An Updated Natural History Model of Cervical Cancer: Derivation of Model Parameters

N. G. Campos, E. A. Burger, S. Sy, M. Sharma, M. Schiffman, A. C. Rodriguez, A. Hildesheim, R. Herrero, and J. J. Kim

WEB APPENDIX

This supplementary appendix provides additional details on methods and results presented in the main manuscript.

Initial parameterization

HPV incidence

Cumulative incidence rates from a longitudinal study (1) of 1,610 sexually active women aged 15–85 years in Bogota, Colombia, who were followed every 6 months for 4.1 years on average, were used to estimate monthly probabilities. We smoothed these step functions to reflect gradual increasing and declining incidence by age, while maintaining the average HPV incidence within each age group. Due to a lack of data on incidence in young girls, we extrapolated incidence among 9- to 14-year-olds to increase gradually and peak at age 19–20 years, consistent with published studies.

Oncogenic HPV progression: Algorithm for statistically determining CIN2+ attribution in the CVT

The relevance of HPV infections for CIN2+ attribution was determined statistically (rather than molecularly) as follows, with the *key visit* defined as the same visit as the CIN2+ procedure, or the last visit with polymerase chain reaction (PCR) results before the CIN2+ procedure: 1) If there was at least one oncogenic type at the key visit that was also present at prior visits, all such types were considered relevant (n = 122); 2) else if there was at least one oncogenic type at the key visit and the key visit was the enrollment visit, all such types were considered relevant (n = 10); 3) else if there was at least one oncogenic type at the visit prior to the key visit which was also present before that, all such types were considered relevant (n = 2); 4) else if there was at least one oncogenic type first detected at either the key visit or the visit prior to the key visit, all such types were considered relevant (n = 7); 5) else there were no relevant types because PCR results from the key visit and the visit prior to the key visit were both oncogenic negative (n = 8).

Based on determination of relevant HPV types, CIN2+ cases were attributed hierarchically as follows: 1) if there was only one relevant type, CIN2+ was attributed fully to that type (n = 57); 2) else if HPV16 was relevant, CIN2+ was attributed fully to HPV16 (n = 64); 3) else if multiple oncogenic types were relevant, CIN2+ was attributed to all relevant types, with the percent attribution set at the proportion of single infected CIN2+ with that type divided by the sum of the percent of single infected CIN2+ with any relevant types present (n = 20); 4) else there were no relevant types and CIN2+ was not attributed (n = 8). Following this type attribution hierarchy, progression rates were calculated based on the cumulative incidence of CIN2+

associated with each type. Due to small numbers, there were some intervals with no progression, in which case rates from the previous interval were carried forward.

Nononcogenic HPV clearance and progression

We derived clearance and progression probabilities for pooled nononcogenic HPV types (**Web Figures 1** and **2**) using data from the Guanacaste cohort study, a large population-based HPV natural history study, in which a risk-stratified cohort of women ages ≥18 years was actively screened every 12 months for 5–7 years, except for women with LSIL or CIN1 at any visit, who were shifted to a 6-month screening interval (2). This analysis was performed comparably to the derivation of clearance rates for oncogenic types, with minor modifications due to differences in design and follow-up of the Guanacaste cohort relative to the CVT. Because of the large number of incident infections and the inclusion of older women in the cohort study, we restricted the analysis of nononcogenic types to incident infections.

Timing of visits in the Guanacaste cohort differed slightly from the CVT. The first interval, "Year 1", included HPV results from the first <15 months of the study, and subsequent 12-month intervals began at 15 months (15 to <27 months, etc); therefore in estimating monthly hazards of clearance or progression, we assumed the length of the "Year 1" interval was 15 months and length of subsequent years was 12 months. Details on imputations made for missing results are discussed below.

Nononcogenic types occasionally progress to CIN2 (3). Based on our type attribution scheme for oncogenic types using CVT data, 8 out of 149 infections remained unattributed after we prioritized infections with persistent and/or oncogenic types. Relative to the 64 CIN2+ attributed to HPV16, 6 cases of CIN2+ were detected with nononcogenic types alone. We thus assumed nononcogenic types progressed to CIN2+ at 10% of the rate of HPV16 infections through month 51, after which nononcogenic progression rates were held constant.

Precancer regression and progression

Transition probabilities related to precancer are presented in **Web Table 1.** We examined the range of CIN2 regression rates from the control arms of published clinical trials, a prospective cohort study, and a screening study that involved watchful waiting of CIN2 to determine a reasonable value for regression (4–8). For ethical reasons, little information is available on the regression of CIN3 lesions. We assumed these more advanced lesions were less likely to clear than CIN2, and set duration-dependent regression rates at 50% of CIN2 rates.

While data on progression to cancer is limited for ethical reasons, a retrospective analysis of historic data provides information on the magnitude of prevalent CIN3 progression rates (9). We assumed these rates were for lesions that had already persisted long enough to be detected (i.e., more than 10 years), and by using the published data as a guide and performing experiments to fit historic cancer incidence in the United States, derived CIN3 progression rates. Based on these experiments, we set CIN2 progression rates at 20% of CIN3 progression rates.

Cancer progression and mortality

Transition probabilities related to cancer are presented in **Web Table 1.** Women with undetected cervical cancer may progress to a more severe stage, face stage-specific cancer mortality rates, or have their cancer detected through presentation of symptoms (**Web Table 1**). Once a cancer is detected, women are assumed to receive stage-specific treatment and are subject to cancer-specific mortality derived from the SEER registry data from years 2000-2009 and conditioned on surviving to months 1, 12, and 36. After 20 years, excess mortality due to cancer is assumed to be zero. To adjust detected cancer mortality for age at diagnosis, we derived age-specific multipliers for each stage of cancer (i.e., local, regional, distant) and applied these to the base cancer-specific mortality rates (**Web Table 1**) for the following age groups: <30 years, 30–39 years, 40–49 years, 50–59 years, 60–69 years, 70–79 years, and ≥80 years (10).

Data imputations for missed visits in the CVT and Guanacaste cohort

Oncogenic HPV infections (CVT)

To standardize time among infections, account for delayed visits, and minimize differences in persistence that could arise from differing follow-up schedules in the CVT, we assigned each study visit to a bin, with bin 0 representing 0 to 3 months following first detection of an HPV infection, bin 0.5 representing 3 to 10 months following first detection date, bin 1 representing 10–22 months following first detection, bin 2 representing 22 to 34 months following first detection, and so on, with subsequent bins representing 12 month intervals. For the purposes of calculating HPV clearance and progression rates, we collapsed bins into approximate yearly intervals that coincided with timing of study visits. The "Year 1 " interval included HPV results from months 3 through <22 of the study (bins 0.5 and 1), and subsequent 12-month intervals began at 22 months (22 to <34 months, etc.; bin 2, etc.). "Year 1" was thus 19 months long, although the majority of infections were associated with visits in the first 15 months of the interval (73% of prevalent infection-related visits; 64% of incident infection-related visits). Thus, for calculating an approximate monthly hazard of clearance or progression (assuming a constant hazard within the interval), we assumed the length of the interval was in fact 15 months rather than 19 months. This assumption allowed us to capture the decreasing (increasing) probability of clearance (progression) over time that is consistent with natural history studies (11), while also adjusting for when the majority of visits actually took place. Subsequent interval hazards were assumed to be constant over 12-month periods, in agreement with actual interval lengths. Infections were censored following detection of CIN2+ (n = 149).

Estimates of viral persistence were based on the following data imputations, made at the bin level: 1) we excluded infections with 2 or more consecutive negative PCR results between positive results; 2) all other missing or negative HPV results flanked by positive results for the same type were recategorized as positive; 3) missing results between positive and negative results were resolved using the following algorithm: if there were 1–3 missing bins, the first was imputed as positive while remaining bins were made negative; if there were 4 missing bins, the

first 2 were imputed as positive while remaining bins were made negative; 4) if there was more than one visit within a bin and HPV results were discordant, the overall result was assumed to be positive.

Nononcogenic HPV infections (Guanacaste cohort)

Estimates of viral persistence were based on the following data imputations, consistent with our assumptions for oncogenic types: 1) missing or negative HPV results flanked by positive results for the same type were recategorized as positive; 2) missing results between positive and negative results were resolved using the following algorithm: if there were 1–3 missing bins, the first was imputed as positive while remaining bins were made negative; if there were 4–5 missing bins, the first 2 were imputed as positive while remaining bins were made negative; and so on; 3) if there was more than one visit within a bin and HPV results were discordant, the overall result was assumed to be positive.

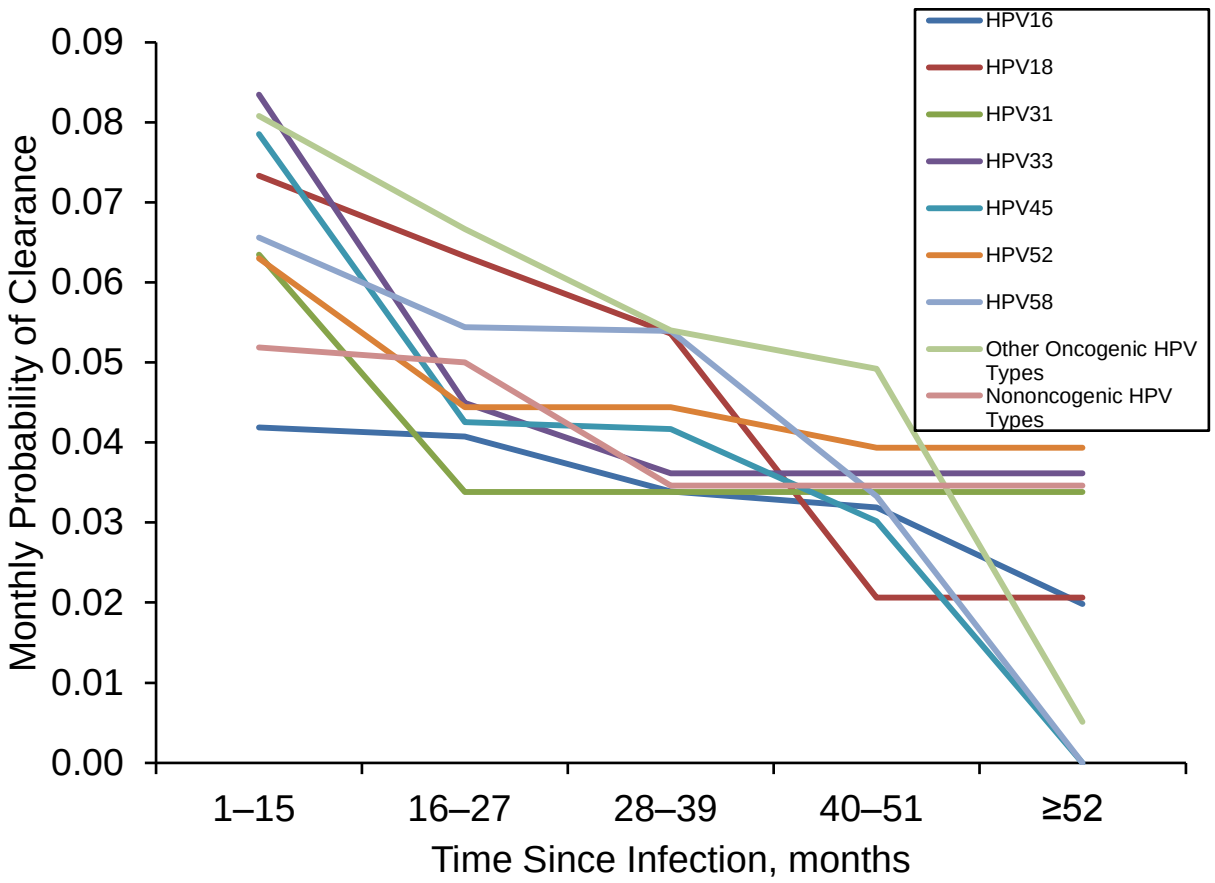
Model calibration process

HPV incidence is known to depend on sexual behavior, which may vary widely across settings and likely contributes to variations in HPV prevalence curves (12). In order to fit HPV prevalence data from New Mexico, we sampled parameter values for type- and age-specific multipliers that were applied to baseline incidence inputs from Colombia (1) to reflect the greater likelihood of HPV infection in New Mexico, particularly in women under 30. Baseline incidence rates and multiplier search ranges are displayed in **Table 2.** With respect to naturally acquired immunity, data on immune response and subsequent protection against future type-specific infection are limited. We assumed that women infected with oncogenic HPV developed at least partial immunity following initial infection and clearance. To explore the uncertainty of this parameter, we sampled parameter values ranging from partial to full immunity. For age- and type-specific HPV incidence, we set plausible search ranges around baseline input values and performed 731,123 model simulations in the absence of any vaccination or screening intervention. For each simulation, we randomly selected one value for each of the uncertain parameters from a uniform distribution over the identified plausible range, creating a unique natural history parameter set. For each parameter set, we scored model outcomes according to their overall fit with calibration targets (**Web Table 2**). We specified likelihood functions for all calibration targets and assumed that prevalence targets (e.g., type-specific HPV prevalence; genotype frequency in CIN2, CIN3, and cancer) followed independent binomial distributions. For each input parameter set, we computed a composite goodness-of-fit score by summing over the individual log likelihood measures of all targets. We selected a sample of 50 good-fitting sets to display calibration results. In forthcoming comparative and cost-effectiveness analyses, a sample of calibrated sets can be used to explore the effect of uncertainty in the natural history parameters as a form of probabilistic sensitivity analysis.

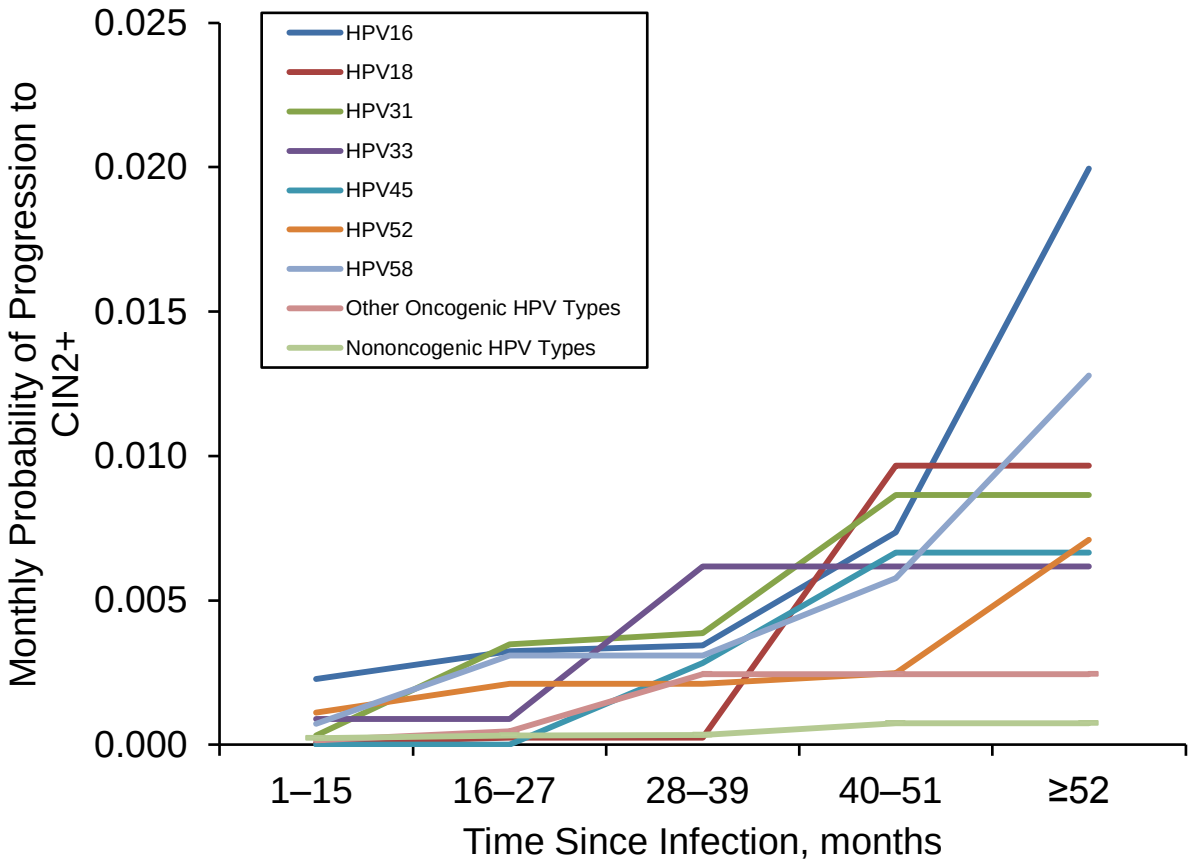
REFERENCES

1. Munoz N, Mendez F, Posso H, et al. Incidence, duration, and determinants of cervical human papillomavirus infection in a cohort of Colombian women with normal cytological results. *J Infect Dis* 2004;190(12):2077–2087.
2. Bratti MC, Rodriguez AC, Schiffman M, et al. Description of a seven-year prospective study of human papillomavirus infection and cervical neoplasia among 10000 women in Guanacaste, Costa Rica. *Rev Panam Salud Publica* 2004;15(2):75–89.
3. Nielsen A, Iftner T, Norgaard M, et al. The importance of low-risk HPV infection for the risk of abnormal cervical cytology/histology in more than 40 000 Danish women. *Sex Transm Infect* 2012;88(8):627–632.
4. Keefe KA, Schell MJ, Brewer C, et al. A randomized, double blind, Phase III trial using oral beta-carotene supplementation for women with high-grade cervical intraepithelial neoplasia. *Cancer Epidemiol Biomarkers Prev* 2001;10(10):1029–1035.
5. Meyskens FL, Jr., Surwit E, Moon TE, et al. Enhancement of regression of cervical intraepithelial neoplasia II (moderate dysplasia) with topically applied all-trans-retinoic acid: a randomized trial. *J Natl Cancer Inst* 1994;86(7):539–543.
6. Moscicki AB, Ma Y, Wibbelsman C, et al. Rate of and risks for regression of cervical intraepithelial neoplasia 2 in adolescents and young women. *Obstet Gynecol* 2010;116(6):1373–1380.
7. Wang SM, Colombara D, Shi JF, et al. Six-year regression and progression of cervical lesions of different human papillomavirus viral loads in varied histological diagnoses. *Int J Gynecol Cancer* 2013;23(4):716–723.
8. Castle PE, Schiffman M, Wheeler CM, et al. Evidence for frequent regression of cervical intraepithelial neoplasia-grade 2. *Obstet Gynecol* 2009;113(1):18–25.
9. McCredie MR, Sharples KJ, Paul C, et al. Natural history of cervical neoplasia and risk of invasive cancer in women with cervical intraepithelial neoplasia 3: a retrospective cohort study. *Lancet Oncol* 2008;9(5):425–434.
10. Surveillance, Epidemiology, and End Results (SEER) Program, *SEER*Stat Database: Incidence – SEER 18 Regs Research Data + Hurricane Katrina Impacted Louisiana Cases Nov 2011 Sub (1973–2009 varying) – Linked to County Attributes – Total U.S., 1969–2010 Counties*. National Cancer Institute, DCCPS, released April 2012, based on the November 2011 submission.
11. Plummer M, Schiffman M, Castle PE, et al. A 2-year prospective study of human papillomavirus persistence among women with a cytological diagnosis of atypical squamous cells of undetermined significance or low-grade squamous intraepithelial lesion. *J Infect Dis* 2007;195(11):1582–1589.
12. Franceschi S, Herrero R, Clifford GM, et al. Variations in the age-specific curves of human papillomavirus prevalence in women worldwide. *Int J Cancer* 2006;119(11):2677–2684.
13. Ries LAG, Melbert D, Krapcho M, et al. eds. *Surveillance, Epidemiology, End Results (SEER) Cancer Statistics Review, 1975–2001.* Bethesda, MD: National Cancer Institute; 2004.
14. McCrory D, Mather D, Bastian L, et al. *Evaluation of Cervical Cytology. (*Evidence Report/Technology Assessment no. 5). Rockville, MD: Agency for Health Care Policy and Research; 1999.
15. Myers ER, McCrory DC, Nanda K, et al. Mathematical model for the natural history of human papillomavirus infection and cervical carcinogenesis. *Am J Epidemiol* 2000;151(12):1158–1171.

16. Wheeler CM, Hunt WC, Cuzick J, et al. A population-based study of human papillomavirus genotype prevalence in the United States: baseline measures prior to mass human papillomavirus vaccination. *Int J Cancer* 2013;132(1):198–207.
17. Wheeler CM, Hunt WC, Joste NE, et al. Human papillomavirus genotype distributions: implications for vaccination and cancer screening in the United States. *J Natl Cancer Inst* 2009;101(7):475–487.
18. Rodriguez AC, Schiffman M, Herrero R, et al. Longitudinal study of human papillomavirus persistence and cervical intraepithelial neoplasia grade 2/3: critical role of duration of infection. *J Natl Cancer Inst* 2010;102(5):315–324.



Web Figure 1. Monthly probability of HPV clearance. Details of derivation are in the Methods section and the Web Appendix. We assumed a constant hazard rate within each interval. For HPV31, 33, 52, and nononcogenic types we carried forward the lowest probability of clearance such that clearance was less likely with infections of increasing duration, in keeping with the literature. As described in the Web Appendix, the probability of clearance for nononcogenic infections was derived from the Guanacaste Cohort Study (18); due to longer study follow-up, we were able to derive clearance probabilities through 75 months; the monthly probability of clearance in this interval was 0.0286. HPV, human papillomavirus.



Web Figure 2. Monthly probability of progression to CIN2+. Details of derivation are in the Methods section and the Web Appendix. We assumed a constant hazard rate within each interval. To avoid intermediate intervals with zero probability of progression (due to small numbers of CIN2+ in the study), we held rates constant from previous intervals. Histology results from the Costa Rica Vaccine Trial were not yet available for CIN2 vs. CIN3, so we assumed that for HPV16 infections, 65% of those progressed to CIN2, while 35% progressed to CIN3; for non-HPV16 oncogenic infections, 80% progressed to CIN2, while 20% progressed to CIN3.

For nononcogenic infections, we derived progression rates as described in the Web Appendix; 90% of those that progressed transitioned to CIN2, while 10% went to CIN3. CIN, cervical intraepithelial neoplasia; HPV, human papillomavirus.

Web Table 1. Transition Probabilities Related to Precancer and Cancer[a]

Variable	Parameter Value
Regression of CIN2[b] (4–8)	
All HPV types	0.0087
Regression of CIN3[b] (4–8)	
All HPV types	0.00435
Progression of CIN2 to invasive cancer[c] (9)	
Years 1 – 5	0.000012
Years 6 – 10	0.000013
Years 11 – 20	0.000238
Years 21 – 29	0.000464
Years 30 – 34	0.000557
Years 35 – 39	0.000603
Years 40 – 44	0.001299
Years 45 – 49	0.001392
Years 50+	0.001485
Progression of CIN3 to invasive cancer[c] (9)	
Years 1 – 5	0.000061
Years 6 – 10	0.000066
Years 11 – 20	0.00119
Years 21 – 29	0.00232
Years 30 – 34	0.00278
Years 35 – 39	0.00302
Years 40 – 44	0.00650
Years 45 – 49	0.00696
Years 50+	0.00742
Progression of invasive cancer stages (13–15)	
Local to regional	0.020
Regional to distant	0.025
Invasive cancer mortality[d] (10)	
Local	
Year 1	0.0016
Years 2–3	0.0014
Years 4–20	0.0009
Regional	
Year 1	0.0095
Years 2–3	0.0078
Years 4–20	0.0036
Distant	
Year 1	0.0293
Years 2–3	0.0195
Years 4–20	0.0076

Probability of symptom detection (15)	
Local	0.0174
Regional	0.0735
Distant	0.1746

[a] Monthly probabilities, unless otherwise noted. CIN2, cervical intraepithelial neoplasia grade 2; CIN3, cervical intraepithelial neoplasia grade 3; HPV, human papillomavirus.

[b] We assume that the monthly CIN3 regression probability is 50% of CIN2 regression, since CIN3 — a fairly reliable morphological diagnosis for which regression data are not readily available — is less likely to regress than CIN2. Of CIN2 and CIN3 lesions that regress, 50% regress to type-specific HPV-infected health states and 50% regress to the Normal health state.

[c] Precancer progression probabilities increase by time (in years) since lesion onset and are constant across oncogenic HPV types. CIN2 progression is set at 20% of CIN3 progression (for oncogenic types only).

[d] Invasive cancer mortality decreases by time since diagnosis. Cancer mortality was also adjusted for age at diagnosis by applying stage-specific multipliers to the baseline probabilities that ranged from 0.30 to 7.39 for local cancer; 0.39 to 1.30 for regional cancer; and 0.002 to 15.16 for distant cancer. For undetected cancers, women faced the year 1 probability of cancer mortality until diagnosis via symptom detection.

Web Table 2. Calibration target data[a]

Prevalence of HPV Infection in Women, by Age (16)	Mean Value, % (95% CI)
HPV 16	
<21 years	9.6 (9.0, 10.3)
21–24 years	8.2 (7.6, 8.7)
25–29 years	5.2 (4.8, 5.6)
30–49 years	2.2 (1.9, 2.5)
≥50 years	1.3 (1.0, 1.6)
HPV 18	
<21 years	3.0 (2.7, 3.4)
21–24 years	3.0 (2.6, 3.3)
25–29 years	1.7 (1.5, 2.0)
30–49 years	0.7 (0.6, 0.9)
≥50 years	0.6 (0.4, 0.8)
HPV 31	
<21 years	4.1 (3.7, 4.6)
21–24 years	4.4 (4.0, 4.8)
25–29 years	2.8 (2.5, 3.1)
30–49 years	1.4 (1.1, 1.6)
≥50 years	0.6 (0.4, 0.8)
HPV 33	
<21 years	1.3 (1.1, 1.6)
21–24 years	1.0 (0.8, 1.2)
25–29 years	0.7 (0.6, 0.9)
30–49 years	0.3 (0.2, 0.4)
≥50 years	0.2 (0.1, 0.3)
HPV 45	
<21 years	2.4 (2.1, 2.8)
21–24 years	2.3 (2.0, 2.6)
25–29 years	1.5 (1.2, 1.7)
30–49 years	0.7 (0.6, 0.9)
≥50 years	0.6 (0.4, 0.8)
HPV 52	
<21 years	4.5 (4.0, 4.9)
21–24 years	4.3 (3.9, 4.8)
25–29 years	3.0 (2.6, 3.3)
30–49 years	1.5 (1.2, 1.7)
≥50 years	0.7 (0.5, 0.9)
HPV 58	
<21 years	3.7 (3.3, 4.1)
21–24 years	3.0 (2.7, 3.4)
25–29 years	1.6 (1.3, 1.8)
30–49 years	0.7 (0.5, 0.9)
≥50 years	0.4 (0.3, 0.6)

HPV genotype frequencies in CIN2 (17)	
16	41.7 (38.9, 44.5)
18	5.2 (4.0, 6.6)
31	15.5 (13.5, 17.6)
33	3.7 (2.7, 4.9)
45	2.9 (2.0, 4.0)
52	8.6 (7.0, 10.3)
58	9.0 (7.4, 10.7)
HPV genotype frequencies in CIN3 (17)	
16	57.6 (55.0, 60.1)
18	4.8 (3.8, 6.0)
31	15.0 (13.2, 16.9)
33	5.8 (4.7, 7.1)
45	2.1 (1.5, 3.0)
52	7.9 (6.6, 9.4)
58	6.1 (5.0, 7.5)
HPV genotype frequencies in SCC (17)	**Mean value, % (95% CI)**
16	58.0 (54.1, 61.8)
18	9.8 (7.7, 12.4)

[a] CI, confidence interval; CIN, cervical intraepithelial neoplasia (grade 2 or 3); HPV, human papillomavirus; SCC, squamous cell carcinoma.