Preventive Medicine 130 (2020) 105927

Contents lists available at ScienceDirect

Preventive Medicine

journal homepage: www.elsevier.com/locate/ypmed

Risk prediction of cervical abnormalities: The value of sociodemographic and lifestyle factors in addition to HPV status

Daniëlle van der Waal[a,b,*], Ruud L.M. Bekkers[c], Stèfanie Dick[d], Charlotte H. Lenselink[e], Leon F.A.G. Massuger[c], Willem J.G. Melchers[f], Channa E. Schmeink[c], Albert G. Siebers[g], Mireille J.M. Broeders[a,b]

[a] Radboud University Medical Center, Radboud Institute for Health Sciences, Nijmegen, the Netherlands
[b] Dutch Expert Centre for Screening, Nijmegen, the Netherlands
[c] Department of Obstetrics and Gynaecology, Radboud University Medical Center, Nijmegen, the Netherlands
[d] Amsterdam UMC, Vrije Universiteit Amsterdam, Pathology, Cancer Center Amsterdam, Amsterdam, the Netherlands
[e] Department of Obstetrics and Gynaecology, Deventer Hospital, Deventer, the Netherlands
[f] Department of Medical Microbiology, Radboud University Medical Center, Nijmegen, the Netherlands
[g] Department of Pathology, Radboud University Nijmegen Medical Center, Nijmegen, the Netherlands

ARTICLE INFO

Keywords:
Cervical abnormalities
Cervical cancer
Risk prediction
Risk-based screening

ABSTRACT

High-risk human papillomavirus (hrHPV) assessment as a primary screening test improves sensitivity but decreases specificity. Determining risk for cervical abnormalities and adapting policy accordingly may improve the balance between screening benefits and harms. Our aim is to assess the value of factors other than HPV in prediction of cervical abnormalities. Data from a Dutch prospective cohort were used. Women aged 18–29 years, not yet eligible for screening, were included in 2007. Data collection consisted of a questionnaire and a cervicovaginal self-sample. Linkage with PALGA (pathology database) was performed in 2017. The analyses included 1483 women. The full model, including sociodemographic and lifestyle factors, was compared to the null model, including baseline HPV only. The outcome of interest was cervical intraepithelial neoplasia 2 or worse (CIN2+). There were 86 women with CIN2+. Baseline hrHPV status was an important predictor (OR = 5.20, 95%CI = 3.27–8.27). The area under the ROC curve (AUC) of the null model was 0.67 (95%CI = 0.61–0.72). The full model had a slightly higher AUC of 0.73 (95%CI = 0.67–0.79). Bootstrap validation indicated that overfitting was present. This exploratory study has confirmed that a single hrHPV measurement is a strong predictor of cervical abnormalities, and additional risk factors in young women appeared to have limited added value. However, prediction based on hrHPV only does leave room for improvement. Future studies should therefore focus on women in the screening age range and search for other predictors to further enhance risk prediction. Adapting policy based on risk may eventually help optimise screening performance.

1. Introduction

Cervical cancer is a common cancer type and cause of cancer death among women worldwide, with an estimated 569,847 new cases and 311,365 deaths in 2018 (Bray et al., 2018). Screening programmes based on cervical smears at regular time intervals have proven to be very effective in reducing disease burden by the detection and subsequent treatment of premalignant disease stages (Peirson et al., 2013). The identification of persistent infection with high-risk human papillomavirus (hrHPV) as a necessary cause of cervical cancer created opportunities to further increase the effectiveness of screening (Walboomers et al., 1999; Cogliano et al., 2005; von Karsa et al., 2015). In the Dutch screening programme, which invites women aged 30–60 years, assessment of hrHPV status has been the primary screening test since 2017. Cytological evaluation is only used as a triage method in hrHPV-positive women. Screening frequency will also be adapted based on hrHPV status. However, the lifetime risk of acquiring a hrHPV infection is estimated at 80% (Bekkers et al., 2004). Since the vast majority of these infections will not result in cervical cancer, other factors must influence pathogenesis. Adapting policy based on risk profiles including these other factors and predictors of future persistent HPV infection, rather than current hrHPV status only, may improve

* Corresponding author at: Radboud University Medical Center, Radboud Institute for Health Sciences (Department for Health Evidence, Mailbox 133), PO Box 9101, 6500 HB Nijmegen, the Netherlands.
E-mail address: Danielle.vanderWaal@radboudumc.nl (D. van der Waal).

https://doi.org/10.1016/j.ypmed.2019.105927
Received 6 March 2019; Received in revised form 21 October 2019; Accepted 18 November 2019
Available online 19 November 2019

screening effectiveness.

Using hrHPV assessment as a primary screening test is known to improve screening sensitivity, albeit with a decrease in specificity (von Karsa et al., 2015; Arbyn et al., 2012; Cuzick et al., 2006; Health Council of the Netherlands, 2011; Rijkaart et al., 2012; Ronco et al., 2010; Ronco et al., 2014). The addition of cytology triage improves specificity, but there is still room for further improvement in terms of unnecessary referrals and follow-up testing. Creating risk profiles with other risk factors, in addition to hrHPV status, could help create a better balance between benefits and harms of screening. A risk model may, for example, inform the screening interval after a negative hrHPV test or, combined with cytology, determine the need for subsequent testing after a positive hrHPV test. There could still be variation in risk of cervical abnormalities within these groups based on risk factors other than HPV status, and the individual benefit-to-harm ratio of screening may be improved if screening is adapted accordingly.

There are several phases that can be identified in cervical carcinogenesis: hrHPV infection, persistence of hrHPV infection, and transition to (pre)malignant stage (Schiffman et al., 2007). Previous studies have observed factors that potentially influence the risk of occurrence of these phases. Sexual behaviour, including age at sexual debut and lifetime number of sexual partners, is associated with HPV positivity (Vaccarella et al., 2006a; Vaccarella et al., 2006b; Vaccarella et al., 2008). Factors that influence persistence and transition, i.e. co-factors, are more difficult to identify. Number of children, previous sexually transmitted infections (STI), smoking, and use of oral contraceptives all appear to be associated with disease occurrence (Bekkers et al., 2004; Castellsagué et al., 2006; Castle et al., 2005; International Collaboration of Epidemiological Studies of Cervical Cancer, 2006; Luhn et al., 2013; Moreno et al., 2002; Muñoz et al., 2002; Plummer et al., 2003; Roura et al., 2014; Roura et al., 2016; Wang et al., 2009; International Collaboration of Epidemiological Studies of Cervical Cancer, 2007). These factors may be combined into a prediction model to accurately estimate the risk of cervical abnormalities.

The aim of this exploratory study is to evaluate what the added value is of epidemiological risk factors, in addition to HPV status, compared to HPV status alone in the prediction of cervical intraepithelial neoplasia (CIN) 2 diagnosis or worse (CIN2+). The current study is performed in a cohort of young unvaccinated women in the Netherlands, in which extensive risk information was collected using questionnaires and HPV self-sample tests. This cohort was chosen for our pilot because no detailed risk factor information is available for the Dutch screening population, and therefore this cohort currently provides the best available information for the development of a prediction model.

2. Methods

2.1. Study population and data collection

Data were used from an observational prospective cohort study (Lenselink et al., 2008; Schmeink et al., 2011). This study focused on the natural history of HPV infection in young unscreened women. Women were recruited June–September 2007. The online study advertisements could be accessed by women throughout the country (e.g. message boards of women's magazines, social media), whereas more active recruitment took place in the eastern part of the country (e.g. posters in fitness centres, pubs, GP offices; recruitment sites at schools and universities). The age range was 18–29 years at baseline. Exclusion criteria consisted of pregnancy or no understanding of the Dutch language. There were 2297 women who initially responded, of whom 2065 eventually participated in the study (Lenselink et al., 2008; Schmeink et al., 2011). Baseline data collection included an extensive questionnaire and a cervicovaginal self-sample kit, consisting of a brush (Rovers Vibabrush®, Rovers Medical Devices Oss, the Netherlands) and a collection tube (SurePath™, Tripath Imaging®, Inc., Burlington NC, USA). Details of sample collection have been described elsewhere (Lenselink et al., 2008). The questionnaire consisted of questions on sociodemographic characteristics and sexual behaviour. All participants provided written informed consent. The study was approved by the local medical ethics committee. The privacy committee of the National Pathology Database (PALGA) consented to the linkage of the cohort data with anonymized cervical cytology and histology data.

2.2. HPV assessment

HPV detection and genotyping has previously been described in detail (Lenselink et al., 2008). In summary, the HPV SPF_{10} PCR-DEIA-$LiPA_{25}$ system (v1; Labo Biomedical Products, Rijswijk, The Netherlands) was used for HPV DNA amplification, detection, and genotype identification. The following HPV types were defined as low-risk types (lrHPV): 6, 11, 34, 40, 42, 43, 44, 53, 54, 70, 74, and X. HPV X refers to samples that were positive for HPV DNA on the DNA enzyme immunoassay (DEIA), but negative on the LiPA and thus were considered positive for HPV but not for hrHPV. This indicates that that specific HPV genotype was not included on the LiPA strip. The high-risk strains (hrHPV) included in the HPV test were: 16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59, 66, and 68/73. HPV type 66 has both been included as a low-risk strain and a high-risk strain in other studies, and it has previously been labelled by IARC as 'possibly carcinogenic' (Bouvard et al., 2009). Reclassification of HPV type 66 to the 'low-risk strains' did not alter our conclusions.

2.3. Outcome assessment

The cohort dataset was linked to PALGA, which contains all cervical cytology and histology data of the participants. This includes the outcome of screening tests as well as the diagnostic test results on medical indication. Linkage was performed in April 2017, resulting in a follow-up time of approximately ten years. Data were anonymized before analysis. No outcome was available for women who had never had a cervical smear, for example because they had not reached the screening age yet at the end of follow-up. There were 1539 participants with baseline data who had been tested during follow-up and could be linked to a record in the PALGA database. There was, however, no outcome included in the PALGA record of some of these participants ($n = 3$). They were excluded from the analyses. In addition, women were excluded when their cervical smear was considered to be of insufficient quality and this smear had not been repeated ($n = 5$). Finally, many potential predictors relate to sexual behaviour, and women who report no lifetime sexual partners are already known to have a very low risk of developing CIN2+ in the following years. Women were therefore only included in the analyses if they reported at least one sexual partner in their lifetime. This resulted in the further exclusion of 48 women.

The following histological confirmed results were included as an outcome of interest: CIN2, CIN3, adenocarcinoma in situ (AIS), and cervical carcinoma (Werkgroep CIN, AIS, en VAIN [Working group CIN, AIS, and VAIN], 2015).

2.4. Statistical analyses

Parts of the questionnaire were repeated at several time points, but we focused on the baseline questions in order to develop a model that more realistically reflects the potential introduction of risk-based screening practice. Although future prediction models may include longitudinal data to improve predictions, a single baseline risk assessment is likely to form the basis at the onset of the programme. All the variables that were measured at baseline are described in Supplemental Table 1. Potential predictors were selected based on review of the literature, clinical relevance, and evaluation of descriptive statistics (Steyerberg, 2009). This selection was discussed with medical experts. The following variables were not identified as strong predictors in the

literature and were therefore excluded: medication use, current treatment, age at menarche, and childhood warts. In addition, some variables both appeared to be of limited importance based on literature and seemed to be non-informative based on overall study population distribution: nationality and/or population group (96.2% of Dutch descent), living with their parents (86.7% not living with their parents), and actively religious (94.9% not actively religious). These variables were excluded as well. Finally, different variables relating to sexual behaviour were measured. Here, the factors that were more commonly identified as predictors in the literature were selected. The final predictor selection consisted of: age at baseline, educational attainment, marital status, smoking status, use of oral contraceptives, age at first sexual contact, number of lifetime sexual partners, and history of STI.

The four categories of marital status were collapsed into 'married or living together' and 'single or in a relationship but not living together', to avoid categories with low numbers. Similarly, educational attainment was dichotomized into: 1) 'low or middle' = primary school or lower, lower vocational or lower secondary school, and intermediate vocational or intermediate/higher secondary school; and 2) 'high' = higher vocational or university. For history of STI, we combined the categories 'one' and 'multiple' previous infection(s). Age, number of lifetime sexual partners, and age at first sexual intercourse were included as continuous linear predictors.

There were few missing values for the potential predictors (Table 1). Multiple imputation, with baseline variables (Supplemental Table 1) as well as the outcome included in the imputation model, was used to replace these missing values. We first performed univariable logistic analyses with all potential predictors selected for analysis. The full multivariable model, consisting of all potential predictors, was then compared to the null model, which consisted of HPV status at baseline only.

We report the Nagelkerke's R^2 of the models, which describes the variability in the outcome that can be explained by the predictors. The discriminatory value of the models was assessed using the area under the receiver operating characteristics curve (AUC). Model discrimination refers to the ability of a model to correctly differentiate between participants with and without an event. AUC values range between 0.5 and 1.0, with a value closer to 1.0 reflecting better discrimination. The change in AUC was calculated as: $((AUC_{full} - 0.5) - (AUC_{null} - 0.5)) / (AUC_{null} - 0.5)$. Calibration refers to the degree to which the estimated probabilities agree with the observed frequencies of the outcome. This was evaluated using calibration plots and calibration slopes, with a value closer to 1 reflecting better calibration. We performed an internal validation using bootstrapping to give optimism-corrected estimates of the AUC and the calibration slope (20 multiple imputation sets, 5 bootstrap samples of each set) (Wahl et al., 2016). The analyses were performed in R 3.5.1.

3. Results

3.1. Study population characteristics

A total of 1483 women filled out the baseline questionnaire, had had a cervical smear after baseline data collection (2007) and before the end of follow-up (2017), and were sexually active at baseline. The baseline characteristics of this population are described in Supplemental Table 1. At the end of the 10-year follow-up period, 86 women had been diagnosed with CIN2+ (Table 2). Mild (CIN1) or unspecified CIN was observed in 23 women. Furthermore, an additional 137 women were followed up based on abnormal cytology, but no histology result was obtained.

3.2. Univariable analyses

At baseline, 16.2% (n = 240) of the women tested positive for hrHPV (Table 1). A clear difference was observed in prevalence of hrHPV infection between women who eventually developed CIN2+ during the 10-year follow-up period (45.3% hrHPV; n = 39) and women who did not (14.4% hrHPV; n = 201). Women with CIN2+ also tended to be lower educated (30.2% vs. 21.0%) and more often had a history of STI (19.8% vs. 10.6%). The number of lifetime sexual partners was higher in women with CIN2+ (median 7 vs. 4). The univariable analyses showed that a baseline infection with hrHPV (OR = 5.20, 95%CI = 3.27–8.27) indeed resulted in a significantly increased risk of CIN2+ (Table 3). A higher number of lifetime sexual partners (OR = 1.06, 95%CI = 1.04–1.09) and a history of STI (OR = 2.08, 95%CI = 1.19–3.64) were also significantly associated with an increased risk, whereas an inverse association with educational attainment was observed (OR = 0.62, 95%CI = 0.38–0.99).

The descriptive results in Table 1 also appeared to suggest that the women who developed CIN2+ may have smoked more often (27.9% vs. 20.6%), more often used oral contraceptives (74.4% vs. 70.9%), were less often married or living together (29.1% vs. 34.0%), were slightly older at baseline (median 25 vs. 24 years), and were somewhat younger at sexual debut (median 16 vs. 17 years). The univariable analyses showed, however, that none of these factors were significantly associated with CIN2+ in the study population (Table 3).

A sensitivity analysis showed that excluding women who were 18–19 years at baseline, i.e. may not have had a screening opportunity during follow-up, had little effect on the association between age and

Table 1
Baseline predictors according to outcome.

	Normal smear (n = 1397)	CIN2+ (n = 86)	Total (n = 1483)
HPV status at baseline, n (%)			
No HPV	1098 (78.6)	41 (47.7)	1139 (76.8)
lrHPV	96 (6.9)	6 (7.0)	102 (6.9)
hrHPV	201 (14.4)	39 (45.3)	240 (16.2)
Missing[a]	2 (0.1)	0 (0.0)	2 (0.1)
Age at baseline in y, median (IQR)	24 (22–27)	25 (23–27)	24 (22–27)
Missing, n (%)	0 (0.0)	0 (0.0)	0 (0.0)
Educational attainment, n (%)			
High[b]	1099 (78.7)	60 (69.8)	1159 (78.2)
Low + Middle[b]	293 (21.0)	26 (30.2)	319 (21.5)
Missing	5 (0.4)	0 (0.0)	5 (0.3)
Living together or married, n (%)			
Yes	475 (34.0)	25 (29.1)	500 (33.7)
No	916 (65.6)	60 (69.8)	976 (65.8)
Missing	6 (0.4)	1 (1.2)	7 (0.5)
Current smoking, n (%)			
Yes	288 (20.6)	24 (27.9)	312 (21.0)
No	1100 (78.7)	62 (72.1)	1162 (78.4)
Missing	9 (0.6)	0 (0.0)	9 (0.6)
Oral contraceptive use, n (%)			
Yes	990 (70.9)	64 (74.4)	1054 (71.1)
No	405 (29.0)	22 (25.6)	427 (28.8)
Missing	2 (0.1)	0 (0.0)	2 (0.1)
Age first sexual contact in y, median (IQR)	17 (16–18)	16 (15–18)	16 (16–18)
Missing, n (%)	3 (0.2)	0 (0.0)	3 (0.2)
Number of lifetime sexual partners, median (IQR)	4 (2–7)	7 (4–12)	4 (2–8)
Missing, n (%)	6 (0.4)	0 (0.0)	6 (0.4)
History of STI, n (%)			
Yes	148 (10.6)	17 (19.8)	165 (11.1)
No	1246 (89.2)	68 (79.1)	1314 (88.6)
Missing	3 (0.2)	1 (1.2)	4 (0.3)

IQR = interquartile range; STI = sexually transmitted infection.

[a] These women did return a self-sample test at baseline, but the control beta-globin test was negative.

[b] 'Low or middle' = primary school or lower; lower vocational or lower secondary school; and intermediate vocational or intermediate/higher secondary school; 'high' = higher vocational or university.

Table 2
Outcome according to HPV status at baseline.

HPV status at baseline[a]	Highest reported outcome							
	Cytology	Histology						
	ASC-US +	CIN1	CIN2	CIN3	AIS	Unspecified dysplasia	Cervical carcinoma	Total CIN2 +
No HPV (n = 1139)	96	11	17	22	0	3	2	41
lrHPV (n = 102)	15	1	3	3	0	0	0	6
hrHPV (n = 240)	26	7	17	18	3	1	1	39
Total	137	19	37	43	3	4	3	86

[a] HPV status was not available for two women who did return the self-sample test: the control beta-globin test was negative.

Table 3
Univariable analyses CIN2 or worse.

Characteristics	OR (95% CI)
HPV status at baseline	
No HPV	1 (reference)
lrHPV	1.67 (0.69–4.04)
hrHPV	5.20 (3.27–8.27)
Age in y	1.07 (0.99–1.15)
High educational level	0.62 (0.38–0.99)
Married or living together	0.81 (0.50–1.30)
Current smoker	1.48 (0.91–2.41)
Current use of oral contraceptives	1.19 (0.72–1.96)
Age at first sexual contact in y	0.90 (0.80–1.01)
Number of lifetime sexual partners	1.06 (1.04–1.09)
History of STI	2.08 (1.19–3.64)

STI = sexually transmitted infection.

CIN2 + (OR = 1.09, 95%CI = 1.00–1.19). Other univariable and multivariable analyses gave very similar results compared to the main analyses (results not shown).

3.3. Multivariable analyses

In the multivariable full model (Table 4), only hrHPV (OR = 4.49, 95%CI = 2.69–7.48) and educational attainment (OR = 0.58, 95%CI = 0.35–0.98) were significantly associated with CIN2 + . Marital status (OR = 0.97, 95%CI = 0.56–1.69) and smoking status (OR = 0.96, 95%CI = 0.56–1.66) did not appear to have any effect on the outcome, with ORs close to one and broad confidence intervals. For age (OR = 1.06, 95%CI = 0.97–1.16), age at first sexual contact (OR = 0.93, 95%CI = 0.82–1.06), and number of lifetime sexual partners (OR = 1.02, 95%CI = 0.99–1.06), it should be noted that the steps (i.e. 1 year or one partner) are relatively small and thus the expected change in risk is also small.

Table 4
Multivariable analyses CIN2 or worse.

	Beta coefficient (SE)	OR (95% CI)
Null model		
Intercept	−3.2892 (0.1591)	
lrHPV	0.5151 (0.4499)	1.67 (0.69–4.04)
hrHPV	1.6487 (0.2365)	5.20 (3.27–8.27)
AUC (95% CI)	0.67 (0.61–0.72)	
Full model		
Intercept	−3.5514 (1.3759)	
lrHPV	0.3246 (0.4585)	1.38 (0.56–3.40)
hrHPV	1.5008 (0.2610)	4.49 (2.69–7.48)
Age in y	0.0590 (0.0451)	1.06 (0.97–1.16)
High educational level	−0.5404 (0.2646)	0.58 (0.35–0.98)
Married or living together	−0.0259 (0.2807)	0.97 (0.56–1.69)
Current smoker	−0.0372 (0.2764)	0.96 (0.56–1.66)
Current use of oral contraceptives	0.3524 (0.2737)	1.42 (0.83–2.43)
Age at first sexual contact in y	−0.0700 (0.0649)	0.93 (0.82–1.06)
Number of lifetime sexual partners	0.0240 (0.0163)	1.02 (0.99–1.06)
History of STI	0.2710 (0.3081)	1.31 (0.72–2.40)
AUC (95% CI)	0.73 (0.67–0.79)	
Optimism-corrected AUC	0.70	

AUC = area under the ROC curve; STI = sexually transmitted infection.

3.4. Model performance

The Nagelkerke's R^2 increased from 8.3% for the null model to 11.1% for the full model. The AUC of the null model with hrHPV status only was 0.67 (95%CI = 0.61–0.72). The AUC of the full model (0.73, 95%CI = 0.67–0.79) was higher compared to the model with hrHPV only. This constitutes a 39% increase in AUC (using non-rounded values): ((0.731–0.5) − (0.666–0.5)) / (0.666–0.5). The lower optimism-corrected AUC estimate for the full model (AUC = 0.70), however, indicates overfitting. The calibration slope of the full model estimated with bootstrap validation was 0.87, which suggests that shrinkage of the regression coefficients is required (Steyerberg, 2009).

Sensitivity analyses (results not shown) indicated that a model based on sociodemographic and lifestyle factors only, excluding HPV status, had an AUC similar to the model based on HPV status only: 0.67 (95%CI = 0.61–0.73).

4. Discussion

Baseline hrHPV status was a strong predictor of cervical abnormalities in the current study, as expected. The AUC of the model based on HPV only – which tells us whether we can distinguish between women who will and women who will not develop CIN2 + − does indicate, however, that there is still room for improvement. In the current study, we focused on the addition of sociodemographic and lifestyle factors. This resulted in a slightly higher AUC. However, it is important to note that our ultimate aim is to study the added value of applying a prediction model in an older age group in the context of risk-based screening for cervical cancer. Future studies should be performed to investigate if model performance can be improved in the screening population.

Although not all factors were significantly associated with the outcome, the direction of the effects in our univariable analyses was in line with previous literature (Bekkers et al., 2004; Castellsagué et al., 2006; Castle et al., 2005; International Collaboration of Epidemiological Studies of Cervical Cancer, 2006; Luhn et al., 2013; Moreno et al., 2002; Muñoz et al., 2002; Plummer et al., 2003; Roura et al., 2014; Roura et al., 2016; Wang et al., 2009; International Collaboration of Epidemiological Studies of Cervical Cancer, 2007). The impact of these factors, their potential relevance as a predictor, and type of association (e.g. linear or not) may, however, differ between populations. Furthermore, it has been suggested that co-factors can only be identified in studies in hrHPV-positive women (Plummer et al., 2003), which raises the question whether we can expect them to add sufficiently to risk prediction. Potentially, more additional factors will have to be identified in terms of genetics and (circulating) biomarkers by means of

transcriptomics, metabolomics and/or microbiomics (Ebisch et al., 2016).

There are no established prediction models for cervical abnormalities, but previous efforts to develop similar models have been reported (Charlton et al., 2013; Lee et al., 2015; Rothberg et al., 2018; Sengupta et al., 1999). Lee et al. developed a CIN2+ prediction model – including age, (passive) smoking, age at first sexual contact, and hrHPV DNA load – which had an AUC of 0.866 (Lee et al., 2015). This was an improvement compared to the model without age and hrHPV DNA (AUC = 0.676). The study of Rothberg and colleagues had an aim similar to ours, using data from a cohort of women (n = 99,319) aged 30 years or older in the USA. The AUC of their CIN2+ prediction model was 0.81, an increase compared to HPV status only (AUC = 0.71) (Rothberg et al., 2018). The model included age, race, marital status, insurance, smoking, income, and HPV status. The results of Charlton et al. were less promising, but their study had a different aim: they developed a model in women with ASC-US or low grade squamous intraepithelial lesions (LSIL) (Charlton et al., 2013). The differences between the results of these studies and our study may be explained by differences in design and setting (e.g. general population vs. women with ASC-US/LSIL, age group), availability of predictors, coding of predictors, sensitivity/specificity of HPV test, and chance findings due to limited sample sizes.

Studies have shown that hrHPV assessment as a primary screening test decreases the number of cervical cancer cases and deaths due to increased sensitivity, but the number of (potentially false-positive) referrals is increased (von Karsa et al., 2015; Health Council of the Netherlands, 2011; Rijkaart et al., 2012; Ronco et al., 2010; Ronco et al., 2014). Information about other (risk) factors could further optimise screening: tailoring screening based on risk might result in less false-positive and false-negative results as well as less overdiagnosis. The minimally required AUC is still unknown, but studies on prediction models in related fields have suggested that an AUC of at least 0.8 is needed in more risk-based strategies (Gail and Pfeiffer, 2018). The AUC of a model based on hrHPV assessment only, using the SPF_{10} PCR-DEIA-$LiPA_{25}$, was 0.67. Although we were not able to compare different HPV tests and the subsequent effect on the AUC, we believe that this result does indicate that other factors need to be added to achieve a sufficiently high discriminatory ability for risk-based screening.

Ultimately, we recommend to study the hrHPV-positive and hrHPV-negative group separately in future research. If further risk stratification is possible in the hrHPV-negative group, then different scenarios based on varying screening intervals could be evaluated in a modelling study, for example. In the hrHPV-positive group, on the other hand, risk stratification may also inform the need for subsequent testing. Here, different combinations of factors – e.g. also including cytology, HPV type, and other biomarkers – have to be considered to further optimise triage. Furthermore, it is also important to eventually distinguish between short-term and long-term risk.

Our study had some limitations. There were relatively few events compared to the number of potential predictors. Data-driven decisions have been reported to lead to overfitting in small studies (Collins et al., 2015; Steyerberg et al., 2003; Steyerberg et al., 2000), and we have therefore decided to limit the number of modelling steps. A larger study sample and additional predictors are needed for the development of a model that can be used in screening. Furthermore, the baseline age differs from the target age of the existing screening programme. In this study, HPV status was determined at an age at which hrHPV screening is not advised due to the high hrHPV prevalence. Both the low baseline age and relatively homogenous study population may have contributed to weaker associations, although they generally pointed in the direction we had anticipated. We emphasize the need for a larger follow-up study in the screening age group, including a more heterogeneous population to improve generalisability to the screening population. This initial exploratory study provides essential leads for such a study.

The cohort consisted of women who had not been vaccinated against HPV. Prophylactic HPV vaccination was implemented in the Netherlands in 2009, starting with a catch-up phase in which girls from the birth cohorts 1993–1997 (13–16 years) were invited. The vaccination coverage was only 50% (Rondy et al., 2010). The vaccination programme has now been fully implemented, inviting girls in the year they turn 13. Vaccination uptake is, however, still low (2004 cohort: 45.5%) (Rijksinstituut voor Volksgezondheid en Milieu (RIVM) [National Institute for Public Health and the Environment], 2019). The three registered vaccines in the Netherlands (Gardasil®, Gardasil 9®, and Cervarix®) protect against several hrHPV types, including the most prevalent strains (16/18). Although these women are thus at a significantly lower risk of developing cervical cancer, they will be advised to attend screening based on current guidelines. Furthermore, the low vaccination uptake indicates that screening is still very important for the years to come. Vaccine developments may change the need for screening in the future, as newer vaccines protect against more HPV types and vaccination coverage potentially increases, but many women could still benefit from screening in the next few decades. Research into more effective screening strategies, optimising the benefit-to-harm ratio, is thus still warranted. A risk model that includes vaccination status may be useful in the future as well, to determine policy in the different vaccination groups before the first screening test has been performed.

In conclusion, this exploratory study has confirmed that a single baseline HPV measurement is a strong predictor of cervical abnormalities and found that additional risk factors in young women appeared to have limited added value. For future use in the screening setting, however, risk prediction would need to be improved by including more factors rather than hrHPV only. The addition of epidemiological risk factors is an interesting option, considering they are relatively easy to measure in a screening population (e.g. questionnaires) and several cofactors have previously been identified. Risk prediction models may both help optimise the initial screening policy, based on vaccination status, or the screening policy after the first test, based on hrHPV status. Large prospective cohort studies are needed to develop and validate risk prediction models for the screening setting.

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ypmed.2019.105927.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Declaration of competing interest

None.

References

Arbyn, M., Ronco, G., Anttila, A., Meijer, C.J.L.M., Poljak, M., Ogilvie, G., et al., 2012. Evidence regarding human papillomavirus testing in secondary prevention of cervical cancer. Vaccine 30 (Suppl. 5), F88–F99.

Bekkers, R.L.M., Massuger, L.F.A.G., Bulten, J., Melchers, W.J.G., 2004. Epidemiological and clinical aspects of human papillomavirus detection in the prevention of cervical cancer. Rev. Med. Virol. 14, 95–105.

Bouvard, V., Baan, R., Straif, K., Grosse, Y., Secretan, B., El Ghissassi, F., et al., 2009. A review of human carcinogens—part B: biological agents. Lancet Oncol. 10, 321–322.

Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R.L., Torre, L.A., Jemal, A., 2018. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J. Clin. 68, 394–424.

Castellsagué, X., Díaz, M., de Sanjosé, S., Muñoz, N., Herrero, R., Franceschi, S., et al., 2006. Worldwide human papillomavirus etiology of cervical adenocarcinoma and its cofactors: implications for screening and prevention. J. Natl. Cancer Inst. 98, 303–315.

Castle, P.E., Walker, J.L., Schiffman, M., Wheeler, C.M., 2005. Hormonal contraceptive use, pregnancy and parity, and the risk of cervical intraepithelial neoplasia 3 among oncogenic HPV DNA-positive women with equivocal or mildly abnormal cytology. Int. J. Cancer 117, 1007–1012.

Charlton, B.M., Carwile, J.L., Michels, K.B., Feldman, S., 2013. A cervical abnormality

risk prediction model: can we use clinical information to predict which patients with ASCUS/LSIL Pap tests will develop CIN 2/3 or AIS? J Low Genit Tract Dis 17, 242–247.
Cogliano, V., Baan, R., Straif, K., Grosse, Y., Secretan, B., El Ghissassi, F., et al., 2005. Carcinogenicity of human papillomaviruses. Lancet Oncol. 6, 204.
Collins, G.S., Reitsma, J.B., Altman, D.G., Moons, K.G.M., 2015. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. BMJ 350, g7594.
Cuzick, J., Clavel, C., Petry, K.U., Meijer, C.J.L.M., Hoyer, H., Ratnam, S., et al., 2006. Overview of the European and North American studies on HPV testing in primary cervical cancer screening. Int. J. Cancer 119, 1095–1101.
Ebisch, R.M.F., Siebers, A.G., Bosgraaf, R.P., Massuger, L.F.A.G., Bekkers, R.L.M., Melchers, W.J.G., 2016. Triage of high-risk HPV positive women in cervical cancer screening. Expert. Rev. Anticancer. Ther. 16, 1073–1085.
Gail, M.H., Pfeiffer, R.M., 2018. Breast cancer risk model requirements for counseling, prevention, and screening. J. Natl. Cancer Inst. 110, 994–1002.
Health Council of the Netherlands, 2011. Population screening for cervical cancer, publication no. 2011/07 [Screening op Baarmoederhalskanker, publicatienr 2011/07]. Available from: https://www.gezondheidsraad.nl/documenten/adviezen/2011/05/24/screening-op-baarmoederhalskanker.
International Collaboration of Epidemiological Studies of Cervical Cancer, 2006. Cervical carcinoma and reproductive factors: collaborative reanalysis of individual data on 16,563 women with cervical carcinoma and 33,542 women without cervical carcinoma from 25 epidemiological studies. Int. J. Cancer 119, 1108–1124.
International Collaboration of Epidemiological Studies of Cervical Cancer, 2007. Comparison of risk factors for invasive squamous cell carcinoma and adenocarcinoma of the cervix: collaborative reanalysis of individual data on 8,097 women with squamous cell carcinoma and 1,374 women with adenocarcinoma from 12 epidemiological studies. Int. J. Cancer 120, 885–891.
von Karsa, L., Arbyn, M., De Vuyst, H., Dillner, J., Dillner, L., Franceschi, S., et al., 2015. European guidelines for quality assurance in cervical cancer screening. Summary of the supplements on HPV screening and vaccination. Papillomavirus Res 1, 22–31.
Lee, C.H., Peng, C.Y., Li, R.N., Chen, Y.C., Tsai, H.T., Hung, Y.H., et al., 2015. Risk evaluation for the development of cervical intraepithelial neoplasia: development and validation of risk-scoring schemes. Int. J. Cancer 136, 340–349.
Lenselink, C.H., Melchers, W.J.G., Quint, W.G.V., Hoebers, A.M.J., Hendriks, J.C.M., Massuger, L.F.A.G., Bekkers, R.L.M., 2008. Sexual behaviour and HPV infections in 18 to 29 year old women in the pre-vaccine era in the Netherlands. PLoS One 3, e3743.
Luhn, P., Walker, J., Schiffman, M., Zuna, R.E., Dunn, S.T., Gold, M.A., et al., 2013. The role of co-factors in the progression from human papillomavirus infection to cervical cancer. Gynecol. Oncol. 128, 265–270.
Moreno, V., Bosch, F.X., Muñoz, N., Meijer, C.J.L.M., Shah, K.V., Walboomers, J.M.M., et al., 2002. Effect of oral contraceptives on risk of cervical cancer in women with human papillomavirus infection: the IARC multicentric case-control study. Lancet 359, 1085–1092.
Muñoz, N., Franceschi, S., Bosetti, C., Moreno, V., Herrero, R., Smith, J.S., et al., 2002. Role of parity and human papillomavirus in cervical cancer: the IARC multicentric case-control study. Lancet 359, 1093–1101.
Peirson, L., Fitzpatrick-Lewis, D., Ciliska, D., Warren, R., 2013. Screening for cervical cancer: a systematic review and meta-analysis. Syst Rev 2, 35.
Plummer, M., Herrero, R., Franceschi, S., Meijer, C.J.L.M., Snijders, P., Bosch, F.X., et al., 2003. Smoking and cervical cancer: pooled analysis of the IARC multi-centric case-control study. Cancer Causes Control 14, 805–814.
Rijkaart, D.C., Berkhof, J., van Kemenade, F.J., Coupe, V.M.H., Hesselink, A.T., Rozendaal, L., et al., 2012. Evaluation of 14 triage strategies for HPV DNA-positive women in population-based cervical screening. Int. J. Cancer 130, 602–610.
Rijksinstituut voor Volksgezondheid en Milieu (RIVM) [National Institute for Public Health and the Environment], 2019. Vaccinatiegraad en jaarverslag Rijksvaccinatieprogramma Nederland 2018 [Immunisation coverage and annual report National Immunisation Programme in the Netherlands 2018]. Available from. https://www.rivm.nl/publicaties/vaccinatiegraad-en-jaarverslag-rijksvaccinatieprogramma-nederland-2018.
Ronco, G., Giorgi-Rossi, P., Carozzi, F., Confortini, M., Dalla Palma, P., Del Mistro, A., et al., 2010. Efficacy of human papillomavirus testing for the detection of invasive cervical cancers and cervical intraepithelial neoplasia: a randomised controlled trial. Lancet Oncol. 11, 249–257.
Ronco, G., Dillner, J., Elfström, K.M., Tunesi, S., Snijders, P.J., Arbyn, M., et al., 2014. Efficacy of HPV-based screening for prevention of invasive cervical cancer: follow-up of four European randomised controlled trials. Lancet 383, 524–532.
Rondy, M., van Lier, A., van de Kassteele, J., Rust, L., de Melker, H., 2010. Determinants for HPV vaccine uptake in the Netherlands: a multilevel study. Vaccine 28, 2070–2075.
Rothberg, M.B., Hu, B., Lipold, L., Schramm, S., Jin, X.W., Sikon, A., Taksler, G.B., 2018. A risk prediction model to allow personalized screening for cervical cancer. Cancer Causes Control 29, 297–304.
Roura, E., Castellsagué, X., Pawlita, M., Travier, N., Waterboer, T., Margall, N., et al., 2014. Smoking as a major risk factor for cervical cancer and pre-cancer: results from the EPIC cohort. Int. J. Cancer 135, 453–466.
Roura, E., Travier, N., Waterboer, T., de Sanjosé, S., Bosch, F.X., Pawlita, M., et al., 2016. The influence of hormonal factors on the risk of developing cervical cancer and pre-cancer: results from the EPIC cohort. PLoS One 11, e0147029.
Schiffman, M., Castle, P.E., Jeronimo, J., Rodriguez, A.C., Wacholder, S., 2007. Human papillomavirus and cervical cancer. Lancet 370, 890–907.
Schmeink, C.E., Melchers, W.J.G., Siebers, A.G., Quint, W.G.V., Massuger, L.F.A.G., Bekkers, R.L.M., 2011. Human papillomavirus persistence in young unscreened women, a prospective cohort study. PLoS One 6, e27937.
Sengupta, S., Chaudhuri, S.K., Das, P., Raichaudhuri, B., Pramanik, R., 1999. Integrated approach to prediction of cervical intra-epithelial neoplasia status through a case-control study. Int. J. Cancer 84, 69–73.
Steyerberg, E.W., 2009. Clinical Prediction Models: A Practical Approach to Development, Validation, and Updating. Springer, New York.
Steyerberg, E.W., Eijkemans, M.J.C., Harrell, F.E., Habbema, J.D.F., 2000. Prognostic modelling with logistic regression analysis: a comparison of selection and estimation methods in small data sets. Stat. Med. 19, 1059–1079.
Steyerberg, E.W., Bleeker, S.E., Moll, H.A., Grobbee, D.E., Moons, K.G.M., 2003. Internal and external validation of predictive models: a simulation study of bias and precision in small samples. J. Clin. Epidemiol. 56, 441–447.
Vaccarella, S., Franceschi, S., Herrero, R., Muñoz, N., Snijders, P.J.F., Clifford, G.M., et al., 2006a. Sexual behavior, condom use, and human papillomavirus: pooled analysis of the IARC human papillomavirus prevalence surveys. Cancer Epidemiol. Biomark. Prev. 15, 326–333.
Vaccarella, S., Herrero, R., Dai, M., Snijders, P.J.F., Meijer, C.J.L.M., Thomas, J.O., et al., 2006b. Reproductive factors, oral contraceptive use, and human papillomavirus infection: pooled analysis of the IARC HPV prevalence surveys. Cancer Epidemiol. Biomark. Prev. 15, 2148–2153.
Vaccarella, S., Herrero, R., Snijders, P.J.F., Dai, M., Thomas, J.O., Hieu, N.T., et al., 2008. Smoking and human papillomavirus infection: pooled analysis of the International Agency for Research on Cancer HPV prevalence surveys. Int. J. Epidemiol. 37, 536–546.
Wahl, S., Boulesteix, A.L., Zierer, A., Thorand, B., Avan de Wiel, M., 2016. Assessment of predictive performance in incomplete data by combining internal validation and multiple imputation. BMC Med. Res. Methodol. 16, 144.
Walboomers, J.M.M., Jacobs, M.V., Manos, M.M., Bosch, F.X., Kummer, J.A., Shah, K.V., et al., 1999. Human papillomavirus is a necessary cause of invasive cervical cancer worldwide. J. Pathol. 189, 12–19.
Wang, S.S., Zuna, R.E., Wentzensen, N., Dunn, S.T., Sherman, M.E., Gold, M.A., et al., 2009. Human papillomavirus cofactors by disease progression and human papillomavirus types in the study to understand cervical cancer early endpoints and determinants. Cancer Epidemiol. Biomark. Prev. 18, 113–120.
Werkgroep CIN, AIS, en VAIN [Working group CIN, AIS, and VAIN]. CIN, AIS en VAIN Landelijke richtlijn (Versie: 1.0) [CIN, AIS, and VAIN National guideline (Version 1.0)]. 2015. Available from: https://www.oncoline.nl/cin-ais-en-vain.