VALGENT : a protocol for clinical validation of human papillomavirus assays

Title: VALGENT: a protocol for clinical validation of human papillomavirus assays

Author: Marc Arbyn Christophe Depuydt Ina Benoy Johannes Bogers Kate Cuschieri Markus Schmitt Michael Pawlita Daan Geraets Isabelle Heard Tarik Gheit Massimo Tommasino Mario Poljak Jesper Bonde Wim Quint

PII: S1386-6532(15)00683-6
DOI: http://dx.doi.org/doi:10.1016/j.jcv.2015.09.014
Reference: JCV 3442

To appear in: *Journal of Clinical Virology*

Received date: 10-8-2015
Revised date: 25-9-2015
Accepted date: 30-9-2015

Please cite this article as: Arbyn Marc, Depuydt Christophe, Benoy Ina, Bogers Johannes, Cuschieri Kate, Schmitt Markus, Pawlita Michael, Geraets Daan, Heard Isabelle, Gheit Tarik, Tommasino Massimo, Poljak Mario, Bonde Jesper, Quint Wim.VALGENT: a protocol for clinical validation of human papillomavirus assays.*Journal of Clinical Virology* http://dx.doi.org/10.1016/j.jcv.2015.09.014

VALGENT: a protocol for clinical validation of human papillomavirus assays

Marc Arbyn[a], Christophe Depuydt[b], Ina Benoy[b], Johannes Bogers[b], Kate Cuschieri[c], Markus Schmitt[d], Michael Pawlita[d], Daan Geraets[e], Isabelle Heard[f], Tarik Gheit[g], Massimo Tommasino[g], Mario Poljak[h], Jesper Bonde[i], Wim Quint[e]

a) Unit of Cancer Epidemiology & Belgian Cancer Centre, Scientific Institute of Public Health, Brussels, Belgium
b) Department of Molecular Pathology, AML Laboratory, Sonic Healthcare, Antwerp, Belgium
c) Scottish HPV Reference Laboratory, Royal Infirmary of Edinburgh, Edinburgh, Scotland, UK
d) Division Molecular Diagnostics of Oncogenic Infections, Research Program Infection, Inflammation and Cancer, German Cancer Research Center (DKFZ), Heidelberg, Germany
e) DDL Diagnostic Laboratory, Rijswijk, the Netherlands
f) French HPV Reference Laboratory, Institut Pasteur, Paris, France
g) International Agency for Research on Cancer, Lyon, France
h) Institute of Microbiology and Immunology, Faculty of Medicine, University of Ljubljana, Ljubljana, Slovenia
i) Department of Pathology and Clinical Research Center, Copenhagen University Hospital, Hvidovre, Denmark

Contact address: M. Arbyn, Belgian Cancer Centre/Unit of Cancer Epidemiology, Scientific Institute of Public Health, J. Wytsmanstreet 14, B1050, Brussels, Belgium. Tel: 0032 2 6425021; Fax: 0032 2 6425410; Email: marc.arbyn@wiv-isp.be

Word count: 3696

This invited review is part of themed section on HPV of a special issue of the *J Clin Virol*, Guest Editor: Mario Poljak

VALGENT is a framework for test comparison and validation of HPV assays. This paper describes the study protocol and results from 2 VALGENT rounds.

Highlights

- Only clinically validated HPV tests should be used in cervical cancer screening
- VALGENT is a robust protocol for validation of hrHPV and HPV genotyping assays
- In VALGENT-2 sufficient clinical accuracy of three HPV assays was demonstrated
- VALGENT is producing a comprehensive study resource for HPV test comparisons
- Integration of VALGENT into the global WHO reference laboratory network is planned

Abstract

Background: Testing for high-risk HPV is more effective in primary cervical cancer screening than the cytological examination of a Pap smear. Separate genotyping may be useful for triage in both HPV-based and cytology-based screening. Only clinically validated tests should be used in clinical practice.

Objectives: VALGENT is a study framework for test comparison and validation of HPV assays in general and HPV genotyping tests in particular according to clinically relevant outcomes and for clinical applications endorsed by scientific evidence.

Study design: VALGENT involves the collation of fresh or archived cervical cell specimen from women attending routine screening supplemented with cytologically abnormal samples. Multiple aliquots of residual material are sent from a central laboratory to participating laboratories for testing with novel HPV assays with limited, extended or full genotyping capacity. Outcomes are derived from screening and pathology registries. Each VALGENT panel includes an assay already validated for screening. A series of accuracy and concordance statistics were generated.

Results: Currently, two VALGENT study rounds were completed from laboratories in Antwerp (Belgium) and Edinburgh (Scotland). Two new assays (G5+/6+ PCR-LMNX and Xpert HPV) were validated for screening by showing similar accuracy for cervical precancer as the standard comparator test. For two other tests (BD Onclarity, PapilloCheck) validation was confirmed. Inter-test agreement was high although certain type-specific discordances were observed which warrant further analysis.

Conclusion: VALGENT extends current guidelines for high-risk HPV test validation in cervical cancer screening and has produced a large study resource for test comparison. More robust procedures of sample selection and handling and integration with the global WHO reference laboratory network focusing on analytical accuracy, may result in the generation of an international standard and a formalised system for clinical validation of HPV assays and quality control in HPV-based screening.

Abbreviations

EIA: enzyme immunoassay
ASC-US: atypical squamous cervical cells of undetermined significance
CIN: cervical intraepithelial neoplasia
LSIL: low-grade squamous intraepithelial lesion
HPV: human papillomavirus
Hr: high-risk
HSIL: high-grade squamous intraepithelial lesion
SHPVRL: Scottish HPV Reference Laboratory
VALGENT: VALidation of HPV GENotyping tests

Keywords: human papillomavirus; cervical cancer; cervical cancer screening; diagnostic test accuracy; test validation; quality control

1. Background

At least 12 high-risk (hr) human papillomavirus (HPV) types (group 1 carcinogens) are causally linked with the development of cervical cancer and some of them also with several other ano-genital cancers as well as certain head and neck cancers [1,2]. The recognition of the strong etiological link between persistent infection with hr HPV types and cervical cancer has prompted the development of prophylactic vaccines and HPV tests usable for cervical cancer screening. Evidence on the clinical utility of HPV testing is applied widely for triage of borderline/equivocal cervical cytology and for follow-up after treatment of cervical cancer precursors [3]. More recently, randomised trials have demonstrated that HPV-based screening is more effective than cytological screening in reducing the incidence of invasive squamous and adeno-carcinoma of the cervix uteri [3,4].

Molecular hrHPV detection assays operate at various levels. The 1st generation HPV assays detected hr-HPV in aggregate with a variety of cocktails of DNA or RNA probes reporting a presence/absence result with or without a quantitative signal related to the amount of virus in a sample [5]. Yet now, newer generation HPV assays allow for more detailed reporting with either limited genotyping (e.g. reporting of HPV16 or HPV18 separately, the remaining hrHPV as a bulk result), extended genotyping where HPV16, HPV18 and certain other hrHPV genotypes are reported separately plus one or more bulk results, or finally, full genotyping assays allowing for individual reporting of the hrHPV types and sometimes also certain potentially high-risk and low-risk genotypes [5].

Guidelines have been developed to define the minimal requirements for hrHPV tests potentially usable in primary cervical cancer screening [6,7].

Recent cohort studies (screening, triage) indicate different carcinogenic potential of certain types. HPV16 is a particularly potent carcinogenic agent [8]. Also HPV18 and to a lesser extent HPV45 are more strongly linked to development of cervical cancer compared to the other hr-types [9-11]. This has led to the proposal that genotyping can be used in clinical practice to risk stratify screening HPV positive women. Certain practice guidelines have included HPV16/18 in triage of hrHPV positive women or as a second step in the triage of women with minor abnormal cytology and a positive hrHPV result [3,12].

However although several genotyping systems exist, there is limited assay performance evidence available, particularly in clinical settings. Therefore, to address this lack of evidence, we propose to undertake a comprehensive evaluation of the performance of a variety of HPV assays, particularly those that have genotyping capability. To conduct this research we have

established a multicentre collaboration between microbiologists, experienced in the development and evaluation of HPV assays, and clinical epidemiologists with recognized skills in diagnostic test accuracy assessment.

2. Objectives

2.1 Overall objectives

The VALGENT (acronym for *VALidation of HPV GENotyping tests*) project aims: to explore the potential role for HPV genotyping tests in clinical practice; and to develop a tool for evaluation and comparison of different HPV genotyping assays.

The VALGENT objectives are in agreement with priorities defined by the World Health Organisation [13] and new European Guidelines in Quality Assurance in Cervical Cancer Screening, which are currently in press [6,14,15].

2.2 Specific objectives

The VALGENT study should answer the following questions:

a) What are potential clinical indications of HPV genotyping tests?
b) If clinical indications for HPV genotyping can be defined: what are minimal requirements justifying the use of genotyping tests in clinical practice?
c) Which HPV genotyping tests fulfil these requirements and can be recommended in clinical practice guidelines?

In addition to these three specific questions, VALGENT also allows assessment of which high-risk HPV DNA assays fulfil requirements for primary screening [6,7]. Distinction will be made between tests that allow for limited (HPV16, HPV18 or HPV45), extended or full identification of hrHPV types.

Three specific objectives can be distinguished:

1. To develop a well-designed evaluation tool to compare virological (analytical detection threshold, type concordance) and clinical test performance (sensitivity and specificity for high-grade cervical intraepithelial neoplasia or worse [CIN2+]).
2. To establish evidence on the accuracy of HPV genotyping assays to identify cervical cancer precursors in primary screening or triage of screen-positive women by conducting systematic reviews.
3. To define minimal requirements of HPV genotyping assays for use in cervical cancer screening and to identify the tests which fulfil these requirements.

3. Study design

3.1. Study populations

The protocol foresees including a continuous series of cervical cell specimen (which can be archived or fresh samples) from women participating in the local cervical cancer screening programme. In addition, the collection is enriched with 300 cytologically abnormal samples (100ASC-US, 100 LSIL and 100 HSIL) to comprise a total of 1,300 samples. Depending on the prevalence of cervical cytological abnormalities across settings, between 50 and 120 abnormal specimens (ASC-US or worse) are expected to be detected among the 1,000 screened women. Altogether it was anticipated that the study population would contain between 350 and 420 cytologically abnormal samples. Through the routinely indicated follow-up and management (i.e. colposcopy and/or biopsy) of the 350-420 women with abnormal cytology, approximately 90-150 histologically confirmed cases of high-grade cervical intraepitheial neoplasia or worse (CIN2+) are expected. This population of 90-150 CIN2+ cases is used to compute the clinical sensitivity, accepting the histology result as the gold standard outcome. Among the 880-950 cytologically normal women, approximately 800 or more may have subsequent normal Pap smears from either the subsequent or previous screening round. This group will constitute the denominator for the computation of clinical specificity. In addition, specificity is also computed on a larger group of women (≥800 with double negative cytology + women with non-normal cytology but with ≤CIN1 histology outcome). The great majority of women included in VALGENT studies, will have long-term passive follow-up through the national or regional screening and/or pathology registries.

3.2. Evaluated tests and participating laboratories

Thus far, two VALGENT studies have occurred. The samples from VALGENT-1 were provided by the AML laboratory (Antwerp, Belgium) and those from VALGENT-2 by the Scottish HPV Reference Laboratory (SHPVRL, Edinburgh, Scotland).

The 4 full genotyping assays used for testing the samples from Antwerp (VALGENT-1) were:

1. Multiplex RT-qPCR, a real-time quantitative PCR targeting E6 or E7 genes of 17 HPV types [16,17], performed in the AML laboratory, Antwerp, Belgium;
2. BSG5+/6+-PCR/MPG assay with quantitative Luminex-based identification of type L1 sequences of 51 HPV types [18,19], performed in the Department of Genome Modifications and Carcinogenesis, DKFZ, Heidelberg, Germany;

3. GP5+/6+ PCR-based enzyme immunoassay (EIA) kit HPV GP HR (EIA; Diassay, Rijswijk, the Netherlands) for simultaneous identification of 14 hrHPV types [20], and LMNX Genotyping Kit GP HR (LMNX; Diassay) for full genotyping of 18 types [21] performed by DDL Diagnostic Laboratory, Rijswijk, The Netherlands[1];
4. TS-E7-MPG, a multiplex ultrasensitive type-specific PCR targeting E7 genes and bead-based identification of 19 HPV types [22], performed by IARC, Lyon, France.

In VALGENT-2, the first and the third tests assessed in VALGENT-1, were evaluated again in the same laboratories and three new commercial assays were added:

1. BD Onclarity HPV assay (BD Diagnostics, Sparks, MD, USA), PCR targeting E6/E7 HPV genes and identifying six HPV types separately (HPV16,18,31,45,51,52) and the other hr-types in groups of two or three types (=extended genotyping) [23], performed by SHRL, Edinburgh, Scotland;
2. XPert HPV (Cepheid, Sunnyvale, CA, USA), PCR targeting E6 and E7 genes of HPV16 and 18/45 separately, and 11 other hr-types in group (=limited genotyping) [24], by SHRL, Edinburgh, Scotland;
3. PapilloCheck HPV-screening test (Greiner Bio-One, Frickenhausen, Germany), a PCR targeting E1 HPV genes of 24 separate HPV types (=full genotyping) [25], performed by the French HPV Reference Laboratory, Institut Pasteur, Paris, France.

Currently new VALGENT-3 and VALGENT-4 panels are being generated by the Laboratory for Molecular Microbiology, Faculty of Medicine, University of Ljubljana, Slovenia and the Department of Pathology, Copenhagen University Hospital, Hvidovre, Denmark. At least one assay already evaluated in a previous VALGENT-panel will be included also in a next iteration in order to allow for inter-test comparisons of assays included in different panels.

3.3. Outcomes

3.3.1. HPV type concordance of the evaluated genotyping assays: positive concordance, negative concordance, overall concordance, kappa values (test 1 vs test 2, test1 vs test 3, …)

- hrHPV positivity (considered for 13 [HPV16,18,31,33,35,39,45,51,52,56,58,59,68] or 14 [+HPV66] hrHPV types)
- type-specific concordance

[1] The first test (GP5+/6+ PCR amplification, followed by enzyme immunoassay identification of 14 hrHPV types as an aggregate) will be abbreviated in this paper as "GP5+/6+-EIA". The second test (GP5+/6+ PCR amplification, followed by separate Luminex-based identification of 14 hrHPV types + HPV types 26,53,73,82) will be abbreviated in this paper as "GP5+/6+ - LMNX".

  - HPV16 presence yes/no (irrespective of single/multiple infections)
  - HPV18 presence yes/no (irrespective of single/multiple infections)
  - HPV16/18 presence yes/no (irrespective of single/multiple infections)
  - hrHPV other than 16/18 presence yes/no (irrespective of single/multiple infections)
  - Presence of each separate type yes/no (irrespective of single/multiple infections)
  - Presence of each separate type in single infections
  - Presence of each separate type in multiple infections
- Number of hrHPV infections
- For tests with viral load result: influence of load on concordance statistics

3.3.2. Viro-cytological correlation, assessed on all tested samples.

3.3.3. Clinical accuracy of each assay in primary screening:

- sensitivity for CIN2+ (assessed on approximately 90-140 cases of CIN2+, ~halve of them being CIN3/AIS+) (allowing verification of cross-sectional equivalence criteria as defined by Meijer *et al.* [6]
- specificity 1: for women with double negative cytology (assessed on approximately 800 cases)
- specificity 2: for women with double negative cytology + women with ≤ CIN1 outcome (allowing verification of cross-sectional equivalence criteria as defined by Meijer *et al.*, [6]
- for tests with type-specific viral load results: influence of viral load on test accuracy and proposition of best viral load cutoff.

3.3.4. For women with ASC-US and LSIL cytology:

- accuracy of hrHPV triage (considering the 13 or 14 hrHPV types as a pool) for CIN2+ and CIN3+
- accuracy of triage using HPV16, HPV16/18 and, if possible, HPV16/18/45 genotyping for CIN2+ and CIN3+.

3.3.5. Potential of triage by HPV genotyping among women being hrHPV-positive. For women with positive cytology, outcomes will be generated through the usual follow-up. For women with normal cytology, outcomes will be generated through the next screening round.

3.4. Inter-test matrix:

From the VALGENT studies, a matrix of HPV genotype test-positivity stratified by degree of cervical epithelial abnormality for each couple of assays. This matrix will be of help in

international comparisons from HPV prevalence and HPV vaccination impact surveys where different tests were used.

3.5. The following covariates will be taken into account as far as possible: age of the woman, age of the specimen, preservation medium in which the specimens were collected (PreservCyt, SurePath, other), test material aliquoted (part of whole residual sample, proteinase K digest, extracted DNA, …), processing of test material, test procedures, viral load or intensity of the signal, viral genes targeted by the test, type of follow-up.

4. Results

Both VALGENT-1 and -2 have generated several peer-reviewed papers [7,19,27-30] and were the object of public workshops in the framework of the International Conferences of the Papillomavirus Society (Berlin 2011, San Juan 2013, Seattle 2014 and Lisbon 2015). Several new manuscripts are in preparation. In the current paper, a restricted series of results are presented, related to outcome 3.3.1 (inter-test reproducibility), 3.3.3 (clinical accuracy to detect CIN2+) and 3.4. (hrHPV prevalence in the Scottish screening population assessed with 5 assays). For more detailed results, we refer to the separate published VALGENT reports and VALGENT proceedings.

4.1. Inter-test reproducibility

In table 1, reproducibility statistics (percentage agreement and kappa values) are shown for the results of testing with five HPV assays (GP5+/6+-EIA, GP5+/6+-LMNX, BD Onclarity, PapilloCheck and Xpert HPV) applied on the VALGENT-2 panel collated in Scotland. Concordance was assessed the type-specific level, the aggregated level for 13 or 14 hrHPV types, or for combinations of HPV16 and HPV18 or HPV18 and HPV45. Concordance was excellent for the majority of test comparisons ($\kappa > 0.8$). However, certain exceptions were noted: concordance was good (κ in the range 0.61-0.80) for HPV52 assessed with BD Onclarity and GP5+/6+-LMNX and for HPV58 assessed with PapilloCheck and GP5+/6+-LMNX, whereas concordance was only fair (κ in the range 0.21-0.40) for HPV68 assessed with PapilloCheck and GP5+/6+-LMNX.

In table 2, kappa values and their 95% confidence intervals are shown for the same assay comparisons for hrHPV testing and HPV16 genotyping, separated for four different subpopulations of women included in VALGENT-2. A tendency of lower concordance was observed in the screening population compared to the enrichment population and in the group

with two consecutive negative Pap smears compared to the group with histologically confirmed CIN2+.

4.2. Clinical accuracy for CIN2+

The absolute accuracy for outcome CIN2+ of three HPV genotyping tests (qPCR(E6/E7, BSG5+/6+-PCR/MPG, TS-E7-MPG) assessed in VALGENT-1 is shown in Table 3. Test positivity was defined by presence of at least one of 14 hrHPV types. The sensitivity was uniformly high for all the three tests. For all the three tests, specificity was increased substantially while still maintaining a high sensitivity, by increasing the cut-off (based on the quantitative signal).

Table 4 also displays the clinical sensitivity and specificity for CIN2+, for five other HPV assays evaluated in VALGENT-2. The specificity was clearly higher when using denominator A (=number of women with two consecutive negative Pap smears) compared to denominator B (=denominator A + number with abnormal cytology but ≤CIN1 outcome). VALGENT included testing with GP5+/6+-EIA which is an accepted standard comparator used for validation of other HPV assays applicable in cervical cancer screening [3,6,7]. In table 5, the relative accuracy for CIN2+ of four assays are compared to the GP5+/6+-EIA. Both relative sensitivity and specificity did not differ significantly from unity. In spite of the obvious differences in absolute specificity computed with denominator A and B, the relative specificity estimates were very similar for both denominators.

4.3. Prevalence of hrHPV infection in the screening population

Figure 1 shows the change in prevalence, by five year age group, of hrHPV infection in the Scottish screening population assessed with five assays evaluated in VALGENT-2. The prevalence typically decreased by age: 37.2% in age group 20-24 to 5.9% in age group 60-64, when measuring with GP5+/6+-EIA. The prevalence amplitude by test (prevalence observed with the test yielding the highest prevalence minus prevalence observed with the test yielding the lowest prevalence) varied between 1.5% and 6.9% in age groups 25-29 and 45-49, respectively. The overall hrHPV prevalence ratios (prevalence observed in one test vs prevalence observed with GP5+/6+) in age group 30-64 were 0.99 (95% CI: 0.98-1.03), 1.08 (95% CI: 0.98-1.20), 1.08 (95% CI: 0.95-1.25), 1.01 (95% CI: 0.87-1.18) for index tests GP5+/6+-LMNX, BD Onclarity, PapilloCheck, Xpert HPV, respectively.

5. Discussion

5.1. Comments on presented results

Data from VALGENT-1 (results being reported elsewhere) and VALGENT-2 (see Table 1) show that inter-test type-specific concordance was generally excellent. However, agreement percentages are not relevant since there are usually very high when the majority of cases are test-negative. Kappa statistics, adjusting for agreement by chance, are therefore more relevant. A few unexpected discordances were observed. For instance, BD Onclarity found 27 cases of HPV52 not identified by GP5+/6+-LMNX. It is unclear whether this is due to false-positive results (cross-reactions) or differences in analytical sensitivity for this type. Correlation with the semi-quantitative MFI signal generated by the GP5+/6+-LMNX test revealed that discordant cases have on average a lower viral load than concordant cases, an observation also made in the Danish HORIZON study [31]. Comparison with other tests and finally sequencing will provide further insight into the nature and clinical relevance of type-specific discordances.

By comparison with the GP5+/6+-EIA assay, VALGENT-2 has demonstrated attainment of the accuracy for use in cervical cancer screening of two new HPV assays: GP5+/6+-LMNX [26] and Xpert HPV [Cuschieri: 2015, 30th International Papilloma-virus Conference, Lisbon, Valgent Workshop]. Moreover, VALGENT-2 has corroborated prior findings regarding the validation of the BD Onclarity HPV assay [23,30] and the Papillocheck [25] [Heard 2015, 30th International Papillomavirus Conference, Lisbon, Valgent Workshop]. More new assays will be assessed in Valgent-3 and -4.

VALGENT will generate matrices which can be used to understand contrasts in prevalence studies conducted with different tests. In VALGENT-2, the prevalence of hrHPV varied in the age group 30-64 between 11.0% (observed with GP5+/6+-LMNX) to 12.2% (observed with BD Onclarity and PapilloCheck). The prevalence of hrHPV was 8% higher if measured with the latter two tests compared with the GP5+/6+ PCR.

5.2. Clinical and analytical accuracy

The global WHO HPV reference laboratory network has established guidance for HPV testing to measure reliable effects of HPV vaccination in trials and to monitor the impact of vaccination programmes [32]. The focus of the WHO HPV reference laboratory network was on analytical accuracy with respect to measuring presence or absence of single or multiple specific HPV genotypes at low or high concentrations in human specimens. The Network

has developed proficiency panels consisting of purified plasmids (full genomic cDNA sequences) of different calibrated compositions and concentrations of HPV genotypes. These panels are distributed to laboratories performing HPV testing with their own particular assay. Laboratories identifying correctly 50 international units (IU) of HPV16 and HPV18 and 500 IU of other hrHPV genotypes are considered proficient with respect to analytical accuracy. Successive proficiency studies have demonstrated improved proficiency among participating laboratories over time [33,34].

In cervical cancer screening, however, the question is to detect clinically relevant infections which may induce precancerous lesions or cancer and therefor clinical accuracy for present or incipient high-grade cervical intraepithelial neoplasia or worse is the main issue [35].

5.3. VALGENT vs Meijer guidelines

VALGENT comprises evaluation of non-inferiority of new hrHPV assays compared to the standard comparator tests (HC2 and GP5+/6+-EIA) which together with demonstration of high reproducibility allows validation for use in cervical cancer screening [6,7]. Moreover VALGENT extends the objectives for validation hrHPV DNA assays towards HPV genotyping assays. The choice of cases of CIN2+ or CIN3+ in the validation panel and, in particular, the selection of non-cases (women with <=CIN1) influences the assessment of absolute sensitivity and specificity, respectively. Although, the recommendation is to include only cases and non-cases from a screening population, this instruction is not always followed in validation studies following the Meijer protocol [7]. In VALGENT, the composition of study subjects is more strictly defined and incorporates two options to compute specificity (A and B). As expected, the absolute specificity of HPV assays was higher in option A (restricted to women with two consecutive negative Pap smears) than in option B (when also women with abnormal cytology but negative outcome are included, see table 4). This finding corroborates observations from earlier meta-analyses indicating that the specificity varies substantially with the spectrum of disease (screening, follow-up of ASC-US or LSIL) [3,36]. However, the relative specificity of assays compared to the standard comparator test was similar in both situations (see table 5), underlying the robustness of relative accuracy as validation parameter, a phenomenon also observed in other assessments of diagnostic test accuracy [37].

5.4. Weaknesses and strengths of VALGENT

VALGENT started as an informal collaboration between dedicated HPV laboratory experts and clinical epidemiologists without external funding other than a contribution from manufacturers of three commercially available assays in VALGENT-2. Certain errors were made during expedition and/or handling of specimens resulting in contamination and the exclusion of certain test evaluations. After running two VALGENT rounds, the need is felt to develop standardised operational procedures in sample selection and processing to substantiate internationally approved recommendations for HPV test validation. Formal collaboration with the WHO reference laboratory network will help in reaching a higher level of standardisation and possible approval from regulatory agencies.

A strength of VALGENT, in particular VALGENT-2 and subsequent VALGENT rounds is that the sample collection is nested within an organised screening programme, using well-annotated specimens, freshly collected or archived from existing cervical cytology banks, linked to comprehensive screening and pathology registries [38]. By recruiting continuous samples from women attending screening with controlled enrichment of abnormal samples, both statistical power and representativeness are assured with possibility to weight sampling fractions which may allow computing predictive values. Moreover by linkage with outcomes from subsequent screening rounds, final disease status maybe obtained allowing for longitudinal assessments.

5.5 Future work

5.5.1. Evidence base for HPV genotyping

Simultaneously with the VALGENT studies, systematic reviews are being conducted on the accuracy and efficacy of triage of women with a minor cytological lesions or a positive hrHPV result using HPV genotyping tests as well as other markers. These reviews are supported by the European Commission and other funders (see acknowledgements). Once these systematic reviews are completed and translated in practice guidelines, VALGENT may play a major role in test validation in applications for which high-quality evidence is demonstrated.

5.5.2. Network meta-analysis

The VALGENT studies will be completed with systematic reviews of studies where test performance of at least 2 HPV genotyping tests were evaluated in primary cervical cancer

screening or in the management of screen-positive women. Network meta-analysis will be used to generate networks of paired test comparisons that allow ranking of the clinical performance of multiple tests from observed and indirect comparisons [39]. By including at least one assay already evaluated in VALGENT in subsequent VALGENT studies, bridging data will be generated allowing for indirect comparisons.

5.5.3. Adaptation of the VALGENT protocols

Several countries have recently switched towards HPV-based screening or are in the process of planning this change [15,40,41]. Enrolment of study subjects will therefore include a consecutive series of women participating in HPV-based cervical cancer screening enriched with HPV positive women. HPV vaccination status will also be an additional covariate which requires assessment since growing cohorts of vaccinated young women have reach screening age. VALGENT-like protocols may be used for validation of other markers than viral nucleic acids.

5.6. Conclusion

VALGENT provides a comprehensive framework for comparison and clinical validation of HPV tests (including those which offer genotyping capability) which is gaining increasing momentum and interest from laboratory experts in the field of HPV virology and developers or manufacturers of HPV assays. New national and international evidence-based recommendations propose use of HPV essays instead of or in combination with cytology. Therefore, and given the rapid development and marketing of novel HPV assays, guidelines for test validation and quality assurance of virological testing in cervical cancer screening are needed. Initiatives such as VALGENT may play an important role to realise this.

Conflict of Interest Declaration

KC's institution received research grants from Abbott, Hologic, Cepheid and Becton Dickinson.
CD has received speaker's fees from Abbott, BD and Innogenetics.
DG has no conflicts of interest to declare
IH has no conflicts of interest to declare
MPa's institution received research grants from Roche and Qiagen and has licensed a HPV genotyping system to Diamex.
MPo's institution received research grants from Abbott Molecular.

The current paper has not been screened by commercial companies

Acknowledgements
M. Arbyn and M. Poljak are supported by the seventh framework program of DG Research of the European Commission, through the COHEAHR Network (grant No 603019). M. Arbyn was also supported by the Joint Action CANCON which has received funding from the European Union in the framework of the Health Programme (2008-13). Other funding was received from the German Guideline Program in Oncology (German Cancer Aid project # 110163) and from the National Institute of Public Health and the Environment (Bilthoven, the Netherlands).

In VAGENT-2 commercial assays were allowed to participate under the condition of a financial contribution, which was received from BD Diagnostics, Greiner Bio-One and Cepheid. Those companies also provided test kits to the laboratories who performed HPV testing. No commercial influence was accepted regarding publication of study results.

We thank J Dillner (Department of Laboratory Medicine and Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Center for Cervical Cancer Prevention, and WHO HPVLabNet Global Reference Laboratory, Karolinska Hospital, Stockholm, Sweden) for discussion on future collaboration and suggestions for the establishment for a formal international standard for quality control for HPV testing in cervical cancer screening in collaboration with the WHO HPV Reflab network.

Reference List

[1] IARC Monograph Working Group. IARC Monographs on the Evaluation of Carcinogenic Risks to Humans. Vol 90: Human Papillomaviruses. Lyon: IARCPress, 2007.

[2] Bouvard V, Baan R, Straif K, et al. A review of human carcinogens--Part B: biological agents. Lancet Oncol 2009;10:321-2.

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

[4] Ronco G, Dillner J, Elfstrom KM, et al. Efficacy of HPV-based screening for prevention of invasive cervical cancer: follow-up of four European randomised controlled trials. Lancet 2014;383:524-32.

[5] Poljak M, Cuzick J, Kocjan BJ, Iftner T, Dillner J, Arbyn M. Nucleic acid tests for the detection of human papillomaviruses. Vaccine 2012;30 (Suppl 5):F100-F106.

[6] Meijer CJLM, Castle PE, Hesselink AT, et al. Guidelines for human papillomavirus DNA test requirements for primary cervical cancer screening in women 30 years and older. Int J Cancer 2009;124:516-20.

[7] Arbyn M, Snijders PJ, Meijer CJ, et al. Which high-risk HPV assays fulfil criteria for use in primary cervical cancer screening? Clin Microbiol Infect 2015;in-press.

[8] Khan MJ, Castle PE, Lorincz AT, et al. The elevated 10-year risk of cervical precancer and cancer in women with human papillomavirus (HPV) type 16 or 18 and the possible utility of type-specific HPV testing in clinical practice. J Natl Cancer Inst 2005;97:1072-9.

[9] Li N, Franceschi S, Howell-Jones R, Snijders PJ, Clifford GM. Human papillomavirus type distribution in 30,848 invasive cervical cancers worldwide: Variation by geographical region, histological type and year of publication. Int J Cancer 2011;128:927-35.

[10] Bzhalava D, Guan P, Franceschi S, Dillner J, Clifford G. A systematic review of the prevalence of mucosal and cutaneous human papillomavirus types. Virology 2013;445:224-31.

[11] Arbyn M, Tommasino M, Depuydt C, Dillner J. Are twenty human papillomavirus types causing cervical cancer? J Pathol 2014;234:431-5.

[12] Huh WK, Ault KA, Chelmow D, et al. Use of primary high-risk human papillomavirus testing for cervical cancer screening: interim clinical guidance. Obstet Gynecol 2015;125:330-7.

[13] Unger E, Dillner J. Human papillomavirus laboratory manual. 1st ed. Geneva: WHO (Quality, Safety and Standards (QSS) team of the Department of Immunization, Vaccines and Biologicals), 2010.

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

[15] von Karsa L, Arbyn M, De Vuyst H, et al. European guidelines for quality assurance in cervical cancer screening. Summary of the supplements on HPV screening and vaccination. Papillomavir Res 2015.

[16] Micalessi IM, Boulet GA, Bogers JJ, Benoy IH, Depuydt CE. High-throughput detection, genotyping and quantification of the human papillomavirus using real-time PCR. Clin Chem Lab Med 2012;50:655-61.

[17] Depuydt CE, Benoy IH, Beert JF, Criel AM, Bogers JJ, Arbyn M. Clinical validation of a type-specific real time quantitative human papillomavirus PCR to the performance of Hybrid Capture 2 for the purpose of cervical cancer screening. J Clin Microbiol 2012;50:4073-7.

[18] Schmitt M, Dondog B, Waterboer T, Pawlita M. Homogeneous amplification of genital human alpha papillomaviruses by PCR using novel broad-spectrum GP5+ and GP6+ primers. J Clin Microbiol 2008;46:1050-9.

[19] Schmitt M, Depuydt CE, Benoy I, et al. Viral load of high-risk human papillomaviruses as reliable clinical predictor for the presence of cervical lesions. Cancer Epidemiol Biomarkers Prev 2013;22:406-14.

[20] Snijders PJ, van den Brule AJ, Jacobs MV, Pol RP, Meijer CJ. HPV DNA detection and typing in cervical scrapes. Methods Mol Med 2005;119:101-14.

[21] Geraets DT, Heideman DA, de Koning MN, et al. High-throughput genotyping of high-risk HPV by the digene HPV Genotyping LQ Test using GP5+/6+-PCR and xMAP technology. J Clin Virol 2009;46 Suppl 3:S21-S26.

[22] Schmitt M, Dondog B, Waterboer T, Pawlita M, Tommasino M, Gheit T. Abundance of multiple high-risk human papillomavirus (HPV) infections found in cervical cells analyzed by use of an ultrasensitive HPV genotyping assay. J Clin Microbiol 2010;48:143-9.

[23] Ejegod DM, Serrano I, Cuschieri K, et al. Clinical Validation of the BD Onclarity™ HPV Assay Using a Non-Inferiority Test. J Med Microbiol Diagn 2014;S3: 003:1-4.

[24] Castle PE, Smith KM, Davis TE, et al. Reliability of the Xpert HPV assay to detect high-risk human papillomavirus DNA in a colposcopy referral population. Am J Clin Pathol 2015;143:126-33.

[25] Hesselink AT, Heideman DA, Berkhof J, et al. Comparison of the clinical performance of PapilloCheck human papillomavirus detection with that of the GP5+/6+-PCR-enzyme immunoassay in population-based cervical screening. J Clin Microbiol 2010;48:797-801.

[26] Geraets D, Cuschieri K, Koning M, et al. Clinical evaluation of a GP5+/6+-based Luminex assay having full high-risk HPV genotyping capability and an internal control. J Clin Microbiol 2014;52:3996-4002.

[27] Arbyn M, Schneider A, Gissmann L, Kaufmann AM. Highlights of the 27th International Papillomavirus Conference and Clinical Workshop (Berlin, 17-22 September 2011) - Part 2: Applied Clinical Science. Future Virol 2012;7:19-24.

[28] Schmitt M, Depuydt C, Benoy I, et al. Multiple HPV infections with high viral loads are associated with cervical lesions but do not differentiate grades of cervical abnormalities. J Clin Microbiol 2013;51:1458-64.

[29] Schmitt M, Depuydt C, Benoy I, et al. Prevalence and viral load of 51 genital human papillomavirus types and 3 subtypes. Int J Cancer 2013;132:2395-403.

[30] Cuschieri K, Geraets D, Moore C, Quint W, Duvall E, Arbyn M. Clinical and analytical performance of the Onclarity HPV assay using the VALGENT framework. J Clin Microbiol 2015;in-press.

[31] Rebolj M, Preisler S, Ejegod DM, Rygaard C, Lynge E, Bonde J. Disagreement between human papillomavirus assays: an unexpected challenge for the choice of an assay in primary cervical screening. PLoS ONE 2014;9:e86835.

[32] Unger E, Dillner J. Human papillomavirus laboratory manual. 1st ed. Geneva: WHO (Quality, Safety and Standards (QSS) team of the Department of Immunization, Vaccines and Biologicals), 2010.

[33] Eklund C, Zhou T, Dillner J. A global proficiency study of Human Papillomavirus genotyping. J Clin Microbiol 2010.

[34] Eklund C, Forslund O, Wallin KL, Dillner J. Global improvement in genotyping of Human Papillomavirus DNA: The 2011 HPV LabNet International Proficiency Study. J Clin Microbiol 2013.

[35] Snijders PJ, van den Brule AJ, Meijer CJLM. The clinical relevance of human papillomavirus testing : relationship between analytical and clinical sensitivity. J Pathol 2003;201:1-6.

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

[37] Arbyn M, Verdoodt F, Snijders PJF, et al. Accuracy of human papillomavirus testing on self-collected versus clinician-collected samples: a meta-analysis. Lancet Oncol 2014;15:172-83.

[38] Arbyn M, Andersson K, Bergeron C, Bogers JP, von Knebel-Doeberitz M, Dillner J. Cervical cytology biobanks as a resource for molecular epidemiology. Methods Mol Biol 2011;675:279-98.

[39] Caldwell DM, Ades AE, Higgins JP. Simultaneous comparison of multiple treatments: combining direct and indirect evidence. BMJ 2005;331:897-900.

[40] Saslow D, Solomon D, Lawson HW, et al. American Cancer Society, American Society for Colposcopy and Cervical Pathology, and American Society for Clinical Pathology screening guidelines for the prevention and early detection of cervical cancer. CA Cancer J Clin 2012;62:147-72.

[41] WHO,. WHO Guidelines for Screening and Treatment of Precancerous Lesions for Cervical Cancer Prevention. In: WHO, Broutet N, editors.Geneva, World Health Organisation, 2013: p. 1-40.

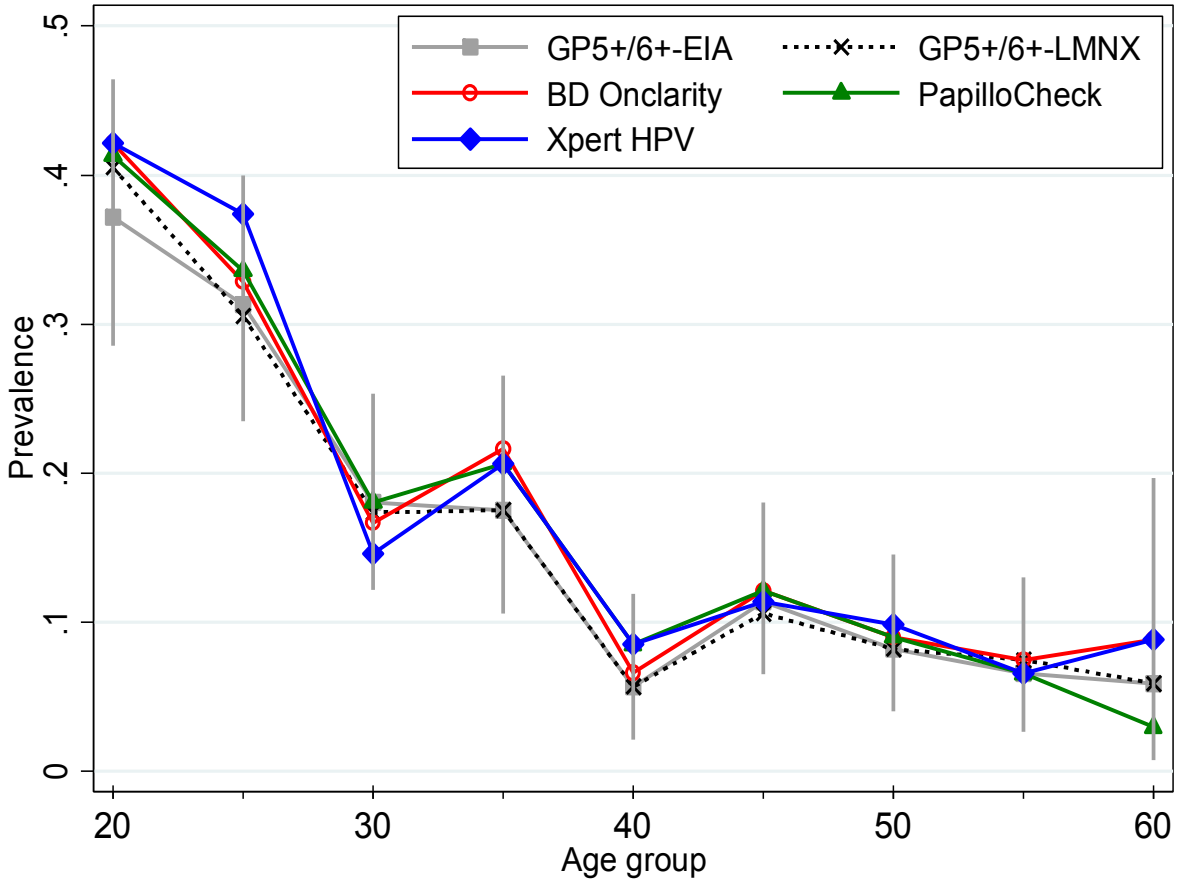
[42] Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977;33:159-74.

[43] Fleiss JL. Statistical Methods for Rates and Proportions. 2nd ed. New York: John Wiley & Sons, 1981.

[44] Tang NS, Tang ML, Chan IS. On tests of equivalence via non-unity relative risk for matched-pair design. Stat Med 2003;22:1217-33.

FIGURES

Figure 1. Prevalence of hrHPV infection by 5-year age group in the Scottish screening population included in VALGENT-2, established with 5 assays. Vertical lines represent 95% confidence intervals around the prevalence measured with the GP5+/6+-EIA assay.


GP5+/6+-EIA
GP5+/6+-LMNX
BD Onclarity
PapilloCheck
Xpert HPV
Prevalence
0
.1
.2
.3
.4
.5
20
30
40
50
60
Age group

TABLES

Table 1. Agreement (% concordance and kappa values) of hrHPV testing between four assays and the GP5/6+ PCR with Luminex-based identification of types , considered at type-specific or aggregated level), assessed in VALGENT-2.

	GP5+/6+-EIA vs GP5+/6+-LMNX		BD Onclarity vs GP5+/6+-LMNX		PapilloCheck vs GP5+/6+-LMNX		Xpert HPV vs GP5+/6+-LMNX	
HPV type	**Con-cordance**	**Kappa**	**Con-cordance**	**Kappa**	**Con-cordance**	**Kappa**	**Con-cordance**	**Kappa**
14 hr types*	98.7%	0.969	96.5%	0.917	94.7%	0.875	93.6%	0.850
13 hr types†	-	-	-	-	92.8%	0.823	-	-
HPV16	-	-	99.3%	0.956	99.0%	0.936	98.5%	0.910
HPV18	-	-	99.6%	0.933	99.0%	0.801	-	-
HPV16/18	-	-	99.1%	0.953	98.2%	0.906	-	-
HPV18/45	-	-	99.4%	0.937	98.5%	0.836	99.0%	0.901
HPV31	-	-	99.5%	0.936	99.5%	0.943	-	-
HPV33	-	-	-	-	99.5%	0.896	-	-
HPV35	-	-	-	-	99.7%	0.887	-	-
HPV39	-	-	-	-	99.0%	0.858	-	-
HPV45	-	-	99.6%	0.913	99.5%	0.891	-	-
HPV51	-	-	99.2%	0.868	98.8%	0.826	-	-
HPV52	-	-	98.4%	0.776	99.4%	0.899	-	-
HPV56	-	-	-	-	99.1%	0.870	-	-
HPV58	-	-	-	-	99.1%	0.695	-	-
HPV59	-	-	-	-	99.1%	0.808	-	-
HPV66	-	-	-	-	98.8%	0.826	-	-
HPV68	-	-	-	-	97.3%	0.279	-	-

hrHPV: high-risk human papillomavirus. High-risk positivity was defined as presence of at least one of 14 HPV types (12 group 1 carcinogens + one group2A carcinogen (HPV68) as defined by IARC [2]). Often a 14th HPV type (HPV66, considered as a group2B carcingoen) is added to the group of high-risk types.

* Thirteen hrHPV types: HPV16,18,31,33,35,39,45,51,52,56,58,59 and 68.

† Fourteen hrHPV types: same as defined in the panel of 13 hrHPV types + HPV66.

Colour legend (adapted from Landis, Biometrics 1977 [42]: dark green ($1.00 \geq \kappa > 0.80$): excellent;
light green ($0.80 \geq \kappa > 0.60$): good; yellow ($0.60 \geq \kappa > 0.40$): moderate; orange ($0.40 \geq \kappa > 0.20$): fair;
red ($0.20 \geq \kappa > 0.00$): poor.

Table 2. Kappa values (and 95% confidence intervals*) expressing the agreement in hrHPV testing and HPV16 genotyping between four HPV assays and GP5+/6+-LMNX according to the type of population (screening/enrichment, consecutive negative cytology/women with histologically confirmed population). Derived from VALGENT-2.

	hrHPV (14)			
Comparison	Screening	Enrichment	Consecutive (-) cytology	CIN2+
GP5+/6+-EIA vs GP5+/6+-LMNX	0.96 (0.93-0.98)	0.96 (0.92-0.99)	0.96 (0.93-0.99)	0.79 (0.51-1.00)
BD Onclarity vs GP5+/6+-LMNX	0.88 (0.84-0.92)	0.90 (0.85-0.96)	0.82 (0.75-0.89)	1.00 (1.00-1.00)
PapilloCheck vs GP5+/6+-LMNX	0.81 (0.76-0.85)	0.90 (0.84-0.95)	0.70 (0.61-0.79)	1.00 (1.00-1.00)
Xpert HPV vs GP5+/6+-LMNX	0.77 (0.72-0.82)	0.86 (0.80-0.93)	0.62 (0.52-0.72)	0.71 (0.39-1.00)

	HPV16			
BD Onclarity vs GP5+/6+-LMNX	0.90 (0.82-0.98)	0.98 (0.95-1.00)	0.78 (0.59-0.97)	0.98 (0.94-1.00)
PapilloCheck vs GP5+/6+-LMNX	0.88 (0.80-0.97)	0.95 (0.91-0.99)	0.82 (0.64-0.99)	1.00 (1.00-1.00)
Xpert HPV vs GP5+/6+-LMNX	0.85 (0.75-0.94)	0.93 (0.88-0.97)	0.74 (0.54-0.93)	0.94 (0.88-1.00)

* confidence intervals are computed as proposed by Fleiss, 1981 [43]

Table 3. Absolute clinical sensitivity and specificity to detect CIN2+ of three HPV genotyping assays (defined for 14 hrHPV types) included in VALGENT-1 for two different definitions of the test positivity threshold.

Test	**Sensitivity (95% CI)‡**	**Specificity (A)† (95% CI) ‡**
Lowest detection cut-off		
qPCR(E6/E7)	100% (95.3-100)	91.1% (88.9-93.0)
BSG5+/6+-PCR/MPG	100% (95.3-100)	79.8% (76.8-82.5)
TS-E7-MPG	98.2% (90.6-100)	80.6% (77.3-83.5)
Optimised detection cut-off		
qPCR(E6/E7)*	100% (95.3-99.7)	95.7% (94.1-97.0)
BSG5+/6+-PCR/MPG**	98.2% (90.8-99.7)	96.2% (94.7-97.4)
TS-E7-MPG***	94.7% (85.4-98.9)	86.4% (83.5-88.9)

‡ Exact binomial 95% confidence intervals.

* The optimised cut-off for test positivity of qPCR was defined at 0.46 copies per cell [19].

** The optimised cut-off for test positivity of BSG5+/6+-PCR/MPG was defined at 0.0007 units [19].

*** The optimised cut-off for test positivity of TS-E7-MPG was defined at 55 units. Contamination with HPV31 took place during sample preparation throughout the whole VALGEN1-1 panel to be tested at IARC-Lyon; moreover multiple type contaminations occurred in a batch of 92 samples (specimen numbers 406-492). For this reason all specimen containing HPV31 as well as the batch of 92 contaminated samples were removed from the analysis.

† Specificity-A was computed for women having negative Pap smears at two consecutive screening rounds.

Table 4. Absolute clinical sensitivity or specificity of HPV assays‡ included in VALGENT-2 studies to detect CIN2+.

Test	**Sensitivity (95% CI)***	**Specificity A (95% CI)***	**Specificity B (95% CI)***
GP5+/6+ PCR-EIA	94.1% (87.6-97.8)	90.3% (88.0-92.4)	83.3% (80.6-85.7)
GP5+/6+ PCR-LMNX	96.1% (90.3-98.9)	90.8% (88.4-92.7)	83.6% (81.0-86.1)
BD Onclarity	96.1% (90.3-98.9)	89.1% (86.7-91.3)	81.4% (78.6-84.0)
PapilloCheck	96.1% (90.3-98.9)	89.7% (87.3-91.8)	82.7% (80.0-85.2)
Xpert HPV	94.1% (87.5-97.8)	90.3% (88.0-92.4)	82.7% (79.9-85.2)

* Exact binomial 95% confidence intervals.

Table 5. Relative accuracy of four novel HPV assays compared to GP5+6+–EIA, evaluated in VALGENT-2, to identify underlying CIN2+.

Comparison	**Outcome**	**Relative sensitivity** (90% CI)*	**Relative specificity A** (90% CI)*	**Relative Specificity B** (90% CI)*
GP5+/6+-LMNX / GP5+/6+ EIA	CIN2+	1.02 (0.97-1.08)	1.00 (0.98-1.03)	1.00 (0.97-1.04)
BD Onclarity / GP5+/6+ EIA	CIN2+	1.02 (0.97-1.08)	0.99 (0.96-1.02)	1.00 (0.97-1.04)
PapilloCheck / GP5+/6+ EIA	CIN2+	1.02 (0.97-1.08)	0.99 (0.97-1.02)	0.99 (0.96-1.03)
Xpert HPV / GP5+/6+ EIA	CIN2+	1.00 (0.94-1.06)	1.00 (0.97-1.03)	0.99 (0.96-1.03)

CI, confidence interval; EIA, enzyme immunoassay; HPV, human papillomavirus.
90% confidence intervals are computed which correspond approximately with one-side non-inferiority testing for matched tests with alpha=0.025 [7,44].