Methods for Risk Estimation, Recommended Management, and Validation

The main data source used to develop risk-based management was an update of the Kaiser Permanente of Northern California (KPNC) study that was previously used to inform the 2012 guidelines.[3,4] The KPNC membership is demographically similar to that of the US census-enumerated population in the Bay Area Metropolitan Statistical Area, except for lacking representation of extremes in income, and is considered a well-screened population with risk of cervical cancer that is lower than the national average.[10] To ensure portability of risk-based management derived from KPNC to different US settings, we validated the risk estimates using 3 cohorts: (1) the Centers for Disease Control and Prevention's (CDC) National Breast and Cervical Cancer Early Detection Program (NBCCEDP),[11] (2) the Onclarity HPV Trial,[12] and (3) the Addressing the Need for Advanced HPV Diagnostic (ATHENA) study.[13] In particular, the CDC's NBCCEDP allowed us to validate risk estimates in a cohort of low-income patients of which 31% reported at study entry that they were never or rarely screened. We had also previously compared KPNC risk estimates to those from the New Mexico HPV Pap Registry.[14] For more details on these cohorts, see the section titled: Validation of Risk-Based Guidelines.

The design of the KPNC study has been extensively described.[15,16] In brief, from January 1, 2003, to December 31, 2017, more than 1.5 million patients aged 25–65 years were screened using combined HPV and cytology cotesting. In addition, more than 200,000 patients aged 21–24 years were primarily screened with cytology. Biopsy outcomes of patients undergoing colposcopy for abnormal test results were complete through the end of 2016 with partial data available through September 31, 2017. The KPNC cohort documented 16,222 CIN 2, 9,712 CIN 3 and adenocarcinoma in situ (AIS), and 796 cancer histologic diagnoses. In comparison with the database available up to 2010 that was used for the 2012 guidelines, the currently available database had more than 500,000 additional screening participants, 7 additional years of follow-up, and approximately 150% more CIN 3+. These additional data allowed us to estimate more precisely the risks for different combinations of clinical scenarios and past/current test results.

Human papillomavirus genotyping conducted on residual aliquots was available for almost 19,000 patients at KPNC who were also at NCI-KPNC Persistence and Progression (PaP) study conducted from 2006 to 2013.[17] Selection occurred through a 2-phase stratified sampling design. In the first phase of sampling, patients at KPNC with a cotest within the PaP enrollment window (2006–2011) were selected for storage of residual HPV testing specimens; the collection included approximately 45,000 HPV-positive patients (representing >3/4 of all HPV positive patients in that period) and a random group of approximately 10,000 HPV-negative patients. Approximately 8% of patients opted out, leaving 44,340 patients enrolled in the study. In the second phase of sampling, residual aliquots of patients in the PaP cohort were selected for genotyping using a complex stratified sampling design based on HPV results and histopathology outcomes (as of 2014) to maximize information yield.

Typing focused primarily on HPV-positive patients, which included a random draw plus all unselected patients who were diagnosed with cancer or AIS, 500 unselected patients with CIN 3, and 500 unselected patients with CIN 2. Retesting by research typing assays of HPV-negative patients was restricted to a random group of 500 plus all unselected patients with rare outcomes suggesting elevated cancer risks (i.e., CIN 2+ histopathology or high-grade cytology). Both the KPNC and NCI-KPNC PaP studies have been reapproved yearly by both KPNC and NCI institutional review board committees.

Screening Tests and Clinical Management in the KPNC

Since 2001, patients at the KPNC have been tested with HPV to triage cytology results of atypical squamous cells of undetermined significance (ASC-US). Beginning in 2003, patients 30 years or older underwent screening with concurrent HPV and cytology cotests. In 2013, cotesting was extended to ages 25–29 years.

The clinical HPV testing was conducted using Hybrid Capture 2 (HC2; Qiagen, Germantown, MD) according to manufacturer's instructions. It reports HPV status as negative versus positive for infection with any of the 13 high-risk HPV types (16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59, and 68) and also inadvertently detects through cross-reaction a percentage of closely related low-risk HPV types (e.g., 53, 66, 67, 70, 82, and 82).[18]

From 2003 to 2009, conventional cytology specimens were manually reviewed after processing by the BD FocalPoint Slide Profiler (BD Diagnostics, Burlington, NC) primary screening and directed quality control system, which automatically scores and sorts slides based on the probability of squamous abnormalities.[19] In 2009, the KPNC transitioned to liquid-based cytology using BD SurePath (BD Diagnostics). Cytology results were reported according to the 2001 Bethesda system[20] as shown in Table 1.

TABLE 1:
Cytology Results According to the Bethesda System

Patients with cytologic abnormalities were referred to colposcopy per Kaiser guidelines, which closely resembled US consensus recommendations.[3,21,22] However, those without abnormalities, specifically patients who tested HPV negative/NILM (cotest negative) were asked to return for testing in 3 years. We note that this differs from the 5-year testing interval that has been recommended nationally since 2012. However, sensitivity analyses restricting to patients who returned from a negative cotest in 5 or more years had risk estimates not much different from those at 3 years. Patients with HPV-negative/ASC-US or HPV-positive/NILM results were monitored annually and referred to colposcopy for either a subsequent abnormal cytology result or HPV positivity (after 2005).

Histopathology results were reported in order of increasing severity as normal, CIN grades 1, 2, and 3, AIS, or cancer. We use the terms <CIN 2 or <CIN 3 to indicate histopathology results less severe than CIN 2 or CIN 3, respectively. Similarly, the terms CIN 2+ or CIN 3+ will refer to histopathology results at least as severe as CIN 2 or CIN 3, respectively.

Three Distinct Clinical Scenarios

Using the KPNC and PaP studies, we created analytic cohorts to address each round of testing in the following clinical scenarios: (1) prior to colposcopic (precolposcopy) referral, (2) after colposcopic findings of <CIN 2 histopathology (postcolposcopy), and (3) after treatment for CIN 2 or CIN 3 histopathology (posttreatment) (see Table 2). The goal was to develop risk-based management for most of the (relatively) common decision points that occur in cervical screening programs including the management of abnormalities.

TABLE 2:
Analysis Cohorts Used to Estimate Risk

Precolposcopy Scenario

We restricted analyses in the precolposcopy scenario to 1,546,462 patients with no known history of CIN 2+ or hysterectomy who were not missing HPV or cytology results at the initial screen (insufficient and noncervical cytology results were also excluded). Risks were estimated for 4 testing rounds. Risk estimated from the initial HPV-based screen was used to inform precolposcopic management for patients whose histories of HPV status were unknown. Beginning with the second testing round, patients in a precolposcopy scenario had an important biomarker for risk stratification—their HPV-based test result from the previous round of testing. We estimated risks using the second round to inform precolposcopic management after immediately prior test results of HPV negative, HPV positive/NILM, cotest negative, HPV negative/ASC-US, or HPV negative/LSIL. We did not provide risk estimates for other prior abnormal test results (e.g., HPV positive/ASC-US, HPV positive/LSIL) that typically lead to colposcopy referral. The third and fourth testing rounds were primarily used to determine whether patients with multiple negative HPV or cotest results after an HPV-positive/NILM, HPV-negative/ASC-US, or HPV-negative/LSIL test result had sufficiently low subsequent risks to permit extension of their testing intervals. Management after more complicated screening histories were not included in these guidelines as we could not precisely estimate risks for some combinations of past/current test results because of the smaller sample sizes.

Postcolposcopy Scenario

Patients identified as high risk in the precolposcopy scenario are recommended to colposcopy or expedited treatment. If the colposcopy does not detect the presence of CIN 2+, then these patients previously defined as high risk are in a postcolposcopy scenario. In the analysis of the postcolposcopy scenario, we restricted to patients with no known history of CIN 2+ or hysterectomy, who were referred to colposcopy for the first time in the KPNC, and whose colposcopy results were normal/CIN 1. We dichotomized the prior level of risk by whether patients were referred to colposcopy after low-grade abnormalities (i.e., HPV positive/NILM, ASC-US, or LSIL) or after a high-grade cytology result (i.e., ASC-H, AGC, or HSIL+). Risk estimates were used to determine management immediately after the <CIN 2 colposcopy and at the first postcolposcopy follow-up surveillance visit. The second and third surveillance rounds were used to determine when patients with multiple negative HPV or cotests results after a <CIN 2 colposcopy could safely extend their testing intervals.

Posttreatment Scenario

If the colposcopy findings are CIN 2+, patients are generally referred to treatment, which at the KPNC meant loop electrosurgical excision procedure (LEEP), also known as large loop excision of the transformation zone (LLETZ). In the posttreatment scenario, we estimated risks after treatment for histopathology findings of CIN 2 and CIN 3. Sample sizes were too small to determine based on risk estimates the appropriate management of rare serious outcomes considered separately, e.g., for AIS. Rare end points required separate consideration and expert opinion. In addition, we did not have data to inform management of patients immediately after the treatment procedure; rather, the risks we estimated pertain to management of patients after they returned to posttreatment surveillance visits. In the posttreatment analysis, we restricted to patients with findings of CIN 2 or CIN 3, who were treated with excision procedures, and who returned for subsequent follow-up visits (first posttreatment surveillance round). Because some laboratories may not distinguish between CIN 2 and CIN 3, treated CIN 3 was ultimately chosen, to emphasize caution, to guide all post-LEEP management guidelines.[23] The second and third follow-up rounds were used to determine when patients with multiple negative HPV or cotests results after treatment for CIN 2 or CIN 3 could safely extend their testing intervals.

Risk Estimation and Definition of Intervals Containing Time of CIN 3+ Onset

For each scenario and combination of past/current test results, we produced a risk profile of CIN 2+, CIN 3+, and cancer risks for yearly time points from the time of the current test (immediate risk) to 5 years after the current test (5-year risk). The CIN 2+ and CIN 3+ risks represent precancers and cancers that are detectable through colposcopy with multiple biopsies (not just targeted to the single worst appearing region) and do not include latent precancers that may be too small to be observed clinically. We focused on estimating risk of CIN 3+ because CIN 3 defines a more likely true precancerous state than CIN 2, whose removal can prevent future cancers.[24] However, we also applied the same methods to estimate risks of CIN 2+

and cancer. In particular, high cancer risks can indicate special situations where more aggressive management may be necessary for safety.

We estimated risks of CIN 3+ using a prevalence-incidence mixture model.[25,26] These models jointly estimate a logistic-regression model for CIN 3+ prevalent at the time of the current test (so called "left censoring" because, on a timeline, we do not know what happened before or to the left of the current test) and a proportional hazards model for incident CIN 3+ detectable at future visits. These models were specifically designed to address data features of cervical cancer screening and management: (1) some of the CIN 3+ detected in screening was already present and undetected for an unknown time in the past (left censoring); (2) the timing and availability of histopathologic outcomes depends on colposcopic referral algorithms and patient adherence; and (3) the actual time of onset of incident CIN 3+ occurs between 2 visits (interval censoring). Extensive simulation studies suggest that prevalence-incidence models are superior to standard statistical methods for risk estimation that do not explicitly account for left censoring and interval censoring.[25,27] In particular, methods that use the time of diagnosis as a proxy for the time of onset, such as standard Kaplan-Meier methods or Cox models, substantially underestimate absolute risks at early time points and overestimate risks at later time points.[25,27] We show this underestimation and overestimation from using standard Kaplan-Meier methods through a simple illustrative example in the Appendix, Part A, https://links.lww.com/LGT/A160. Details on the prevalence-incidence models are given in the Appendix, Part B, https://links.lww.com/LGT/A160.

Prevalence-incidence models require knowledge of each patient's history of test results to determine at which visits they had <CIN 3, CIN 3+, or an unknown <CIN 3/CIN 3+ status. We define prevalent CIN 3+ as CIN 3+ detected after colposcopic referral because of the baseline (current) test results. We define a patient as having <CIN 3 at a visit if the diagnosis was either (1) directly confirmed through histopathology or, in the absence of histopathologic outcomes, (2) if neither HPV nor cytology tests were positive (i.e., HPV negative/NILM or HPV negative/ASC-US). Condition (2) is useful to shorten intervals during which CIN 3+ onset could have plausibly occurred (thus increasing statistical efficiency of model estimates) and cause negligible bias because of the very high probability of an HPV negative with NILM or ASC-US cytologic result indicating concurrently histology <CIN 3. In contrast, patients with positive test results (HPV-positive or ≥LSIL cytology results), but without histopathologic outcomes or a future <CIN 3 to rule out CIN 3+, are considered to have unknown (in statistical terms, "latent") histologic status at that screen.

These considerations allow us to define the shortest time intervals in which onset of detected CIN 3+ is likely to have occurred: (1) prevalent (left-censored at the current test), (2) incident (interval-censored between the last result <CIN 3 and the time of CIN 3+ detection), or (3) unknown prevalent or incident (CIN 3+ detected in follow-up but cannot rule out disease prevalence). Patients who never had CIN 3+

detected were considered right censored at the time of the last <CIN 3 visit. Individuals who were never defined as <CIN 3 or CIN 3+ at any time points did not contribute to risk estimation. Some illustrative examples of how patients' visit histories are used to define intervals are shown in the Appendix, Part C, https://links.lww.com/LGT/A160.

We considered genotyping results as a categorical variable with the following hierarchical levels: type 16 positive, else type 18 positive, and else positive for other high-risk types. An alternative grouping of genotyping that we explored used the following levels: type 16 positive; else type 18 positive; else type 45 positive; else positive for 31, 33, 52, or 58; and else positive for 35, 39, 51, 56, 59, 66, or 68. We calculated survey weights (i.e., as the inverse of sampling fractions) for each patient in the PaP cohort with genotyping results to reconstitute the larger KPNC first and second rounds of testing in the precolposcopy clinical scenario. We then estimated genotype-specific CIN 3+ risks by applying extensions of the prevalence-incidence model for survey data.[26]

We did not use variables outside of current HPV and cytology results, current genotyping results, and history of screening/histopathology results in risk estimation. Other cofactors, including age, race, hormonal contraceptives use, smoking, and income, had minimal or clinically important effects on CIN 3+ risk in the KPNC.

Defining Risk Thresholds for Clinical Management

Representatives from 19 organizations attended the initial guidelines meeting and convened 7 working groups (including treatment, colposcopy, and surveillance working groups) to determine the consensus clinical action thresholds used to define risk-based clinical management. The 6 clinical management options considered include 3 levels of immediate intervention as follows: (a) expedited treatment (i.e., without preceding colposcopy/biopsy) is preferred, (b) either treatment or colposcopy/biopsy is acceptable, (c) colposcopy/biopsy is recommended; 2 levels of shortened testing intervals as followed: (d) retest in 1 year, (e) retest in 3 years; or (f) continue with/return to routine testing in 5 years as per US Preventive Services Task Force[28] and American Cancer Society screening guidelines.[29] The treatment working group defined 2 risk levels in which expedited treatment without colposcopy could be used. The surveillance working group limited the options of shortened testing intervals to the 1- and 3-year testing intervals used in the 2012 guidelines.[3]

The immediate CIN 3+ risk (i.e., probability of having clinically detectable CIN 3+ if referred to colposcopy at the time of the current test) is most relevant in determining which, if any, clinical intervention is currently needed. For patients whose immediate risks fall below the thresholds for clinical intervention, the long-term cumulative risk is most relevant in determining when there is enough CIN 3+ in a population to warrant retesting. Although 1-, 3-, and 5-year risks may be relevant to determining who should return in 1, 3, or 5 years, the lines of most risk curves in a clinical

setting/round did not cross. Therefore, the management decisions would be similar regardless of which time interval was used to compare risks after different test combinations. For simplicity and to maximize the chance that any CIN 3+ cases were found if present, the surveillance working group chose 5-year CIN 3+ risks as the relevant measure for determining when to return a patient for further testing.

Risk profiles (CIN 3+ and cancer risks from the time of the current test to 5 years after the current test) were translated to risk-based management using consensus clinical action thresholds[23] as followed:

- (a) expedited treatment is preferred [treatment]: immediate CIN 3+ risk of at least 60%,
- (b) either treatment or colposcopy/biopsy is acceptable [treatment/colposcopy]: immediate CIN 3+ risk of at least 25% but less than 60%,
- (c) colposcopy/biopsy is recommended [colposcopy]: immediate CIN 3+ risk of at least 4% but less than 25%,
- (d) retest in 1 year: 5-year CIN 3+ risk at least 0.55% but immediate CIN 3+ risk less than 4%,
- (e) retest in 3 years: 5-year CIN 3+ risk of at least 0.15% but less than 0.55%, and
- (f) retest in 5 years: 5-year CIN 3+ risk less than 0.15%.

The risk of cancer was used to help identify special situations associated with increased cancer risks that are not reflected by the CIN 3+ estimates. These were then considered separately based on consensus opinion.

The choice of clinical action thresholds were consensus decisions specific to the US setting. To inform clinical action thresholds for treatment and colposcopy/biopsy, we used the frequency and immediate risk of each cotest in the initial screening setting to project the following measures of benefits, harms, and efficiency per 1 million patients screened: the number of patients sent to colposcopy/treatment, the number of patients with CIN 3+ timely detected/treated (number of test-positive for the referral criterion), the number of patients with CIN 3+ detected/treated (number of true positives), the number of patients with <CIN 2 undergoing unnecessary procedures (number of false positives), the number of patients with CIN 3+/cancers for whom detection/treatment would be delayed (number of false negatives), and the number of colposcopies/treatments that would be needed to detect/treat one CIN 3+ (efficiency).

For those below the immediate risk threshold for colposcopy/biopsy or treatment, the surveillance working group determined the clinical action thresholds for retesting in 1, 3, or 5 years by using the 5-year CIN 3+ risks after NILM and HPV-negative results as a benchmark. To inform the clinical action thresholds, we produced a range of projected risks of NILM, because the risk of NILM depends on the risk of HPV positive/NILM, the risk of HPV negative/NILM, and the fraction of each in a population.

Uncertainty Estimates for the Recommended Risk-Based Management

Confidence intervals for each of the risk estimates were calculated using a normal approximation (for large sample sizes) or exact methods for the binomial distribution (for prevalent risks and small sample sizes). However, a more relevant measure of uncertainty for risk-based management is the probability that given a random sample of similar size and censoring as the observed sample, the estimated risks for a particular scenario and combination of past/current test results in the KPNC would fall between the clinical action thresholds for that management. This measure, which we call the recommendation confidence score, includes the statistical precision and how close the risk estimates fall to the clinical action thresholds. The calculation also accounts for both immediate and 5-year risks being used to determine the appropriate management. Note that a relatively precise risk estimate could still lead to incorrect management by chance; all observational data are sampled and have random error. If the estimate of a risk happens to fall at the extreme of its sampling distribution, it could cross a clinical action threshold by "bad luck."

To illustrate that precision in risk estimation does not always equal precision in management, we present 2 scenarios. (1) A particular combination of past/current test results is relatively rare, resulting in a risk estimate that has a large confidence intervals. However, both the lower and upper bounds of the confidence interval fall within the same clinical action thresholds. Despite the uncertainty in the risk estimate, the probability of having the same risk-based management from another random sample of the same size is very high. (2) A particular combination of past/current test results is very common, resulting in a very precise risk estimate. However, the estimated risk is very close to a clinical action threshold so that one of the bounds of the confidence interval crosses the threshold. In this case, despite the precisely estimated risks, the probability of having the same risk-based management from another random sample of the same size might not be particularly high. Mathematical details on estimating the recommendation confidence scores are presented in the Appendix, Part D, https://links.lww.com/LGT/A160.

Validation of Risk-Based Guidelines

A previous validation effort in the New Mexico HPV Pap Registry found risks there to be largely similar to that of KPNC.[14] However, we further validated risk-based management for the initial screen using 3 cohorts: (1) the CDC NBCCEDP, (2) the Onclarity Human Papillomavirus Trial, and (3) the ATHENA study.

The CDC's NBCCEDP was established to provide low-income patients (defined as family income at or less than 250% of the federal poverty level) access to screening and diagnostic examinations for breast and cervical cancer.[11] The NBCCEDP analysis cohort consisted of 363,544 patients 30 years and older with nonmissing HPV and cytology test results and no history of CIN 2+ histology. This population includes both patients undergoing screening with cotesting and patients undergoing primary

cytology screening who had an HPV test to triage an ASC-US result. Patients also reported whether they had previously been screened. Using estimated risks from the CDC NBCCEDP,[30] we conducted separate validation studies for patients who reported having a prior cytology test in the past 5 years (well-screened) versus those who reported that their prior cytology test occurred more than 5 years ago, that they were never screened, or that they did not know when they were last screened (screened rarely/never/unknown). Because most patients in the cohort had only a single screen, only the immediate risk of positive test results (i.e., HPV positive/NILM, HPV positive/ASC-US, or LSIL+) was available.

We also conducted validation studies using estimated immediate and 3-year risks of test results in the Onclarity trial[31] and in the ATHENA study.[32] The Onclarity trial was designed to obtain FDA-approval in the US for use of high-risk HPV pooled detection, with individual identification of types 16, 18, and 45. The analysis cohort consisted of 29,513 patients 25 years and older who were tested with cytology and Onclarity at baseline.[12] All patients with abnormal cytology or high-risk HPV-positive results were referred to colposcopy at baseline and entered into the longitudinal phase of the study. In addition, random selections of patients with negative cotests were referred to colposcopy at baseline (approximately 5%) or entered into the longitudinal phase of the study (approximately 10%). Biopsies and endocervical curettage (ECC) samples underwent a blinded review by 2 pathologists with third pathologist review used for adjudication. The ongoing longitudinal phase consists of 3 additional rounds of annual testing with cytology and Onclarity. In the longitudinal phase of the study, patients with ASC-US or greater cytology results are referred to colposcopy and censored from the study if CIN 2+ was detected. In addition, all patients are also referred to colposcopy at the final study visit.

The ATHENA study was a multicenter trial designed to evaluate primary screening with the cobas test, which tests for high-risk HPV and individually identifies genotypes 16 and 18.[13] For this effort, the analysis cohort consisted of 40,871 patients 25 years and older who were tested with cytology, the cobas test, and earlier generation Roche HPV tests at baseline. Patients with abnormal cytology or who were positive by earlier generation Roche HPV tests were referred to colposcopy, along with a random sample of cotest-negative patients. Biopsies and ECC samples were reviewed by a panel of 3 pathologists who were masked to the screening test results. Patients who underwent colposcopy and did not have CIN 2+ were eligible for the longitudinal phase of the study, which consisted of annual cotesting with cytology and the cobas test. In the longitudinal phase of the study, patients with ASC-US or greater are referred to colposcopy and censored from the study if CIN 2+ was detected. At the final visit, patients were invited to have an exit colposcopy.

We assessed portability of both the KPNC-derived estimated risks and the recommended management. To assess portability of risk estimates, we calculated the ratio of observed to expected CIN 3+ risk (O/E) for each test result, where the KPNC-estimated risk is the expected and the study-specific estimated risk is the

observed. Using estimated risks instead of raw observed numbers account for the various features of cervical cancer screening data (i.e., left, interval, and right censoring). Variance estimates were derived using the Delta method, and confidence intervals were estimated using asymptotic normality of O/E on the logarithm scale. Portability of the recommended risk-based management was assessed by noting agreements/disagreements and in the case of disagreements, reporting a p value for the null hypothesis of the estimated risk falling into the recommended management. Further details are given in the Appendix, Part E, https://links.lww.com/LGT/A160.

RESULTS

The risk tables and the accompanying risk-based management for each clinical scenario are presented in Egemen et al.[1] Risk tables and risk-based management for the use of HPV genotyping are presented in DeMarco et al.[2] In this section, we focus on the tables used to inform the consensus clinical action thresholds and the validation results examining portability to other settings of the risk estimates and risk-based recommendations.

Informing Clinical Action Thresholds

Table 3 gives the projected benefit, harms, and efficiency from screening 1 million patients with different treatment and colposcopy clinical action thresholds. For example, the risk threshold for colposcopy/biopsy was set at 4%. Under this threshold (ignoring special situations), patients with HPV-positive/ASC-US+ or HPV-negative/HSIL+ results have immediate CIN 3+ risks of at least 4% and would receive a recommended management of colposcopy/biopsy (or expedited treatment if risks also exceed treatment thresholds). Patients with HPV-negative (except HSIL+) or HPV-positive/NILM results have immediate CIN 3+ risks less than 4% and would receive a recommended management of repeat testing at a later date. For every 1 million patients who have an initial screen, 40,784 patients would receive a colposcopy/biopsy recommendation. If all 40,784 patients undergo colposcopy/biopsy, 3,614 would have CIN 3+ detected (for an efficiency of 11 colposcopies per CIN 3+ detected). Among the approximately 959,216 patients who were not referred to colposcopy/biopsy, 958 would have CIN 3+ (90 with cancers) that would have detection delayed, with most likely detected after a 1-year retest (only HPV negative/ASC-US and HPV/negative/NILM would have testing intervals greater than 1 year).

TABLE 3:
Projected Outcomes From Different Treatment or Colposcopy Referral Thresholds Per 1 Million Patients Screened

To inform the 0.55% 5-year CIN 3+ risk threshold for a 3-year return, the surveillance group used the 5-year CIN 3+ risk of NILM as a benchmark (because 2012 guidelines

recommended a 3-year return for NILM). We projected the 5-year CIN 3+ risks after an NILM result for different percentages of HPV positivity. If the proportion of HPV positive among NILM is 4.5% (KPNC is 4.4%), the risk of NILM would be 0.33%. If the proportion of HPV positive among NILM is 7% (CDC is 6.9% and 7.1% for well screened and screened rarely/never/unknown populations, respectively), the risk of NILM would be 0.45%. The 0.55% 5-year CIN 3+ risk corresponds to an HPV positivity of 9% among NILM cytology results and is close to the 0.52% risk estimated in the New Mexico Pap Registry. Similarly, to inform the 0.15% clinical action thresholds for a 5-year return, the surveillance group used the 5-year CIN 3+ risk of HPV negative at the KPNC as a benchmark, which is 0.14% with a 95% upper confidence limit of 0.15%.

Validation Results—CDC NBCCEDP

Table 4 compares the well-screened and rarely/never/unknown screening populations of the CDC NBCCEDP to that of the KPNC population at the initial screen. Even after accounting for extra ASC-US in the CDC NBCCEDP well-screened population (from including HPV triage of ASC-US results), the CDC NBCCEDP well-screened population was substantially more likely to be HPV positive (10% of all patients after excluding ASC-US) and cytology positive (4.5% of all patients after excluding ASC-US) than KPNC (6% HPV positive and 2.6% cytology positive after excluding ASC-US). As a result, the overall immediate CIN 3+ risk in the CDC NBCCEDP well-screened population was greater than that of KPNC (0.78% vs 0.46%, p < .001). However, once stratified by HPV and cytology results, risks for the well-screened population of the CDC NBCCEDP were largely similar to that of KPNC (O/Es 0.91–1.44) and fell within the bounds of the recommended management. The lone exception was HPV negative/AGC (O/E = 2.73; 95% CI = 1.74–4.28) for which the risks implied a 1-year return (KPNC risks also imply a 1-year return but colposcopy/biopsy was recommended because of elevated cancer risks).

TABLE 4:
Portability of KPNC Risks and Risk-Based Management to the CDC Cohort

The overall immediate CIN 3+ risk among the CDC NBCCEDP never/rarely/unknown screened population was 1.23%. In contrast to the well-screened population, risk stratified by HPV and cytology results from the never/rarely/unknown screened population of the CDC NBCCEDP remained significantly higher than in the KPNC cohort (O/Es 1.09–6.36). However, the elevated risks of the CDC NBCCEDP never/rarely/unknown screened population still largely fell within the recommended management. The lone exception was patients who tested HPV positive/HSIL+ for which the estimated immediate CIN 3+ risk of 64.1% exceeds the risk range of 25% to 60% for recommended management based on KPNC data of treatment/colposcopy (p = .016). Because of this finding and the increased risk of not returning for colposcopy, the guidelines recommend expedited

treatment for patients testing HPV positive/HSIL+ who were rarely or never screened.[23]

Validation Results—Onclarity Trial

Compared with the KPNC cohort, the Onclarity trial had greater overall immediate CIN 3+ risk (0.72% vs 0.46%, p < .001) and greater proportions of patients who tested HPV positive/NILM (8.7% vs 4.1%, p < .001) and HPV negative/ASC-US (3.6% vs 1.6%, p < .001) at the initial screen. Because the Onclarity trial performed colposcopy on all abnormal tests at enrollment, the trial had very few CIN 3+ that could not be classified as either prevalent or incident. Despite these differences, risks and risk-based management in the KPNC and the Onclarity trial largely agreed (see Table 5), with only HPV negative/ASC-US appearing to confer risks in the Onclarity trial that might lead to a different recommended management if the trial was observed for 5 years (3-year CIN 3+ risk: 0.54%; 3-year O/E: 2.25 [95% CI = 0.88–5.78]). When HPV-positive results are further stratified by genotyping groups and cytology (not shown), the estimated O/Es ranged from 0.29 to 2.29 with large confidence intervals, but the corresponding risk-based management implied by the estimated risks still largely agreed.

TABLE 5:
Portability of KPNC Risks and Risk-Based Management to BD Onclarity Cohort

Validation Results—ATHENA Study

Compared with the KPNC cohort, the ATHENA study had greater overall immediate CIN 3+ risks (0.79% vs 0.46%, p < .001) and greater proportions of patients who tested HPV positive/NILM (7.7% vs 4.1%, p < .001) at the initial screen. When stratified by HPV and cytology results (see Table 6), risks for HPV-positive results in ATHENA remained significantly higher than that of KPNC (O/Es from 1.3–2.2. Despite the increased risks, the implied risk-based management in the KPNC and ATHENA largely agreed, with the exception of HPV positive/NILM (immediate CIN 3+ risk: 4.6%; O/E: 2.2 [95% CI = 1.8–2.6]). When HPV-positive/NILM results were further stratified by genotyping results of 16, else 18, else other high-risk types, the implied management using risks from ATHENA agreed with that of KPNC (e.g., colposcopy referral for types 16 and 18 and a 1-year return for non-16/18 HPV-positive NILM results).

TABLE 6:
Portability of KPNC Risks and Risk-Based Management to the ATHENA Trial

We also compared HPV-negative results and found CIN 3+ risks in ATHENA to be higher than in KPNC (not shown). However, a previous post-hoc analysis

re-examined the HPV-negative CIN 3+ cases with Linear Array and Amplicor to detect false-negative HPV and with immunostaining with p16 to identify false-positive CIN 3+.[33] Their analysis did not identify any true CIN 3+ not attributable to HPV.