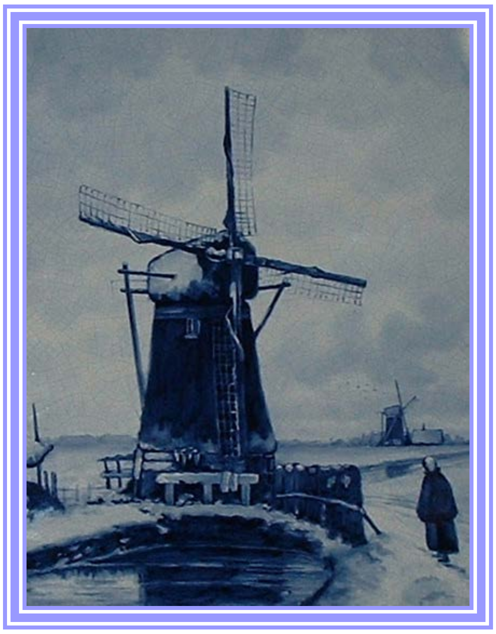
Evidence Review for the S3 guideline

“Prevention of Cervical Cancer”



Kleijnen Systematic Reviews Ltd

4 December 2014

The development of this evidence review was funded by the German Guideline Program in Oncology (GGPO)

Kleijnen Systematic Reviews Ltd
Unit 6, Escrick Business Park
Riccall Road
Escrick
York
YO19 6FD
United Kingdom

Richard Birnie
Robert Wolff
Diana Hilmer
Steven Duffy
Jos Kleijnen

Telephone: +44 (0)1904 727980
Fax: +44 (0)1904 720429
Email: jos@systematic-reviews.com
Website: www.systematic-reviews.com

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ACS	American Cancer Society
ASCCP	American Society for Colposcopy and Cervical Pathology
ASCP	American Society for Clinical Pathology
ASCUS	Atypical Squamous Cells of Undetermined Significance
AWMF	Association of the Scientific Medical Societies in Germany (German: Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften)
CDSR	Cochrane Database of Systematic Reviews
CENTRAL	Cochrane Central Register of Controlled Trials
CI	Confidence Interval
CIN	Cervical Intraepithelial Neoplasia
CIS	Carcinoma in situ
DARE	Database of Abstracts of Reviews of Effects
FDA	US Food and Drug Administration
GHQ	General Health Questionnaire
GIN	Guideline International Network
GRADE	Grading of Recommendations Assessment, Development and Evaluation
HCII	Hybrid Capture II
HPV	Human Papillomavirus
hrHPV	High Risk Human Papillomavirus
HR	Hazard Ratio
HSROC	Hierarchical Summary Receiver Operating Characteristic
HTA	Health Technology Assessment
INAHTA	International Network of Agencies for Health Technology Assessment
IQWiG	Institute for Quality and Efficiency in Health Care (German: Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen)
lrHPV	Low Risk Human Papillomavirus
LBC	Liquid Based Cytology
NGC	National Guidelines Clearinghouse
NICE	National Institute for Health and Care Excellence
NIHR	National Institute for Health Research
PCR	Polymerase Chain Reaction
RCT	Randomised Controlled Trial
RR	Relative Risk
SD	Standard Deviation
SE	Standard Error
SoF	Summary of Findings
SROC	Summary Receiver Operating Characteristic
UK	United Kingdom
USA	United States of America
USPSTF	US Preventative Services Task Force

1. EXECUTIVE SUMMARY

1.1 BACKGROUND

The incidence of cervical cancer was estimated at 530,000 cases worldwide in 2008. Incidence ranges from <6 per 100,000 in North America, Western Asia and Australia up to >30 per 100,000 in parts of Africa. Persistent infection with certain types of Human Papillomavirus (HPV) known as high risk HPV (hrHPV) is now believed to be a major causal factor in the development of cervical cancer.

Screening programs for cervical cancer are now well established in many parts of the developed world, particularly in Europe and the USA. Previously, screening has been based on cytological testing to look for cellular abnormalities, commonly known as the smear test or Pap test. More recently, there has been a rapid increase in the development of testing systems to detect the presence of hrHPV DNA.

1.2 PURPOSE

To prepare a systematic review of the clinical utility of including HPV testing in population screening for cervical cancer.

1.3 METHODS

This review followed the guidance published by the Centre for Reviews and Dissemination and the Cochrane Collaboration.

Comprehensive searches were undertaken to identify randomised controlled trials (RCT) and observational studies of cervical cancer screening methods in four databases (Medline, Medline In-Process, Embase, CENTRAL). Searches for systematic reviews were also performed in 11 databases (CDSR, DARE, HTA, NIHR, TRIP, INAHTA, AWMF, IQWiG, NICE Guidance, GIN, NGC).

Randomised trials and observational studies were included if they compared any HPV test alone or in combination with cytology against cytology alone in women >20 years old undergoing primary cervical cancer screening. Positive screening test results must have been confirmed by a combination of colposcopy and histology. Studies must have reported at least one outcome of interest (Overall survival, Disease Specific Survival, Incidence of Cervical Cancer, Incidence of CIN3, Incidence of CIN3+, Incidence of CIN2+ or screening related harm).

The quality of evidence was assessed using the Cochrane Risk of Bias Tool for individual studies. The collective evidence base for each outcome was assessed using the GRADE system of quality assessment for guideline development. Meta-analysis for each outcome was performed using the Mantel-Haenszel method with a random effects model. Statistical heterogeneity was quantified using the I^2 statistic. Where pooling of data was not appropriate a narrative summary was reported.

1.4 RESULTS

Inclusion screening identified six relevant randomised controlled trials. One of these trials was reported as two separate studies. The majority of these were conducted in developed countries (UK, Finland, Italy, Netherlands, and Sweden) with the exception of one study which was conducted in rural India. In total, 462,096 women were included across all studies with the number of participants per study ranging from 12,527 to 203,425. The age of participants ranged from 20 to 65 years although only one study included participants <25 years old. The interval between screening rounds was either three years or five years with all studies reporting either one or two screening rounds.

In light of the absence of data from RCTs on the effects of age or screening intervals the review was extended to include controlled observational studies for these specific questions. There were no relevant controlled observational studies identified.

There were no studies which reported overall survival, only the single study conducted in rural India reported disease specific survival as an outcome. The aim of this study was to assess the clinical effectiveness of a single lifetime test for cervical cancer by either HPV testing alone or cervical cytology. This study showed that a single lifetime HPV test significantly reduced the risk of death from cervical cancer compared to a single cytology test (RR 0.59, 95%CI 0.39 to 0.91).

The evidence showed that HPV testing + cytology detected more cases of CIN3+ in the first screening round based on 6 studies compared to cytology alone although the difference was not statistically significant (RR 1.23, 95%CI 0.91 to 1.67). At the second screening round cytology alone detected significantly more cases of CIN3+ than the HPV containing regime (RR 0.52, 95%CI 0.35 to 0.76).

Similar results were observed for the outcomes CIN3 and CIN2+. The HPV containing method detected more cases than cytology alone for both outcomes at the first screening round. At the second screening round the result was reversed, cytology alone detected more cases than HPV testing + cytology for both outcomes.

Incidence of invasive cervical cancer was reported in five studies. These studies showed a reduction in relative risk of cervical cancer for participants screened with an HPV containing regime compared to cytology alone in the first screening round (RR 0.89, 95%CI 0.45 to 1.75). The difference was non-significant and the quality of evidence was very low.

1.5 DISCUSSION

These results indicate that HPV based screening may provide better protection against cervical cancer than cytology alone through improved detection of premalignant disease in the first screening round prior to progression. This is supported by the reduced detection of CIN3 in the second screening round. In the second screening round HPV testing detects only new incident cases that have arisen since the first round whereas cytology testing detects both new incident cases and those cases that were not detected in the first round but have since progressed.

1.6 CONCLUSIONS

The combined use of HPV testing + cytology resulted in fewer participants diagnosed with cervical cancer compared to cytology alone as a result of increased detection of earlier, premalignant stages of disease, however, this effect was not statistically significant.

2. BACKGROUND

Cervical cancer is currently the third most common cancer in women with an estimated incidence of 530,000 cases worldwide in 2008. Incidence ranges from <6 per 100,000 in North America, Western Asia and Australia up to >30 per 100,000 in parts of Africa[1]. Previously identified risk factors for cervical cancer include smoking, oral contraceptive use >5 years, diagnosis with HIV/AIDS, organ transplant and having a first degree relative with cervical cancer[2]. More recently, a strong causal relationship has been shown between cervical cancer and persistent infection with certain types of human papilloma virus (HPV)[3]. This is now believed to account for the largest proportion of the risk of developing cervical cancer.

There are over 100 different types of HPV which are broadly divided into high risk (hrHPV) and low risk types (lrHPV). High risk types are those where persistent infection is associated with an increased risk of developing cervical cancer[2]. The prevalence of HPV infection has been shown to vary widely between different geographic regions and between countries within those regions. Prevalence also varies according to age with the peak prevalence occurring at approximately 25 years old then declining with increasing age. Although the prevalence of HPV infection varies geographically this age distribution is conserved around the developed world. This geographic variation in HPV prevalence is likely to have a significant impact on the accuracy of HPV DNA testing methods when applied in different countries[4].

Screening programs for cervical cancer are now established in many parts of the world, particularly in developed countries in Europe and the USA. Conventionally, screening has been based on cytological testing, commonly known as the smear test or Pap test. In conventional cytology, cellular material is sampled from the cervix using a spatula or brush and smeared directly onto a glass slide for evaluation by a cytologist. Liquid-based cytology (LBC) is a commonly used alternative whereby the sampled material is deposited in a preservative solution and transferred to a laboratory where the slide is prepared for evaluation[5].

The use of LBC has been reported to improve the so-called adequacy of the sample; i.e. the proportion of samples that can be successfully evaluated by the cytologist, and to reduce the sample interpretation time. The relative performance of the different techniques in terms of patient relevant outcomes such as cancer incidence and mortality is still the subject of much investigation[5].

There has been a rapid increase in HPV testing systems in the last few years. A recent review identified 125 distinct tests[6]. These tests can be broadly divided into five main categories:

1. hrHPV tests
2. hrHPV DNA tests with partial genotyping
3. HPV DNA full genotyping tests

4. HPV DNA type- or –group specific genotyping tests
5. hrHPV E6/E7 mRNA tests.

hrHPV DNA tests detect high risk oncogenic HPV types in aggregate and do not allow the distinction of individual HPV types. hrHPV tests with partial genotyping detect high risk oncogenic HPV types in aggregate but also permit identification of the most common oncogenic types, typically HPV16 and HPV18. HPV DNA full genotyping tests allow the distinction of the 12 major hrHPV types. HPV DNA type and group specific tests allow the identification of only a limited subset of the major hrHPV types. hrHPV mRNA tests detect transcripts of the viral oncogenes E6 and E7. Only a small proportion of the available tests have ever been rigorously validated. Currently there are nine HPV tests of varying types that are either approved by the US FDA or are considered to be clinically validated according to published guidelines[6, 7]. Hybrid Capture (HCII) is the longest established and most widely used system for HPV detection. The HCII system targets 13 high risk HPV types and five low risk types, however, only the high risk types are widely tested in practice.

Current European guidelines recommend that women should be invited for cervical cancer screening every 3-5 years between the ages of 25 and 65[8]. There is significant variation in how these guidelines are implemented by individual countries with some countries initiating screening as early as 15 while others continue up to age 69. In Europe, the majority of countries adhere to the recommended screening interval of 3-5 years, however, in some countries (Germany, Austria, and Luxembourg) screening is carried out every year. The UK uses different screening intervals for different age groups where screening is recommended every three years for women aged 25-49 but every five years for women aged 50-64 years old[5].

3. OBJECTIVES OF THE PROJECT

The objective of this project is to provide a systematic review of the evidence on the clinical utility of including HPV testing alone or in combination with cytology in population screening for cervical cancer in order to facilitate the development of an S3 guideline on this topic.

4. RESEARCH QUESTIONS

1) What is the clinical effectiveness of HPV testing (alone or in combination with cytology), compared to cytology alone, in population screening for cervical cancer?
   a) At what age should cervical cancer screening (using HPV testing +/- cytology) start and stop?
   b) What is the optimal HPV screening interval?

5. METHODS

5.1 LITERATURE SEARCHES

Searches for evidence syntheses

Searches for evidence syntheses were conducted to identify systematic reviews, health technology assessments, guidelines and guidance. The following resources were searched with no date limits:

- Cochrane Database of Systematic Reviews (CDSR) (Wiley)
- Database of Abstracts of Reviews of Effects (DARE) (Wiley)
- Health Technology Assessment database (HTA) (Wiley)
- NIHR Health Technology Assessment Programme (Internet)
  http://www.hta.ac.uk/
- TRIP (Internet)
  http://www.tripdatabase.com/
- International Network of Agencies for Health Technology Assessment (INAHTA) (Internet)
  http://www.inahta.org/
- AWMF (Internet)
  http://www.awmf.org/
- IQWiG (Internet)
  https://www.iqwig.de/
- NICE Guidance (Internet)
  http://guidance.nice.org.uk/
- Guideline International Network (GIN) (Internet)
  http://www.g-i-n.net/
- National Guidelines Clearinghouse (NGC) (Internet)
  http://www.guideline.gov/

Structured literature searching

Comprehensive searches were undertaken to identify randomised controlled trials (RCT) of cervical cancer screening methods.

The search strategies (keywords) were developed specifically for each database and a variety of synonyms for HPV and cervical cancer were utilised. Specific current validated filters for randomised controlled trials were used for Medline and Embase. Only studies conducted in humans were sought and no date limits were applied. The Embase search strategy can be found in Appendix 1.

The following databases were searched with no date or language limit:

- Medline (OvidSP)
- Medline In-Process Citations & Daily Update (OvidSP)
- Embase (OvidSP)
- Cochrane Central Register of Controlled Trials (CENTRAL) (Wiley)

The initial searches for RCTs did not identify any relevant information about screening age or screening intervals, so the searches were widened to include observational studies. The strategies used to search for observational studies included some minor amendments to the original Medline and Cochrane Library search strategies for RCTs. This enabled an update of the original RCT searches alongside the searches for observational studies. The updated Embase search strategy can be found in Appendix 1.

Reference checking

The bibliographies of identified relevant research and review articles were checked for studies.

Handling of citations

Identified references were downloaded into Endnote software (version X7, Thomson Reuters, USA) for further assessment and handling. Rigorous records were maintained as part of the searching process. Individual records within the Endnote reference libraries were tagged with searching information, such as searcher, date searched, database host, database searched, strategy name and iteration, theme or search question. This enables the origin of each individual database record and its progress through the screening and review process to be tracked.

Quality assurance within the search process

The main Embase strategy for each set of searches was independently peer reviewed by a second Information Specialist, using the CADTH checklist[9].

5.2 INCLUSION CRITERIA

Screening was carried out based on the inclusion criteria outlined in Table 1. The selection of HPV tests to include was based on those tests identified as US FDA-approved or clinically validated in a recent review of HPV tests[6]. Composite outcomes were included if the study reported data for the composite outcomes or if each of the component outcomes was reported, e.g. the composite outcome CIN3+ included both CIN3 and invasive cervical cancer. Similarly, the composite outcome CIN2+ included the components CIN2, CIN3 and invasive cervical cancer.

Table 1: Inclusion criteria

Question	**What is the clinical effectiveness of HPV testing (alone or in combination with cytology), compared to cytology alone, in population screening for cervical cancer?**
Participants:	Women older than 20 years undergoing HPV testing as part of primary cervical cancer screening
Interventions (index test):	Any test for the detection of HPV DNA or RNA; e.g. HCII, PCR assay, EIA kit HPV GP HR, Cervista HPV HR Test, CareHPV Test, Cervista HPV 16/18 Test, Abbott RT hrHPV, Cobas-4800, Papillocheck, qPCR (E6/E7), APTIMA Any combination of a cytology test (Pap test or LBC) with a direct test for HPV as above
Comparators:	Any cytology test (Pap test or LBC) used in the absence of an HPV test
Reference standard:	For screening test positives: Combination of colposcopy and histology. If colposcopy was normal a histology result was not required to confirm the absence of disease. If colposcopy was abnormal then histology was used to confirm the diagnosis of disease. For screening test negatives: No diagnosis of CIN3 or worse within 3 years or no diagnosis of cervical cancer within 5 years
Outcomes:	*Primary:* Overall mortality, mortality from cervical cancer, incidence of cervical cancer, harm directly or indirectly resulting from screening (e.g. psychological distress, quality of life) *Secondary:* Incidence of cervical intraepithelial dysplasia grade III (CIN3) Incidence of CIN3 or worse disease (CIN3+). Incidence of CIN2 or worse disease (CIN2+)
Study design[‡]:	Randomised Clinical or population-based trials[†] Controlled Observational studies All studies of any design must have a minimum follow-up of at least 12 months

‡ Inclusion was not restricted by language.

† Analysis was limited to the best available evidence. Lower quality study designs (e.g. Observational Cohort studies, Case-Control studies) were included only if insufficient higher quality studies were available.

5.3 STUDY SELECTION

Titles and abstracts identified through electronic database and web searching were independently screened by two reviewers. During this initial phase of the screening process any references which obviously do not meet the inclusion criteria listed previously were excluded. Full paper copies were obtained for all of the remaining references. These were then independently examined in detail by two reviewers in order to determine whether they met the criteria for inclusion in the review. All studies excluded at this second stage of the screening process were documented along with the reasons for exclusion. With respect to both screening stages, any discrepancies between reviewers were resolved through discussion or the intervention of a third reviewer.

5.4 DATA EXTRACTION

The guidelines formulated by the Association of the Scientific Medical Societies in Germany (AWMF) for the preparation of scientific assessments by external evaluators were implemented[10]. Data was extracted by one reviewer using standardised data extraction forms and checked for errors against the original study report by a second reviewer. Any discrepancies were resolved through discussion or through the intervention of a third reviewer.

Studies have been identified by the main study name/identifier. Where this is not available the surname and year of the first author of the main report/publication has been used. To avoid the duplication of data where studies (or study populations) have multiple publications the most recent and complete report has been used as the main reference, but additional details were extracted from the other publications as necessary.

For each study the following general types of information/data have been recorded:

- Endnote ID
- Study ID or name (if reported or otherwise surname of first author)
- Study group (if reported)
- Year of publication
- Other related publications
- Study country(ies)
- Study design/details
- Screening test(s)

Examples of specific details include:

- Type of study
- Sample size
- Location/setting
- Participant demographics
- Screening test methods compared – description of index test, comparators, reference test and associated parameters, e.g. positivity thresholds, screening

interval
- Outcomes assessed (e.g. definition of outcome, when assessed (follow-up), who assessed, methods used to assess outcome(s))
- Results (e.g. numbers, percentages and effect sizes with confidence intervals (where relevant))

5.5 ASSESSMENT OF METHODOLOGICAL QUALITY

The quality of each individual study was assessed in order to ensure that the conclusions and findings of this review are based on the best available evidence and that any potential sources of bias in the data are identified. Quality assessment was undertaken by one reviewer and checked by a second reviewer, any disagreements were resolved by consensus or discussion with a third reviewer. The methodological quality of included RCTs was assessed using the Cochrane Risk of Bias Tool[11]. The collective evidence base was evaluated based on the GRADE system of quality assessment for guideline development[12]. GRADE rates the quality of a complete body of evidence for a specific outcome in a specific population. Quality of evidence was assessed for risk of bias, publication bias, imprecision, inconsistency, indirectness, magnitude of effect, dose-response gradient and the effects of any confounding.

- Risk of bias describes any limitations in the design and execution of a collection of studies, for example failure to properly randomise the participants, failure to blind participants and investigators or selective reporting of outcomes.
- Publication bias is a measure of the degree to which the available published data are skewed by selective publication of trials dependent on their results, e.g. positive trials are more likely to be published than those with negative results.
- Imprecision assesses the degree to which random error influences the interpretation of the results.
- Inconsistency captures the degree of heterogeneity between studies in terms of their PICO elements, i.e. how comparable are the studies to each other.
- The remaining GRADE criteria can be used to rate up the quality of evidence if there is a very large effect of intervention, if there is evidence of a dose response or if the effects of any confounding would reduce rather than increase any observed effects.

Each of the GRADE criteria is described in detail in a series of papers published by the GRADE working group[12].

5.6 ANALYSIS

Where a formal meta-analysis was considered unsuitable for some or all of the data identified (e.g. due to the heterogeneity and/or small numbers of studies), then we employed a narrative synthesis method. This involved the use of narrative text and tables to summarise data in order to allow the reader to consider outcomes in the light of differences in study designs and potential sources of bias for each of the studies being reviewed. This involved organising the studies by (as appropriate) intervention, population, or outcomes assessed, summarising the results of the studies, summarising the range and size of the associations these studies report, and describing the most important characteristics of the included studies. A detailed commentary on the major methodological problems or biases that affected the studies is included, together with a description of how this has affected the individual study results.

For outcomes where sufficient studies assessing similar interventions were available then a formal meta-analysis was carried out. All meta-analyses were carried out using the Mantel-Haenszel method with a random effects model. Statistical analyses were performed using the following software: RevMan (version 5)[13] and STATA (version 10, StataCorp, USA).

Clinical effectiveness studies

Estimates of comparative clinical effectiveness were based on direct, within study comparisons. For studies with multiple screening rounds the data from the individual rounds were analysed separately. The following quantitative methods were used:

Dichotomous data were analysed by calculating the relative risk (RR) for each trial and the corresponding 95% confidence intervals.

We anticipated that systematic differences between studies (heterogeneity) were likely. Therefore, the random-effects model was used for the calculation of relative risks or weighted mean differences. Heterogeneity was initially assessed by measuring the degree of inconsistency in the studies' results (I^2)[14]. This measure (I^2) describes the percentage of total variation across studies that is due to inter-study heterogeneity rather than the play of chance. The value of I^2 lies between 0% and 100%, and a simplified categorization of heterogeneity could be low, moderate, and high for I^2 values of 25%, 50%, and 75%.

Where sufficient data were available, clinically relevant subgroup analysis was considered. In particular, we used this approach to explore possible modifying effects of the following pre-specified factors:

- Age (e.g. 20-24, 25-65, >65).
- Screening interval (e.g. <3 years, 3-5 years, >5 years)
- Methodological factors (e.g. Test positivity threshold, study risk of bias)
- Hysterectomy
- Cytology testing method (e.g. conventional cytology, LBC, alternative LBC protocols)
- HPV testing method (e.g. HCII, PCR assay, EIA kit HPV GP HR, Cervista HPV HR Test, CareHPV Test, Cervista HPV 16/18 Test, Abbott RT hrHPV, Cobas-4800, Papillocheck, qPCR (E6/E7), APTIMA)
- Order of tests
- HPV vaccination status

If available data allowed we also considered sensitivity analysis for the effects of the following variables:

- HPV prevalence
- Cervical cancer prevalence
- Developed *versus* developing countries[15]
- Other Risk factors – Women with one or more risk factors other than HPV infection (e.g. Smoking, Oral Contraceptive use >5 years, diagnosis with HIV/AIDS, organ transplant, other immunosuppression, first degree relative with cervical cancer)

Comparative diagnostic accuracy

We planned an analysis of diagnostic accuracy only if there was insufficient data to address the question of clinical effectiveness. Positivity thresholds would have been extracted from primary studies whenever possible. For the HCII HPV test the FDA-approved positivity threshold of >1pg/ml would have been applied. For tests or combinations of tests with no established positivity threshold a consensus threshold would have been identified from the primary studies if possible. Combination tests using both cytology and HPV testing would have been stratified into four risk groups for the development of cervical cancer and the results compared between groups:

1. HPV + / cytology + → high risk,
2. HPV + / cytology - → intermediate risk,
3. HPV - / cytology + → low risk,
4. HPV - / cytology - → very low risk[16].

If available data allowed, summary estimates of the sensitivity and specificity together with 95% confidence intervals (CIs) and prediction regions of HPV tests and cytological tests, used singly or in combination would have been calculated. In addition we would have used the bivariate/hierarchical summary receiver operating characteristic (HSROC) random effects model to generate summary estimates and an SROC curve[17, 18]. Estimates of relative effectiveness would have been derived from direct, within study comparisons. Depending on the availability of suitable data we planned to consider subgroup analysis of the following variables:

- Cytology testing method (e.g. conventional cytology, LBC, alternative LBC protocols)
- HPV testing method (e.g. HCII, PCR)
- PCR primer set for PCR based HPV tests
- DNA sample preparation method (e.g. Dedicated sample, excess sample from LBC)

5.7 AMENDMENTS TO PROTOCOL

After inclusion screening we considered that there were sufficient RCT data to adequately address the clinical effectiveness question. Therefore, the question of diagnostic test accuracy was not included in the data extraction or analysis.

It is recommended that formal meta-analysis should not be carried out if the included studies have high heterogeneity. Typically, I^2 >50% represents substantial heterogeneity while an I^2 >75% represents high heterogeneity[19]. In analyses where heterogeneity was moderate or high we considered that it is preferable to provide the results of a meta-analysis but to downgrade the quality of evidence in the GRADE assessment. This is based on the principle of providing the guideline group with the ‘best available’ evidence to support decision making even if the quality of that evidence is very low. We implemented the following rule in order to facilitate this in the GRADE assessment: if I^2 >50% rate inconsistency as serious (-1), if I^2 >75% rate inconsistency as very serious (-2).

We performed sensitivity analyses for including the results of the second round of the POBASCAM study (see Table 12 and Figure 8).

6. RESULTS

6.1 LITERATURE SEARCHING AND INCLUSION ASSESSMENT

The initial literature search for primary studies (Medline, Medline in process, Embase, CENTRAL; searched 18 October 2013) yielded in 1,401 references. After de-duplication, 983 references were available for screening of titles and abstracts (see Figure 1).

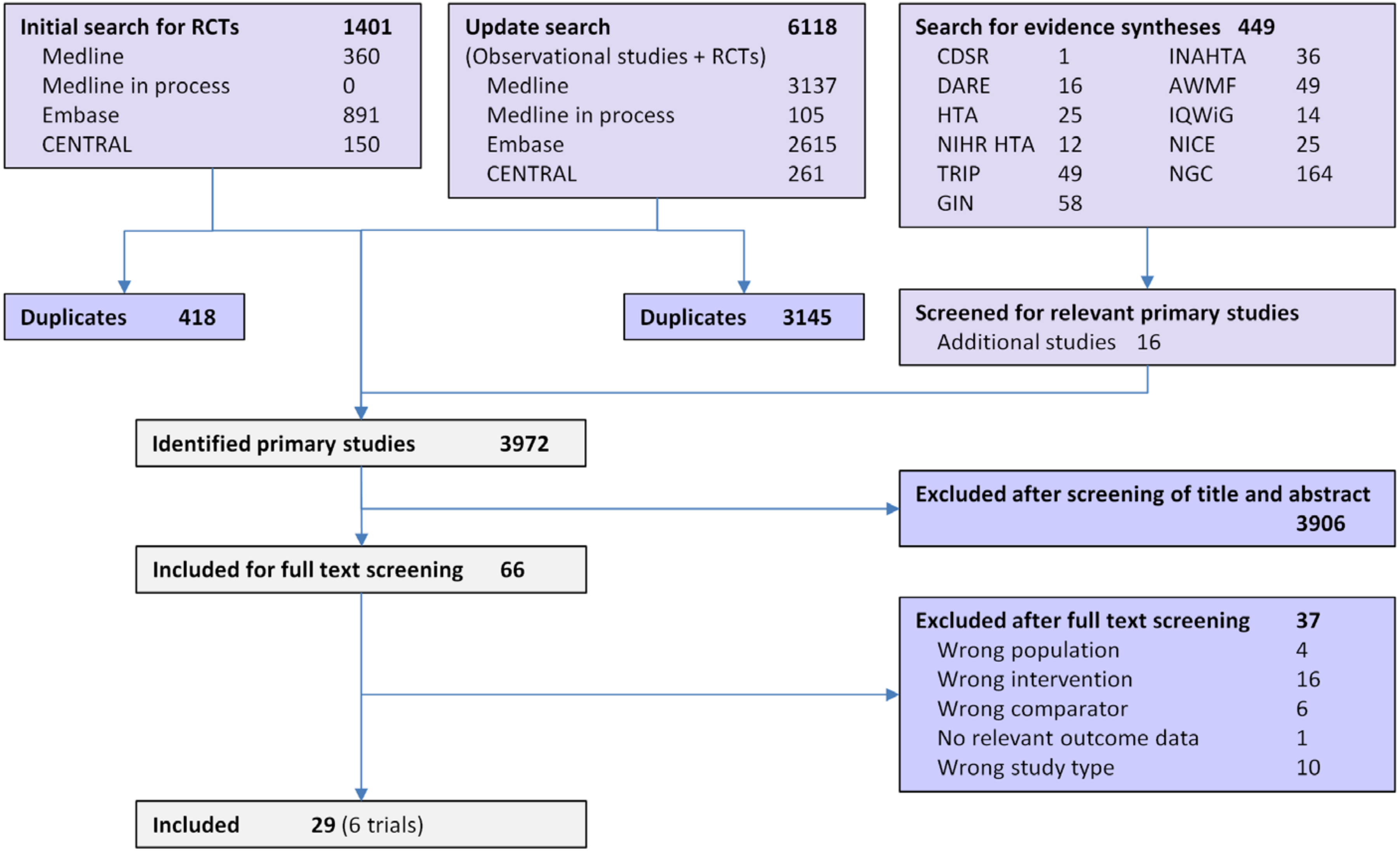
Furthermore, searches were undertaken to identify relevant systematic reviews, technology appraisals, guidelines and guidance (CDSR, DARE, HTA, NIHR HTA, TRIP, INAHTA, AWMF, IQWiG, NICE, GIN, NGC, searched between 21 and 23 October 2013). These searches retrieved a total number of 449 hits. In order to identify further studies which may provide information on screening intervals or the age at which to start and stop, reference lists of relevant evidence syntheses were screened. Overall, 16 studies were identified (see Figure 1).

The majority of established cervical cancer screening programmes focus on women aged 25-65 years old. One of the objectives of this review was to consider the issue of whether screening should start earlier or finish later. In order to address this we searched for studies that reported data in participants 20-24, 25-65 or >65 years. Although several studies reported age stratified outcomes (ARTISTIC, NTCC-I, NTCC-II, Leinonen 2012, Sankaranarayanan 2009) the specific age groups varied between studies and these groups did not correspond to those specified in the review protocol. The ARTISTIC study was the only one that included participants <25 years old.

As the initial searches for RCTs identified only one study that reported data relevant to the age groups specified in the protocol and no studies that included alternative screening intervals, the searches were widened to include observational studies (searches conducted 13 January 2014). As detailed in section 5.1, these searches also updated the initial literature searches for RCTs. The searches retrieved 6,118 records; after removal of duplicates there were 2,973 records remaining (see Figure 1).

Titles and abstracts of 3,972 references were screened and 47 potentially relevant papers ordered as full texts. Of these, 29 (six trials) were included (see Figure 1). Details on included studies are presented in Table 2 while details on the 37 excluded studies are given in Table 3.

Figure 1: Flow chart of study searches and inclusion



6.2 OVERVIEW OF INCLUDED STUDIES

Characteristics of included studies

Inclusion screening identified six relevant randomised controlled trials. A summary of the screening regimes, outcomes and demographic characteristics of the included studies is reported in Table 2.

The majority of studies were conducted in developed countries (UK, Finland, Italy, Netherlands, and Sweden) with the exception of Sankaranarayanan 2009 which was conducted in rural India. In total, 462,096 participants were included across all studies. The number of participants per study ranged from 12,527 in the Swedescreen study to 203,425 in the study by Leinonen 2012. The age of participants ranged from 20 to 65 across all studies although the majority of studies only initiated screening at age 25 years. The Swedescreen study focused on a restricted group of participants aged 32-38 based on the argument that age-specific incidence of cervical cancer peaks at around 40 years of age, therefore, screening should be most effective when performed in women aged 30-40.

Tests assessed in the included studies

The majority of included trials (five out of six) compared some combination of HPV testing plus cytology against cytology alone. The NTCC study included two separate recruitment phases and applied different index tests in the two phases. Phase I compared HPV testing + cytology against cytology only whereas phase II compared HPV testing only against cytology. The study authors considered the two phases to be statistically homogeneous between phases prior to pooling data from the two phases and reported that no heterogeneity was observed. The testing method was not explicitly reported, however, it is well known that tests for heterogeneity are frequently underpowered and substantial heterogeneity may still exist even in the absence of a significant result[19]. It is also necessary to consider clinical heterogeneity as distinct from statistical heterogeneity. For the purposes of this review the two phases of the NTCC study were considered to be clinically heterogeneous on the grounds that the index tests differed significantly and were therefore included as separate studies (referred to as NTCC-I and NTCC-II). This increased the total number of studies from six to seven. HPV testing in combination with cytology was compared with cytology alone in four out of these seven studies (ARTISTIC, NTCC-I, POBASCAM and Swedescreen). Leinonen 2012 compared HPV testing with cytology triage of positive cases against cytology alone. NTCC-II and Sankaranarayanan 2009 compared HPV testing alone against cytology alone. The Sankaranarayanan 2009 study had a different objective to the other included studies. This trial aimed to compare HPV testing versus cytology in the context of a single lifetime cervical cancer test in a rural Indian population. The other studies aimed to compare the relative effectiveness of different screening methods in the context of an ongoing national screening program. HPV testing was performed using the Hybrid Capture II system in five out of seven studies, the two remaining studies used the GP5+/6+ PCR system. Liquid based cytology was used in three out of seven studies whereas conventional cytology was used in the remaining four studies.

Screening processes in the included studies

The interval between screening rounds was either three years (four studies: ARTISTIC, NTCC-I, NTCC-II, Swedescreen) or five years (one study: POBASCAM) for those studies with two screening rounds. The follow-up period for those studies with a single round was five years (Leinonen 2012) or eight years (Sankaranarayanan 2009). Data were extracted and analysed separately for each screening round in those studies with two screening rounds. In all studies with two screening rounds the second round included only those women who were screen test negative in the first round with those who were screen test positive being referred for further follow up.

Although the POBASCAM study reported two screening rounds only data from the first round were extracted and analysed in this report. In the second screening round both study groups received HPV testing + cytology as a combined test which effectively merged the two groups into a single group. The main objective of this review was to compare the effectiveness of HPV testing alone or in combination with cytology relative to cytology alone. In the absence of a cytology alone group the second round of the POBASCAM trial cannot provide data that are relevant to this objective.

Outcomes assessed in the included studies

Disease specific survival was only reported in a single study by Sankaranarayanan 2009. Screening related harm was also reported in only one study (ARTISTIC) which looked at differences in General Health Questionnaire scores between study groups. Incidence of invasive cervical cancer was reported in five out of seven studies. Incidence of CIN3 was reported in five out of seven studies. Incidence of CIN3+ and incidence of CIN2+ were both reported in all seven studies.

Table 2: Characteristics of included studies

Study	Population	Index Test[$]	Comparator	Reference	Outcomes	Related Publications[*]
ARTISTIC Interval: 3 years Rounds: 2 Study start: 2001 Study end: 2007	40052 women Age: 20-64 Country: UK	HPV + Cytology HPV test: HCII HPV threshold: >1 RLU Cytology: ThinPrep LBC Cytology Threshold: Borderline Dyskaryoisis	Cytology ThinPrep LBC Cytology Threshold: Borderline Dyskaryoisis	Colposcopy + Histology	Incidence CIN3+ Incidence CIN2+ Screening related Harm (General Health Questionnaire, GHQ)	Kitchener 2006[20] Kitchener 2008[21] Kitchener 2009a[22] *Kitchener 2009b[23]* Sargent 2010[24] Kitchener 2011[25]
Leinonen 2012 Interval: 5 years Rounds: 1 Study start: 2003 Study end: 2007	203425 women Age: 25-65 Country: Finland	HPV with cytology triage HPV test: HCII HPV threshold: >1 RLU Cytology: Conventional Cytology Threshold: Pap II or ASCUS	Cytology Conventional Cytology Threshold: Pap II or ASCUS	Colposcopy + Histology	Incidence of Invasive Cervical Cancer Incidence CIN3 Incidence CIN3+ Incidence CIN2+	Antilla 2010[26] Kotaniemi-Talonen 2005[27] Kotaniemi-Talonen 2008[28] Leinonen 2009[29] *Leinonen 2012[30]* Malila 2013[31]
NTCC-I Interval: 3 years Rounds: 2 Study start: Feb-2002 Study end: Nov-2008	45774 women Age: 25-60 Country: Italy	HPV + Cytology HPV test: HCII HPV threshold: >1 RLU Cytology: ThinPrep LBC Cytology Threshold: ASCUS	Cytology ThinPrep LBC Cytology Threshold: ASCUS	Colposcopy + Histology	Incidence of Invasive Cervical Cancer Incidence CIN3 Incidence CIN3+ Incidence CIN2+	Giorgi-Rossi 2007[32] Ronco 2006a[33] Ronco 2006b[34] Ronco 2007a[35] Ronco 2007b[36] Ronco 2008[37] *Ronco 2010[38]*
NTCC-II Interval: 3 years Rounds: 2 Study start: Feb-2002 Study end: Nov-2008	49196 women Age: 25-60 Country: Italy	HPV only HPV test: HCII HPV threshold: >1 RLU	Cytology ThinPrep LBC Cytology Threshold: ASCUS	Colposcopy + Histology	Incidence of Invasive Cervical Cancer Incidence CIN3 Incidence CIN3+ Incidence CIN2+	

Study	Population	Index Test[$]	Comparator	Reference	Outcomes	Related Publications[*]
POBASCAM Interval: 5 years Rounds: 2[‡] Study start: 1999 Study end: 2005	44938 women Age: 30-57 Country: Netherlands	HPV + Cytology HPV test: GP5+/6+ PCR HPV threshold: Not Reported Cytology: Conventional Cytology Threshold: ≥Moderate Dyskaryosis	Cytology Conventional Cytology Threshold: ≥Moderate Dyskaryosis	Colposcopy + Histology	Incidence of Invasive Cervical Cancer Incidence CIN3 Incidence CIN3+ Incidence CIN2+	Budenholzer 2012[39] Bulkmans 2004[40] Bulkmans 2006[41] Bulkmans 2007[42] *Rijkaart 2012*[43]
Sankaranarayanan 2009 Interval: 7 years Rounds: 1 Study start: Jan-2000 Study end: Dec-2007	66184 women Age: 30-59 Country: India	HPV only HPV test: HCII HPV threshold: >1RLU	Cytology Conventional Threshold: ASCUS	Colposcopy + Histology	Disease Specific Survival Incidence of Invasive Cervical Cancer Incidence CIN3 Incidence CIN3+ Incidence CIN2+	Sankaranarayanan 2005[44] *Sankaranarayanan 2009*[45]
Swedescreen Interval: 3 years Rounds: 2 Study start: May-1997 Study end: Aug-2005	12527 women Age: 32-38 Country: Sweden	HPV + Cytology HPV test: GP5+/6+ PCR HPV threshold: Not Reported Cytology: Conventional Cytology Threshold: ASCUS	Cytology Conventional Cytology Threshold: ASCUS	Colposcopy + Histology	Incidence CIN3+ Incidence CIN2+	Elfgren 2005[46], *Naucler 2007*[47]

$ HPV threshold >1 RLU = relative light units. Equivalent to 1pg/ml of HPV DNA

* Publications highlighted in italic indicate the main trial report

‡ Only 1 screening round provided data suitable for inclusion in analysis. In the second screening round both groups received HPV + Cytology as a combined test.

6.3 OVERVIEW OF EXCLUDED STUDIES

A further seven RCTs (10 publications) were excluded after full text screening as they did not meet the specified inclusion criteria. One study was not randomised. Three studies were excluded due to inappropriate interventions or comparators. Two studies were excluded because the follow up was <12 months. One study was not conducted in a primary screening population, i.e. all participants had a diagnosis of ASCUS or worse at the time of enrolment (Table 3).

We were unable to identify any RCTs which were relevant to the sub-questions, detailed in section 4. Therefore, we searched the reference lists of relevant systematic reviews and guidelines to identify further studies which may provide information on screening intervals or the age at which to start and stop[48-51]. This identified 16 studies, all of which were subsequently excluded as cytology was the only screening method in these studies (Table 3)[52-67].

For this reason we performed a further search to identify controlled observational studies. There were 11 potentially relevant studies identified from title and abstract screening. Full text screening showed that none of these studies were relevant to the sub-questions of this review. There were six studies excluded due the absence of a comparator group. Two studies were excluded because they did not report relevant outcome data. Three studies were excluded because they were not conducted in a primary screening population (Table 3).

Table 3: Excluded studies

Study	Country	Reason for Exclusion	Related publications*
RCT			
SHENCCAST I	China	Wrong study type (Not an RCT)	*Wu 2010*[68]
SHENCCAST II	China	Wrong comparator (Comparison of two different HPV tests. No cytology component)	*Belinson 2011*[69]
CCCaST	Canada	Wrong comparator (Intervention/Comparator: both arms received HPV + Cytology but in a different order. Outcomes: Diagnostic Accuracy measures only; clinical effectiveness not included)	Mayrand 2006[70], *Mayrand 2007*[71]
Cheng 2012	China	Wrong population (Selected population. Women with ASCUS at enrolment)	*Cheng 2012*[72]
FOCAL	Canada	Wrong comparator (Comparator – Control arm received LBC cytology with HPV triage for ASCUS or above)	Ogilvie 2010[73], *Ogilvie 2012*[74], Van Neikerk 2012[75]
MARCH	Mexico	Wrong study type (Follow up <12 months)	*Lazcano-Ponce 2011*[76]
Sancho-Garnier 2013	France	Wrong study type (Follow up <12 months)	*Sancho-Garnier 2013*[77]
Observational			
Bailon Munoz 2009	Multiple	Wrong study type (Uncontrolled. No comparator group or useful outcome data)	*Bailon Munoz 2009*[78]
Carter 2010	NR	Wrong population (Selected population – Women with ASCUS at baseline)	*Carter 2010*[79]
Cuzick 2008	UK	Wrong study type (Uncontrolled. No comparator group)	*Cuzick 2008*[80]
de Vries 2012	USA	Wrong study type (Uncontrolled. No comparator group)	*De Vries 2012*[81]
Grainge 2005	UK	Wrong study type (Case-Control. Both groups received the same screening tests. No relevant outcome data)	*Grainge 2005*[82]
Gyllensten 2012	Sweden	Wrong study type (Uncontrolled. No comparator group)	*Gyllensten 2012*[83]
Herbert 2007	UK	No relevant outcome data	*Herbert 2007*[84]
Inoue 2010	Japan	Wrong population (Selected Population. All patients received cytology. Follow up varied by cytology result. Not all participants received HPV test)	*Inoue 2010*[85]
Kjaer 2006	Denmark	Wrong population (Selected population. Women who were cytology negative hrHPV positive at enrolment)	*Kjaer 2006*[86]
Kjaer 2010	Denmark	Wrong study type (Uncontrolled. No comparator group and no relevant outcome data)	*Kjaer 2010*[87]
Lee 2013	Korea	Wrong study type (Uncontrolled. No comparator group)	*Lee 2013*[88]

Study	Country	Reason for Exclusion	Related publications*
Primary studies identified after checking evidence syntheses (see section 6.3)			
Andrae 2008	Sweden	Wrong intervention (Cytology was the only screening method)	*Andrae 2008*[52]
Herbert 1996	UK	Wrong intervention (Cytology was the only screening method)	*Herbert 1996*[53]
Herrero 1992	Multiple	Wrong intervention (Cytology was the only screening method)	*Herrero 1992*[54]
Hoffmann 2003	South Africa	Wrong intervention (Cytology was the only screening method)	*Hoffmann 2003*[55]
Jiménez-Perez 1999	Multiple	Wrong intervention (Cytology was the only screening method)	*Jiménez-Perez 1999*[56]
Kasinpila 2011	Thailand	Wrong intervention (Cytology was the only screening method)	*Kasinpila 2011*[57]
La Vecchia 1984	Italy	Wrong intervention (Cytology was the only screening method)	*La Vecchia 1984*[58]
Makino 1995	Japan	Wrong intervention (Cytology was the only screening method)	*Makino 1995*[59]
Miller 2003	USA	Wrong intervention (Cytology was the only screening method)	*Miller 2003*[60]
Rebolj 2009	Multiple	Wrong intervention (Cytology was the only screening method)	*Rebolj 2009*[61]
Sasieni 1996	UK	Wrong intervention (Cytology was the only screening method)	*Sasieni 1996*[62]
Sasieni 2003	UK	Wrong intervention (Cytology was the only screening method)	*Sasieni 2003*[63]
Sasieni 2009a	UK	Wrong intervention (Cytology was the only screening method)	*Sasieni 2009a*[64]
Sasieni 2009b	UK	Wrong intervention (Cytology was the only screening method)	*Sasieni 2009b*[65]
Yang 2008	Australia	Wrong intervention (Cytology was the only screening method)	*Yang 2008*[66]
Zappa 2004	Multiple	Wrong intervention (Cytology was the only screening method)	*Zappa 2004*[67]

* Publications highlighted in italic indicate the main trial report

6.4 QUALITY OF EVIDENCE

Risk of bias assessment for individual studies

Quality assessments for all included studies are summarised in Table 4. Cochrane Risk of Bias tables reporting the justifications for each assessment are provided in Appendix 2. Overall, the quality of randomisation was good although allocation concealment was often poorly reported making it difficult to assess the risk of bias. Five studies were assessed as high risk of bias for blinding of patients and study personnel. This was typically because the clinical management of the patients differed according to which test results were available therefore patients and personnel could not be blinded. Five studies had concerns over missing outcome data (ARTISTIC, Leinonen 2012, Sankaranarayanan 2009, NTCC and Swedescreen). Although all five used the intention to treat principle in their analysis a substantial fraction of participants, up to a third in some cases, did not attend screening or had missing data during follow up. The proportion of missing data was similar between groups within each of the five trials. In all five cases there was no imputation or adjustment for missing data explicitly reported in the analysis.

Table 4: Overview of risk of bias assessment

Study	RISK OF BIAS ITEMS						
	Randomisation	Allocation concealment	Patient/ personnel blinding	Outcome assessor blinding	Incomplete outcome data	Selective outcome reporting	Other
ARTISTIC	☺	☺	☹	?	☹	☺	☺
Leinonen 2012	☺	?	☹	☹	?	☺	☹
Sankaranarayanan 2009	☺	?	☹	☺	?	☺	☹
NTCC	?	?	☹	☹	?	☺	☺
POBASCAM	☺	☺	☹	?	☺	☺	☺
Swedescreen	☺	☺	?	☺	?	☺	☺
☺ = low risk of bias, ☹ = high risk of bias, ? = unclear of bias							

GRADE summary of findings

Tables 5-10 show the GRADE Summary of Findings (SoF) table for each of the outcomes: disease specific survival, incidence of cervical cancer, incidence of CIN3, incidence of CIN3+, incidence of CIN2+ and screening related harm. GRADE evidence profiles are also provided in Appendix 3. Each row in the SoF table represents a meta-analysis of the studies identified in the comments column along with the GRADE evaluation of that analysis. Heterogeneity was moderate-high in some analyses as indicated in the footnotes of the respective GRADE tables. The relative risk estimates for these analyses should therefore be interpreted with caution. The detailed results of each meta-analysis including relative risk estimates for the individual studies are presented in Appendix 4. Each analysis may be interpreted as follows:

- **Outcome** – Analyses are described as either 'Outcome-All Studies' or 'Outcome-Subgroup-Group name'. Analyses marked 'All Studies' include all studies which reported that outcome. Analyses marked 'Subgroup' explore the effect of different elements of the screening process on the reported relative effect estimate. In some cases the outcome field contains multiple subgroup names; e.g. Incidence of Cervical Cancer - Subgroup - LBC comparator, three year interval. This indicates that these subgroups contained the same studies and could not be analysed separately. The subgroups for which there was sufficient data to perform the analysis for at least one outcome were:
    - Screening interval – three year interval versus ≥4 year interval
    - Screening test method – HPV + Cytology combined test versus HPV testing only. The effect of excluding Leinonen 2012 was also investigated as this study used HPV testing with cytology triage for participants with positive results rather than as a combined test.
    - Cytology method – Comparing the effect of using different cytology methods as the comparator test. Liquid based cytology versus conventional cytology.
- **Illustrative comparative risks** – The risk of the reported outcome in the control and intervention groups respectively.
- **Relative effect** – Reports the risk ratio and 95% confidence intervals for the intervention group relative to control.
- **No. of participants** – The total number of participants included in the analysis and the number of studies from which they were pooled. For studies that included multiple screening rounds the results of each round were included in the analysis separately, therefore, the number of studies reported in this column may not equal the number of studies reported in the Comments column.
- **Quality of Evidence** – Summary of the GRADE quality of evidence assessment. The final grade is indicated in bold type on the scale: High-Moderate-Low-Very Low. The reasons for the reported grade are indicated by the footnotes.
- **Comments** – Shows the studies that were pooled for this analysis

Some pre-planned subgroup/sensitivity analyses could not be carried out due to insufficient data. Specifically: the importance of hysterectomy, order of tests, HPV test positivity threshold, HPV vaccination status and other risk factors could not be investigated as these were not reported in any of the included studies. Study risk of bias could not be investigated as all included studies were rated as either unclear or high risk of bias. Some subgroup analyses were modified based on the available data. For the analysis of screening interval we were unable to assess the effect of a screening interval <3 years due to an absence of data, however, we were able to compare the effects of a three year interval versus an interval of ≥4 years for the first screening round.

Overall survival was not reported in any of the included studies. Summary of Findings tables are provided for both disease specific survival (Table 5) and screening related harm (Table 10). However, it should be emphasised that these outcomes were only reported in a single study each. GRADE quality assessment has limited utility in this context. Inconsistency, imprecision and publication bias in particular cannot be meaningfully assessed for a single study.

Sankaranarayanan 2009 showed that a single HPV test significantly reduced the relative risk of death from cervical cancer (RR 0.59, 95%CI 0.39 to 0.91) compared to a single cytology test. In the group which received a single HPV test, 34/34126 (0.1%) participants died of cervical cancer compared to 54/32058 (0.17%) participants in the group which received a single cytology test (Figure 2).

The ARTISTIC study reported screening related harm in terms of the proportion of participants with a General Health Questionnaire (GHQ-28) score ≥4[21]. GHQ-28 measures generalized psychological distress. The aim of this analysis was to investigate the increased psychological distress associated with receiving an HPV test compared to cytology only. This study showed that there was almost no difference in the proportion of participants with GHQ≥4 between participants screened by HPV testing in combination with cytology compared to those screened by cytology only (RR 0.98, 95%CI 0.87 to 1.11). In the group which received the HPV + Cytology combined test 37.6% of participants (223/593) had GHQ≥4 compared to 38.3% of participants (717/1872) in the cytology only group.

In studies where first and second round data were available separately the two screening rounds were analysed independently. Only women who were screen test negative at the first round were included in the second round of these studies. The number of participants diagnosed with cervical cancer was lower when screening with an HPV based method compared to cytology alone. This observation was consistent across both screening rounds, however the effect was not significant in either round. It was notable that the results of the different screening rounds favoured different tests for the outcomes incidence of CIN3, incidence of CIN3+, incidence of CIN2+. In each case, evidence from the first screening round showed that the HPV containing regimen detected more cases of CIN3, CIN3+ or CIN2+ respectively. In contrast, data from the second screening round showed that cytology detected more cases than the HPV containing regimen for the same outcomes (Tables 7-9 and Appendix 4: Figures 4-6). This observation was consistent across all subgroups analysed for these three outcomes. The magnitude of the effect varied for different outcomes and the effects were not necessarily statistically significant.

Table 5: GRADE summary of findings – disease specific survival

HPV test compared to conventional cytology - disease specific survival for cervical cancer

Patient or population: patients with cervical cancer
Settings:
Intervention: HPV test
Comparison: Conventional Cytology - Disease specific survival

Outcomes	Illustrative comparative risks* (95% CI)		Relative effect (95% CI)	No of Participants (studies)	Quality of the evidence (GRADE)	Comments
	Assumed risk	Corresponding risk				
	Conventional Cytology - Disease specific survival	**HPV test**				
Disease Specific Survival Follow-up: mean 8 years	**168 per 100000**	**99 per 100000** (66 to 153)	**RR 0.59** (0.39 to 0.91)	66184 (1 study)	⊕⊕⊝⊝ **low**[1,2,3,4]	Only one study reported this outcome (Sankranarayanan 2009). GRADE quality assessment has limited value in this context.

*The basis for the **assumed risk** (e.g. the median control group risk across studies) is provided in footnotes. The **corresponding risk** (and its 95% confidence interval) is based on the assumed risk in the comparison group and the **relative effect** of the intervention (and its 95% CI).

CI: Confidence interval; **RR:** Risk ratio

GRADE Working Group grades of evidence
High quality: Further research is very unlikely to change our confidence in the estimate of effect.
Moderate quality: Further research is likely to have an important impact on our confidence in the estimate of effect and may change the estimate.
Low quality: Further research is very likely to have an important impact on our confidence in the estimate of effect and is likely to change the estimate.
Very low quality: We are very uncertain about the estimate.

[1] Blinding not reported, allocation concealment unclear. Missing data - about 30% of women did not attend screening
[2] Inconsistency cannot be graded for a single study
[3] Study conducted in a rural Indian population whereas the target population for the guideline is Germany
[4] Difficult to detect publication bias with a single study. Funnel plot based methods are not applicable

Table 6: GRADE summary of findings – incidence of cervical cancer

HPV test compared to cytology - incidence of cervical cancer for cervical cancer

Patient or population: patients with cervical cancer
Settings:
Intervention: HPV test
Comparison: Cytology - Incidence of Cervical cancer

Outcomes	Illustrative comparative risks* (95% CI)		Relative effect (95% CI)	No of Participants (studies)	Quality of the evidence (GRADE)	Comments
	Assumed risk	Corresponding risk				
	Cytology - Incidence of Cervical cancer	**HPV test**				
Incidence of Cervical Cancer - 1st Screening round Follow-up: 3-5 years	**29 per 100000**	**25 per 100000** (1 to 5)	**RR 0.89** (0.45 to 1.75)	362710 (5 studies)	⊕⊖⊖⊖ **very low**[1,2,3,4,5]	ARTISTIC, Leinonen 2012, NTCC-I, NTCC-II, POBASCAM
Incidence of Cervical Cancer - Subgroup - HPV + Cytology - 1st round Follow-up: 3-5 years	**32 per 100000**	**25 per 100000** (1 to 5)	**RR 0.77** (0.38 to 1.58)	313514 (4 studies)	⊕⊖⊖⊖ **very low**[1,2,3,5,6]	ARTISTIC, Leinonen 2012, NTCC-I, POBASCAM
Incidence of Cervical Cancer - Subgroup - LBC comparator, 3 year interval - 1st round Follow-up: mean 3 years	**25 per 100000**	**15 per 100000** (4 to 53)	**RR 0.63** (0.18 to 2.19)	119180 (3 studies)	⊕⊖⊖⊖ **very low**[2,3,4,6]	ARTISTIC, NTCC-I, NTCC-II
Incidence of Cervical Cancer - 2nd Screening round Follow-up: mean 3 years	**18 per 100000**	**5 per 100000** (1 to 40)	**RR 0.29** (0.04 to 2.26)	108315 (3 studies)	⊕⊖⊖⊖ **very low**[2,3,5]	ARTISTIC, NTCC-I, NTCC-II
Incidence of Cervical Cancer - Subgroup - HPV + Cytology - 2nd round Follow-up: mean 3 years	**23 per 100000**	**10 per 100000** (0 to 277)	**RR 0.42** (0.01 to 12.07)	59965 (2 studies)	⊕⊖⊖⊖ **very low**[2,3,5,6]	ARTISTIC, NTCC-I

*The basis for the **assumed risk** (e.g. the median control group risk across studies) is provided in footnotes. The **corresponding risk** (and its 95% confidence interval) is based on the assumed risk in the comparison group and the **relative effect** of the intervention (and its 95% CI).

CI: Confidence interval; **RR:** Risk ratio

GRADE Working Group grades of evidence
High quality: Further research is very unlikely to change our confidence in the estimate of effect.
Moderate quality: Further research is likely to have an important impact on our confidence in the estimate of effect and may change the estimate.

Low quality: Further research is very likely to have an important impact on our confidence in the estimate of effect and is likely to change the estimate.
Very low quality: We are very uncertain about the estimate.

[1] Allocation concealment was unclear in some studies
[2] Blinding of participants and personnel absent or inadequate in several studies
[3] Missing outcome data not accounted for in several studies
[4] Effect estimates go in different directions in different studies
[5] Widely differing estimates each with wide confidence intervals
[6] I-squared = 50% Effect estimates in differing directions

Table 7: GRADE summary of findings – incidence of CIN3

HPV test compared to Cytology - Incidence of CIN3/CIS for cervical cancer

Patient or population: patients with cervical cancer
Settings:
Intervention: HPV test
Comparison: Cytology - Incidence of CIN3/CIS

Outcomes	Illustrative comparative risks* (95% CI)		Relative effect (95% CI)	No of Participants (studies)	Quality of the evidence (GRADE)	Comments
	Assumed risk	Corresponding risk				
	Cytology - Incidence of CIN3/CIS	**HPV test**				
Incidence of CIN3/CIS - 1st Screening round Follow-up: 3-5 years	**246 per 100000**	**371 per 100000** (265 to 516)	**RR 1.51** (1.08 to 2.10)	338500 (4 studies)	⊕⊖⊖⊖ **very low**[1,2,3]	Leinonen 2012, NTCC-I, NTCC-II, POBASCAM
Incidence of CIN3/CIS - LBC comparator, 3 year interval, 1st round Follow-up: 3 years	**175 per 100000**	**351 per 100000** (166 to 738)	**RR 2.01** (0.95 to 4.23)	94970 (2 studies)	⊕⊖⊖⊖ **very low**[1,2,3]	NTCC-I, NTCC-II
Incidence of CIN3/CIS - Conventional Cytology comparator, >=4 year interval Follow-up: 4-5	**273 per 100000**	**333 per 100000** (284 to 394)	**RR 1.22** (1.04 to 1.44)	243530 (2 studies)	⊕⊕⊕⊖ **moderate**[2,4]	Leinonen 2012, POBASCAM
Incidence of CIN3/CIS - HPV + Cytology: all tests, 1st round Follow-up: 3-5 years	**266 per 100000**	**333 per 100000** (290 to 378)	**RR 1.25** (1.09 to 1.42)	289304 (3 studies)	⊕⊕⊕⊖ **moderate**[1,2,4]	Leinonen 2012, NTCC-I, POBASCAM
Incidence of CIN3/CIS - HPV + Cytology: Combined tests - exclude triage, 1st round Follow-up: 3-4 years	**458 per 100000**	**541 per 100000** (444 to 655)	**RR 1.18** (0.97 to 1.43)	85879 (2 studies)	⊕⊕⊕⊖ **moderate**[1,2]	NTCC-I, POBASCAM
Incidence of CIN3/CIS - 2nd Screening round Follow-up: 3 years	**71 per 100000**	**37 per 100000** (9 to 144)	**RR 0.52** (0.13 to 2.04)	92773 (2 studies)	⊕⊖⊖⊖ **very low**[1,2,5]	NTCC-I, NTCC-II

*The basis for the **assumed risk** (e.g. the median control group risk across studies) is provided in footnotes. The **corresponding risk** (and its 95% confidence interval) is based on the assumed risk in the comparison group and the **relative effect** of the intervention (and its 95% CI).

CI: Confidence interval; **RR:** Risk ratio

GRADE Working Group grades of evidence
High quality: Further research is very unlikely to change our confidence in the estimate of effect.

Moderate quality: Further research is likely to have an important impact on our confidence in the estimate of effect and may change the estimate.
Low quality: Further research is very likely to have an important impact on our confidence in the estimate of effect and is likely to change the estimate.
Very low quality: We are very uncertain about the estimate.

[1] Allocation concealment unclear in some studies
[2] Blinding of participants and personnel inadequate in some studies
[3] I-squared >75%
[4] Possible missing outcome data, unclear
[5] Widely differing estiamtes each with wide confidence intervals

Table 8: GRADE summary of findings – incidence of CIN3+

HPV test compared to cytology - incidence of CIN3+ for cervical cancer

Patient or population: patients with cervical cancer
Settings:
Intervention: HPV test
Comparison: Cytology - Incidence of CIN3+

Outcomes	Illustrative comparative risks* (95% CI) Assumed risk	Corresponding risk	Relative effect (95% CI)	No of Participants (studies)	Quality of the evidence (GRADE)	Comments
	Cytology - Incidence of CIN3+	**HPV test**				
Incidence of CIN3+ - 1st screening round Follow-up: 3-5 years	**360 per 100000**	**443 per 100000** (328 to 602)	**RR 1.23** (0.91 to 1.67)	375537 (6 studies)	⊕⊖⊖⊖ **very low**[1,2,3,4,5]	ARTISTIC, Leinonen 2012, NTCC-I, NTCC-II, POBASCAM, Swedescreen
Incidence of CIN3+ - 3 year interval - 1st round Follow-up: 3 years	**387 per 100000**	**558 per 100000** (352 to 871)	**RR 1.44** (0.91 to 2.25)	132007 (4 studies)	⊕⊖⊖⊖ **very low**[2,4]	ARTISTIC, NTCC-I, NTCC-II, Swedescreen
Incidence of CIN3+ - >=4 year interval Follow-up: 4-5 years	**347 per 100000**	**330 per 100000** (236 to 469)	**RR 0.95** (0.68 to 1.35)	243530 (2 studies)	⊕⊖⊖⊖ **very low**[2,4]	Leinonen 2012, POBASCAM
Incidence of CIN3+ - HPV + Cytology: all tests, 1st round Follow-up: 3-5 years	**396 per 100000**	**412 per 100000** (340 to 499)	**RR 1.04** (0.86 to 1.26)	326341 (5 studies)	⊕⊕⊖⊖ **low**[2,3,6]	ARTISTIC, Leinonen 2012, NTCC-I, POBASCAM, Swedescreen
Incidence of CIN3+ - HPV + Cytology: Combined tests - exclude triage, 1st round Follow-up: 3-4 years	**631 per 100000**	**707 per 100000** (6 to 8)	**RR 1.12** (0.97 to 1.28)	122916 (4 studies)	⊕⊕⊕⊖ **moderate**[2,3]	ARTISTIC, NTCC-I, POBASCAM, Swedescreen
Incidence of CIN3+ - LBC comparator, 1st round Follow-up: 3 years	**329 per 100000**	**491 per 100000** (260 to 922)	**RR 1.49** (0.79 to 2.8)	119480 (3 studies)	⊕⊖⊖⊖ **very low**[2,3,4]	ARTISTIC, NTCC-I, NTCC-II
Incidence of CIN3+ - Conventional Cytology Comparator, 1st round Follow-up: 3-5 years	**373 per 100000**	**388 per 100000** (287 to 523)	**RR 1.04** (0.77 to 1.4)	256057 (3 studies)	⊕⊖⊖⊖ **very low**[2,3,4]	Leinonen 2012, POBASCAM, Swedescreen

Incidence of CIN3+ - 2nd screening round	**159 per 100000**	**82 per 100000** (56 to 121)	**RR 0.52** (0.35 to 0.76)	120652 (4 studies)	⊕⊕⊕⊝ **moderate**[2,3]	ARTISTIC, NTCC-I, NTCC-II, Swedescreen
Incidence of CIN3+ - HPV + Cytology: exclude single HPV tests, 2nd round Follow-up: 3 years	**207 per 100000**	**122 per 100000** (87 to 176)	**RR 0.59** (0.42 to 0.85)	72302 (3 studies)	⊕⊕⊕⊝ **moderate**[2,3]	ARTISTIC, NTCC-I, Swedescreen

*The basis for the **assumed risk** (e.g. the median control group risk across studies) is provided in footnotes. The **corresponding risk** (and its 95% confidence interval) is based on the assumed risk in the comparison group and the **relative effect** of the intervention (and its 95% CI).

CI: Confidence interval; **RR:** Risk ratio

GRADE Working Group grades of evidence
High quality: Further research is very unlikely to change our confidence in the estimate of effect.
Moderate quality: Further research is likely to have an important impact on our confidence in the estimate of effect and may change the estimate.
Low quality: Further research is very likely to have an important impact on our confidence in the estimate of effect and is likely to change the estimate.
Very low quality: We are very uncertain about the estimate.

[1] Allocation concealment unclear in some studies
[2] Blinding of participants and personnel inadequate or unclear in some studies
[3] Missing outcome data in some studies
[4] I-squared >75%
[5] Effect estimates go in different directions in different studies
[6] I-squared >50%

Table 9: GRADE summary of findings – incidence of CIN2+

HPV test compared to cytology - incidence of CIN2+ for cervical cancer

Patient or population: patients with cervical cancer
Settings:
Intervention: HPV test
Comparison: Cytology - Incidence of CIN2+

Outcomes	Illustrative comparative risks* (95% CI)		Relative effect (95% CI)	No of Participants (studies)	Quality of the evidence (GRADE)	Comments
	Assumed risk **Cytology - Incidence of CIN2+**	Corresponding risk **HPV test**				
Incidence of CIN2+ - 1st screening round Follow-up: 3-5 years	**605 per 100000**	**914 per 100000** (732 to 1150)	**RR 1.51** (1.21 to 1.9)	375537 (6 studies)	⊕⊝⊝⊝ **very low**[1,2,3]	ARTISTIC, Leinonen 2012, NTCC-I, NTCC-II, POBASCAM, Swedescreen
Incidence of CIN2+ - 3 year interval, 1st round Follow-up: 3 years	**626 per 100000**	**1027 per 100000** (7 to 16)	**RR 1.64** (1.07 to 2.51)	132007 (4 studies)	⊕⊝⊝⊝ **very low**[1,2,3]	ARTISTIC, NTCC-I, NTCC-II, Swedescreen
Incidence of CIN2+ - >=4 year interval Follow-up: 4-5 years	**595 per 100000**	**803 per 100000** (720 to 898)	**RR 1.35** (1.21 to 1.51)	243530 (2 studies)	⊕⊕⊕⊝ **moderate**[1]	Leinonen 2012, POBASCAM
Incidence of CIN2+ - HPV + Cytology: all tests, 1st round	**653 per 100000**	**869 per 100000** (778 to 961)	**RR 1.33** (1.19 to 1.47)	326341 (5 studies)	⊕⊕⊕⊝ **moderate**[1,2]	ARTISTIC, Leinonen 2012, NTCC-I, POBASCAM, Swedescreen
Incidence of CIN2+ - HPV + Cytology: Combined tests - exclude triage, 1st round Follow-up: 3-4 years	**935 per 100000**	**1197 per 100000** (1047 to 1375)	**RR 1.28** (1.12 to 1.47)	122916 (4 studies)	⊕⊕⊕⊝ **moderate**[1,2]	ARTISTIC, NTCC-I, POBASCAM, Swedescreen
Incidence of CIN2+ - LBC comparator, 1st round Follow-up: 3 years	**557 per 100000**	**942 per 100000** (529 to 1683)	**RR 1.69** (0.95 to 3.02)	119480 (3 studies)	⊕⊝⊝⊝ **very low**[1,2,3]	ARTISTIC, NTCC-I, NTCC-II
Incidence of CIN2+ - Conventional Cytology comparator, 1st round Follow-up: 3-5 years	**625 per 100000**	**857 per 100000** (788 to 938)	**RR 1.37** (1.26 to 1.5)	256057 (3 studies)	⊕⊕⊕⊝ **moderate**[1,2]	Leinonen 2012, POBASCAM, Swedescreen

Incidence of CIN2+ - 2nd screening round Follow-up: 3 years	**250 per 100000**	**143 per 100000** (105 to 938)	**RR 0.57** (0.42 to 0.77)	120652 (4 studies)	⊕⊕⊕⊝ **moderate**[1,2]	ARTISTIC, NTCC-I, NTCC-II, Swedescreen
Incidence of CIN2+ - HPV + Cytology: all tests, 2nd round Follow-up: 3 years	**321 per 100000**	**143 per 100000** (157 to 273)	**RR 0.65** (0.49 to 0.85)	72302 (3 studies)	⊕⊕⊕⊝ **moderate**[1,2]	ARTISTIC, NTCC-I, Swedescreen
Incidence of CIN2+ - LBC comparator, 2nd round Follow-up: 3 years	**196 per 100000**	**110 per 100000** (71 to 170)	**RR 0.56** (0.36 to 0.87)	108315 (3 studies)	⊕⊕⊝⊝ **low**[1,2,4]	ARTISTIC, NTCC-I, NTCC-II

*The basis for the **assumed risk** (e.g. the median control group risk across studies) is provided in footnotes. The **corresponding risk** (and its 95% confidence interval) is based on the assumed risk in the comparison group and the **relative effect** of the intervention (and its 95% CI).

CI: Confidence interval; **RR:** Risk ratio

GRADE Working Group grades of evidence
High quality: Further research is very unlikely to change our confidence in the estimate of effect.
Moderate quality: Further research is likely to have an important impact on our confidence in the estimate of effect and may change the estimate.
Low quality: Further research is very likely to have an important impact on our confidence in the estimate of effect and is likely to change the estimate.
Very low quality: We are very uncertain about the estimate.

[1] Blinding of participants and personnel inadequate or unclear in some studies
[2] Missing outcome data in some studies
[3] I-squared >75%
[4] I-squared >50%

Table 10: GRADE summary of findings – screening related harm

HPV test compared to cytology - screening related harm for cervical cancer

Patient or population: patients with cervical cancer
Settings:
Intervention: HPV test
Comparison: Cytology - Screening related harm

Outcomes	Illustrative comparative risks* (95% CI)		Relative effect (95% CI)	No of Participants (studies)	Quality of the evidence (GRADE)	Comments
	Assumed risk	Corresponding risk				
	Cytology - Screening related harm	**HPV test**				
Screening related harm Follow-up: 3 years	**38301 per 100000**	**37535 per 100000** (33322 to 42514)	**RR 0.98** (0.87 to 1.11)	2465 (1 study)	⊕⊕⊕⊝ **moderate**[1,2]	This outcome was only reported for a single study (ARTISTIC). GRADE quality assessment has limited value in this context

*The basis for the **assumed risk** (e.g. the median control group risk across studies) is provided in footnotes. The **corresponding risk** (and its 95% confidence interval) is based on the assumed risk in the comparison group and the **relative effect** of the intervention (and its 95% CI).

CI: Confidence interval; **RR:** Risk ratio

GRADE Working Group grades of evidence
High quality: Further research is very unlikely to change our confidence in the estimate of effect.
Moderate quality: Further research is likely to have an important impact on our confidence in the estimate of effect and may change the estimate.
Low quality: Further research is very likely to have an important impact on our confidence in the estimate of effect and is likely to change the estimate.
Very low quality: We are very uncertain about the estimate.

[1] Blinding of participants and personnel inadequate
[2] Inconsistency cannot be graded for a single study

7. DISCUSSION

The majority of included RCTs assessed the clinical effectiveness of different screening methods for cervical cancer in terms of the incidence of diagnosis with cervical cancer or the premalignant stages CIN2 or CIN3. The primary objective of cervical cancer screening is to detect CIN3 early enough that it can be treated to prevent the development of cancer. The evidence presented in this review shows that HPV containing screening methods detected more cases of CIN3, CIN3+ and CIN2+ in the first screening round compared to cytology alone (Tables 7-9). In the second screening round, cytology alone detected more cases of CIN3, CIN3+ or CIN2+ than HPV containing methods. The increased detection of these outcomes in the first screening round reflects the higher sensitivity of the HPV based method compared to cytology alone, however, a proportion of these cases would likely never progress to cancer. The observation that in the second round HPV testing detects a smaller proportion of CIN3+ cases suggests that at least some of these cases would have progressed. The increased detection of CIN3+ in the second round by cytology indicates that the cases detected in the second round are either new incident cases or progressive cases that were missed in the first round. The combined interpretation of these results is that HPV based screening detects a greater proportion of these CIN3+ cases in the first round at the expense of also detecting a proportion of non-progressive cases. This is consistent with the lower risk of being diagnosed with cervical cancer observed in those participants screened with an HPV containing regime compared to those screened by cytology alone.

One possible explanation is that this is a consequence of a higher sensitivity of HPV testing compared to cytology. If HPV based screening detects more cases of premalignant disease at the first screening round then only the newly arisen incident cases will be detected in the second round. In comparison, if cytology detects fewer cases of premalignant disease in the first screening round then the second round will detect new incident cases in addition to any residual cases from the first screening round.

There were no studies which reported overall survival as an outcome and only one study (Sankaranarayanan 2009) that reported disease specific survival.

7.1 COMPARISON WITH OTHER REVIEWS

The results of this review are comparable with a previous review of this subject published by IQWiG in 2011[89] which reported that HPV testing alone or in combination with cytology led to a reduction in CIN3+. This review was consistent with our approach in considering the NTCC study as two distinct trials. This review reported heterogeneous findings without a recognisable direction of effect for the incidence of invasive cervical cancer and weak evidence in favour of HPV alone or HPV + cytology for the incidence of CIN3. The IQWiG review reported heterogeneous findings for the diagnosis of CIN2+. The IQWiG authors did not conduct a formal meta-analysis of the first round data due to high heterogeneity between studies. HPV testing alone or in combination with cytology detected more cases of CIN2+ than cytology alone. Meta-analysis of data from the second screening round showed a significant advantage in favour of HPV based regimes (RR 0.53, 95%CI 0.41-0.68). An update of the report[90] was published in June 2014 which included four additional publications on POBASCAM[39, 43] and Leinonen 2012[30, 31]. All studies have already been included in this report. Compared to the earlier IQWiG report, no changes were reported for the aforementioned outcomes.

A recent systematic review for the US preventative services task force (USPSTF)[48] found that primary HPV testing detected more cases of CIN3 or cancer in women older than 30 years compared to cytology which is consistent with the findings of this review. The same review also showed mixed results for the use of HPV + cytology cotesting compared to cytology alone. HPV + cytology combined testing did not detect more CIN3+ than cytology alone, however, cotesting did detect more CIN2+. This review led the USPSTF to recommend cervical cancer screening by cytology every three years for women aged 21-65 years. For women aged 30-65 years who wish to extend their screening interval beyond three years then screening with a combination of HPV + cytology was recommended as an acceptable alternative[49].

A similar joint guideline by the American Cancer Society (ACS), American Society for Colposcopy and Cervical Pathology (ASCCP) and American Society for Clinical Pathology (ASCP) recommended that women aged 30-65 years should be screened with combined HPV + cytology testing every five years as the preferred option while cytology alone every three years was considered an acceptable alternative[50]. This recommendation was based on an evidence review which showed that addition of HPV testing to cytology resulted in increased detection of CIN3 in the initial screening round with a corresponding decrease in the detection of CIN3+ in later screening rounds. This translates to a lower risk following a negative screening result thus allowing for an extended screening interval.

A new study was published during the preparation of this review which presented a pooled analysis of the individual records from four of the trials included in this review with extended follow-up (ARTISTIC, NTCC, POBASCAM, Swedescreen)[91]. The main findings of that study are summarised in Table 11. The overall conclusion of the study by Ronco et al is broadly consistent with the conclusions of this report in that the evidence favours HPV containing regimens over cytology alone for reducing the incidence of cervical cancer. Ronco 2013 reported an overall *rate* ratio of 0.60 (95%CI 0.40-0.89) whereas our analysis reported a *risk* ratio of 0.89 (95%CI 0.45-1.75). Ronco 2013 does include some methodological concerns that may account for the difference between their estimate of the risk and our estimate.

The four studies included in Ronco 2013 all followed participants through two screening rounds. In their meta-analysis Ronco et al defined follow-up in terms of person-years to account for the same participants being followed over two rounds and analysed cancers identified in either screening round. The use of rate ratios rather than risk ratios means that these results are not directly comparable with the results of this review. In addition, Ronco et al analysed the NTCC and POBASCAM studies differently to the way they were analysed in this review. Ronco 2013 included the NTCC trial in their analysis as a single study. This trial was conducted in two phases with different screening regimes in each phase. The first phase compared HPV testing + cytology against cytology alone whereas the second phase compared HPV testing only against cytology alone. We do not consider these screening regimes to be equivalent, therefore, we included the two phases in our analysis as separate studies designated NTCC-I and NTCC-II in this report. The POBASCAM trial included a change in screening regime between the first and second screening rounds. The first screening round compared HPV + cytology against cytology alone. In the second screening round both groups received HPV + cytology. The interpretation of this study differs depending on the research question being addressed. In this review the objective was to identify the most effective screening method in a comparison of HPV testing alone or in combination with cytology versus cytology alone. In this context, the crossover of the comparator group from cytology alone to HPV + cytology means that the comparison in the second round is between the same test applied in two different populations; women who were screen test negative by HPV + cytology versus women who were screen test negative by cytology. The alternative interpretation is to consider the effectiveness of the whole screening algorithm over multiple rounds. This would be a different research question. In this case the comparison would be between two complex interventions implemented over two screening rounds; i.e. HPV + cytology followed by HPV + cytology versus cytology followed by HPV + cytology. This is a valid comparison per se, however, it is outside the scope of the question posed in this review. For these reasons we excluded data from the second screening round and included only the first round data in our analysis. In addition, Ronco 2013 included data on detection of cervical cancer in the Swedescreen trial which were not included in the original publication of this trial and therefore could not be included in our analysis. The cumulative effect of these differences in analytical approach is likely to account for the increased effect size reported by Ronco 2013 compared to our analysis.

We performed a sensitivity analysis including data from the second screening round of POBASCAM in Table 12. These additional results are mostly in line with the results previously presented, i.e. no changes of direction of effect were observed. However, for two of the outcomes a change of the level of significance was observed (see ‘Comments’ in Table 12). The outcome “incidence of cervical cancer” previously showing a non-significant advantage of HPV testing compared to cytology, now shows a significant advantage while the outcome “incidence of CIN3” changed from a significant to a non-significant advantage of HPV. Forest plots for these sensitivity analyses are presented in Figure 8. It should be noted that these analyses were not pre-specified.

Table 11: Summary of meta-analysis published by Ronco 2013

	Experimental*			Cytology			Median FU (years)	Risk Ratio+ (95% CI)	Negative test at entry RR (95% CI)
	No. of Cancers	N	Total person-years	No. of Cancers	N	Total person-years			
ARTISTIC	10	18386	136223	4	6124	45376	7.5	0.83 (0.26-2.66)	2.06 (0.10-41.19)
NTCC	9	47369	242984	24	47001	241025	5.1	0.37 (0.17-0.80)	0.07 (0.01-0.56)
POBASCAM	20	21996	198525	28	22106	199340	9.0	0.72 (0.40-1.27)	0.36 (0.14-0.91)
Swedescreen	5	6257	75477	7	6270	75465	12.0	0.71 (0.23-2.25)	0.50 (0.09-2.73)
Pooled	44	94008	653209	63	81501	561206	6.5	0.60 (0.40-0.89)	0.30 (0.15-0.60)

Summarised from Figure 1, Table 2 and Table 3 Ronco 2013

FU = Follow up

* Experimental includes any screening regime used in the 4 trials reported that included an HPV testing component

+ Risk ratio is the cancer detection rate in the experimental arm vs control arm

Table 12: GRADE summary of findings – Sensitivity analyses with data from 2nd screening round of POBASCAM

HPV test compared to cytology - Incidence of cervical cancer, CIN3, CIN3+, and CIN2+

Patient or population: Patients with cervical cancer
Intervention: HPV test
Comparison: Cytology - Incidence of cervical cancer, CIN3, CIN3+, and CIN2+

Outcomes	Illustrative comparative risks* (95% CI) Assumed risk: Cytology - Incidence of Cervical cancer	Corresponding risk: HPV test	Relative effect (95% CI)	No of Participants (studies)	Quality of the evidence (GRADE)	Comments
Incidence of Cervical Cancer - 2nd Screening round Follow-up: 3-5 years	**20 per 100000**	**6 per 100000** (2 to 15)	**RR 0.29** (0.11 to 0.73)	147625 (4 studies)	⊕⊖⊖⊖ **very low**[1,2,3]	ARTISTIC, NTCC-I, NTCC-II, POBASCAM Original analysis (Table 6): RR 0.29 (95% CI: 0.04 to 2.26)
Incidence of CIN3 - 2nd Screening round Follow-up: 3-5 years	**212 per 100000**	**138 per 100000** (74 to 257)	**RR 0.65** (0.35 to 1.21)	132083 (3 studies)	⊕⊖⊖⊖ **very low**[1,3,5]	NTCC-I, NTCC-II, POBASCAM Original analysis (Table 7): RR 0.55 (95% CI: 0.31 to 0.98)
Incidence of CIN3+ - 2nd screening round Follow-up: 3-5 years	**277 per 100000**	**164 per 100000** (122 to 222)	**RR 0.59** (0.44 to 0.80)	159962 (5 studies)	⊕⊕⊕⊖ **moderate**[1,2]	ARTISTIC, NTCC-I, NTCC-II, POBASCAM, Swedescreen Original analysis (Table 8): RR 0.52 (95% CI: 0.35 to 0.76)
Incidence of CIN2+ - 2nd screening round Follow-up: 3-5 years	**426 per 100000**	**277 per 100000** (200 to 375)	**RR 0.65** (0.47 to 0.88)	159962 (5 studies)	⊕⊕⊕⊖ **moderate**[1,2]	ARTISTIC, NTCC-I, NTCC-II, POBASCAM, Swedescreen Original analysis (Table 9): RR 0.57 (95% CI: 0.42 to 0.77)
Incidence of CIN2+ - HPV + Cytology: all tests, 2nd round Follow-up: 3-5 years	**553 per 100000**	**431 per 100000** (359 to 514)	**RR 0.78** (0.65 to 0.93)	111612 (4 studies)	⊕⊕⊕⊖ **moderate**[1,2]	ARTISTIC, NTCC-I, POBASCAM, Swedescreen Original analysis (Table 9): 0.65 (95% CI: 0.49 to 0.85)
***Incidence of Cervical Cancer – Subgroup: HPV + Cytology - 2nd round*[§]** *Follow-up: 3-5 years*	***44 per 100000***	***14 per 100000*** *(4 to 56)*	***RR 0.33*** *(0.08 to 1.29)*	*99275 (3 studies)*	⊕⊖⊖⊖ ***very low***[1,2,3,4]	*ARTISTIC, NTCC-I, POBASCAM* *Original analysis (Table 6): RR 0.42 (95% CI: 0.01 to 12.07)*

*The basis for the **assumed risk** (e.g. the median control group risk across studies) is provided in footnotes. The **corresponding risk** (and its 95% confidence interval) is

based on the assumed risk in the comparison group and the **relative effect** of the intervention (and its 95% CI).
§ Please note that this row presents a subgroup analysis of HPV + cytology in the 2nd screening round.

CI: Confidence interval; **RR:** Risk ratio

GRADE Working Group grades of evidence
High quality: Further research is very unlikely to change our confidence in the estimate of effect.
Moderate quality: Further research is likely to have an important impact on our confidence in the estimate of effect and may change the estimate.
Low quality: Further research is very likely to have an important impact on our confidence in the estimate of effect and is likely to change the estimate.
Very low quality: We are very uncertain about the estimate.

[1] Blinding of participants and personnel absent or inadequate in several studies
[2] Missing outcome data not accounted for in several studies
[3] Widely differing estimates each with wide confidence intervals
[4] I-squared = 50% Effect estimates in differing directions
[5] Allocation concealment unclear in some studies
[6] I-squared >50%

7.2 STRENGTHS, LIMITATIONS AND UNCERTAINTIES

This review sought wherever possible to reduce the risk of bias during the review processes and analyses. One of the main strengths of the review is the adherence to accepted standards and methods for systematic reviews, including the Centre for Reviews and Dissemination Guidance for Undertaking Systematic Reviews in Healthcare[92] and the Cochrane Collaboration Handbook[19].

In order to try and identify all of the potentially relevant evidence relating to the review question and reduce the risk of publication bias, an extensive range of resources were searched including electronic databases, guidelines and systematic reviews. Both published and unpublished trials were eligible for inclusion. There were no date or language restrictions.

However, despite all efforts to ensure the risk of bias and error was minimised, the findings of the review may still be subject to limitations and uncertainties. Many of these were beyond our control and many related to the quality and quantity of the available evidence base.

All of the studies included in this review used either the HCII method or the GP5+/6+ PCR method for the HPV testing component. These methods were among the earliest developed for HPV testing in cervical cancer screening. Numerous alternative methods have since been developed[6]. We did not identify any published RCTs evaluating the clinical effectiveness of these newer methods either alone or in combination with cytology. A set of published guidelines for the evaluation of HPV tests in cervical cancer screening suggest that such trials are unlikely to be forthcoming[7]. These guidelines propose that HPV tests should be validated by demonstrating non-inferiority in terms of specificity and sensitivity on samples derived from a population based screening cohort previously tested with HCII alone or in combination with cytology. This suggests that any future review in this area will need to consider alternative study designs in order to capture validation studies of newer HPV testing methods.

As detailed in table 1, this report used either a combination of colposcopy and histology (for screening test positives) or follow-up (screening test negatives) as a reference test. This approach is widely reported in the literature, however, it has been argued that colposcopy is an imperfect reference test and performs better when used in a diagnostic rather than a screening setting due to a relatively high sensitivity and lower specificity[93]. Some authors suggested using a multi-biopsy standard instead of a colposcopic-directed biopsy[94] or to combine a multi-biopsy test with endocervical curettage[95]. Another approach is to address potential imperfections in the reference standard through latent class analysis[96]. While latent class analyses are accepted and widely used, no clinical definition is made which impacts on the interpretability of results[97]. Readers should be aware of this limitation while alternative reference tests which could be used in future research should be explored.

There were few studies that directly reported screening related harm. Only the ARTISTIC study explicitly reported on quality of life in relation to screening. This trial reported General Health Questionnaire (GHQ) scores associated with receiving an HPV test in addition to cytology. Future reviews could consider test-positivity rates and referral rates as a surrogate for the burden of follow-up.

There is a lack of high quality data pertaining to the questions of what is the most effective screening interval and at what age to start or stop screening. One of the main objectives of this review was to consider the question of when screening should start or stop. Since the majority of national screening programs focus on women aged 25-65 years we elected to search for evidence that screening is effective in women outside this core age range; i.e. in women aged 20-24 or >65 years. There was only one RCT (ARTISTIC) which reported data on women in either of these groups therefore we extended the searches to include observational studies for the questions of age and screening interval. These extended searches still did not identify any studies that considered the age groups specified in the review protocol. There were 4 RCTs (ARTISTIC, Leinonen 2012, NTCC-I and NTCC-II) that reported age stratified data for some outcomes, however, the age groups varied between trials making it difficult to compare results. There were no studies which specifically addressed the effect of different screening intervals when using HPV testing alone or in combination with cytology. The only studies which considered screening intervals used cytology as the only screening method; i.e. there was no HPV testing component in these studies (Table 3). The intervals currently used in national screening programs appear to have been derived from these earlier studies based on cytology screening prior to the incorporation of HPV testing. It does not necessarily follow that the same intervals will give optimal results with HPV testing given the increased sensitivity and high negative predictive value associated with HPV testing.

The potential risk of bias of the six included studies should be noted when interpreting the results of this review (Tables 4-10). As can be seen from Table 4, five studies were rated as having a high risk of bias regarding blinding of patients and study personnel while the sixth study (Swedescreen) provided insufficient information to judge this item. Although lack of blinding is a common problem in screening trials, it should not be ignored as it could potentially impact on the results of the study. In theory, some of these problems could be avoided, e.g. by a delayed unblinding of patients and personnel. While this approach might decrease the risk of bias, it might not reflect clinical practice, i.e. could influence the applicability of study results. Therefore, it is important to identify these issues and address them for each individual research project. Furthermore, a quarter of the questions regarding the risk of bias could not be answered due to insufficient information provided in the study reports (Table 4).

8. RECOMMENDATIONS FOR FUTURE RESEARCH

There is a clear need for trials that specifically address when to start or stop screening. The majority of RCTs in this area are conducted within existing national screening programs. Since most national screening programs focus on women aged 25-65 years there is a lack of data on the clinical effectiveness of cervical cancer screening in women outside this age range. There is a similar need for trials that consider the effect of screening intervals in the context of HPV testing. The existing studies of screening interval used cytology as the only screening method. An interval which gives optimal results for cytology screening will not necessarily give optimal results for HPV testing. The increased sensitivity, high negative predictive value and long latency between HPV infection and the development of cervical pathology all indicate that a longer interval may be more appropriate for HPV based screening, however, this still needs to be formally tested in a randomized controlled trial. Screening related harm is often neglected in screening trials. Only one study in this review included any assessment of the negative effects of screening (ARTISTIC). Test-positivity rates and referral rates may provide surrogate measures for the burden of follow-up but the analysis of this data is often not clearly reported. Future clinical trials should follow the guidance of the CONSORT statement for the reporting of clinical trials[98, 99] and should include the assessment of more recently developed screening methods. In addition these trials should include explicit assessment and reporting of harm as well as benefit.

9. CONCLUSION

The use of HPV testing in combination with cytology resulted in fewer participants diagnosed with cervical cancer compared to cytology alone. This is likely to be a consequence of the increased detection of earlier, premalignant stages of disease at the first screening round by combined HPV + cytology testing. The quality of evidence in this area is generally low or very low

.

10. REFERENCES

[1] Arbyn M, Castellsague X, de Sanjose S, Bruni L, Saraiya M, Bray F, et al. Worldwide burden of cervical cancer in 2008. *Ann Oncol* 2011;22(12):2675-86.

[2] IARC Working Group. *Human Papillomaviruses [Internet]*. Lyon: International Agency for Research on Cancer, 2007 [accessed 24.9.13]. Available from: http://monographs.iarc.fr/ENG/Monographs/vol90/mono90.pdf

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

[4] Bruni L, Diaz M, Castellsague X, Ferrer E, Bosch FX, de Sanjose S. Cervical human papillomavirus prevalence in 5 continents: meta-analysis of 1 million women with normal cytological findings. *J Infect Dis* 2010;202(12):1789-99.

[5] Group IW. *Cervix Cancer Screening [Internet]*. Lyon: International Agency for Research on Cancer, 2005 [accessed 18.9.13]. Available from: http://www.iarc.fr/en/publications/pdfs-online/prev/handbook10/HANDBOOK10.pdf

[6] Poljak M, Cuzick J, Kocjan BJ, Iftner T, Dillner J, Arbyn M. Nucleic acid tests for the detection of alpha human papillomaviruses. *Vaccine* 2012;30 Suppl 5:F100-6.

[7] Meijer CJ, Berkhof J, Castle PE, Hesselink AT, Franco EL, Ronco G, 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.

[8] 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):448-58.

[9] Canadian Agency for Drugs and Technologies in Health. *CADTH peer review checklist for search strategies [Internet]*. Ottawa: CADTH, 2013 [accessed 17.7.13]. 3p. Available from: http://www.cadth.ca/en/resources/finding-evidence-is

[10] Muche-Borowski C, Selbmann HK, Nothacker M, Muller W, Kopp I. *AWMF guidance manual and rules for guideline development: preliminary English version*. Marburg: AWMF Institute for Medical Knowledge Management, 2012 [accessed 18.9.13] Available from: http://www.awmf.org/fileadmin/user_upload/Leitlinien/AWMF-Regelwerk/AWMF-Guidance_2013-pre.pdf

[11] 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. *Br Med J* 2011;343:d5928.

[12] Guyatt G, Oxman AD, Akl EA, Kunz R, Vist G, Brozek J, et al. GRADE guidelines: 1. Introduction—GRADE evidence profiles and summary of findings tables. *J Clin Epidemiol* 2011;Elsevier(64):4.

[13] *Review Manager (RevMan) [Computer program]*. ver. 5.2. Copenhagen: The Nordic Cochrane Centre, The Cochrane Collaboration, 2012.

[14] Higgins JP, Thompson SG. Quantifying heterogeneity in a meta-analysis. *Stat Med* 2002;21(11):1539-58.

[15] Arbyn M, Ronco G, Anttila A, Meijer CJ, Poljak M, Ogilvie G, et al. Evidence regarding human papillomavirus testing in secondary prevention of cervical cancer. *Vaccine* 2012;30 Suppl 5:F88-99.

[16] Katki HA, Kinney WK, Fetterman B, Lorey T, Poitras NE, Cheung L, et al. Cervical cancer risk for women undergoing concurrent testing for human papillomavirus and cervical cytology: a population-based study in routine clinical practice. *Lancet Oncol* 2011;12(7):663-72.

[17] Harbord RM, Deek JJ, Egger M, Whiting P, Sterne JA. A unification of models for meta-analysis of diagnostic accuracy studies. *Biostatistics* 2007;8(2):239-51.

[18] Harbord RM, Whiting P, Sterne JAC, Egger M, Deeks JJ, Shang A, et al. An empirical comparison of methods for meta-analysis of diagnostic accuracy showed hierarchical models are necessary. *J Clin Epidemiol* 2008;61:1095-103.

[19] Higgins JPT, Green S, eds. *Cochrane handbook for systematic reviews of interventions [Internet]*. Version 5.1.0 [updated March 2011]: The Cochrane Collaboration, 2011 [accessed 23.3.11]. Available from: http://www.cochrane-handbook.org/

[20] Kitchener HC, Almonte M, Wheeler P, Desai M, Gilham C, Bailey A, et al. HPV testing in routine cervical screening: cross sectional data from the ARTISTIC trial. *Br J Cancer* 2006;95(1):56-61.

[21] Kitchener HC, Fletcher I, Roberts C, Wheeler P, Almonte M, Maguire P. The psychosocial impact of human papillomavirus testing in primary cervical screening - A study within a randomized trial. *International Journal of Gynecological Cancer* 2008;18(4):743-748.

[22] Kitchener HC, Almonte M, Thomson C, Wheeler P, Sargent A, Stoykova B, et al. HPV testing in combination with liquid-based cytology in primary cervical screening (ARTISTIC): a randomised controlled trial.[Erratum appears in Lancet Oncol. 2009 Aug;10(8):748]. *Lancet Oncol* 2009;10(7):672-82.

[23] Kitchener HC, Almonte M, Gilham C, Dowie R, Stoykova B, Sargent A, et al. ARTISTIC: a randomised trial of human papillomavirus (HPV) testing in primary cervical screening. *Health Technol Assess* 2009;13(51):1-150, iii-iv.

[24] Sargent A, Bailey A, Turner A, Almonte M, Gilham C, Baysson H, et al. Optimal threshold for a positive hybrid capture 2 test for detection of human papillomavirus: data from the ARTISTIC trial. *J Clin Microbiol* 2010;48(2):554-8.

[25] Kitchener HC, Gilham C, Sargent A, Bailey A, Albrow R, Roberts C, et al. A comparison of HPV DNA testing and liquid based cytology over three rounds of primary cervical screening: extended follow up in the ARTISTIC trial. *Eur J Cancer* 2011;47(6):864-71.

[26] Anttila A, Kotaniemi-Talonen L, Leinonen M, Hakama M, Laurila P, Tarkkanen J, et al. Rate of cervical cancer, severe intraepithelial neoplasia, and adenocarcinoma in situ in primary HPV DNA screening with cytology triage: Randomised study within organised screening programme. *BMJ* 2010;340(7754):1014.

[27] Kotaniemi-Talonen L, Nieminen P, Anttila A, Hakama M. Routine cervical screening with primary HPV testing and cytology triage protocol in a randomised setting. *Br J Cancer* 2005;93(8):862-867.

[28] Kotaniemi-Talonen L, Anttila A, Malila N, Tarkkanen J, Laurila P, Hakama M, et al. Screening with a primary human papillomavirus test does not increase detection of cervical cancer and intraepithelial neoplasia 3. *Eur J Cancer* 2008;44(4):565-571.

[29] Leinonen M, Nieminen P, Kotaniemi-Talonen L, Malila N, Tarkkanen J, Laurila P, et al. Age-specific evaluation of primary human papillomavirus screening vs conventional cytology in a randomized setting. *J Natl Cancer Inst* 2009;101(23):1612-1623.

[30] Leinonen MK, Nieminen P, Lönnberg S, Malila N, Hakama M, Pokhrel A, et al. Detection rates of precancerous and cancerous cervical lesions within one screening round of primary human papillomavirus DNA testing: prospective randomised trial in Finland. *BMJ* 2012;345:e7789.

[31] Malila N, Leinonen M, Kotaniemi-Talonen L, Laurila P, Tarkkanen J, Hakama M. The HPV test has similar sensitivity but more overdiagnosis than the Pap test--a randomised health services study on cervical cancer screening in Finland. *Int J Cancer* 2013;132(9):2141-7.

[32] Giorgi-Rossi P, Segnan N, Zappa M, Naldoni C, Zorzi M, Confortini M, et al. The impact of new technologies in cervical cancer screening: results of the recruitment phase of a large randomised controlled trial from a public health perspective. *Int J Cancer* 2007;121(12):2729-34.

[33] Ronco G, Segnan N, Giorgi-Rossi P, Zappa M, Casadei GP, Carozzi F, et al. Human papillomavirus testing and liquid-based cytology: results at recruitment from the new technologies for cervical cancer randomized controlled trial. *J Natl Cancer Inst* 2006;98(11):765-74.

[34] Ronco G, Giorgi-Rossi P, Carozzi F, Dalla Palma P, Del Mistro A, De Marco L, 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(7):547-55.

[35] Ronco G, Cuzick J, Pierotti P, Cariaggi MP, Palma PD, Naldoni C, et al. Accuracy of liquid based versus conventional cytology: Overall results of new technologies for cervical cancer screening: Randomised controlled trial. *BMJ* 2007;335(7609):28-31.

[36] 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;43(3):476-480.

[37] Ronco G, Giorgi-Rossi P, Carozzi F, Confortini M, Palma PD, Del Mistro A, et al. Results at recruitment from a randomized controlled trial comparing human papillomavirus testing alone with conventional cytology as the primary cervical cancer screening test. *J Natl Cancer Inst* 2008;100(7):492-501.

[38] Ronco G, Giorgi-Rossi P, Carozzi F, Confortini M, Dalla Palma P, Del Mistro A, et al. Efficacy of human papillomavirus testing for the detection of invasive cervical cancers and cervical intraepithelial neoplasia: a randomised controlled trial. *Lancet Oncol* 2010;11(3):249-57.

[39] Budenholzer B. Adding HPV testing to cytology screening reduced >= grade 3 cervical intraepithelial neoplasia at 5 years. *Ann Intern Med* 2012;157(2):JC2-6.

[40] Bulkmans NWJ, Rozendaal L, Snijders PJF, Voorhorst FJ, Boeke AJP, Zandwijken GRJ, et al. POBASCAM, a population-based randomized controlled trial for implementation of high-risk HPV testing in cervical screening: design, methods and baseline data of 44,102 women. *Int J Cancer* 2004;110(1):94-101.

[41] Bulkmans NWJ, Bulk S, Ottevanger MS, Rozendaal L, Hellenberg SM, van Kemenade FJ, et al. Implementation of human papillomavirus testing in cervical screening without a concomitant decrease in participation rate. *J Clin Pathol* 2006;59(11):1218-20.

[42] Bulkmans NWJ, Berkhof J, Rozendaal L, van Kemenade FJ, Boeke AJP, Bulk S, 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(9601):1764-72.

[43] Rijkaart DC, Berkhof J, Rozendaal L, van Kemenade FJ, Bulkmans NWJ, Heideman DAM, et al. Human papillomavirus testing for the detection of high-grade cervical intraepithelial neoplasia and cancer: Final results of the POBASCAM randomised controlled trial. *The Lancet Oncology* 2012;13(1):78-88.

[44] Sankaranarayanan R, Nene BM, Dinshaw KA, Mahe C, Jayant K, Shastri SS, et al. A cluster randomized controlled trial of visual, cytology and human papillomavirus screening for cancer of the cervix in rural India. *Int J Cancer* 2005;116(4):617-23.

[45] Sankaranarayanan R, Nene BM, Shastri SS, Jayant K, Muwonge R, Budukh AM, et al. HPV screening for cervical cancer in rural India. *N Engl J Med* 2009;360(14):1385-1394.

[46] Elfgren K, Rylander E, Rådberg T, Strander B, Strand A, Paajanen K, et al. Colposcopic and histopathologic evaluation of women participating in population-based screening for human papillomavirus deoxyribonucleic acid persistence. *Am J Obstet Gynecol* 2005;193(3 Pt 1):650-7.

[47] Naucler P, Ryd W, Tornberg S, Strand A, Wadell G, Elfgren K, et al. Human papillomavirus and Papanicolaou tests to screen for cervical cancer.[Erratum appears in N Engl J Med. 2008 Oct 9;359(15):1637 Note: Johansson, Bo [added]]. *N Engl J Med* 2007;357(16):1589-97.

[48] Whitlock EP, Vesco KK, Eder M, Lin JS, Senger CA, Burda BU. Liquid-based cytology and human papillomavirus testing to screen for cervical cancer: a systematic review for the U.S. Preventive Services Task Force. *Ann Intern Med* 2011;155(10):687-697, W214-215.

[49] Quality AAfHR. *Screening for cervical cancer: U.S. Preventive Services Task Force recommendation statement. U.S. Preventive Services Task Force. NGC:009010*, 2012. Available from: http://www.guideline.gov/content.aspx?id=36624

[50] Quality AAfHR. *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. American Cancer Society. American Society for Clinical Pathology. American Society for Colposcopy and Cervical Pathology. NGC:009048*, 2012. Available from: http://www.guideline.gov/content.aspx?id=36834

[51] Peirson L, Fitzpatrick-Lewis D, Ciliska D, Warren R. Screening for cervical cancer: a systematic review and meta-analysis (Structured abstract). *Syst Rev*, 2013.

[52] Andrae B, Kemetli L, Sparen P, Silfverdal L, Strander B, Ryd W, et al. Screening-preventable cervical cancer risks: evidence from a nationwide audit in Sweden. *J Natl Cancer Inst* 2008;100(9):622-9.

[53] Herbert A, Stein K, Bryant TN, Breen C, Old P. Relation between the incidence of invasive cervical cancer and the screening interval: is a five year interval too long? *J Med Screen* 1996;3(3):140-5.

[54] Herrero R, Brinton LA, Reeves WC, Brenes MM, de Britton RC, Gaitan E, et al. Screening for cervical cancer in Latin America: a case-control study. *Int J Epidemiol* 1992;21(6):1050-6.

[55] Hoffman M, Cooper D, Carrara H, Rosenberg L, Kelly J, Stander I, et al. Limited Pap screening associated with reduced risk of cervical cancer in South Africa. *Int J Epidemiol* 2003;32(4):573-7.

[56] Jiménez-Perez M, Thomas DB. Has the use of pap smears reduced the risk of invasive cervical cancer in Guadalajara, Mexico? *Int J Cancer* 1999;82(6):804-9.

[57] Kasinpila C, Promthet S, Vatanasapt P, Sasieni P, Parkin DM. Evaluation of the nationwide cervical screening programme in Thailand: a case-control study. *J Med Screen* 2011;18(3):147-53.

[58] La Vecchia C, Franceschi S, Decarli A, Fasoli M, Gentile A, Tognoni G. "Pap" smear and the risk of cervical neoplasia: quantitative estimates from a case-control study. *Lancet* 1984;2(8406):779-82.

[59] Makino H, Sato S, Yajima A, Komatsu S, Fukao A. Evaluation of the effectiveness of cervical cancer screening: a case-control study in Miyagi, Japan. *Tohoku J Exp Med* 1995 Tohoku J Exp Med(3):171-8.

[60] Miller MG, Sung HY, Sawaya GF, Kearney KA, Kinney W, Hiatt RA. Screening interval and risk of invasive squamous cell cervical cancer. *Obstet Gynecol* 2003;101(1):29-37.

[61] Rebolj M, van Ballegooijen M, Lynge E, Looman C, Essink-Bot ML, Boer R, et al. Incidence of cervical cancer after several negative smear results by age 50: prospective observational study. *Br Med J* 2009;338:b1354.

[62] Sasieni PD, Cuzick J, Lynch-Farmery E. Estimating the efficacy of screening by auditing smear histories of women with and without cervical cancer. The National Co-ordinating Network for Cervical Screening Working Group. *Br J Cancer* 1996;73(8):1001-5.

[63] 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(1):88-93.

[64] Sasieni P, Castanon A, Cuzick J. Effectiveness of cervical screening with age: population based case-control study of prospectively recorded data. *Br Med J* 2009;339:b2968.

[65] Sasieni P, Castanon A, Cuzick J. Screening and adenocarcinoma of the cervix. *Int J Cancer* 2009;125(3):525-9.

[66] Yang B, Morrell S, Zuo Y, Roder D, Tracey E, Jelfs P. A case-control study of the protective benefit of cervical screening against invasive cervical cancer in NSW women. *Cancer Causes Control* 2008;19(6):569-76.

[67] Zappa M, Visioli CB, Ciatto S, Iossa A, Paci E, Sasieni P. 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(9):1784-6.

[68] 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. *International Journal of Gynecological Cancer* 2010;20(8):1411-4.

[69] Belinson JL, Wu R, Belinson SE, Qu X, Yang B, Du H, et al. A population-based clinical trial comparing endocervical high-risk HPV testing using hybrid capture 2 and Cervista from the SHENCCAST II Study. *Am J Clin Pathol* 2011;135(5):790-5.

[70] Mayrand M-H, Duarte-Franco E, Coutlee F, Rodrigues I, Walter SD, Ratnam S, et al. Randomized controlled trial of human papillomavirus testing versus Pap cytology in the primary screening for cervical cancer precursors: design, methods and preliminary accrual results of the Canadian cervical cancer screening trial (CCCaST). *Int J Cancer* 2006;119(3):615-23.

[71] Mayrand MH, Duarte-Franco E, Rodrigues I, Walter SD, Hanley J, Ferenczy A, et al. Human papillomavirus DNA versus Papanicolaou screening tests for cervical cancer. *N Engl J Med* 2007;357(16):1579-88.

[72] Cheng JX, Yao LL, Yuan M, Liu CM, Guzhalinuer, Zhang Y. HPV testing in diversion management of patients with ASCUS in cervical cytology. [Chinese]. *Journal of Practical Oncology* 2012;27(6):630-633.

[73] Ogilvie GS, van Niekerk DJ, Krajden M, Martin RE, Ehlen TG, Ceballos K, et al. A randomized controlled trial of Human Papillomavirus (HPV) testing for cervical cancer screening: trial design and preliminary results (HPV FOCAL Trial). *BMC Cancer* 2010;10:111.

[74] Ogilvie GS, Krajden M, van Niekerk DJ, Martin RE, Ehlen TG, Ceballos K, et al. Primary cervical cancer screening with HPV testing compared with liquid-based cytology: results of round 1 of a randomised controlled trial -- the HPV FOCAL Study. *Br J Cancer* 2012;107(12):1917-24.

[75] Van Niekerk D, Ogilvie G, Krajden M, Martin R, Stuart G, Ceballos K, et al. hrHPV DNA testing in cervical cancer screening. First round results From the HPV for CervicAL screening (HPV FOCAL) trial. In:Histopathology. Conference: 29th Congress of the International Academy of Pathology Cape Town South Africa. Conference Start: 20120930 Conference End: 20121005. Conference Publication: (var.pagings). 61 (pp 57), 2012. Date of Publication: October 2012., 2012.

[76] Lazcano-Ponce E, Lorincz AT, Cruz-Valdez A, Salmerón J, Uribe P, Velasco-Mondragón E, et al. Self-collection of vaginal specimens for human papillomavirus testing in cervical cancer prevention (MARCH): a community-based randomised controlled trial. *Lancet* 2011;378(9806):1868-73.

[77] Sancho-Garnier H, Tamalet C, Halfon P, Leandri FX, Retraite LL, Djoufelkit K, et al. HPV self-sampling or the Pap-smear: A randomized study among cervical screening nonattenders from lower socioeconomic groups in France. *Int J Cancer* 2013;133(11):2681-2687.

[78] Bailon Munoz E. Cervical cancer screening interval can be extended to six years after a negative HPV test: European multicentric study. [Spanish]. *FMC Formacion Med Continuada Aten Primaria* 2009;16(9):606.

[79] Carter JS, Sage YH, Vragovic O, Rosen L, Stier EA. Effect of patient age on outcomes and compliance in women with minimally abnormal pap tests. *J Reprod Med Obstetrician Gynecologist* 2010;55(7-8):351-356.

[80] Cuzick J, Szarewski A, Mesher D, Cadman L, Austin J, Perryman K, et al. Long-term follow-up of cervical abnormalities among women screened by HPV testing and cytology-Results from the Hammersmith study. *Int J Cancer* 2008;Journal international du cancer. 122(10):2294-2300.

[81] de Vries CE, Shen R, Stephens J, Suarez AA. Equivocal and weakly positive hybrid capture 2 tests in women aged 50 and older. *Diagn Cytopathol* 2012;40(8):708-12.

[82] Grainge MJ, Seth R, Coupland C, Guo L, Rittman T, Vryenhoef P, et al. Human papillomavirus infection in women who develop high-grade cervical intraepithelial neoplasia or cervical cancer: A case-control study in the UK. *Br J Cancer* 2005;92(9):1794-1799.

[83] Gyllensten U, Gustavsson I, Lindell M, Wilander E. Primary high-risk HPV screening for cervical cancer in post-menopausal women. *Gynecol Oncol* 2012;125(2):343-5.

[84] Herbert A, Best JM, Chana P, Ktori E, Nowicki M, Dunsmore H, et al. Human papillomavirus testing with conventional Pap smear screening in three inner London community clinics. *J Fam Plann Reprod Health Care* 2007;33(3):171-176.

[85] Inoue M, Okamura M, Hashimoto S, Tango M, Ukita T. Adoption of HPV testing as an adjunct to conventional cytology in cervical cancer screening in Japan. *International Journal of Gynaecology & Obstetrics* 2010;111(2):110-4.

[86] Kjaer S, Hogdall E, Frederiksen K, Munk C, van den Brule A, Svare E, et al. The absolute risk of cervical abnormalities in high-risk human papillomavirus-positive, cytologically normal women over a 10-year period. *Cancer Res* 2006;66(21):10630-6.

[87] Kjaer SK, Frederiksen K, Munk C, Iftner T. Long-term absolute risk of cervical intraepithelial neoplasia grade 3 or worse following human papillomavirus infection: Role of persistence. *J Natl Cancer Inst* 2010;102(19):1478-1488.

[88] Lee S, Park HT, Hong JH, Song JY, Lee JK, Lee NW, et al. Assessment of cervical cancer screening policy in Korea for women over age 65. *Journal of Geriatric Oncology* 2013;4(4):382-387.

[89] Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen. Nutzenbewertung eines HPV-Tests im Primärscreening des Zervixkarzinoms: Abschlussbericht; S10-01. *IQWiG-Berichte; Nr. 106*. Köln: IQWiG, 2011.

[90] Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen. Nutzenbewertung eines HPV-Tests im Primärscreening des Zervixkarzinoms - Aktualisierung; S13-03. *IQWiG-Berichte; Nr. 222*. Köln: IQWiG, 2014.

[91] Ronco G, Dillner J, Elfström 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;Nov 1. pii:S0140-6736(13)62218-7. [Epub ahead of print].

[92] Centre for Reviews and Dissemination. *Systematic Reviews: CRD's guidance for undertaking reviews in health care [Internet]*. York: University of York, 2009 [accessed 23.3.11] Available from: http://www.york.ac.uk/inst/crd/SysRev/!SSL!/WebHelp/SysRev3.htm

[93] Cantor SB, Cárdenas-Turanzas M, Cox DD, Atkinson EN, Nogueras-Gonzalez GM, Beck JR, et al. Accuracy of colposcopy in the diagnostic setting compared with the screening setting. *Obstet Gynecol* 2008;111(1):7-14.

[94] Pretorius RG, Bao YP, Belinson JL, Burchette RJ, Smith JS, Qiao YL. Inappropriate gold standard bias in cervical cancer screening studies. *Int J Cancer* 2007;121(10):2218-24.

[95] Cagle AJ, Hu SY, Sellors JW, Bao YP, Lim JM, Li SM, et al. Use of an expanded gold standard to estimate the accuracy of colposcopy and visual inspection with acetic acid. *Int J Cancer* 2010;126(1):156-61.

[96] Gaffikin L, McGrath JA, Arbyn M, Blumenthal PD. Visual inspection with acetic acid as a cervical cancer test: accuracy validated using latent class analysis. *BMC Med Res Methodol* 2007;7:36.

[97] Rutjes AW, Reitsma JB, Coomarasamy A, Khan KS, Bossuyt PM. Evaluation of diagnostic tests when there is no gold standard. A review of methods. *Health Technol Assess* 2007;11(50):iii, ix-51.

[98] Schulz KF, Altman DG, Moher D, CONSORT Group. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. *BMJ* 2010;340:c332.

[99] Moher D, Hopewell S, Schulz KF, Montori V, Gøtzsche PC, Devereaux PJ, et al. CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. *BMJ* 2010;340:c869.

APPENDIX 1: SEARCH STRATEGIES

EMBASE SEARCH STRATEGY FOR RCTS

Embase (OvidSP): 1974-2013/10/17
Searched 18.10.13

1 (papillomavirus$ or papilloma-virus$ or alphapapillomavirus$ or alpha-papillomavirus$ or alpha papilloma-virus$ or papillomaviridae).ti,ab,ot,hw. (39010)
2 HPV$.ti,ab,ot,hw. (32362)
3 or/1-2 (45592)
4 exp papillomavirus infection/ (18046)
5 exp Papilloma virus/ (38117)
6 or/4-5 (48669)
7 3 or 6 (57390)
8 ((cervix$ or cervical or services or cervico$ or endocervix$ or endocervical or endocervices or endocervico$ or endo-cervix$ or endo-cervical or endo-cervices or endo-cervico$ or ectocervix$ or ectocervical or ectocervices or ectocervico$ or ecto-cervix$ or ecto-cervical or ecto-cervices or ecto-cervico$ or cervix-uteri) adj5 (cancer$ or neoplas$ or dysplas$ or oncolog$ or malignan$ or tumo?r$ or carcinoma$ or adenocarcinoma$ or metasta$ or meta sta$ or sarcoma$ or adenoma$ or lesion$ or dyskaryos?s or squamous or SCC or SqCC or rhabdomyosarcoma$ or neuroendocrin$ or neuro-endocrin$ or cin or cin1 or cin2 or cin3 or "cin-1" or "cin-2" or "cin-3" or "cin 1" or "cin 2" or "cin 3" or CINI or CINII or CINIII or cis or ascus or "asc-us" or "asc us" or "asc-h" or "asc h" or asch or lgsil or lsil or hgsil or hsil or "AGC-NOS" or "AGC NOS" or AGCNOS or AGC-neoplas$ or AIS or "atypical glandular cell$")).ti,ab,ot,hw. (112507)
9 ((cervix$ or cervical or services or cervico$ or endocervix$ or endocervical or endocervices or endocervico$ or endo-cervix$ or endo-cervical or endo-cervices or endo-cervico$ or ectocervix$ or ectocervical or ectocervices or ectocervico$ or ecto-cervix$ or ecto-cervical or ecto-cervices or ecto-cervico$ or cervix-uteri) adj5 ((precancer$ adj3 cell$) or (pre-cancer$ adj3 cell$) or (preneoplas$ adj3 cell$) or (pre-neoplas$ adj3 cell$) or (premalignan$ adj3 cell$) or (pre-malignan$ adj3 cell$) or (precancer$ adj3 change$) or (pre-cancer$ adj3 change$) or (preneoplas$ adj3 change$) or (pre-neoplas$ adj3 change$) or (premalignan$ adj3 change$) or (pre-malignan$ adj3 change$) or (abnormal adj3 cell$))).ti,ab,ot,hw. (295)
10 exp uterine cervix tumor/ (83696)
11 uterine cervix dysplasia/ (3943)
12 uterine cervix carcinoma in situ/ (11048)
13 or/8-12 (112548)
14 7 and 13 (22668)
15 (screen$ or test$ or cytolog$ or histocytodiagnos$ or cytodiagnos$ or cyto-diagnos$ or histocytochemi$ or cytochemi$ or cyto-chemi$ or pap or papanicolaou or smear$ or swab$ or scrap$).ti,ab,ot,hw. (4292258)
16 exp mass screening/ (154814)
17 cytology/ or exp cytochemistry/ or exp cytodiagnosis/ (523146)
18 or/15-17 (4333279)
19 14 and 18 (12400)
20 Random$.tw. or placebo$.mp. or double-blind$.tw. (1074690)
21 19 and 20 (893)

22 animal/ (1890937)
23 animal experiment/ (1721607)
24 (rat or rats or mouse or mice or murine or rodent or rodents or hamster or hamsters or pig or pigs or porcine or rabbit or rabbits or animal or animals or dogs or dog or cats or cow or bovine or sheep or ovine or monkey or monkeys).ti,ab,ot,hw. (5828979)
25 or/22-24 (5828979)
26 exp human/ (15032575)
27 human experiment/ (317393)
28 or/26-27 (15034016)
29 25 not (25 and 28) (4644866)
30 21 not 29 (891)

Trials filter:

Wong SS, Wilczynski NL, Haynes RB. Developing optimal search strategies for detecting clinically sound treatment studies in EMBASE (optimised sensitivity/specificity). J Med Libr Assoc 2006;94(1):41-7.

EMBASE SEARCH STRATEGY FOR OBSERVATIONAL STUDIES (AND UPDATING RCTS SEARCH)

Embase (OvidSP): 1974-2014/week 2
Searched 13.01.14

1 (papillomavirus$ or papilloma-virus$ or alphapapillomavirus$ or alpha-papillomavirus$ or alpha papilloma-virus$ or papillomaviridae).ti,ab,ot,hw. (39768)
2 HPV$.ti,ab,ot,hw. (33123)
3 or/1-2 (46527)
4 exp papillomavirus infection/ (18481)
5 exp Papilloma virus/ (38741)
6 or/4-5 (49523)
7 3 or 6 (58474)
8 ((cervix$ or cervical or services or cervico$ or endocervix$ or endocervical or endocervices or endocervico$ or endo-cervix$ or endo-cervical or endo-cervices or endo-cervico$ or ectocervix$ or ectocervical or ectocervices or ectocervico$ or ecto-cervix$ or ecto-cervical or ecto-cervices or ecto-cervico$ or cervix-uteri) adj5 (cancer$ or neoplas$ or dysplas$ or oncolog$ or malignan$ or tumo?r$ or carcinoma$ or adenocarcinoma$ or metasta$ or meta sta$ or sarcoma$ or adenoma$ or lesion$ or dyskaryos?s or squamous or SCC or SqCC or rhabdomyosarcoma$ or neuroendocrin$ or neuro-endocrin$ or cin or cin1 or cin2 or cin3 or "cin-1" or "cin-2" or "cin-3" or "cin 1" or "cin 2" or "cin 3" or CINI or CINII or CINIII or cis or ascus or "asc-us" or "asc us" or "asc-h" or "asc h" or asch or lgsil or lsil or hgsil or hsil or "AGC-NOS" or "AGC NOS" or AGCNOS or AGC-neoplas$ or AIS or "atypical glandular cell$")).ti,ab,ot,hw. (114109)
9 ((cervix$ or cervical or services or cervico$ or endocervix$ or endocervical or endocervices or endocervico$ or endo-cervix$ or endo-cervical or endo-cervices or endo-cervico$ or ectocervix$ or ectocervical or ectocervices or ectocervico$ or ecto-cervix$ or ecto-cervical or ecto-cervices or ecto-cervico$ or cervix-uteri) adj5 ((precancer$ adj3 cell$) or (pre-cancer$ adj3 cell$) or (preneoplas$ adj3 cell$) or (pre-neoplas$ adj3 cell$) or (premalignan$ adj3 cell$) or (pre-malignan$ adj3 cell$) or (precancer$ adj3 change$) or (pre-cancer$ adj3 change$) or (preneoplas$ adj3 change$) or (pre-neoplas$ adj3 change$)

or (premalignan$ adj3 change$) or (pre-malignan$ adj3 change$) or (abnormal adj3 cell$))).ti,ab,ot,hw. (302)
10 exp uterine cervix tumor/ (84888)
11 uterine cervix dysplasia/ (3979)
12 uterine cervix carcinoma in situ/ (11195)
13 or/8-12 (114151)
14 7 and 13 (23067)
15 (screen$ or test$ or cytolog$ or histocytodiagnos$ or cytodiagnos$ or cyto-diagnos$ or histocytochemi$ or cytochemi$ or cyto-chemi$ or pap or papanicolaou or smear$ or swab$ or scrap$).ti,ab,ot,hw. (4363933)
16 exp mass screening/ (157729)
17 cytology/ or exp cytochemistry/ or exp cytodiagnosis/ (526287)
18 or/15-17 (4405474)
19 14 and 18 (12639)
20 Random$.tw. or placebo$.mp. or double-blind$.tw. (1097839)
21 Clinical study/ (97157)
22 Case control study/ (86508)
23 Family study/ (10238)
24 Longitudinal study/ (67139)
25 Retrospective study/ (356527)
26 Prospective study/ (259838)
27 Randomized controlled trials/ (44329)
28 26 not 27 (258613)
29 Cohort analysis/ (167278)
30 (Cohort adj (study or studies)).mp. (112772)
31 (Case control adj (study or studies)).tw. (75082)
32 (follow up adj (study or studies)).tw. (46667)
33 (observational adj (study or studies)).tw. (61425)
34 (epidemiologic$ adj (study or studies)).tw. (75548)
35 (cross sectional adj (study or studies)).tw. (82518)
36 or/21-25,28-35 (1219852)
37 19 and (20 or 36) (2618)
38 animal/ (1896593)
39 animal experiment/ (1738993)
40 (rat or rats or mouse or mice or murine or rodent or rodents or hamster or hamsters or pig or pigs or porcine or rabbit or rabbits or animal or animals or dogs or dog or cats or cow or bovine or sheep or ovine or monkey or monkeys).ti,ab,ot,hw. (5880438)
41 or/38-40 (5880438)
42 exp human/ (15253431)
43 human experiment/ (320062)
44 or/42-43 (15254873)
45 41 not (41 and 44) (4677573)
46 37 not 45 (2615)

Trials filter:

Wong SS, Wilczynski NL, Haynes RB. Developing optimal search strategies for detecting clinically sound treatment studies in EMBASE (optimised sensitivity/specificity). J Med Libr

Assoc 2006;94(1):41-7.

Observational Studies Filter:

Scottish Intercollegiate Guidelines Network (SIGN). Search filters: observational studies [Embase (OvidSP)]. Edinburgh: SIGN, Last modified 26/04/13 Available from: http://www.sign.ac.uk/methodology/filters.html#obs

APPENDIX 2: COCHRANE RISK OF BIAS ASSESSMENT

ARTISTIC

Domain	Judgement	Criteria	Supporting text
Selection bias			
Random sequence generation	**Low risk of bias**	The investigators describe a random component in the sequence generation process such as: • Referring to a random number table • Using a computer random number generator • Coin tossing • Shuffling cards or envelopes • Throwing dice • Drawing of lots • Minimization* *Minimization may be implemented without a random element, and this is considered to be equivalent to being random	Computer generated random numbers
Allocation concealment	**Low risk of bias**	Participants and investigators enrolling participants could not foresee assignment because one of the following, or an equivalent method, was used to conceal allocation: • Central allocation (including telephone, web-based and pharmacy-controlled randomization) • Sequentially numbered drug containers of identical appearance • Sequentially numbered, opaque, sealed envelopes	Centralised allocation
Performance bias			
Blinding of participants and personnel *Assessments should be made for each main outcome (or class of outcomes).*	**High risk of bias**	Any one of the following: • No blinding or incomplete blinding, and the outcome is likely to be influenced by lack of blinding • Blinding of key study participants and personnel attempted, but likely that the blinding could have been broken, and the outcome is likely to be influenced by lack of blinding	Test results have to be reported back to inform clinical management. Knowing which test results are available indicates which group the patient is in. Could influence referral decisions
Detection bias			
Blinding of outcome assessment *Assessments*	**Unclear risk of bias**	Any one of the following: • Insufficient information to permit judgement of 'Low risk' or 'High risk' • The study did not address this outcome	Outcome assessment was blinded. Test results have to be reported back to inform clinical management. Knowing which test results are available indicates which group

Domain	Judgement	Criteria	Supporting text
should be made for each main outcome (or class of outcomes).			the patient is in. Could influence referral decisions. Colposcopists were aware of HPV and Cytology results
Attrition bias			
Incomplete outcome data *Assessments should be made for each main outcome (or class of outcomes).*	**High risk of bias**	Any one of the following: • Reason for missing outcome data likely to be related to true outcome, with either imbalance in numbers or reasons for missing data across intervention groups • For dichotomous outcome data, the proportion of missing outcomes compared with observed event risk enough to induce clinically relevant bias in intervention effect estimate • For continuous outcome data, plausible effect size (difference in means or standardized difference in means) among missing outcomes enough to induce clinically relevant bias in observed effect size • 'As-treated' analysis done with substantial departure of the intervention received from that assigned at randomization • Potentially inappropriate application of simple imputation	Describe the completeness of outcome data for each main outcome, including attrition and exclusions from the analysis. State whether attrition and exclusions were reported, the numbers in each intervention group (compared with total randomized participants), reasons for attrition/exclusions where reported, and any re-inclusions in analyses performed by the review authors.
Reporting bias			
Selective reporting.	**Low risk of bias**	Any of the following: • The study protocol is available and all of the study's pre-specified (primary and secondary) outcomes that are of interest in the review have been reported in the pre-specified way • The study protocol is not available but it is clear that the published reports include all expected outcomes, including those that were pre-specified (convincing text of this nature may be uncommon)	Trial protocol is available. Specified outcomes are reported although scattered across multiple publications
Other bias			
Other sources of bias.	**Low risk of bias**	The study appears to be free of other sources of bias.	The study appears to be free of other sources of bias.

LEINONEN 2012

Domain	Judgement	Criteria	Supporting text
Selection bias			
Random sequence generation	**Low risk of bias**	The investigators describe a random component in the sequence generation process such as: • Referring to a random number table • Using a computer random number generator • Coin tossing • Shuffling cards or envelopes • Throwing dice • Drawing of lots • Minimization* *Minimization may be implemented without a random element, and this is considered to be equivalent to being random	Computer generated random numbers
Allocation concealment	**Unclear risk of bias**	Insufficient information to permit judgement of 'Low risk' or 'High risk'. This is usually the case if the method of concealment is not described or not described in sufficient detail to allow a definite judgement – for example if the use of assignment envelopes is described, but it remains unclear whether envelopes were sequentially numbered, opaque and sealed.	Allocation concealment not described
Performance bias			
Blinding of participants and personnel *Assessments should be made for each main outcome (or class of outcomes).*	**High risk of bias**	Any one of the following: • No blinding or incomplete blinding, and the outcome is likely to be influenced by lack of blinding • Blinding of key study participants and personnel attempted, but likely that the blinding could have been broken, and the outcome is likely to be influenced by lack of blinding	No blinding. Clinical management differed according to test results
Detection bias			
Blinding of outcome assessment *Assessments should be made for each main outcome (or*	**High risk of bias**	Any one of the following: • No blinding of outcome assessment, and the outcome measurement is likely to be influenced by lack of blinding • Blinding of outcome assessment, but likely that the blinding could have been broken and the outcome measurement are likely to be influenced by lack of blinding	No Blinding. Clinical management and therefore likelihood of detecting disease differed according to test results.

Domain	Judgement	Criteria	Supporting text
class of outcomes).			
Attrition bias			
Incomplete outcome data *Assessments should be made for each main outcome (or class of outcomes).*	**Unclear risk of bias**	Any one of the following: • Insufficient reporting of attrition/exclusions to permit judgement of 'Low risk' or 'High risk' (e.g. number randomized not stated, no reasons for missing data provided) • The study did not address this outcome	Analysis was by intention to treat. About a third of randomised women not attending
Reporting bias			
Selective reporting.	**Low risk of bias**	Any of the following: • The study protocol is available and all of the study's pre-specified (primary and secondary) outcomes that are of interest in the review have been reported in the pre-specified way • The study protocol is not available but it is clear that the published reports include all expected outcomes, including those that were pre-specified (convincing text of this nature may be uncommon)	No obvious evidence of missing outcomes although reporting is limited
Other bias			
Other sources of bias.	**High risk of bias**	There is at least one important risk of bias. For example, the study: • Had a potential source of bias related to the specific study design used or • Has been claimed to have been fraudulent or • Had some other problem	About a third of randomised women not attending. No informed consent asked from participants.

SANKARANARAYANAN 2009

Domain	Judgement	Criteria	Supporting text
Selection bias			
Random sequence generation	**Low risk of bias**	The investigators describe a random component in the sequence generation process such as: • Referring to a random number table • Using a computer random number generator • Coin tossing • Shuffling cards or envelopes • Throwing dice • Drawing of lots • Minimization* *Minimization may be implemented without a random element, and this is considered to be equivalent to being random	Method of randomisation was not reported in the original paper. IQWiG 2011 reports the randomisation as adequate based on correspondence with the authors
Allocation concealment	**Unclear risk of bias**	Insufficient information to permit judgement of 'Low risk' or 'High risk'. This is usually the case if the method of concealment is not described or not described in sufficient detail to allow a definite judgement – for example if the use of assignment envelopes is described, but it remains unclear whether envelopes were sequentially numbered, opaque and sealed.	Not Reported
Performance bias			
Blinding of participants and personnel *Assessments should be made for each main outcome (or class of outcomes).*	**High risk of bias**	Any one of the following: • No blinding or incomplete blinding, and the outcome is likely to be influenced by lack of blinding • Blinding of key study participants and personnel attempted, but likely that the blinding could have been broken, and the outcome is likely to be influenced by lack of blinding	No blinding. Clinical management differed according to test results
Detection bias			
Blinding of outcome assessment *Assessments should be made for each main outcome (or*	**Low risk of bias**	Any one of the following: • No blinding of outcome assessment, but the review authors judge that the outcome measurement is not likely to be influenced by lack of blinding • Blinding of outcome assessment ensured, and unlikely that the blinding could have been broken	Outcome assessment performed by Cancer registry personnel

Domain	Judgement	Criteria	Supporting text
class of outcomes).			
Attrition bias			
Incomplete outcome data *Assessments should be made for each main outcome (or class of outcomes).*	**Unclear risk of bias**	Any one of the following: • Insufficient reporting of attrition/exclusions to permit judgement of 'Low risk' or 'High risk' (e.g. number randomized not stated, no reasons for missing data provided) • The study did not address this outcome	Analysis was by intention to treat with cluster as the unit of analysis. About 30% of women did not attend screening
Reporting bias			
Selective reporting.	**Low risk of bias**	Any of the following: • The study protocol is available and all of the study's pre-specified (primary and secondary) outcomes that are of interest in the review have been reported in the pre-specified way • The study protocol is not available but it is clear that the published reports include all expected outcomes, including those that were pre-specified (convincing text of this nature may be uncommon)	No obvious evidence of missing outcomes although reporting is limited
Other bias			
Other sources of bias.	**High risk of bias**	There is at least one important risk of bias. For example, the study: • Had a potential source of bias related to the specific study design used or • Has been claimed to have been fraudulent or • Had some other problem	About 30% of randomised women not attending Not clear whether cluster randomisation resulted in comparable groups

NTCC

Domain	Judgement	Criteria	Supporting text
Selection bias			
Random sequence generation	**Unclear risk of bias**	Insufficient information about the sequence generation process to permit judgement of 'Low risk' or 'High risk'	Computer generated random numbers in two locations, sealed numbered envelopes in remaining centres. Not clear how the random sequence was generated before being placed in the envelopes
Allocation concealment	**Unclear risk of bias**	Insufficient information to permit judgement of 'Low risk' or 'High risk'. This is usually the case if the method of concealment is not described or not described in sufficient detail to allow a definite judgement – for example if the use of assignment envelopes is described, but it remains unclear whether envelopes were sequentially numbered, opaque and sealed.	Computer generated random numbers or sealed numbered envelopes
Performance bias			
Blinding of participants and personnel *Assessments should be made for each main outcome (or class of outcomes).*	**High risk of bias**	Any one of the following: • No blinding or incomplete blinding, and the outcome is likely to be influenced by lack of blinding • Blinding of key study participants and personnel attempted, but likely that the blinding could have been broken, and the outcome is likely to be influenced by lack of blinding	No blinding. Clinical management differed according to test results
Detection bias			
Blinding of outcome assessment *Assessments should be made for each main outcome (or class of outcomes).*	**High risk of bias**	Any one of the following: • No blinding of outcome assessment, and the outcome measurement is likely to be influenced by lack of blinding • Blinding of outcome assessment, but likely that the blinding could have been broken and the outcome measurement are likely to be influenced by lack of blinding	No blinding. Colposcopists had access to all screening test results
Attrition bias			
Incomplete outcome data	**Unclear risk of bias**	Any one of the following: • Insufficient reporting of attrition/exclusions to permit judgement of 'Low risk' or 'High risk' (e.g.	Analysis by intention to treat. Missing data unbalanced between groups

Domain	Judgement	Criteria	Supporting text
Assessments should be made for each main outcome (or class of outcomes).		number randomized not stated, no reasons for missing data provided) • The study did not address this outcome	
Reporting bias			
Selective reporting.	**Low risk of bias**	Any of the following: • The study protocol is available and all of the study's pre-specified (primary and secondary) outcomes that are of interest in the review have been reported in the pre-specified way • The study protocol is not available but it is clear that the published reports include all expected outcomes, including those that were pre-specified (convincing text of this nature may be uncommon)	No evidence of missing outcomes
Other bias			
Other sources of bias.	**Low risk of bias**	The study appears to be free of other sources of bias.	The study appears to be free of other sources of bias.

POBASCAM

Domain	Judgement	Criteria	Supporting text
Selection bias			
Random sequence generation	**Low risk of bias**	The investigators describe a random component in the sequence generation process such as: • Referring to a random number table • Using a computer random number generator • Coin tossing • Shuffling cards or envelopes • Throwing dice • Drawing of lots • Minimization* *Minimization may be implemented without a random element, and this is considered to be equivalent to being random	Computer generated random numbers
Allocation concealment	**Low risk of bias**	Participants and investigators enrolling participants could not foresee assignment because one of the following, or an equivalent method, was used to conceal allocation: • Central allocation (including telephone, web-based and pharmacy-controlled randomization) • Sequentially numbered drug containers of identical appearance • Sequentially numbered, opaque, sealed envelopes	Randomised after the cervical specimen had been taken and administrative data entered into the central study database.
Performance bias			
Blinding of participants and personnel *Assessments should be made for each main outcome (or class of outcomes).*	**High risk of bias**	Any one of the following: • No blinding or incomplete blinding, and the outcome is likely to be influenced by lack of blinding • Blinding of key study participants and personnel attempted, but likely that the blinding could have been broken, and the outcome is likely to be influenced by lack of blinding	Technicians performing the tests were blinded to group assignment. Patients and clinicians could not be blinded as clinical management was based on available test results
Detection bias			
Blinding of outcome assessment *Assessments should be made for each main*	**Unclear risk of bias**	Any one of the following: • Insufficient information to permit judgement of 'Low risk' or 'High risk' • The study did not address this outcome	Not described whether colposcopists were aware of screening test results

Domain	Judgement	Criteria	Supporting text
outcome (or class of outcomes).			
Attrition bias			
Incomplete outcome data *Assessments should be made for each main outcome (or class of outcomes).*	**Low risk of bias**	Any one of the following: • No missing outcome data • Reasons for missing outcome data unlikely to be related to true outcome (for survival data, censoring unlikely to be introducing bias) • Missing outcome data balanced in numbers across intervention groups, with similar reasons for missing data across groups • For dichotomous outcome data, the proportion of missing outcomes compared with observed event risk not enough to have a clinically relevant impact on the intervention effect estimate • For continuous outcome data, plausible effect size (difference in means or standardized difference in means) among missing outcomes not enough to have a clinically relevant impact on observed effect size • Missing data have been imputed using appropriate methods	Analysis by intention to treat. Missing data approximately balanced between groups and missing for the same reasons
Reporting bias			
Selective reporting.	**Low risk of bias**	Any of the following: • The study protocol is available and all of the study's pre-specified (primary and secondary) outcomes that are of interest in the review have been reported in the pre-specified way • The study protocol is not available but it is clear that the published reports include all expected outcomes, including those that were pre-specified (convincing text of this nature may be uncommon)	Trial protocol is available. Specified outcomes reported in main trial report
Other bias			
Other sources of bias.	**Low risk of bias**	The study appears to be free of other sources of bias.	The study appears to be free of other sources of bias.

SWEDESCREEN

Domain	Judgement	Criteria	Supporting text
Selection bias			
Random sequence generation	**Low risk of bias**	The investigators describe a random component in the sequence generation process such as: • Referring to a random number table • Using a computer random number generator • Coin tossing • Shuffling cards or envelopes • Throwing dice • Drawing of lots • Minimization* *Minimization may be implemented without a random element, and this is considered to be equivalent to being random	Computer generated random numbers
Allocation concealment	**Low risk of bias**	Participants and investigators enrolling participants could not foresee assignment because one of the following, or an equivalent method, was used to conceal allocation: • Central allocation (including telephone, web-based and pharmacy-controlled randomization) • Sequentially numbered drug containers of identical appearance • Sequentially numbered, opaque, sealed envelopes	Central Allocation
Performance bias			
Blinding of participants and personnel *Assessments should be made for each main outcome (or class of outcomes).*	**Unclear risk of bias**	Any one of the following: • Insufficient information to permit judgement of 'Low risk' or 'High risk' • The study did not address this outcome	Unclear reporting. Clinical management appears to be based on test results but paper claims that clinicians and participants were unaware of HPV test results and group assignment. Blinding discontinued 3 years after completion of enrolment.
Detection bias			
Blinding of outcome assessment *Assessments should be made for each main*	**Low risk of bias**	Any one of the following: • No blinding of outcome assessment, but the review authors judge that the outcome measurement is not likely to be influenced by lack of blinding • Blinding of outcome assessment ensured, and unlikely that the blinding could have been broken	Histological results were obtained from registry data. All histological samples with abnormal diagnosis and all biopsies from study colposcopy were confirmed by an independent pathologist unaware of group assignment

Domain	Judgement	Criteria	Supporting text
outcome (or class of outcomes).			
Attrition bias			
Incomplete outcome data *Assessments should be made for each main outcome (or class of outcomes).*	**Unclear risk of bias**	Any one of the following: • Insufficient reporting of attrition/exclusions to permit judgement of 'Low risk' or 'High risk' (e.g. number randomized not stated, no reasons for missing data provided) • The study did not address this outcome	Analysis by intention to treat. A number of participants not following protocol or missing data at each step
Reporting bias			
Selective reporting.	**Low risk of bias**	Any of the following: • The study protocol is available and all of the study's pre-specified (primary and secondary) outcomes that are of interest in the review have been reported in the pre-specified way • The study protocol is not available but it is clear that the published reports include all expected outcomes, including those that were pre-specified (convincing text of this nature may be uncommon)	No evidence of missing outcomes
Other bias			
Other sources of bias.	**Low risk of bias**	The study appears to be free of other sources of bias.	The study appears to be free of other sources of bias.

APPENDIX 3: GRADE EVIDENCE PROFILES

Table 13: GRADE evidence profile – disease specific survival

Author(s): Richard Birnie, Robert Wolff, Jos Kleijnen
Date: 2014-02-07
Question: HPV test vs Conventional Cytology - Disease specific survival for cervical cancer
Settings:
Bibliography: Clinical effectiveness of HPV testing (alone or in combination with cytology), compared to cytology alone, in population screening for cervical cancer

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	**Design**	**Risk of bias#**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**HPV test**	**Conventional Cytology - Disease specific survival**	**Relative (95% CI)**	**Absolute**		
Disease Specific Survival (follow-up mean 8 years)[a]												
1	randomised trials	serious[1]	no serious inconsistency[2]	serious[3]	no serious imprecision	none[4]	34/34126 (0.1%)	54/32058 (0.17%)	RR 0.59 (0.39 to 0.91)	1 fewer per 1000 (from 0 fewer to 1 fewer)	⊕⊕OO LOW	

Detailed risk of bias assessments for individual studies are presented in appendix 2
Included studies: *a –Sankaranarayanan 2009*
[1] Blinding not reported, Allocation concealment unclear. Missing data - about 30% of women did not attend screening
[2] Inconsistency cannot be graded for a single study
[3] Study conducted in a rural Indian population whereas the target population for the guideline is Germany
[4] Difficult to detect publication bias with a single study. Funnel plot based methods are not applicable

Table 14: GRADE evidence profile – incidence of cervical cancer

Author(s): Richard Birnie, Robert Wolff, Jos Kleijnen
Date: 2014-02-07
Question: HPV test vs Cytology - Incidence of Cervical cancer for cervical cancer
Settings:
Bibliography: Clinical effectiveness of HPV testing (alone or in combination with cytology), compared to cytology alone, in population screening for cervical cancer

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias[#]	Inconsistency	Indirectness	Imprecision	Other considerations	HPV test	Cytology - Incidence of Cervical cancer	Relative (95% CI)	Absolute		
Incidence of Cervical Cancer - 1st Screening round (follow-up 3-5 years)[a]												
5	randomised trials	very serious[1,2,3]	serious[4]	no serious indirectness	serious[5]	none	49/187732 (0.03%)	50/174978 (0.03%)	RR 0.89 (0.45 to 1.75)	0 fewer per 1000 (from 0 fewer to 0 more)	⊕OOO VERY LOW	
Incidence of Cervical Cancer - Subgroup - HPV + Cytology - 1st round (follow-up 3-5 years)[b]												
4	randomised trials	very serious[1,2,3]	serious[6]	no serious indirectness	serious[5]	none	44/163071 (0.03%)	48/150443 (0.03%)	RR 0.77 (0.38 to 1.58)	0 fewer per 1000 (from 0 fewer to 0 more)	⊕OOO VERY LOW	
Incidence of Cervical Cancer - Subgroup - LBC comparator, 3 year interval - 1st round (follow-up mean 3 years)[c]												
3	randomised trials	serious[2,3]	serious[6]	no serious indirectness	serious[4]	none	12/66055 (0.02%)	13/53125 (0.02%)	RR 0.63 (0.18 to 2.19)	0 fewer per 1000 (from 0 fewer to 0 more)	⊕OOO VERY LOW	
Incidence of Cervical Cancer - 2nd Screening round (follow-up mean 3 years)[d]												
3	randomised trials	serious[2,3]	serious[5]	no serious indirectness	serious[5]	none	3/57747 (0.005%)	9/50568 (0.02%)	RR 0.29 (0.04 to 2.26)	0 fewer per 1000 (from 0 fewer to 0 more)	⊕OOO VERY LOW	

Incidence of Cervical Cancer - Subgroup - HPV + Cytology - 2nd round (follow-up mean 3 years)[e]												
2	randomised trials	serious[2,3]	serious[6]	no serious indirectness	serious[5]	none	3/33769 (0.009%)	6/26196 (0.02%)	RR 0.42 (0.01 to 12.07)	0 fewer per 1000 (from 0 fewer to 3 more)		

Detailed risk of bias assessments for individual studies are presented in appendix 2

Included Studies: *a* – ARTISTIC, Leinonen 2012, NTCC-I, NTCC-II, POBASCAM; *b* – ARTISTIC, Leinonen 2012, NTCC-I, POBASCAM; *c* – ARTISTIC, NTCC-I, NTCC-II; *d* – ARTISTIC, NTCC-I, NTCC-II; *e* – ARTISTIC, NTCC-I

[1] Allocation concealment was unclear in some studies

[2] Blinding of participants and personnel absent or inadequate in several studies

[3] Missing outcome data not accounted for in several studies

[4] Effect estimates go in different directions in different studies

[5] Widely differing estimates each with wide confidence intervals

[6] I-squared = 50% Effect estimates in differing directions

Table 15: GRADE evidence profile – incidence of CIN3

Author(s): Richard Birnie, Robert Wolff, Jos Kleijnen
Date: 2014-02-07
Question: HPV test vs Cytology - Incidence of CIN3 for cervical cancer
Settings:
Bibliography: Clinical effectiveness of HPV testing (alone or in combination with cytology), compared to cytology alone, in population screening for cervical cancer

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	**Design**	**Risk of bias[#]**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**HPV test**	**Cytology - Incidence of CIN3/CIS**	**Relative (95% CI)**	**Absolute**		
Incidence of CIN3 - 1st Screening round (follow-up 3-5 years)[a]												
4	randomised trials	serious[1,2]	very serious[3]	no serious indirectness	no serious imprecision	none	572/169646 (0.34%)	415/168854 (0.25%)	RR 1.51 (1.08 to 2.10)	1 more per 1000 (from 0 more to 3 more)	⊕OOO VERY LOW	
Incidence of CIN3 - LBC comparator, 3 year interval, 1st round (follow-up 3 years)[b]												
2	randomised trials	serious[1,2]	very serious[3]	no serious indirectness	no serious imprecision	none	165/47969 (0.34%)	82/47001 (0.17%)	RR 2.01 (0.95 to 4.23)	2 more per 1000 (from 0 fewer to 6 more)	⊕OOO VERY LOW	
Incidence of CIN3 - Conventional Cytology comparator, >=4 year interval (follow-up 4-5)[c]												
2	randomised trials	serious[2,4]	no serious inconsistency	no serious indirectness	no serious imprecision	none	407/121677 (0.33%)	333/121853 (0.27%)	RR 1.22 (1.04 to 1.44)	1 more per 1000 (from 0 more to 1 more)	⊕⊕⊕O MODERATE	
Incidence of CIN3 - HPV + Cytology: all tests, 1st round (follow-up 3-5 years)[d]												
3	randomised trials	serious[1,2,4]	no serious inconsistency	no serious indirectness	no serious imprecision	none	480/144985 (0.33%)	384/144319 (0.27%)	RR 1.25 (1.09 to 1.42)	1 more per 1000 (from 0 more to 1 more)	⊕⊕⊕O MODERATE	

Incidence of CIN3 - HPV + Cytology: Combined tests - exclude triage, 1st round (follow-up 3-4 years)[e]												
2	randomised trials	serious[1,2]	no serious inconsistency	no serious indirectness	no serious imprecision	none	232/43307 (0.54%)	195/42572 (0.46%)	RR 1.18 (0.97 to 1.43)	1 more per 1000 (from 0 fewer to 2 more)	⊕⊕⊕O MODERATE	
Incidence of CIN3 - 2nd Screening round (follow-up 3 years)[f]												
2	randomised trials	serious[1,2]	very serious[5]	no serious indirectness	serious[5]	none	18/46071 (0.04%)	33/46702 (0.07%)	RR 0.52 (0.13 to 2.04)	0 fewer per 1000 (from 1 fewer to 1 fewer)	⊕OOO VERY LOW	

Detailed risk of bias assessments for individual studies are presented in appendix 2

Included studies: *a* – Leinonen 2012, NTCC-I, NTCC-II, POBASCAM; *b* – NTCC-I, NTCC-II; *c* – Leinonen 2012, POBASCAM; *d* – Leinonen 2012, NTCC-I, POBASCAM; *e* – NTCC-I, POBASCAM; *f* – NTCC-I, NTCC-II

[1] Allocation concealment unclear in some studies

[2] Blinding of participants and personnel inadequate in some studies

[3] I-squared >75%

[4] Possible missing outcome data, unclear

[5] Widely differing estimates each with wide confidence intervals

Table 16: GRADE evidence profile – incidence of CIN3+

Author(s): Richard Birnie, Robert Wolff, Jos Kleijnen
Date: 2014-02-07
Question: HPV test vs Cytology - Incidence of CIN3+ for cervical cancer
Settings:
Bibliography: Clinical effectiveness of HPV testing (alone or in combination with cytology), compared to cytology alone, in population screening for cervical cancer

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias[#]	Inconsistency	Indirectness	Imprecision	Other considerations	HPV test	Cytology - Incidence of CIN3+	Relative (95% CI)	Absolute		
Incidence of CIN3+ - 1st screening round (follow-up 3-5 years)[a]												
6	randomised trials	serious[1,2,3]	very serious[4]	no serious indirectness	serious[5]	none	873/194358 (0.45%)	653/181179 (0.36%)	RR 1.23 (0.91 to 1.67)	1 more per 1000 (from 0 fewer to 2 more)	⊕OOO VERY LOW	
Incidence of CIN3+ - 3 year interval - 1st round (follow-up 3 years)[b]												
4	randomised trials	serious[2]	very serious[4]	no serious indirectness	no serious imprecision	none	482/72612 (0.66%)	230/59395 (0.39%)	RR 1.44 (0.91 to 2.25)	2 more per 1000 (from 0 fewer to 5 more)	⊕OOO VERY LOW	
Incidence of CIN3+ - >=4 year interval (follow-up 4-5 years)[c]												
2	randomised trials	serious[2]	very serious[4]	no serious indirectness	no serious imprecision	none	391/121746 (0.32%)	423/121784 (0.35%)	RR 0.95 (0.68 to 1.35)	0 fewer per 1000 (from 1 fewer to 1 more)	⊕OOO VERY LOW	
Incidence of CIN3+ - HPV + Cytology: all tests, 1st round (follow-up 3-5 years)[d]												
5	randomised trials	serious[2,3]	serious[6]	no serious indirectness	no serious imprecision	none	776/169697 (0.46%)	620/156644 (0.4%)	RR 1.04 (0.86 to 1.26)	0 more per 1000 (from 1 fewer to 1 more)	⊕⊕OO LOW	

Incidence of CIN3+ - HPV + Cytology: Combined tests - exclude triage, 1st round (follow-up 3-4 years)[e]												
4	randomised trials	serious[2,3]	no serious inconsistency	no serious indirectness	no serious imprecision	none	556/67950 (0.82%)	347/54966 (0.63%)	RR 1.12 (0.97 to 1.28)	1 more per 1000 (from 0 fewer to 2 more)	⊕⊕⊕O MODERATE	
Incidence of CIN3+ - LBC comparator, 1st round (follow-up 3 years)[f]												
3	randomised trials	serious[2,3]	very serious[4]	no serious indirectness	no serious imprecision	none	410/66355 (0.62%)	175/53125 (0.33%)	RR 1.49 (0.79 to 2.8)	2 more per 1000 (from 1 fewer to 6 more)	⊕OOO VERY LOW	
Incidence of CIN3+ - Conventional Cytology Comparator, 1st round (follow-up 3-5 years)[g]												
3	randomised trials	serious[2,3]	very serious[4]	no serious indirectness	no serious imprecision	none	463/128003 (0.36%)	478/128054 (0.37%)	RR 1.04 (0.77 to 1.4)	0 more per 1000 (from 1 fewer to 1 more)	⊕OOO VERY LOW	
Incidence of CIN3+ - 2nd screening round[h]												
4	randomised trials	serious[2,3]	no serious inconsistency	no serious indirectness	no serious imprecision	none	66/63890 (0.1%)	90/56762 (0.16%)	RR 0.52 (0.35 to 0.76)	1 fewer per 1000 (from 0 fewer to 1 fewer)	⊕⊕⊕O MODERATE	
Incidence of CIN3+ - HPV + Cytology: exclude single HPV tests, 2nd round (follow-up 3 years)[i]												
3	randomised trials	serious[2,3]	no serious inconsistency	no serious indirectness	no serious imprecision	none	61/39912 (0.15%)	67/32390 (0.21%)	RR 0.59 (0.42 to 0.85)	1 fewer per 1000 (from 0 fewer to 1 fewer)	⊕⊕⊕O MODERATE	

Detailed risk of bias assessments for individual studies are presented in appendix 2

Included studies: *a* – ARTISTIC, Leinonen 2012, NTCC-I, NTCC-II, POBASCAM, Swedescreen; *b* – ARTISTIC, NTCC-I, NTCC-II, Swedescreen; *c* – Leinonen 2012, POBASCAM; *d* – ARTISTIC, Leinonen 2012, NTCC-I, POBASCAM, Swedescreen; *e* – ARTISTIC, NTCC-I, POBASCAM, Swedescreen; *f* – ARTISTIC, NTCC-I, NTCC-II; *g* - Leinonen 2012, POBASCAM, Swedescreen; *h* – ARTISTIC, NTCC-I, NTCC-II, Swedescreen; *i* – ARTISTIC, NTCC-I, Swedescreen

[1] Allocation concealment unclear in some studies

[2] Blinding of participants and personnel inadequate or unclear in some studies

[3] Missing outcome data in some studies

[4] I-squared >75%

[5] Effect estimates go in different directions in different studies
[6] I-squared >50%

Table 17: GRADE evidence profile – incidence of CIN2+

Author(s): Richard Birnie, Robert Wolff, Jos Kleijnen
Date: 2014-02-07
Question: HPV test vs Cytology - Incidence of CIN2+ for cervical cancer
Settings:
Bibliography: Clinical effectiveness of HPV testing (alone or in combination with cytology), compared to cytology alone, in population screening for cervical cancer

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias[#]	Inconsistency	Indirectness	Imprecision	Other considerations	HPV test	Cytology - Incidence of CIN2+	Relative (95% CI)	Absolute		
Incidence of CIN2+ - 1st screening round (follow-up 3-5 years)[a]												
6	randomised trials	serious[1,2]	very serious[3]	no serious indirectness	no serious imprecision	none	1907/194289 (0.98%)	1097/181248 (0.61%)	RR 1.51 (1.21 to 1.9)	3 more per 1000 (from 1 more to 5 more)	⊕OOO VERY LOW	
Incidence of CIN2+ - 3 year interval, 1st round (follow-up 3 years)[b]												
4	randomised trials	serious[1,2]	very serious[3]	no serious indirectness	no serious imprecision	none	922/72612 (1.3%)	372/59395 (0.63%)	RR 1.64 (1.07 to 2.51)	4 more per 1000 (from 0 more to 9 more)	⊕OOO VERY LOW	
Incidence of CIN2+ - >=4 year interval (follow-up 4-5 years)[c]												
2	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	no serious imprecision	none	985/121677 (0.81%)	725/121853 (0.59%)	RR 1.35 (1.21 to 1.51)	2 more per 1000 (from 1 more to 3 more)	⊕⊕⊕O MODERATE	
Incidence of CIN2+ - HPV + Cytology: all tests, 1st round[d]												
5	randomised trials	serious[1,2]	no serious inconsistency	no serious indirectness	no serious imprecision	none	1689/169628 (1%)	1024/156713 (0.65%)	RR 1.33 (1.19 to 1.47)	2 more per 1000 (from 1 more to 3 more)	⊕⊕⊕O MODERATE	

Incidence of CIN2+ - HPV + Cytology: Combined tests - exclude triage, 1st round (follow-up 3-4 years)[e]												
4	randomised trials	serious[1,2]	no serious inconsistency	no serious indirectness	no serious imprecision	none	971/67950 (1.4%)	514/54966 (0.94%)	RR 1.28 (1.12 to 1.47)	3 more per 1000 (from 1 more to 4 more)	⊕⊕⊕O MODERATE	
Incidence of CIN2+ - LBC comparator, 1st round (follow-up 3 years)[f]												
3	randomised trials	serious[1,2]	very serious[3]	no serious indirectness	no serious imprecision	none	808/66355 (1.2%)	296/53125 (0.56%)	RR 1.69 (0.95 to 3.02)	4 more per 1000 (from 0 fewer to 11 more)	⊕OOO VERY LOW	
Incidence of CIN2+ - Conventional Cytology comparator, 1st round (follow-up 3-5 years)[g]												
3	randomised trials	serious[1,2]	no serious inconsistency	no serious indirectness	no serious imprecision	none	1099/127934 (0.86%)	801/128123 (0.63%)	RR 1.37 (1.26 to 1.5)	2 more per 1000 (from 2 more to 3 more)	⊕⊕⊕O MODERATE	
Incidence of CIN2+ - 2nd screening round (follow-up 3 years)[h]												
4	randomised trials	serious[1,2]	no serious inconsistency	no serious indirectness	no serious imprecision	none	124/63890 (0.19%)	142/56762 (0.25%)	RR 0.57 (0.42 to 0.77)	1 fewer per 1000 (from 1 fewer to 1 fewer)	⊕⊕⊕O MODERATE	
Incidence of CIN2+ - HPV + Cytology: all tests, 2nd round (follow-up 3 years)[i]												
3	randomised trials	serious[1,2]	no serious inconsistency	no serious indirectness	no serious imprecision	none	112/39912 (0.28%)	104/32390 (0.32%)	RR 0.65 (0.49 to 0.85)	1 fewer per 1000 (from 0 fewer to 2 fewer)	⊕⊕⊕O MODERATE	
Incidence of CIN2+ - LBC comparator, 2nd round (follow-up 3 years)[j]												
3	randomised trials	serious[1,2]	serious[4]	no serious indirectness	no serious imprecision	none	99/57747 (0.17%)	99/50568 (0.2%)	RR 0.56 (0.36 to 0.87)	1 fewer per 1000 (from 0 fewer to 1 fewer)	⊕⊕OO LOW	

Detailed risk of bias assessments for individual studies are presented in appendix 2

Included studies: *a* – ARTISTIC, Leinonen 2012, NTCC-I, NTCC-II, POBASCAM, Swedescreen; *b* – ARTISTIC, NTCC-I, NTCC-II; *c* – Leinonen 2012, POBASCAM; *d* – ARTISTIC, Leinonen 2012, NTCC-I, POBASCAM, Swedescreen; *e* – ARTISTIC, NTCC-I, POBASCAM, Swedescreen; *f* – ARTISTIC, NTCC-I, NTCC-II; *g* – Leinonen 2012, POBASCAM, Swedescreen; *h* – ARTISTIC, NTCC-I, NTCC-II, Swedescreen; *i* – ARTISTIC, NTCC-I, Swedescreen; *j* – ARTISTIC, NTCC-I, NTCC-II

[1] Blinding of participants and personnel inadequate or unclear in some studies
[2] Missing outcome data in some studies
[3] I-squared >75%
[4] I-squared >50%

Table 18: GRADE evidence profile – screening related harm

Author(s): Richard Birnie, Robert Wolff, Jos Kleijnen
Date: 2014-02-07
Question: HPV test vs Cytology - Screening related harm for cervical cancer
Settings:
Bibliography: Clinical effectiveness of HPV testing (alone or in combination with cytology), compared to cytology alone, in population screening for cervical cancer

Quality assessment							No of patients		Effect		Quality	Importance
No of studies	**Design**	**Risk of bias[#]**	**Inconsistency**	**Indirectness**	**Imprecision**	**Other considerations**	**HPV test**	**Cytology - Screening related harm**	**Relative (95% CI)**	**Absolute**		
Screening related harm (follow-up 3 years)[a]												
1	randomised trials	serious[1]	no serious inconsistency[2]	no serious indirectness	no serious imprecision	none	223/593 (37.6%)	717/1872 (38.3%)	RR 0.98 (0.87 to 1.11)	8 fewer per 1000 (from 50 fewer to 42 more)	⊕⊕⊕O MODERATE	

Detailed risk of bias assessments for individual studies are presented in appendix 2
Included studies: *a* - ARTISTIC
[1] Blinding of participants and personnel inadequate
[2] Inconsistency cannot be graded for a single study

APPENDIX 4: META-ANALYSES

Figure 2: Forest plot – disease specific survival

Study or Subgroup	HPV test Events	HPV test Total	Conventional Cytology Events	Conventional Cytology Total	Weight	Risk Ratio M-H, Random, 95% CI
Sankaranarayanan 2009	34	34126	54	32058		0.59 [0.39, 0.91]

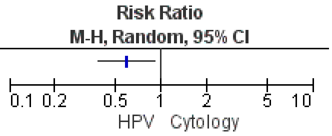
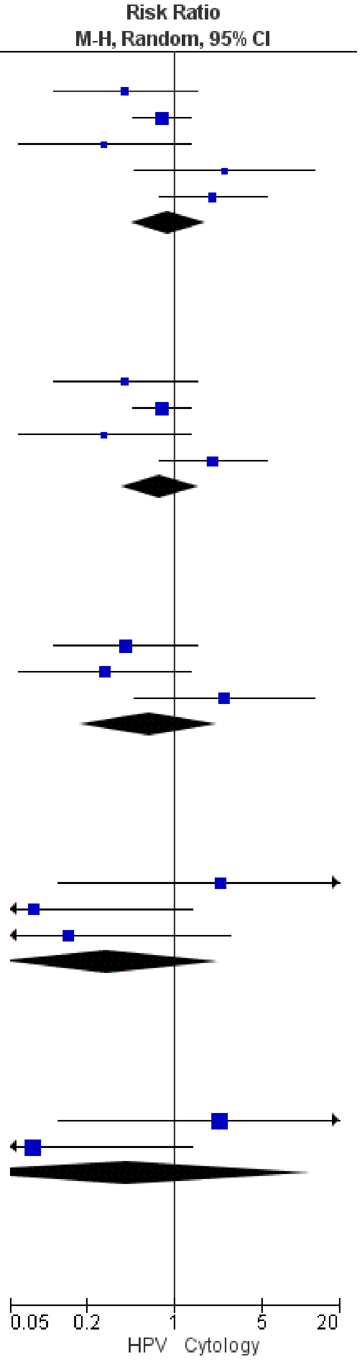


Figure 3: Meta-analysis – incidence of cervical cancer

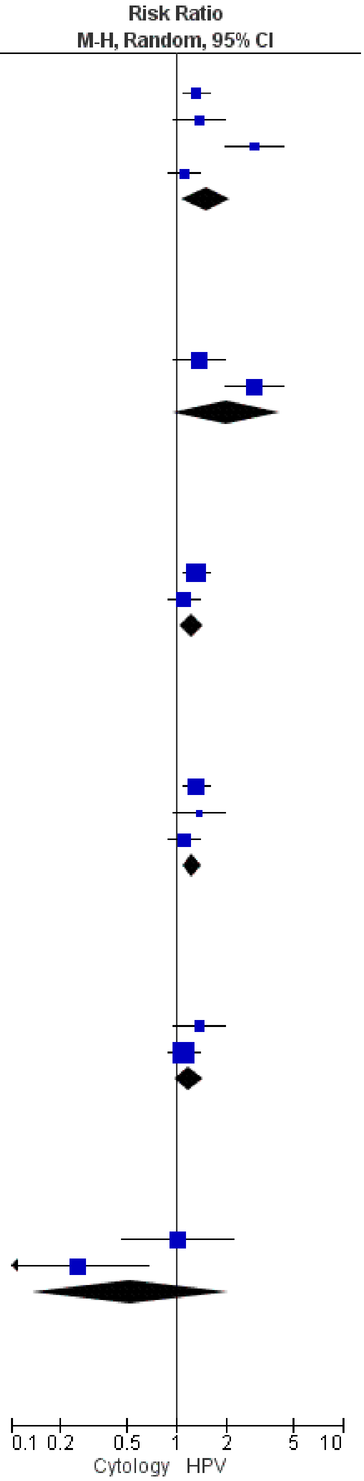
Study or Subgroup	HPV test Events	HPV test Total	Cytology Events	Cytology Total	Weight	Risk Ratio M-H, Random, 95% CI
2.1.1 1st Screening round						
ARTISTIC 1st round	5	18686	4	6124	16.6%	0.41 [0.11, 1.53]
Leinonen 2012	25	101678	31	101747	34.9%	0.81 [0.48, 1.37]
NTCC-I 1st round	2	22708	7	22466	13.1%	0.28 [0.06, 1.36]
NTCC-II 1st round	5	24661	2	24535	12.3%	2.49 [0.48, 12.82]
POBASCAM	12	19999	6	20106	23.0%	2.01 [0.75, 5.36]
Subtotal (95% CI)		**187732**		**174978**	**100.0%**	**0.89 [0.45, 1.75]**
Total events	49		50			
Heterogeneity: Tau² = 0.27; Chi² = 7.69, df = 4 (P = 0.10); I² = 48%						
Test for overall effect: Z = 0.33 (P = 0.74)						
2.1.2 Subgroup - HPV + Cytology - 1st round						
ARTISTIC 1st round	5	18686	4	6124	18.8%	0.41 [0.11, 1.53]
Leinonen 2012	25	101678	31	101747	40.1%	0.81 [0.48, 1.37]
NTCC-I 1st round	2	22708	7	22466	14.8%	0.28 [0.06, 1.36]
POBASCAM	12	19999	6	20106	26.2%	2.01 [0.75, 5.36]
Subtotal (95% CI)		**163071**		**150443**	**100.0%**	**0.77 [0.38, 1.58]**
Total events	44		48			
Heterogeneity: Tau² = 0.26; Chi² = 6.06, df = 3 (P = 0.11); I² = 50%						
Test for overall effect: Z = 0.71 (P = 0.48)						
2.1.3 Subgroup - LBC comparator, 3 year interval - 1st round						
ARTISTIC 1st round	5	18686	4	6124	37.6%	0.41 [0.11, 1.53]
NTCC-I 1st round	2	22708	7	22466	31.9%	0.28 [0.06, 1.36]
NTCC-II 1st round	5	24661	2	24535	30.5%	2.49 [0.48, 12.82]
Subtotal (95% CI)		**66055**		**53125**	**100.0%**	**0.63 [0.18, 2.19]**
Total events	12		13			
Heterogeneity: Tau² = 0.62; Chi² = 4.10, df = 2 (P = 0.13); I² = 51%						
Test for overall effect: Z = 0.73 (P = 0.47)						
2.1.4 2nd Screening round						
ARTISTIC 2nd round	3	11676	0	3866	32.9%	2.32 [0.12, 44.87]
NTCC-I 2nd round	0	22093	6	22330	34.2%	0.08 [0.00, 1.38]
NTCC-II 2nd round	0	23978	3	24372	32.9%	0.15 [0.01, 2.81]
Subtotal (95% CI)		**57747**		**50568**	**100.0%**	**0.29 [0.04, 2.26]**
Total events	3		9			
Heterogeneity: Tau² = 1.03; Chi² = 2.92, df = 2 (P = 0.23); I² = 32%						
Test for overall effect: Z = 1.18 (P = 0.24)						
2.1.5 Subgroup - HPV + Cytology - 2nd round						
ARTISTIC 2nd round	3	11676	0	3866	49.4%	2.32 [0.12, 44.87]
NTCC-I 2nd round	0	22093	6	22330	50.6%	0.08 [0.00, 1.38]
Subtotal (95% CI)		**33769**		**26196**	**100.0%**	**0.42 [0.01, 12.07]**
Total events	3		6			
Heterogeneity: Tau² = 3.68; Chi² = 2.66, df = 1 (P = 0.10); I² = 62%						
Test for overall effect: Z = 0.51 (P = 0.61)						



Test for subgroup differences: Chi² = 1.26, df = 4 (P = 0.87), I² = 0%

Figure 4: Meta-analysis – incidence of CIN3

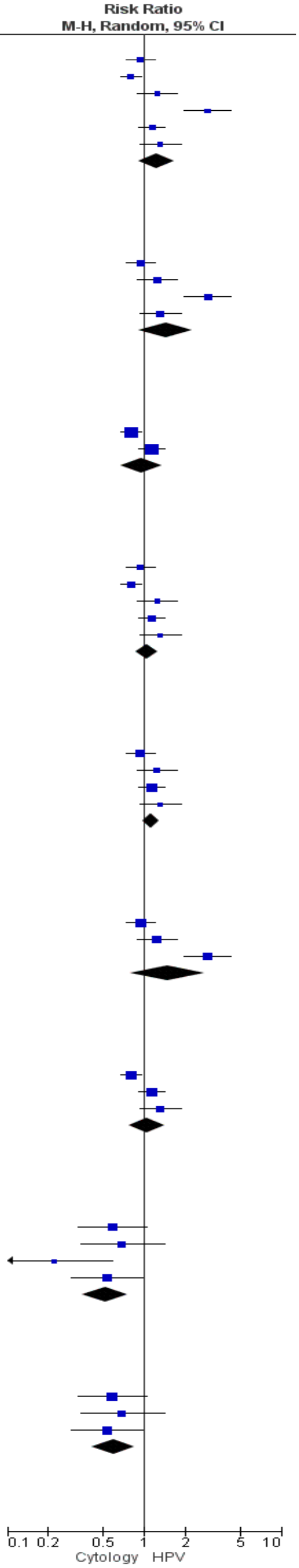
Study or Subgroup	HPV Events	HPV Total	Cytology Events	Cytology Total	Weight	Risk Ratio M-H, Random, 95% CI
3.1.1 1st Screening round						
Leinonen 2012	248	101678	189	101747	28.4%	1.31 [1.09, 1.59]
NTCC-I 1st round	73	23308	51	22466	22.9%	1.38 [0.97, 1.97]
NTCC-II 1st round	92	24661	31	24535	21.3%	2.95 [1.97, 4.43]
POBASCAM	159	19999	144	20106	27.4%	1.11 [0.89, 1.39]
Subtotal (95% CI)		**169646**		**168854**	**100.0%**	**1.51 [1.08, 2.10]**
Total events	572		415			
Heterogeneity: $Tau^2 = 0.09$; $Chi^2 = 17.28$, df = 3 (P = 0.0006); $I^2 = 83\%$						
Test for overall effect: Z = 2.43 (P = 0.01)						
3.1.2 LBC comparator, 3 year interval, 1st round						
NTCC-I 1st round	73	23308	51	22466	50.8%	1.38 [0.97, 1.97]
NTCC-II 1st round	92	24661	31	24535	49.2%	2.95 [1.97, 4.43]
Subtotal (95% CI)		**47969**		**47001**	**100.0%**	**2.01 [0.95, 4.23]**
Total events	165		82			
Heterogeneity: $Tau^2 = 0.25$; $Chi^2 = 7.62$, df = 1 (P = 0.006); $I^2 = 87\%$						
Test for overall effect: Z = 1.83 (P = 0.07)						
3.1.3 Conventional Cytology comparator, >=4 year interval						
Leinonen 2012	248	101678	189	101747	56.8%	1.31 [1.09, 1.59]
POBASCAM	159	19999	144	20106	43.2%	1.11 [0.89, 1.39]
Subtotal (95% CI)		**121677**		**121853**	**100.0%**	**1.22 [1.04, 1.44]**
Total events	407		333			
Heterogeneity: $Tau^2 = 0.00$; $Chi^2 = 1.26$, df = 1 (P = 0.26); $I^2 = 20\%$						
Test for overall effect: Z = 2.40 (P = 0.02)						
3.1.4 HPV + Cytology: all tests, 1st round						
Leinonen 2012	248	101678	189	101747	50.3%	1.31 [1.09, 1.59]
NTCC-I 1st round	73	23308	51	22466	14.1%	1.38 [0.97, 1.97]
POBASCAM	159	19999	144	20106	35.6%	1.11 [0.89, 1.39]
Subtotal (95% CI)		**144985**		**144319**	**100.0%**	**1.25 [1.09, 1.42]**
Total events	480		384			
Heterogeneity: $Tau^2 = 0.00$; $Chi^2 = 1.62$, df = 2 (P = 0.44); $I^2 = 0\%$						
Test for overall effect: Z = 3.21 (P = 0.001)						
3.1.5 HPV + Cytology: Combined tests - exclude triage, 1st round						
NTCC-I 1st round	73	23308	51	22466	28.8%	1.38 [0.97, 1.97]
POBASCAM	159	19999	144	20106	71.2%	1.11 [0.89, 1.39]
Subtotal (95% CI)		**43307**		**42572**	**100.0%**	**1.18 [0.97, 1.43]**
Total events	232		195			
Heterogeneity: $Tau^2 = 0.00$; $Chi^2 = 1.02$, df = 1 (P = 0.31); $I^2 = 2\%$						
Test for overall effect: Z = 1.70 (P = 0.09)						
3.1.6 2nd Screening round						
NTCC-I 2nd round	13	22093	13	22330	52.5%	1.01 [0.47, 2.18]
NTCC-II 2nd round	5	23978	20	24372	47.5%	0.25 [0.10, 0.68]
Subtotal (95% CI)		**46071**		**46702**	**100.0%**	**0.52 [0.13, 2.04]**
Total events	18		33			
Heterogeneity: $Tau^2 = 0.76$; $Chi^2 = 4.79$, df = 1 (P = 0.03); $I^2 = 79\%$						
Test for overall effect: Z = 0.93 (P = 0.35)						



Test for subgroup differences: $Chi^2 = 4.73$, df = 5 (P = 0.45), $I^2 = 0\%$

Figure 5: Meta-analysis – incidence of CIN3+

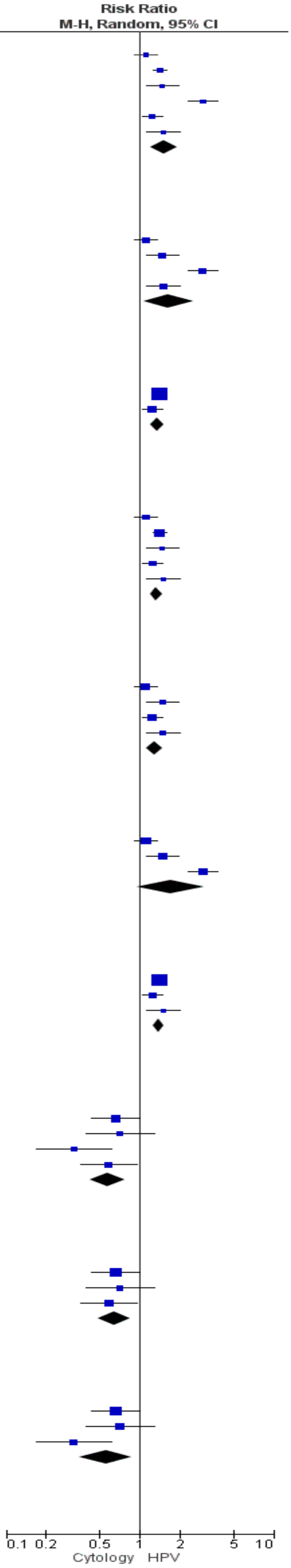
Study or Subgroup	HPV Events	HPV Total	Cytology Events	Cytology Total	Weight	Risk Ratio M-H, Random, 95% CI
4.1.1 1st screening round						
ARTISTIC 1st round	238	18386	84	6124	17.4%	0.94 [0.74, 1.21]
Leinonen 2012	220	101747	273	101678	18.5%	0.81 [0.67, 0.96]
NTCC-I 1st round	75	23308	58	22466	15.8%	1.25 [0.89, 1.76]
NTCC-II 1st round	97	24661	33	24535	14.8%	2.92 [1.97, 4.34]
POBASCAM	171	19999	150	20106	17.9%	1.15 [0.92, 1.43]
Swedescreen 1st round	72	6257	55	6270	15.6%	1.31 [0.93, 1.86]
Subtotal (95% CI)		**194358**		**181179**	**100.0%**	**1.23 [0.91, 1.67]**
Total events	873		653			
Heterogeneity: Tau² = 0.12; Chi² = 38.31, df = 5 (P < 0.00001); I² = 87%						
Test for overall effect: Z = 1.36 (P = 0.17)						
4.1.2 3 year interval - 1st round						
ARTISTIC 1st round	238	18386	84	6124	26.7%	0.94 [0.74, 1.21]
NTCC-I 1st round	75	23308	58	22466	24.9%	1.25 [0.89, 1.76]
NTCC-II 1st round	97	24661	33	24535	23.7%	2.92 [1.97, 4.34]
Swedescreen 1st round	72	6257	55	6270	24.7%	1.31 [0.93, 1.86]
Subtotal (95% CI)		**72612**		**59395**	**100.0%**	**1.44 [0.91, 2.25]**
Total events	482		230			
Heterogeneity: Tau² = 0.18; Chi² = 22.82, df = 3 (P < 0.0001); I² = 87%						
Test for overall effect: Z = 1.57 (P = 0.12)						
4.1.3 >=4 year interval						
Leinonen 2012	220	101747	273	101678	51.7%	0.81 [0.67, 0.96]
POBASCAM	171	19999	150	20106	48.3%	1.15 [0.92, 1.43]
Subtotal (95% CI)		**121746**		**121784**	**100.0%**	**0.95 [0.68, 1.35]**
Total events	391		423			
Heterogeneity: Tau² = 0.05; Chi² = 6.04, df = 1 (P = 0.01); I² = 83%						
Test for overall effect: Z = 0.26 (P = 0.79)						
4.1.4 HPV + Cytology: all tests, 1st round						
ARTISTIC 1st round	238	18386	84	6124	20.9%	0.94 [0.74, 1.21]
Leinonen 2012	220	101747	273	101678	25.2%	0.81 [0.67, 0.96]
NTCC-I 1st round	75	23308	58	22466	15.8%	1.25 [0.89, 1.76]
POBASCAM	171	19999	150	20106	22.6%	1.15 [0.92, 1.43]
Swedescreen 1st round	72	6257	55	6270	15.5%	1.31 [0.93, 1.86]
Subtotal (95% CI)		**169697**		**156644**	**100.0%**	**1.04 [0.86, 1.26]**
Total events	776		620			
Heterogeneity: Tau² = 0.03; Chi² = 11.31, df = 4 (P = 0.02); I² = 65%						
Test for overall effect: Z = 0.42 (P = 0.67)						
4.1.5 HPV + Cytology: Combined tests - exclude triage, 1st round						
ARTISTIC 1st round	238	18386	84	6124	30.3%	0.94 [0.74, 1.21]
NTCC-I 1st round	75	23308	58	22466	15.9%	1.25 [0.89, 1.76]
POBASCAM	171	19999	150	20106	38.5%	1.15 [0.92, 1.43]
Swedescreen 1st round	72	6257	55	6270	15.3%	1.31 [0.93, 1.86]
Subtotal (95% CI)		**67950**		**54966**	**100.0%**	**1.12 [0.97, 1.28]**
Total events	556		347			
Heterogeneity: Tau² = 0.00; Chi² = 3.05, df = 3 (P = 0.38); I² = 2%						
Test for overall effect: Z = 1.59 (P = 0.11)						
4.1.6 LBC comparator, 1st round						
ARTISTIC 1st round	238	18386	84	6124	34.8%	0.94 [0.74, 1.21]
NTCC-I 1st round	75	23308	58	22466	33.1%	1.25 [0.89, 1.76]
NTCC-II 1st round	97	24661	33	24535	32.1%	2.92 [1.97, 4.34]
Subtotal (95% CI)		**66355**		**53125**	**100.0%**	**1.49 [0.79, 2.80]**
Total events	410		175			
Heterogeneity: Tau² = 0.28; Chi² = 22.84, df = 2 (P < 0.0001); I² = 91%						
Test for overall effect: Z = 1.23 (P = 0.22)						
4.1.7 Conventional Cytology Comparator, 1st round						
Leinonen 2012	220	101747	273	101678	37.6%	0.81 [0.67, 0.96]
POBASCAM	171	19999	150	20106	35.2%	1.15 [0.92, 1.43]
Swedescreen 1st round	72	6257	55	6270	27.2%	1.31 [0.93, 1.86]
Subtotal (95% CI)		**128003**		**128054**	**100.0%**	**1.04 [0.77, 1.40]**
Total events	463		478			
Heterogeneity: Tau² = 0.05; Chi² = 9.34, df = 2 (P = 0.009); I² = 79%						
Test for overall effect: Z = 0.27 (P = 0.79)						
4.1.8 2nd screening round						
ARTISTIC 2nd round	32	11676	18	3866	32.2%	0.59 [0.33, 1.05]
NTCC-I 2nd round	13	22093	19	22330	23.8%	0.69 [0.34, 1.40]
NTCC-II 2nd round	5	23978	23	24372	14.1%	0.22 [0.08, 0.58]
Swedescreen 2nd round	16	6143	30	6194	30.0%	0.54 [0.29, 0.99]
Subtotal (95% CI)		**63890**		**56762**	**100.0%**	**0.52 [0.35, 0.76]**
Total events	66		90			
Heterogeneity: Tau² = 0.04; Chi² = 3.87, df = 3 (P = 0.28); I² = 22%						
Test for overall effect: Z = 3.31 (P = 0.0009)						
4.1.9 HPV + Cytology: exclude single HPV tests, 2nd round						
ARTISTIC 2nd round	32	11676	18	3866	38.9%	0.59 [0.33, 1.05]
NTCC-I 2nd round	13	22093	19	22330	26.0%	0.69 [0.34, 1.40]
Swedescreen 2nd round	16	6143	30	6194	35.2%	0.54 [0.29, 0.99]
Subtotal (95% CI)		**39912**		**32390**	**100.0%**	**0.59 [0.42, 0.85]**
Total events	61		67			
Heterogeneity: Tau² = 0.00; Chi² = 0.28, df = 2 (P = 0.87); I² = 0%						
Test for overall effect: Z = 2.84 (P = 0.005)						



Test for subgroup differences: Chi² = 27.29, df = 8 (P = 0.0006), I² = 70.7%

Figure 6: Meta-analysis – incidence of CIN2+

Study or Subgroup	HPV Events	HPV Total	Cytology Events	Cytology Total	Weight	Risk Ratio M-H, Random, 95% CI
5.1.1 1st screening round						
ARTISTIC 1st round	458	18386	137	6124	17.4%	1.11 [0.92, 1.34]
Leinonen 2012	718	101678	510	101747	18.9%	1.41 [1.26, 1.58]
NTCC-I 1st round	132	23308	86	22466	15.4%	1.48 [1.13, 1.94]
NTCC-II 1st round	218	24661	73	24535	15.6%	2.97 [2.28, 3.87]
POBASCAM	267	19999	215	20106	17.6%	1.25 [1.04, 1.49]
Swedescreen 1st round	114	6257	76	6270	15.0%	1.50 [1.13, 2.01]
Subtotal (95% CI)		**194289**		**181248**	**100.0%**	**1.51 [1.21, 1.90]**
Total events	1907		1097			
Heterogeneity: Tau^2 = 0.07; Chi^2 = 38.64, df = 5 (P < 0.00001); I^2 = 87%						
Test for overall effect: Z = 3.59 (P = 0.0003)						
5.1.2 3 year interval, 1st round						
ARTISTIC 1st round	458	18386	137	6124	26.1%	1.11 [0.92, 1.34]
NTCC-I 1st round	132	23308	86	22466	24.7%	1.48 [1.13, 1.94]
NTCC-II 1st round	218	24661	73	24535	24.8%	2.97 [2.28, 3.87]
Swedescreen 1st round	114	6257	76	6270	24.4%	1.50 [1.13, 2.01]
Subtotal (95% CI)		**72612**		**59395**	**100.0%**	**1.64 [1.07, 2.51]**
Total events	922		372			
Heterogeneity: Tau^2 = 0.17; Chi^2 = 35.28, df = 3 (P < 0.00001); I^2 = 91%						
Test for overall effect: Z = 2.27 (P = 0.02)						
5.1.3 >=4 year interval						
Leinonen 2012	718	101678	510	101747	67.0%	1.41 [1.26, 1.58]
POBASCAM	267	19999	215	20106	33.0%	1.25 [1.04, 1.49]
Subtotal (95% CI)		**121677**		**121853**	**100.0%**	**1.35 [1.21, 1.51]**
Total events	985		725			
Heterogeneity: Tau^2 = 0.00; Chi^2 = 1.25, df = 1 (P = 0.26); I^2 = 20%						
Test for overall effect: Z = 5.33 (P < 0.00001)						
5.1.4 HPV + Cytology: all tests, 1st round						
ARTISTIC 1st round	458	18386	137	6124	20.4%	1.11 [0.92, 1.34]
Leinonen 2012	718	101678	510	101747	34.9%	1.41 [1.26, 1.58]
NTCC-I 1st round	132	23308	86	22466	12.0%	1.48 [1.13, 1.94]
POBASCAM	267	19999	215	20106	21.8%	1.25 [1.04, 1.49]
Swedescreen 1st round	114	6257	76	6270	10.9%	1.50 [1.13, 2.01]
Subtotal (95% CI)		**169628**		**156713**	**100.0%**	**1.33 [1.19, 1.47]**
Total events	1689		1024			
Heterogeneity: Tau^2 = 0.00; Chi^2 = 6.16, df = 4 (P = 0.19); I^2 = 35%						
Test for overall effect: Z = 5.23 (P < 0.00001)						
5.1.5 HPV + Cytology: Combined tests - exclude triage, 1st round						
ARTISTIC 1st round	458	18386	137	6124	30.9%	1.11 [0.92, 1.34]
NTCC-I 1st round	132	23308	86	22466	18.9%	1.48 [1.13, 1.94]
POBASCAM	267	19999	215	20106	33.0%	1.25 [1.04, 1.49]
Swedescreen 1st round	114	6257	76	6270	17.2%	1.50 [1.13, 2.01]
Subtotal (95% CI)		**67950**		**54966**	**100.0%**	**1.28 [1.12, 1.47]**
Total events	971		514			
Heterogeneity: Tau^2 = 0.01; Chi^2 = 4.43, df = 3 (P = 0.22); I^2 = 32%						
Test for overall effect: Z = 3.62 (P = 0.0003)						
5.1.6 LBC comparator, 1st round						
ARTISTIC 1st round	458	18386	137	6124	34.1%	1.11 [0.92, 1.34]
NTCC-I 1st round	132	23308	86	22466	32.9%	1.48 [1.13, 1.94]
NTCC-II 1st round	218	24661	73	24535	33.0%	2.97 [2.28, 3.87]
Subtotal (95% CI)		**66355**		**53125**	**100.0%**	**1.69 [0.95, 3.02]**
Total events	808		296			
Heterogeneity: Tau^2 = 0.25; Chi^2 = 35.29, df = 2 (P < 0.00001); I^2 = 94%						
Test for overall effect: Z = 1.77 (P = 0.08)						
5.1.7 Conventional Cytology comparator, 1st round						
Leinonen 2012	718	101678	510	101747	64.3%	1.41 [1.26, 1.58]
POBASCAM	267	19999	215	20106	25.8%	1.25 [1.04, 1.49]
Swedescreen 1st round	114	6257	76	6270	9.9%	1.50 [1.13, 2.01]
Subtotal (95% CI)		**127934**		**128123**	**100.0%**	**1.37 [1.26, 1.50]**
Total events	1099		801			
Heterogeneity: Tau^2 = 0.00; Chi^2 = 1.67, df = 2 (P = 0.43); I^2 = 0%						
Test for overall effect: Z = 6.87 (P < 0.00001)						
5.1.8 2nd screening round						
ARTISTIC 2nd round	68	11676	34	3866	34.5%	0.66 [0.44, 1.00]
NTCC-I 2nd round	19	22093	27	22330	20.8%	0.71 [0.40, 1.28]
NTCC-II 2nd round	12	23978	38	24372	17.7%	0.32 [0.17, 0.61]
Swedescreen 2nd round	25	6143	43	6194	27.0%	0.59 [0.36, 0.96]
Subtotal (95% CI)		**63890**		**56762**	**100.0%**	**0.57 [0.42, 0.77]**
Total events	124		142			
Heterogeneity: Tau^2 = 0.03; Chi^2 = 4.09, df = 3 (P = 0.25); I^2 = 27%						
Test for overall effect: Z = 3.61 (P = 0.0003)						
5.1.9 HPV + Cytology: all tests, 2nd round						
ARTISTIC 2nd round	68	11676	34	3866	45.8%	0.66 [0.44, 1.00]
NTCC-I 2nd round	19	22093	27	22330	22.4%	0.71 [0.40, 1.28]
Swedescreen 2nd round	25	6143	43	6194	31.8%	0.59 [0.36, 0.96]
Subtotal (95% CI)		**39912**		**32390**	**100.0%**	**0.65 [0.49, 0.85]**
Total events	112		104			
Heterogeneity: Tau^2 = 0.00; Chi^2 = 0.27, df = 2 (P = 0.87); I^2 = 0%						
Test for overall effect: Z = 3.07 (P = 0.002)						
5.1.10 LBC comparator, 2nd round						
ARTISTIC 2nd round	68	11676	34	3866	42.0%	0.66 [0.44, 1.00]
NTCC-I 2nd round	19	22093	27	22330	30.6%	0.71 [0.40, 1.28]
NTCC-II 2nd round	12	23978	38	24372	27.4%	0.32 [0.17, 0.61]
Subtotal (95% CI)		**57747**		**50568**	**100.0%**	**0.56 [0.36, 0.87]**
Total events	99		99			
Heterogeneity: Tau^2 = 0.08; Chi^2 = 4.10, df = 2 (P = 0.13); I^2 = 51%						
Test for overall effect: Z = 2.58 (P = 0.010)						



Test for subgroup differences: Chi^2 = 71.75, df = 9 (P < 0.00001), I^2 = 87.5%

Figure 7: Forest plot – screening related harm

Study or Subgroup	HPV Events	HPV Total	Cytology Events	Cytology Total	Weight	Risk Ratio M-H, Random, 95% CI
ARTISTIC 1st round	223	593	717	1872		0.98 [0.87, 1.11]

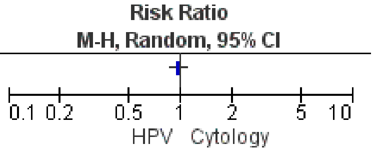


Figure 8: Forest plot – Sensitivity analyses with data from 2nd screening round of POBASCAM

Study or Subgroup	HPV Events	HPV Total	Cytology Events	Cytology Total	Weight	Risk Ratio M-H, Random, 95% CI
8.1.1 Incidence of cervical cancer - 2nd screening round						
ARTISTIC 2nd round	3	11676	0	3866	9.8%	2.32 [0.12, 44.87]
NTCC-I 2nd round	0	22093	6	22330	10.4%	0.08 [0.00, 1.38]
NTCC-II 2nd round	0	23978	3	24372	9.8%	0.15 [0.01, 2.81]
POBASCAM 2nd round	4	19579	14	19731	69.9%	0.29 [0.09, 0.87]
Subtotal (95% CI)		**77326**		**70299**	**100.0%**	**0.29 [0.11, 0.73]**
Total events	7		23			
Heterogeneity: Tau² = 0.00; Chi² = 2.92, df = 3 (P = 0.40); I² = 0%						
Test for overall effect: Z = 2.62 (P = 0.009)						
8.1.2 Incidence of cervical cancer - 2nd screening round - Subgroup HPV + cytology						
ARTISTIC 2nd round	3	11676	0	3866	17.9%	2.32 [0.12, 44.87]
NTCC-I 2nd round	0	22093	6	22330	18.9%	0.08 [0.00, 1.38]
POBASCAM 2nd round	4	19579	14	19731	63.2%	0.29 [0.09, 0.87]
Subtotal (95% CI)		**53348**		**45927**	**100.0%**	**0.33 [0.08, 1.29]**
Total events	7		20			
Heterogeneity: Tau² = 0.46; Chi² = 2.69, df = 2 (P = 0.26); I² = 26%						
Test for overall effect: Z = 1.59 (P = 0.11)						
8.1.3 Incidence of CIN3 - 2nd screening round						
NTCC-I 2nd round	13	22093	13	22330	29.3%	1.01 [0.47, 2.18]
NTCC-II 2nd round	5	23978	20	24372	22.8%	0.25 [0.10, 0.68]
POBASCAM 2nd round	84	19579	108	19731	47.9%	0.78 [0.59, 1.04]
Subtotal (95% CI)		**65650**		**66433**	**100.0%**	**0.65 [0.35, 1.21]**
Total events	102		141			
Heterogeneity: Tau² = 0.19; Chi² = 5.39, df = 2 (P = 0.07); I² = 63%						
Test for overall effect: Z = 1.35 (P = 0.18)						
8.1.4 Incidence of CIN3+ - 2nd screening round						
ARTISTIC 2nd round	32	11676	18	3866	19.0%	0.59 [0.33, 1.05]
NTCC-I 2nd round	13	22093	19	22330	14.1%	0.69 [0.34, 1.40]
NTCC-II 2nd round	5	23978	23	24372	8.4%	0.22 [0.08, 0.58]
POBASCAM 2nd round	88	19579	122	19731	40.8%	0.73 [0.55, 0.96]
Swedescreen 2nd round	16	6143	30	6194	17.7%	0.54 [0.29, 0.99]
Subtotal (95% CI)		**83469**		**76493**	**100.0%**	**0.59 [0.44, 0.80]**
Total events	154		212			
Heterogeneity: Tau² = 0.04; Chi² = 5.95, df = 4 (P = 0.20); I² = 33%						
Test for overall effect: Z = 3.37 (P = 0.0007)						
8.1.5 Incidence of CIN2+ - 2nd screening round						
ARTISTIC 2nd round	68	11676	34	3866	21.8%	0.66 [0.44, 1.00]
NTCC-I 2nd round	19	22093	27	22330	15.5%	0.71 [0.40, 1.28]
NTCC-II 2nd round	12	23978	38	24372	13.7%	0.32 [0.17, 0.61]
POBASCAM 2nd round	160	19579	184	19731	30.5%	0.88 [0.71, 1.08]
Swedescreen 2nd round	25	6143	43	6194	18.6%	0.59 [0.36, 0.96]
Subtotal (95% CI)		**83469**		**76493**	**100.0%**	**0.65 [0.47, 0.88]**
Total events	284		326			
Heterogeneity: Tau² = 0.07; Chi² = 10.01, df = 4 (P = 0.04); I² = 60%						
Test for overall effect: Z = 2.80 (P = 0.005)						
8.1.6 Incidence of CIN2+ - 2nd screening round - HPV + cytology: all tests						
ARTISTIC 2nd round	68	11676	34	3866	18.1%	0.66 [0.44, 1.00]
NTCC-I 2nd round	19	22093	27	22330	9.1%	0.71 [0.40, 1.28]
POBASCAM 2nd round	160	19579	184	19731	60.1%	0.88 [0.71, 1.08]
Swedescreen 2nd round	25	6143	43	6194	12.8%	0.59 [0.36, 0.96]
Subtotal (95% CI)		**59491**		**52121**	**100.0%**	**0.78 [0.65, 0.93]**
Total events	272		288			
Heterogeneity: Tau² = 0.00; Chi² = 3.17, df = 3 (P = 0.37); I² = 5%						
Test for overall effect: Z = 2.77 (P = 0.006)