Y	N	Y	Y	Y	Y	Y	Y	Y	Y	Y
Sarian,2004	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Alonso,2006	Y	Y	Y	Y	N	Y	Y	Y	Y	Y	N	Y	Y	Y
Kreimer,2006	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Verguts,2006	Y	Y	Y	Y	N	Y	Y	Y	Y	Y	Y	Y	Y	Y
Fambrini,2008	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Aerssens,2009	Y	Y	Y	Y	N	Y	Y	Y	Y	Y	Y	Y	Y	Y
Bais,2009	Y	Y	Y	Y	N	Y	Y	Y	Y	Y	Y	Y	Y	Y
Kang,2010	Y	Y	Y	Y	N	Y	Y	Y	?	Y	?	Y	Y	Y
Smart,2010	Y	Y	?	Y	N	Y	Y	Y	Y	Y	?	Y	Y	Y
Heymans,2011	Y	Y	Y	Y	?	Y	Y	Y	Y		N	Y	Y	Y
Trope,2011	Y	Y	?	Y	N	Y	Y	Y	Y	Y	Y	Y	Y	
Ryu,2012	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Torne,2012	Y	Y	Y	Y	N	Y	Y	Y	Y	Y	?	Y	Y	Y

QUADAS items: (P1) acceptable enrolment method, (P2) inappropriate exclusions avoided, (P3) acceptable follow-up time, (T1) pre-specified test cut-off, (T2) results of index and comparator tests blinded towards each other and reference test, (R1) acceptable reference test, (R2) results of reference test blinded towards index and comparator tests, (R3) incorporation bias avoided, (F1) acceptable delay between triage tests and reference test, (F2) partial verification avoided, (F3) differential verification avoided, (F4) withdrawals explained, (F5) uninterpretable results reported for tests, (F6) uninterpretable results reported for reference test. Each quality item is judged with: Y (fulfilled, green), ? (unclear, yellow), N (not fulfilled, red).

9.4.2. Treatment Failure

Treatment failure was expressed in terms of residual or recurrent CIN2+, confirmed by histology during follow-up. Across studies with a prospective design, residual/recurrent disease was detected in 224 out of a total of 2930 women. The proportion of CIN2+, detected during follow-up, ranged from 4%[28] to 16%[29], with a pooled estimate of 7.5% (95% CI: 6.1-8.9%), as shown in Figure 17 .

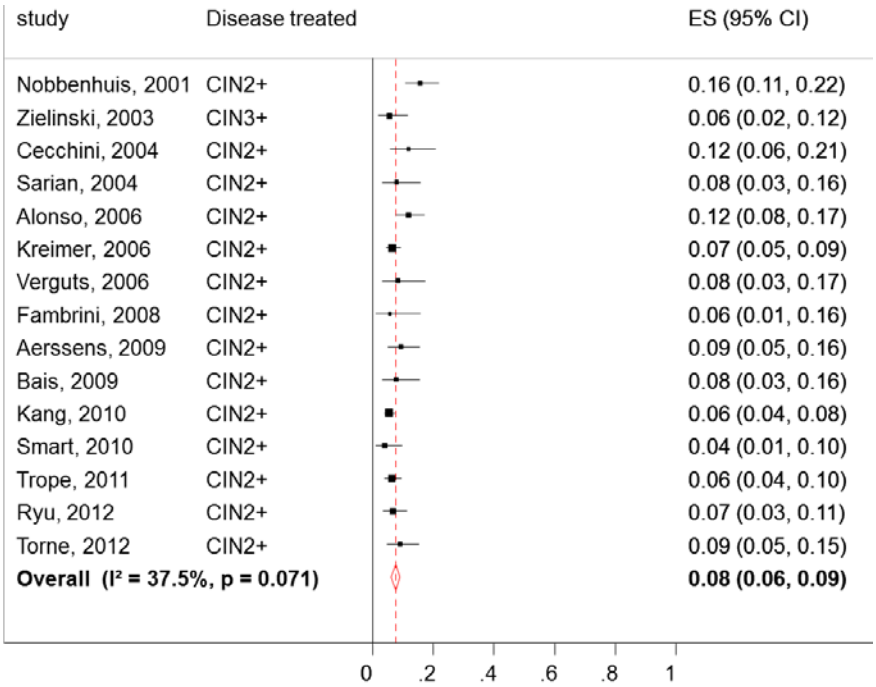


Figure 17: Meta-analysis of the occurrence of residual or recurrent CIN2+ after treatment for CIN2+ or CIN3+.

9.4.3. Absolute accuracy of hrHPV-DNA testing to predict occurrence of CIN2+ after treatment for precancer.

Overall, the pooled sensitivity and specificity of hrHPV-DNA testing to predict occurrence of CIN2+ was 94.3% (95% CI: 88.4-97.3%) and 80.0% (95% CI: 74.2-84.8%), respectively (Figure 18, Figure 19).

Among studies that used HC2[5;19;20;23-28], sensitivity ranged from 83.3% to 100.0%, and specificity ranged from 63.8% to 93. (Figure 19). The pooled sensitivity and specificity of HC2 to detect residual/recurrent CIN2+ was 95.6% (95% CI: 89-7-98.2%) and 81.9% (95% CI: 75.3-87.1%), respectively. Across all eight studies that used a PCR-based test[18;21;22;29-33], sensitivity and specificity ranged from 53.8% to 100.0% and 57.1% to 100.0%, respectively (Figure 19). The pooled sensitivity and specificity for PCR-based assays were 93.1% (95% CI=86.3-99.9) and 76.9% (95% CI=67.8-86.1%), respectively.

In the study of Aerssens et al.[32], the lowest sensitivity for hrHPV testing was observed although the analytically highly sensitive SPF10-PCR assay is used. Performing a sensitivity analysis by excluding this study, results in a sensitivity and specificity of PCR-based testing of 95.8% (95% CI=90.9-1.00%) and 76.5% (95% CI=65.8-87.1%), respectively.

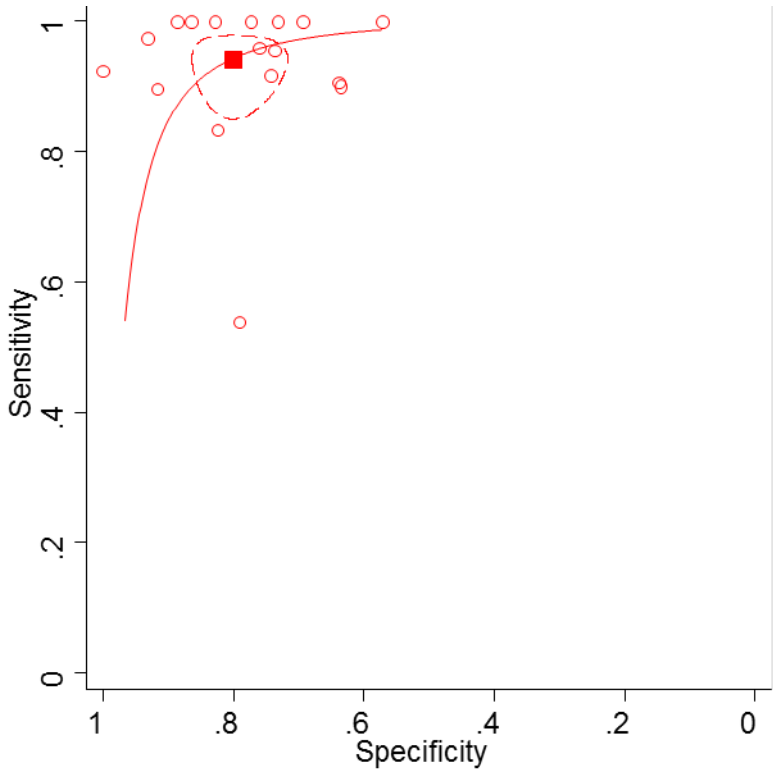


Figure 18: Specificity and sensitivity of hrHPV-DNA testing to predict occurrence of CIN2 or CIN3 after treatment of cervical precancer. Full line= summary ROC curve, Square= pooled summary measure, interrupted line is the95% confidence ellipse around the summary measure.

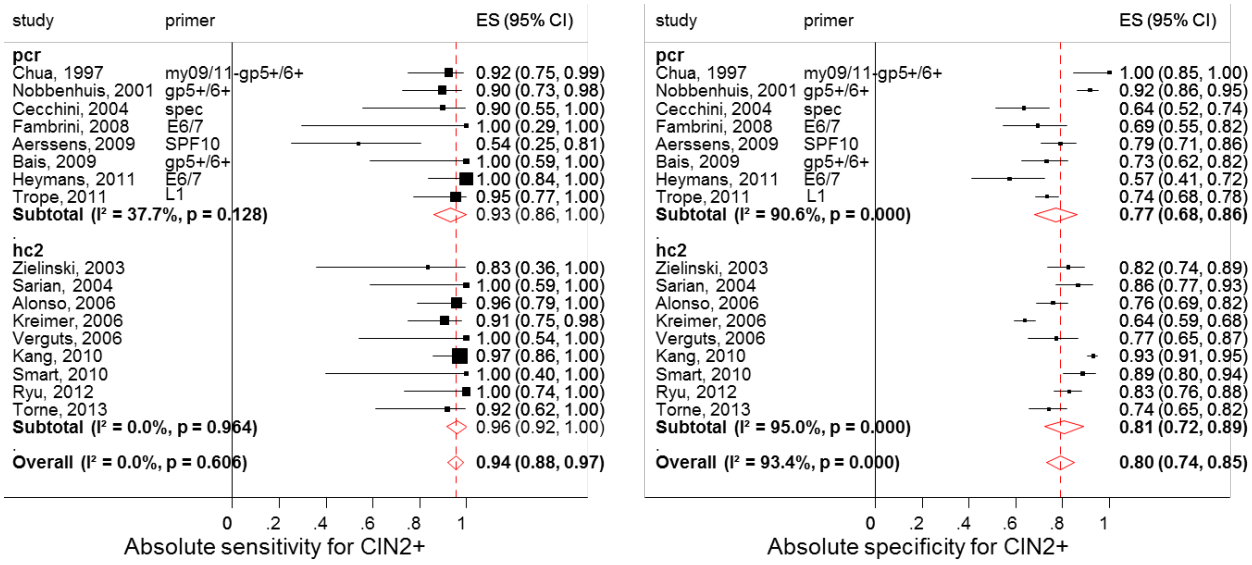


Figure 19: Study-specific and pooled absolute sensitivity (left) and specificity (right) of hrHPV-DNA testing by HC2 and PCR to predict occurrence of CIN2+ after treatment for CIN2+ or CIN3+.

9.4.4. Absolute accuracy of cytology follow-up to predict residual/recurrent CIN2+.

The absolute sensitivity and specificity to predict occurrence of CIN2+ after treatment for CIN2+ was 72.0% (95% CI: 65.6-77.5%; range: 50.0-100.0%) and 84.6% (95% CI: 80.7-87.9%; range: 64.3-91.8%), respectively (Figure 21, Figure 23).

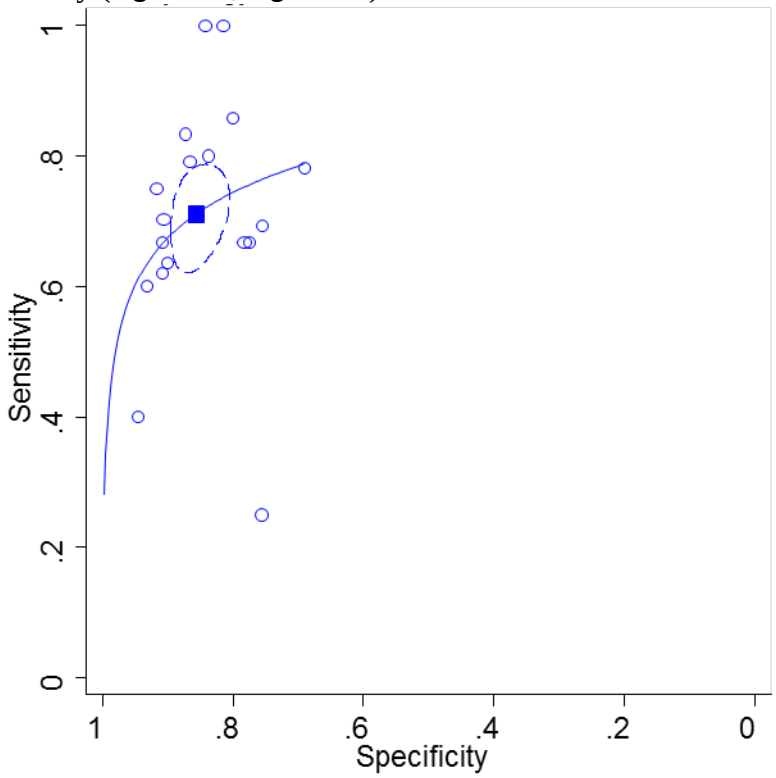


Figure 20: Specificity and sensitivity of cytology (cut-off: ASCUS) to predict occurrence of CIN2 or CIN3 after treatment of cervical precancer. Full line= summary ROC curve, Square= pooled summary measure, interrupted line is the95% confidence ellipse around the summary measure.

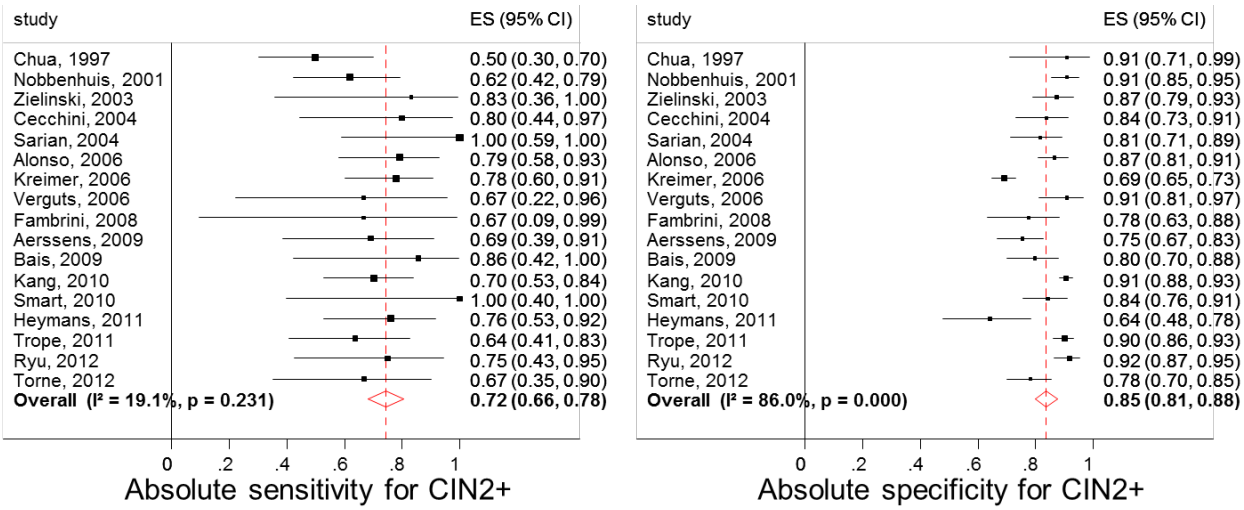


Figure 21: Study-specific and pooled absolute sensitivity (left) and specificity (right) of cytology (cut-off ASC-US) to predict occurrence of CIN2+ after treatment for CIN2+ or CIN3+.

9.4.5. Absolute accuracy of combined hrHPV-DNA and cytology testing to predict occurrence of CIN2+ after treatment for precancer

Ten studies contained accuracy data on combined testing with an HPV-DNA assay and cytology to detect residual/recurrent disease. Pooled values for sensitivity and specificity were 95.3% (95% CI: 88.1-98.2%) and 69.6% (95% CI: 61.7-76.5%), respectively.

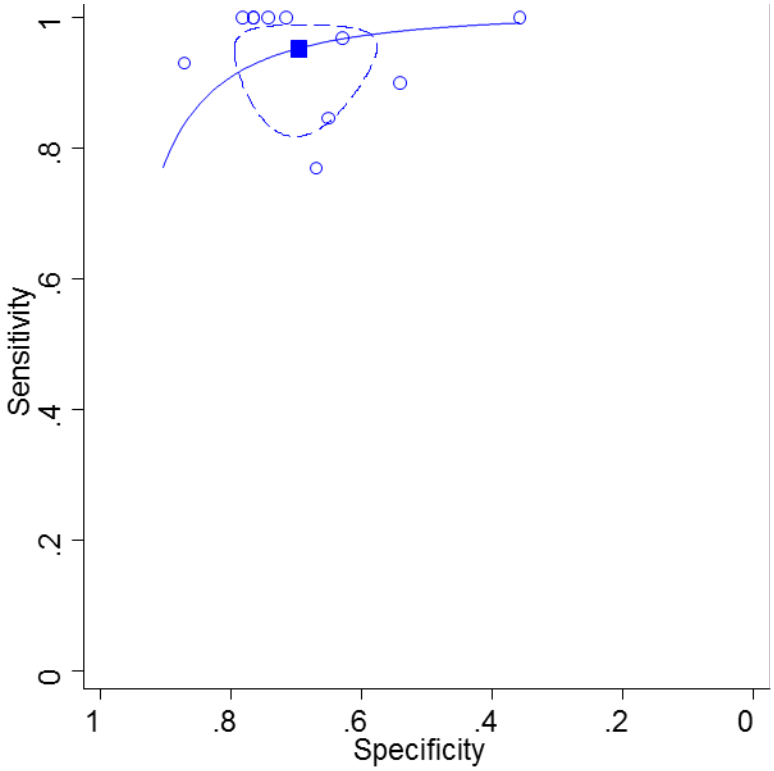


Figure 22: Specificity and sensitivity of combined testing with hrHPV-DNA testing and cytology (cut-off: ASCUS) to predict occurrence of CIN2+ after treatment of cervical precancer. Full line= summary ROC curve, Square= pooled summary measure, interrupted line is the95% confidence ellipse around the summary measure.

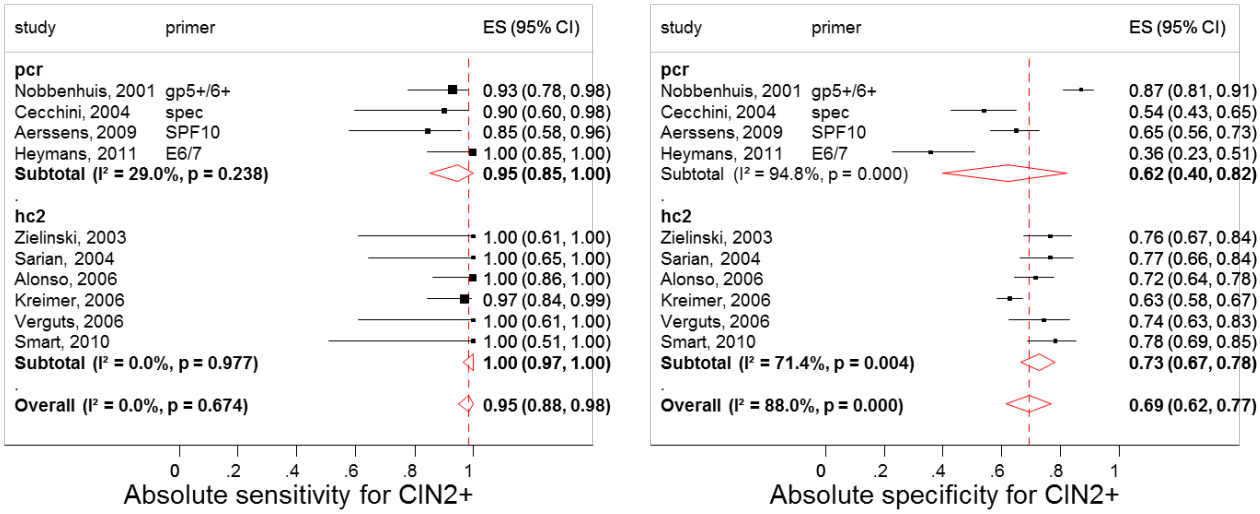


Figure 23: Study-specific and pooled absolute sensitivity (left) and specificity (right) of combined testing with an hrHPV-DNA assay and cytology (cut-off: ASCUS), to predict occurrence of CIN2+ after treatment for CIN2+ or CIN3+.

9.4.6. Relative accuracy of hrHPV-DNA testing versus cytology to predict occurrence of CIN2+ after treatment for precancer.

Compared to cytology, hrHPV-DNA testing at 3-9 months post-treatment was significantly more sensitive, but significantly less specific to predict residual/recurrent CIN2+. The pooled values for relative sensitivity and specificity were 1.29 (95% CI: 1.18-1.40) and 0.94 (95% CI: 0.90-0.99), respectively (Figure 76).

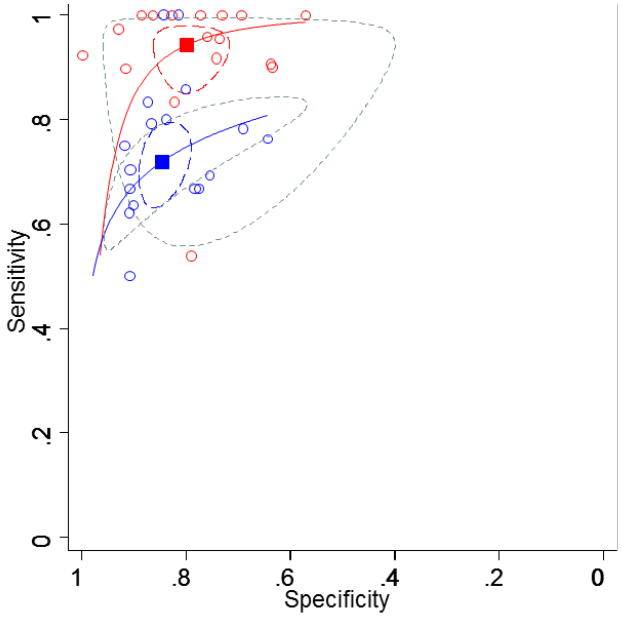


Figure 24: Specificity and sensitivity of hrHPV-DNA testing (red) and cytology at cut-off ASCUS (blue) to predict occurrence of CIN2+ after treatment of cervical precancer. Full line= summary ROC curve, Square= pooled summary measure, interrupted line is the 95% confidence ellipse around the summary measure.

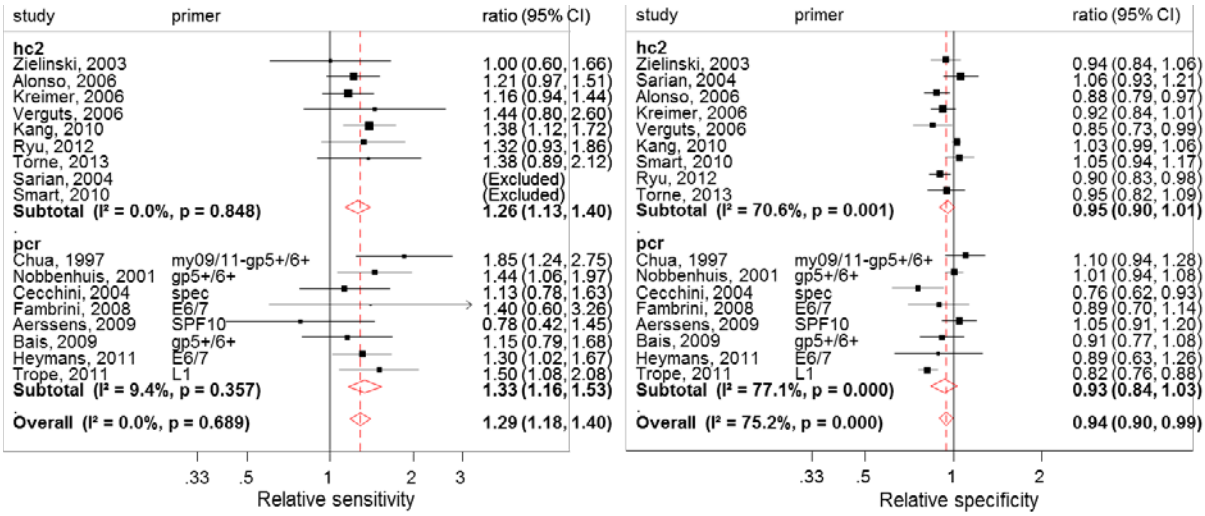


Figure 25: Study-specific and pooled relative sensitivity (left) and specificity (right) of hrHPV-DNA testing by HC2 and PCR versus cytology (cut-off: ASCUS), to predict occurrence of CIN2+ after treatment for CIN2+ or CIN3+.

9.4.7. Relative accuracy of combined hrHPV-DNA and cytology testing versus hrHPV-DNA testing alone to predict occurrence of CIN2+ after treatment for precancer.

Combined testing for hrHPV-DNA and cytology at 3-9 months after treatment was equally sensitive (ratio 1.07, 95% CI: 0.97-1.17) but less specific (ratio 0.93, 95% CI: 0.88-0.97), compared to HPV-testing alone (Figure 77).

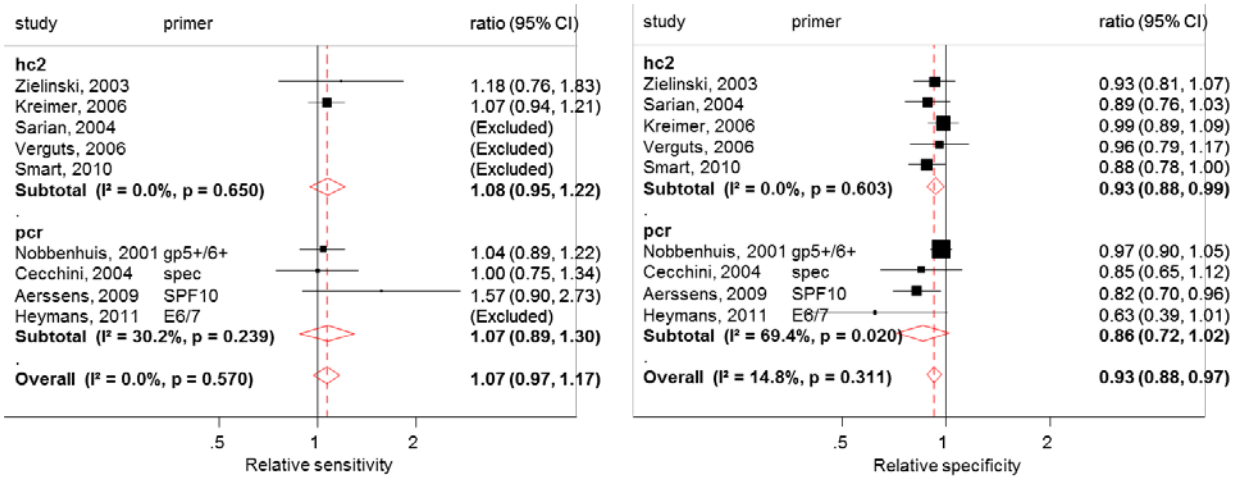


Figure 26: Study-specific and pooled relative sensitivity (left) and specificity (right) of combined testing with an hrHPV-DNA assay and cytology versus hrHPV-DNA testing alone, to predict occurrence of CIN2+ after treatment for CIN2+ or CIN3+.

9.4.8. Age as a risk factor for occurrence of disease after treatment

It has been suggested that higher age at time of treatment might be a risk factor for residual or recurrent disease after conservative treatment for high-grade CIN. However, conclusions are heterogeneous. In this report, a review of findings in literature is presented and discussed. Because different age cut-offs and definitions of disease were used in different studies and general study characteristics were too diverse, meta-analytical pooling could not be performed.

A link between age and risk of disease after treatment is supported by the study of Tropé and colleagues, who demonstrated that the mean age of women who had recurrent CIN2+ (n=22) after treatment was significantly higher than women who remained disease free during 18 months of follow-up (43.2y vs. 37.2y, p<0.001)[18]. Similarly, Verguts et al. reported a significant age difference among women with (n=6) and without CIN2+ during 24 months of follow-up after treatment (51.5v vs. 39.8y, p=0.007)[26]. However, it must be noted that in other studies, although a higher mean age in the group of women with residual or recurrent disease was observed, the difference was not significant[5;27;35].

Secondly, five studies were identified that contained data on the occurrence of disease after treatment, stratified by age[5;19;30;36-38]. The absolute numbers and relative risks are listed in Table 58.

Table 58: Relative risk of occurrence of CIN after treatment for precancer by age.

Study	Definition of disease	Age cut-off	Recurrence young age group (%)	Recurrence old age group (%)	RR (95% CI)
Prato 2008[38]	CIN1+	>35	2/60 (3.3)	10/55 (18.2)	5.45 (1.25-23.80)
Ang 2011[36]	CIN2+	>35	41/1145 (3.6)	16/413 (3.9)	1.08 (0.61-1.91)
Cecchini 2004[30]	CIN2+	>39	8/67 (11.9)	2/17 (11.8)	0.99 (0.23-4.22)
Alonso 2006[5]	CIN1+	>45	26/165 (15.8)	10/38 (26.3)	1.67 (0.88-3.16)
Flannelly 2001[37]	CIN1+	≥50	215/3205 (6.7)	56/221 (25.3)	3.78 (2.91-4.90)
Ryu 2012[19]	CIN2+	≥50	11/165 (6.7)	1/18 (5.6)	0.83 (0.11-6.09)

Valuable data is presented in the study of Flannelly and colleagues, which included 271 cases of CIN1+ and 152 cases of CIN2+ during 35 months of follow-up, among 3426 treated women. In this large study, the risk of post-treatment disease was almost four times higher in women of 50 years and older (relative risk: 3.78 [95% CI=2.91-4.90)] [37]. In the same way, Prato and colleagues showed a clear link between age and risk of post-treatment CIN in women older than 35 (relative risk: 5.45 [95% CI=1.25-23.80]). On the other hand, other studies failed to statistically confirm age as a risk factor for residual or recurrent disease.

Additional valuable data is presented in two long-term surveillance studies. In the study of Melnikow and colleagues, based on 3013 cases of CIN2/3 among 37142 women treated and up to 15 years of surveillance, the authors concluded that women of 40 years or older and those with more severe index disease had higher rates of CIN 2/3 after treatment[7]. In a large Swedish cohort study, comprising 3 148 222 woman years, a trend-analysis was performed, demonstrating a

significant link ($p<0.001$) between older age and risk of cervical cancer and this risk was observed to accelerate in women of 50 years or older[10;39]. Previous observations that link older age to a higher risk of viral persistence after treatment, further support the notion that age is a risk factor of post-treatment disease[40-42].

9.5. Discussion and interpretation

9.5.1. Accuracy of HPV-DNA testing and cytology to predict disease after treatment

The meta-analysis presented here, evaluates the accuracy of testing with a hrHPV-DNA assay and/or cytology for treatment failure in women who were treated for CIN2/3. The results confirm the significantly improved potential to detect treatment failure with hrHPV-DNA testing compared to cytology, as demonstrated by previous meta-analyses[4;12;43-46].
Women treated for CIN2/3 by excisional or ablative procedures have a risk of residual or recurrent CIN2 or worse disease. Therefore, surveillance of treated women is necessary to monitor this risk. The rate of residual or recurrent high-grade CIN within 18 or more post-treatment months, varied among the included studies from 4% to 16%, with a pooled average of 8%. This residual/recurrent disease was predicted by hrHPV-DNA testing with a considerably higher sensitivity (ratio 1.29, 95% CI 1.18-1.40), but a slightly lower specificity (ratio 0.94, 95% CI 0.90-0.99), compared to cytology. Combined testing for hrHPV-DNA and cytology, was slightly and not significantly more sensitive (ratio 1.07, 95% CI: 0.97-1.17) but resulted in a significant drop of the specificity (ratio 0.93, 95% CI0.88-0.98), compared to hrHPV-DNA testing alone.

The heterogeneity which was observed between the included studies, may be explained by their diverse study characteristics. Differences in notation of follow-up time were encountered and a mean, minimal, and maximal duration of follow-up could not always be extracted for each study. Furthermore, variations in time of testing after treatment were observed. For the analysis presented here, only data of tests that had been performed within the 3-9m post-treatment interval were included. Additionally, the age of the participants and the exact type of hrHPV-DNA test (HC2 vs. various PCR-based assays) and cytology test (conventional vs. liquid-based cytology) might account for variation. Nonetheless, the data presented here, demonstrate a largely significant increase in sensitivity when using hrHPV-DNA testing versus cytology. Furthermore, although inclusion criteria are not always identical, our conclusions confirm those of previous systematic reviews and meta-analyses regarding sensitivity. In the current meta-analysis pooled specificity of hrHPV-DNA testing was slightly but significantly lower than cytological surveillance. Previous meta-analyses have shown similar or non-significantly lower specificity of hrHPV-DNA testing compared to cytology (Arbyn2005 Gyn Oncol, Vaccine 2006-12). In this review, the relative specificity is significantly lower than unity. These findings do not constitute a change in conclusions. The statistical significance is mainly due to the accumulation of more studies, since the pooled specificity ratios from previous meta-analyses (0.94 in Arbyn GO 2005; 0.96 in Arbyn Vaccine 2006 and 0.97 in Arbyn Vaccine 2012) are only minimaly different from those in the current systematic review.

Co-testing provided a small but insignificant gain in sensitivity (7%) compared to hrHPV-DNA testing alone but resulted also in a loss of specificity (-7%, 95% CI -2 ot -12%). Given the setting (post-treatment follow-up) where higher priority may be given to sensitivity than specificity, clinicians may prefer cotesting rather that testing only with a hrHPV-DNA assay. TO BE DISCUSSED FURTHER WITH CLINICIANS THROUG THE GRADE DISCUSSION.

9.5.2. Timing of testing and duration of follow-up

Our meta-analysis demonstrated a considerable improvement in the detection of post-treatment disease when using a hrHPV-DNA assay or combined testing with hrHPV-DNA and cytology, compared to cytology alone. Additionaly, the number and timing of follow-up tests is an important issue to assess. It has been documented that one-time testing, even with an hrHPV-DNA assay, might not be sufficient to identify women at risk for residual or recurrent high-grade disease[47;48]. A study addressing the predictive value of consecutive testing for an outcome over 10 years, demonstrated that one negative test at 6 months after treatment resulted in a 10-year risk of CIN3+ of 2.1% (if hrHPV-negative), 2.8% (if cytology-negative) and 1.4% (if combined hrHPV- and cytology-negative)[48]. By adding one or two follow-up tests however, safety could be improved considerably. For example, three consecutive negative cytology results (6, 12 and 24 months) or two negative co-tests (hrHPV-negative and cytology-negative, at 6 months and 12 months) were associated with a low risk of residual CIN3+ of 0.7% and 0.0%, respectively, which was not different from the general population. On the other hand, one positive test at 6 months increased the risk of CIN3+ dramatically: 29%, 13% and 23%, if hrHPV-positive, cytology ASC-US+, or one of both tests positive, respectively[48].

Although a rapid decline in the occurrence of CIN2 and CIN3 is observed the first two years post-treatment (see Figure 1), long-term follow-up data demonstrate that the risk for invasive cancer remains elevated, even up to 20 years[7;39]. Therefore, long-term surveillance of women treated for precancer is justified. Supporting this statement is data in a large US study which demonstrated a lower rate of invasive cancers in women under active surveillance compared with women in the overall cohort of women treated for precancer[7].

9.6. References

(1) Moscicki AB, Schiffman M, Burchell A, Albero G, Giuliano AR, Goodman MT et al. Updating the natural history of human papillomavirus and anogenital cancers. *Vaccine* 2012; 30 Suppl 5:F24-F33.

(2) Paraskevaidis E, Arbyn M, Diakomanolis E, Martin-Hirsch P, Koliopoulos G, Makrydimas G et al. The role of HPV DNA testing in the follow-up period after treatment for CIN: a systematic review of the literature. *Cancer Treat Rev* 2004; 30(2):205-211.

(3) Lacey CJ. Therapy for genital human papillomavirus-related disease. *J Clin Virol* 2005; 32 Suppl 1:S82-S90.

(4) Arbyn M, Ronco G, Anttila A, Meijer CJLM, Poljak M, Ogilvie G et al. Evidence regarding HPV testing in secondary prevention of cervical cancer. *Vaccine* 2012; 30 Suppl 5:F88-F99.

(5) Alonso I, Torne A, Puig-Tintore LM, Esteve R, Quinto L, Campo E et al. Pre- and post-conization high-risk HPV testing predicts residual/recurrent disease in patients treated for CIN 2-3. *Gynecol Oncol* 2006; 103(2):631-636.

(6) Stern PL, van der Burg SH, Hampson IN, Broker TR, Fiander A, Lacey CJ et al. Therapy of human papillomavirus-related disease. *Vaccine* 2012; 30 Suppl 5:F71-F82.

(7) Melnikow J, McGahan C, Sawaya GF, Ehlen T, Coldman A. Cervical Intraepithelial Neoplasia Outcomes After Treatment: Long-term Follow-up From the British Columbia Cohort Study. *J Natl Cancer Inst* 2009; 110(10):721-728.

(8) Soutter WP, Sasieni P, Panoskaltsis T. Long-term risk of invasive cervical cancer after treatment of squamous cervical intraepithelial neoplasia. *Int J Cancer* 2005; 118(8):2048-2055.

(9) Kalliala I, Anttila A, Pukkala E, Nieminen P. Risk of cervical and other cancers after treatment of cervical intraepithelial neoplasia: retrospective cohort study. *BMJ* 2005; 331(7526):1183-1185.

(10) Strander B, Hallgren J, Sparen P. Effect of ageing on cervical or vaginal cancer in Swedish women previously treated for cervical intraepithelial neoplasia grade 3: population based cohort study of long term incidence and mortality. *BMJ* 2014; 348:f7361.

(11) Arbyn M, Kyrgiou M, Gondry J, Petry KU, Paraskevaidis E. Long term outcomes for women treated for cervical precancer. *BMJ* 2014; 348:f7700.

(12) Clinical applications of HPV testing: a summary of meta-analyses. 06 Apr 8; Barcelona: 2006.

(13) Harbord RM, Deeks JJ, Egger M, Whiting P, Sterne JA. A unification of models for meta-analysis of diagnostic accuracy studies. *Biostatistics* 2007; 8:239-251.

(14) Harbord RM, Whiting P. metandi: Meta-analysis of diagnostic accuracy using hierarchical logistic regression. *The Stata Journal* 2009; 9(2):211-229.

(15) Takwoingi Y, Diagnostic Test AccuracyWorking Group. METADAS: A SAS macro for meta-analysis of diagnostic accuracy studies. http://srdta cochrane org/sites/srdta cochrane org/files/uploads/MetaDAS%20Quick%20Reference%20v1 3%20May%202012 pdf [2009 :[-*]

(16) Diagnostic Test Accuracy Working Group. Handbook for Diagnostic Test Accuracy Reviews. Available from: http://srdta.cochrane.org/handbook-dta-reviews. 2011.

(17) Dersimonian R, Laird NM. Meta-analysis in clinical trials. *Controlled Clin Trials* 1986; 7:177-188.

(18) Trope A, Jonassen CM, Sjoborg KD, Nygard M, Dahl FA, Alfsen GC et al. Role of high-risk human papillomavirus (HPV) mRNA testing in the prediction of residual disease after conisation for high-grade cervical intraepithelial neoplasia. *Gynecol Oncol* 2011; 123(2):157-162.

(19) Ryu A, Nam K, Kwak J, Kim J, Jeon S. Early human papillomavirus testing predicts residual/ recurrent disease after LEEP. *J Gynecol Oncol* 2012; 23(4):217-225.

(20) Torne A, Fuste P, Rodriguez-Carunchio L, Alonso I, del PM, Nonell R et al. Intraoperative post-conisation human papillomavirus testing for early detection of treatment failure in patients with cervical intraepithelial neoplasia: a pilot study. *BJOG* 2013; 120(4):392-399.

(21) Chua KL, Hjerpe A. Human papillomavirus analysis as a prognostic marker following conization of the cervix uteri. *Gynecol Oncol* 1997; 66(1):108-113.

(22) Heymans J, Benoy IH, Poppe W, Depuydt CE. Type-specific HPV geno-typing improves detection of recurrent high-grade cervical neoplasia after conisation. *Int J Cancer* 2011; 129(4):903-909.

(23) Zielinski GD, Rozendaal L, Voorhorst FJ, Berkhof J, Snijders PJ, Risse EJ et al. HPV testing can reduce the number of follow-up visits in women treated for cervical intraepithelial neoplasia grade 3. *Gynecol Oncol* 2003; 91(1):67-73.

(24) Sarian LO, Derchain SF, Andrade LA, Tambascia J, Morais SS, Syrjanen KJ. HPV DNA test and Pap smear in detection of residual and recurrent disease following loop electrosurgical excision procedure of high-grade cervical intraepithelial neoplasia. *Gynecol Oncol* 2004; 94(1):181-186.

(25) Kreimer AR, Guido RS, Solomon D, Schiffman M, Wacholder S, Jeronimo J et al. Human papillomavirus testing following loop electrosurgical excision procedure identifies women at risk for posttreatment cervical intraepithelial neoplasia grade 2 or 3 disease. *Cancer Epidemiol Biomarkers Prev* 2006; 15(5):908-914.

(26) Verguts J, Bronselaer B, Donders G, Arbyn M, Van Eldere J, Drijkoningen M et al. Prediction of recurrence after treatment for high-grade cervical intraepithelial neoplasia: the role of human papillomavirus testing and age at conisation. *BJOG* 2006; 113(11):1303-1307.

(27) Kang WD, Jeong OM, Kim SM, Nam JH, Soo PC, Choi HS. Significance of human papillomavirus genotyping with high-grade cervical intraepithelial neoplasia treated by a loop electrosurgical excision procedure. *Am J Obstet Gynecol* 2010; 203(1):72-76.

(28) Smart OC, Sykes P, MacNab H, Jennings L. Testing for high risk human papilloma virus in the initial follow-up of women treated for high-grade squamous intraepithelial lesions. *Aust N Z J Obstet Gynaecol* 2010; 50(20):164-167.

(29) Nobbenhuis MA, Meijer CJ, van den Brule AJ, Rozendaal L, Voorhorst FJ, Risse EK et al. Addition of high-risk HPV testing improves the current guidelines on follow-up after treatment for cervical intraepithelial neoplasia. *Br J Cancer* 2001; 84(6):796-801.

(30) Cecchini S, Carozzi F, Confortini M, Zappa M, Ciatto S. Persistent human papilloma virus infection as an indicator of risk of recurrence of high-grade cervical intraepithelial neoplasia treated by the loop electrosurgical excision procedure. *Tumori* 2004; 90(2):225-228.

(31) Fambrini M, Penna C, Pieralli A, Bussani C, Fallani MG, Andersson KL et al. PCR detection rates of high risk human papillomavirus DNA in paired self-collected urine and cervical scrapes after laser CO2 conization for high-grade cervical intraepithelial neoplasia. *Gynecol Oncol* 2008.

(32) Aerssens A, Claeys P, Beerens E, Garcia A, Weyers S, Van Renterghem L et al. Prediction of recurrent disease by cytology and HPV testing after treatment of cervical intraepithelial neoplasia. *Cytopathology* 2009; 20(1):27-35.

(33) Bais AG, Eijkemans MJ, Rebolj M, Snijders PJ, Verheijen RH, van Ballegooijen M et al. Post-treatment CIN: randomised clinical trial using hrHPV testing for prediction of residual/recurrent disease. *Int J Cancer* 2009; 124(4):889-895.

(34) 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):529-536.

(35) Jeong NH, Lee NW, Kim HJ, Kim T, Lee KW. High-risk human papillomavirus testing for monitoring patients treated for high-grade cervical intraepithelial neoplasia. *J Obstet Gynaecol Res* 2009; 35(4):706-711.

(36) Ang C, Mukhopadhyay A, Burnley C, Faulkner K, Cross PA, Martin-Hirsch P et al. Histological recurrence and depth of loop treatment of the cervix in women of reproductive age: incomplete excision versus adverse pregnancy outcome. *BJOG* 2011; 118(6):685-692.

(37) Flannelly G, Bolger B, Fawzi H, De Lopes AB, Monaghan JM. Follow up after LLETZ: could schedules be modified according to risk of recurrence? *BJOG* 2001; 108(10):1025-1030.

(38) Prato B, Ghelardi A, Gadducci A, Marchetti I, Di Cristofano C, Di Coscio G et al. Correlation of recurrence rates and times with posttreatment human papillomavirus status in patients treated with loop electrosurgical excision procedure conization for cervical squamous intraepithelial lesions. *Int J Gynecol Cancer* 2008; 18(1):90-94.

(39) Strander B, Andersson-Ellstrom A, Milsom I, Sparen P. Long term risk of invasive cancer after treatment for cervical intraepithelial neoplasia grade 3: population based cohort study. *BMJ* 2007; 335(7629):1077.

(40) Castle PE, Schiffman MA, Herrero R, Hildesheim A, Rodriguez AC, Bratti MC et al. A prospective study of age trends in cervical human papillomavirus acquisition and persistence in Guanacaste, Costa Rica. *J Infect Dis* 2005; 191(11):1808-1816.

(41) Aerssens A, Claeys P, Garcia A, Sturtewagen Y, Velasquez R, Vanden Broeck D et al. Natural history and clearance of HPV after treatment of precancerous cervical lesions. *Histopathology* 2008; 52(3):381-386.

(42) Sarian LO, Derchain SF, Pitta DR, Morais SS, Rabelo-Santos SH. Factors associated with HPV persistence after treatment for high-grade cervical intra-epithelial neoplasia with large loop excision of the transformation zone (LLETZ). *J Clin Virol* 2004; 31(4):270-274.

(43) M.D.Anderson Cancer Centre (University of Texas), editor. Clinical utility of HPV DNA detection: triage of equivocal and low-grade cervical lesions, follow-up of women treated for high-grade CIN. An update of pooled evidence. 05 Mar 22; Houston: 2005.

(44) Zielinski GD, Bais AG, Helmerhorst TJ, Verheijen RH, De Schipper FA, Snijders PJ et al. HPV testing and monitoring of women after treatment of CIN 3: review of the literature and meta-analysis. *Obstet Gynecol Surv* 2004; 59(7):543-553.

(45) Chan BK, Melnikow J, Slee CA, Arellanes R, Sawaya GF. Posttreatment human papillomavirus testing for recurrent cervical intraepithelial neoplasia: a systematic review. *Am J Obstet Gynecol* 2009; 200(4):422.e1-422.e9.

(46) Kocken M, Uijterwaal MH, de Vries AL, Berkhof J, Ket JC, Helmerhorst TJ et al. High-risk human papillomavirus testing versus cytology in predicting post-treatment disease in women treated for high-grade cervical disease: A systematic review and meta-analysis. *Gynecol Oncol* 2012; 125(2):500-507.

(47) Strander B, Ryd W, Wallin KL, Warleby B, Zheng B, Milsom I et al. Does HPV-status 6-12 months after treatment of high grade dysplasia in the uterine cervix predict long term recurrence? *Eur J Cancer* 2007; 43(12):1849-1855.

(48) Kocken M, Helmerhorst TJ, Berkhof J, Louwers JA, Nobbenhuis MA, Bais AG et al. Risk of recurrent high-grade cervical intraepithelial neoplasia after successful treatment: a long-term multi-cohort study. *Lancet Oncol* 2011; 12(5):441-450.

9.7. GRADE-Profil

Authors: M. Arbyn, M. Jentschke,

GRADE has built on previous systems to create a highly structured, transparent, and informative system for rating quality of evidence(Guyatt *et al.,* 2008b).

Steps in evidence assessment for making guidelines
1) Formulate a question
2) Identify the PICO(S) components
3) Qualify outcomes as critical, important, not important

1) Questions
Is hrHPV-DNA testing more accurate than follow-up cytology to predict failure or success after treatment of CIN?

2) PICOS
P: Women treated for histologically confirmed CIN2/3 by an excisional (LLETZ, laser conisation) or ablative (cryotherapy, laser ablation, electrocoagulation) procedure.
I: hrHPV-DNA testing, three to nine months post-treatment.
C: Cytology, three to nine months post-treatment.
O: Occurrence of residual or recurrent CIN2/3+ after treatment.
Absolute sensitivity and specificity of hrHPV-DNA testing, cytology testing and co-testing (hrHPV-DNA and cytology) to detect residual/recurrent CIN2+ or CIN3, confirmed by histology[r]. Relative sensitivity and specificity of hrHPV-DNA testing versus cytology and of co-testing versus hrHPV-DNA testing alone.
S: Prospective studies and case-control studies:

3) Importance of outcomes

Outcome:
37. Reduction of mortality from cervical cancer, (quality-adjusted) life-years gained.
38. Reduction of morbidity due to cervical cancer: incidence of cancer (Ib+).
39. Reduction of incidence of cancer (including micro-invasive cancer).
40. Reduction of incidence of CIN3 or worse disease (CIN3+).
41. Increased detection rate of CIN3+ or CIN2+.
42. Increased test positivity with increased, similar or hardly reduced positive predictive value.

Increased detection rate of cervical disease after treatment for CIN.
The following outcome measures were assessed:
- treatment failure rate (occurence of residual/recurrent CIN2+)

[r] From cohort studies predictive values (PPV, NPV and cNPV) can be derived as well but these accuracy indicators will not be pooled for the current systematic review.

- absolute accuracy of hrHPV-DNA testing and cytology to detect CIN2+
- absolute accuracy of combined testing with hrHPV-DNA and cytology to detect CIN2+
- relative accuracy hrHPV-DNA testing versus cytology to detect CIN2+
- relative accuracy of combined testing with hrHPV-DNA and cytology to detect CIN2+

4) Quality of evidence for each outcome in four categories;

- High: ++++
- Moderate: +++
- Low: ++
- Very low: +

Three studies[18-20] were found eligible and were added to the list of studies that had been included in the previously published meta-analysis[4], resulting in a total of 17 studies. Overall, included studies were heterogeneous with regard to their study characteristics. A comprehensive summary of study and test-characteristics is listed in Table1 and Table2.
All included studies were prospective, except for two that had a case-control design[21;22]. High risk hrHPV-DNA testing was performed by the Hybrid Capture 2 assay (HC2) in ten studies[5;19;20;23-28], by a PCR method in eight studies[18;21;22;29-33]. Overall, follow-up data of 3041 women treated for CIN2+ or CIN3+ were included in the analysis.

5 factors that lower the quality of the evidence (Guyatt *et al.*, 2008a)

31. Risk of bias due to limitations in design or execution: see CONSORT, STROBE, QUADAS
32. Inconsistency or heterogeneity: if consistency unexplained, lower quality
33. Indirectness, applicability (relevance of studies for answering the PICPO question)
34. Imprecision: number of studies, width of CI
35. Reporting bias, publication bias.

3 factors that increase the quality

4. Large effect
5. Dose effect gradient
6. Confounders (presence of confounders that have lowered the observed effect, absence of these would have increased the effect) (Guyatt *et al.,* 2008a)

Items downgrading quality of evidence		Downgrading
Bias, design	All included studies were **prospective**, except for two that had a case-control design.	No (-0)
Inconsistency	Overall, included studies were heterogeneous with regard to their study characteristics.	Yes (-1)
Indirectness	No	No (-0)
Imprecision	Compared to cytology, hrHPV-DNA testing at 3-9 months post-treatment was significantly more sensitive, but significantly less specific to predict residual/recurrent CIN2+. The pooled values for relative sensitivity and specificity were 1.29 (95% CI: 1.18-1.40) and 0.94 (95% CI: 0.90-0.99), respectively. Combined testing for hrHPV-DNA and cytology at 3-9 months after treatment was equally sensitive (ratio 1.07, 95% CI: 0.97-1.17) but less specific (ratio 0.93, 95% CI: 0.88-0.97), compared to HPV-testing alone	No (-0)
Publication bias, other	No information	No (-0)
Items upgrading quality of evidence		
Large effect	The pooled values for relative sensitivity and specificity were 1.29 (95% CI: 1.18-1.40)	No (+0)
Dose-effect correlation	No	No (+0)
Confounding factors neutralising effects	No	No (+0)˚

Conclusion: evidence of low quality.

Overall Grade of the quality of evidence is assigned and this is based on the outcome with the lowest quality of evidence given that it is a critical outcome.

For the accuracy we consider the relative sensitivity (outcome CIN3+) and specificity (outcome CIN2+) as critical. The other outcomes: absolute accuracy,

relative sensitivity for CIN2+ and relative specificity for CIN3+ are considered as important.

Table 59 GRADE evidence profile

	Quality of evidence								
# studies (N)	Absence of study limitations	Consistency	Directness (outcome, representativity Germany)	Precision	Absence publication bias	Large effect	Dose-effect relation	Bias lowering effect	Comment
Outcome 1: Relative accuracy of hrHPV-DNA testing versus cytology to predict occurrence of CIN2+ after treatment for precancer. [CRITICAL]									
17	Yes	No	Yes	Yes	Yes	No	No	No	Low

Outcome 2: Relative accuracy of combined hrHPV-DNA and cytology testing versus hrHPV-DNA testing alone to predict occurrence of CIN2+ after treatment for precancer. [CRITICAL]									
17	Yes	No	Yes	Yes	Yes	No	No	No	Low

References

Guyatt G.H., Oxman A.D., Kunz R., Falck-Ytter Y., Vist G.E., Liberati A., & Schunemann H.J. (2008a) Going from evidence to recommendations. *BMJ* **336:** 1049-1051.

Guyatt G.H., Oxman A.D., Kunz R., Vist G.E., Falck-Ytter Y., & Schunemann H.J. (2008b) What is "quality of evidence" and why is it important to clinicians? *BMJ* **336:** 995-998.

Guyatt G.H., Oxman A.D., Vist G.E., Kunz R., Falck-Ytter Y., Alonso-Coello P., & Schunemann H.J. (2008c) GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *BMJ* **336:** 924-926.

Schunemann H.J., Oxman A.D., Brozek J., Glasziou P., Jaeschke R., Vist G.E., Williams J.W., Jr., Kunz R., Craig J., Montori V.M., Bossuyt P., & Guyatt G.H. (2008) GRADE: Grading quality of evidence and strength of recommendations for diagnostic tests and strategies. *BMJ* **336:** 1106-1110.