Mathematical Model for the Natural History of Human Papillomavirus Infection and Cervical Carcinogenesis

Evan R. Myers,1,2 Douglas C. McCrory,2-4 Kavita Nanda,1,2 Lori Bastian,2-4 and David B. Matchar2-4

The authors constructed a Markov model as part of a systematic review of cervical cytology conducted at the Duke University Evidence-based Practice Center (Durham, North Carolina) between October 1997 and September 1998. The model incorporated states for human papillomavirus infection (HPV), low- and high-grade squamous intraepithelial lesions, and cervical cancer stages I–IV to simulate the natural history of HPV infection in a cohort of women from ages 15 to 85 years. The age-specific incidence rate of HPV, and regression and progression rates of HPV and squamous intraepithelial lesions, were obtained from the literature. The effects of varying natural history parameters on cervical cancer incidence were evaluated by using sensitivity analysis. The base-case model resulted in a lifetime cervical cancer risk of 3.67% and a lifetime cervical cancer mortality risk of 1.26%, with a peak incidence of 81/100,000 at age 50 years. Age-specific distributions of precursors were similar to reported data. Lifetime risk of cancer was most sensitive to the incidence of HPV and the probability of rapid HPV progression to high-grade lesions (two- to threefold variations in risk). The model approximates the age-specific incidence of cervical cancer and provides a tool for evaluating the natural history of HPV infection and cervical cancer carcinogenesis as well as the effectiveness and cost-effectiveness of primary and secondary prevention strategies. Am J Epidemiol 2000;151:1158-71.

cervical intraepithelial neoplasia; cervix neoplasms; models, theoretical; papillomavirus, human

Carcinoma of the cervix is one of the most common malignancies of women in many parts of the world. Secondary prevention by using cervical smears to detect preinvasive and early invasive disease has led to significant reductions in both incidence and mortality in many countries (1). In the United States, both incidence and mortality have declined steadily; Surveillance, Epidemiology, and End Results (SEER) registry data show a 43 percent decrease in incidence and a 45.9 percent decrease in mortality from 1973 to 1995 (2). Such reductions have not been observed in countries in which cytologic screening is not widely available (3).

Although there have been no randomized trials of the effectiveness of cervical cytologic screening in preventing mortality from cervical cancer, there is wide consensus based on results of historical series and case-control studies that screening does result in significant decreases in incidence and mortality (4). However, considerable controversy remains about the optimal frequency for such testing, the potential role of adjunctive technologies for improving the sensitivity of screening, and the appropriate management of low-grade lesions that may be preinvasive. In addition, although there is general agreement on the broad outlines of the natural history of cervical cancer (5, 6), uncertainty exists about the specifics of many of the elements that contribute to natural history. These issues have been approached primarily with modeling (7–14).

Comprehensive simulation models enable integration of evidence from a wide variety of sources to evaluate natural history as well as prevention and treatment strategies, and they have been used to evaluate strategies for preventing stroke (15). As part of a comprehensive review of the effectiveness of both conventional cervical cytologic screening and new adjunctive technologies (16) performed at the Duke University Center for Clinical Health Policy Research (Durham, North Carolina), we took a similar approach, using...
simulation modeling to integrate available evidence on strategies for preventing cervical cancer. This review, funded by the US Agency for Health Care Policy and Research as part of its Evidence-based Practice Centers program (a group of 12 sites in the United States and Canada that conduct systematic reviews and data synthesis on topics designated by the agency), included a meta-analysis of studies on the sensitivity and specificity of the conventional cervical smear, an analysis of costs associated with screening and treatment for cervical cancer and preinvasive lesions in the United States, and a cost-effectiveness analysis of screening strategies to prevent cervical cancer. In this paper, we present a model of the natural history of cervical cancer that builds on both the work of previous authors and recent epidemiologic evidence to predict the age-specific incidence of cervical cancer in unscreened populations and can be used to assess the potential impact of preventive strategies.

MATERIALS AND METHODS

We constructed a 19-state Markov model (17) by using DATA 3.0 software (TreeAge Software, Williamstown, Massachusetts). In a Markov model, the conditional distribution of the outcomes given an exposure status depends on prior outcome observations only. Our model follows a simulated cohort of women from ages 15 through 85 years. The probability of moving from one state to another (e.g., from normal to human papillomavirus (HPV) infected) during a given Markov cycle (e.g., a 1-year time period) is determined by the modeler; typically, these probabilities are state and often cycle specific. States and allowed transitions are shown in table 1 and figure 1. The model is described in detail in the final evidence report prepared by Duke University (16), and copies of the software program are available from the authors on request. Because the model generates probabilities, the cohort can be any size; for a person, the model generates lifetime probabilities of being in a given health state. Acquisition of HPV is based on age-specific incidence rates. Regression and progression between the various states is based on published data. Because the topic of our review was suggested to the Agency for Health Care Policy and Research by the American College of Obstetricians and Gynecologists (Washington, DC), we used US data as much as possible for our probability and prevalence estimates.

Assumptions of the model

To produce a model with a manageable number of possible outcomes, some simplifying assumptions were necessary. The following list outlines the main underlying assumptions of the model and our rationale for making them.

1. The model assumes that all cases of cervical cancer begin with HPV infection. We incorporated HPV status into the model for two reasons. First, although a small percentage of cervical cancers do not contain detectable HPV DNA, even with sensitive assays there is consensus that HPV infection is the causative agent for the vast majority of cervical cancers (5, 6, 18, 19). Second, certain HPV types clearly are more likely to progress to cancer than others, and identification of these types in cervical cells may help determine optimal diagnostic and treatment strategies for patients with abnormal cervical smears (20). Estimates of the age-specific incidence of HPV infection were derived from published cohort studies. For our model, the HPV infected state is defined as the presence of detectable HPV DNA with normal cervical cytology. Under the Bethesda System (21), cytologic changes consistent with HPV infection that do not meet the criteria for a diagnosis of cervical intraepithelial neoplasia (CIN) are classified as low-grade squamous intraepithelial lesions (SIL).

2. Studies that used older classification systems, primarily the one for CIN, were converted to the Bethesda System (21) as follows: cytologic evidence of HPV infection and CIN I = low-grade SIL; CIN II, CIN III, and carcinoma in situ = high-grade SIL.

3. Regression of HPV is defined as the inability to detect a previously detected HPV viral type in the same patient by using the same diagnostic techniques. Published regression rates, usually expressed as percentage of infections per time period, and progression rates to low-grade SIL and high-grade SIL were converted to transition probabilities (22).

4. Similarly, regression and progression probabilities for low-grade SIL and high-grade SIL were derived from the literature. Low-grade SIL lesions were allowed to regress to both latent HPV infection and the Well state, and high-grade SIL lesions were allowed to regress to low-grade SIL, HPV, and Well.

5. Base-case estimates for incidence, regression, and progression rates were chosen on the basis of two criteria. First, parameters were adjusted to result in predicted age-specific prevalence rates for HPV, low-grade SIL, and high-grade SIL and age-specific incidence of cervical cancer that were within the range reported in cross-sectional data. Second, because we planned to use the
TABLE 1. States and possible transitions between states: Markov model of human papillomavirus (HPV) infection and cervical carcinogenesis

<table>
<thead>
<tr>
<th>State</th>
<th>Description</th>
<th>Possible transitions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Well</td>
<td>Normal; no HPV infection or cervical dysplasia</td>
<td>Well, HPV, dead from other cause</td>
</tr>
<tr>
<td>HPV</td>
<td>HPV infection, no cytologic abnormality</td>
<td>Well, HPV, low-grade SIL,* high-grade SIL, dead from other cause</td>
</tr>
<tr>
<td>Low-grade SIL</td>
<td>Low-grade SIL (cervical intraepithelial lesion 1)</td>
<td>Well, HPV, low-grade SIL, high-grade SIL, dead from other cause</td>
</tr>
<tr>
<td>High-grade SIL</td>
<td>High-grade SIL (cervical intraepithelial lesion 2–3, including carcinoma in situ)</td>
<td>Well, HPV, low-grade SIL, high-grade SIL, unknown stage I cervical cancer, dead from other cause</td>
</tr>
<tr>
<td>Unknown stage I cervical cancer</td>
<td>Cancer confined to cervix</td>
<td>Unknown stage I cervical cancer, detected stage I cervical cancer, unknown stage II cervical cancer, dead from other cause</td>
</tr>
<tr>
<td>Unknown stage II cervical cancer</td>
<td>Cancer involving upper two-thirds of vagina or parametral tissues but not to pelvic sidewall</td>
<td>Unknown stage II cervical cancer, detected stage II cervical cancer, unknown stage III cervical cancer, dead from other cause</td>
</tr>
<tr>
<td>Unknown stage III cervical cancer</td>
<td>Cancer involving lower one-third of vagina or parametral tissues to pelvic sidewall</td>
<td>Unknown stage III cervical cancer, detected stage III cervical cancer, unknown stage IV cervical cancer, dead from other cause</td>
</tr>
<tr>
<td>Unknown stage IV cervical cancer</td>
<td>Cancer spread outside of the pelvis</td>
<td>Unknown stage IV cervical cancer, detected stage IV cervical cancer, dead from other cause</td>
</tr>
<tr>
<td>Detected stage I–IV cervical cancer (one state per stage of cancer)</td>
<td>Diagnosed cancer, years 1–5 after diagnosis and treatment</td>
<td>Detected Stage I–IV cervical cancer, dead from cervical cancer, cancer survivor, dead from other cause</td>
</tr>
<tr>
<td>Cancer survivor (one state per stage of cancer)</td>
<td>Alive 5 years after detection of cancer</td>
<td>Cancer survivor, dead from other cause</td>
</tr>
<tr>
<td>Hysterectomy</td>
<td>Hysterectomy with removal of the cervix for indications other than SIL or cervical cancer</td>
<td>Dead from other cause</td>
</tr>
<tr>
<td>Dead from cervical cancer</td>
<td>Death due to cervical cancer or complications of therapy</td>
<td>Absorbing state†</td>
</tr>
<tr>
<td>Dead from other cause</td>
<td>Death from cause other than cervical cancer</td>
<td>Absorbing state</td>
</tr>
</tbody>
</table>

* SIL, squamous intraepithelial lesion.
† Subjects remain in this state for the remainder of the simulation.

model as the basis of a cost-effectiveness analysis of screening strategies (16), we chose estimates that resulted in cervical cancer incidences that would bias the cost-effectiveness model in favor of improving screening sensitivity. Thus, our base-case estimates result in peak cervical cancer incidence biased toward higher values at earlier ages.

6. We include a hysterectomy state, since removal of the organ at risk clearly affects calculation of cervical cancer incidence (23, 24). However, we did not correct for hysterectomy in our natural history model, since population-based registries do not make a similar correction. We did test the impact of hysterectomy on our estimates.

Model parameters

Incidence of HPV infection. The natural history of HPV infection is complex, and clearance and persistence of viral DNA, along with progression to SIL, vary depending on the viral type, patient characteris-
FIGURE 1. Depiction of the Markov model for the natural history of human papillomavirus (HPV) infection and cervical carcinogenesis. Boxes represent health states; arrows allowed transitions between states. SIL, squamous intraepithelial lesion.

am J Epidemiol Vol. 151, No. 12, 2000
**TABLE 2. Transition probabilities and incidence rates\* of preinvasive human papillomavirus (HPV) disease: Markov model**

<table>
<thead>
<tr>
<th>Parameter (reference no.)</th>
<th>Base case</th>
<th>Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prevalence of HPV infection, age 15 years</td>
<td>0.10</td>
<td>0–0.25</td>
</tr>
<tr>
<td>Prevalence of low-grade SIL,† age 15 years</td>
<td>0.01</td>
<td>0–0.1</td>
</tr>
<tr>
<td>Age (years)-specific incidence of HPV infection (5, 6, 27, 28)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>0.1</td>
<td>0.5–2 x base estimate</td>
</tr>
<tr>
<td>16</td>
<td>0.1</td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>0.12</td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>0.15</td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>0.17</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>0.15</td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>0.12</td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>0.10</td>
<td></td>
</tr>
<tr>
<td>23</td>
<td>0.10</td>
<td></td>
</tr>
<tr>
<td>24–29</td>
<td>0.05</td>
<td></td>
</tr>
<tr>
<td>30–49</td>
<td>0.01</td>
<td></td>
</tr>
<tr>
<td>≥50</td>
<td>0.005</td>
<td></td>
</tr>
<tr>
<td>Age (years)-specific regression rate, HPV infection (HPV to Well) (26, 28, 31)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>15–24</td>
<td>0.7/18 months</td>
<td>0.6–0.9/18 months</td>
</tr>
<tr>
<td>25–29</td>
<td>0.5/18 months</td>
<td>0.45–0.6/18 months</td>
</tr>
<tr>
<td>≥30</td>
<td>0.15/18 months</td>
<td>0.1–0.2/18 months</td>
</tr>
<tr>
<td>Progression rate (HPV to low-grade SIL) (26, 28, 31)</td>
<td>0.2/36 months</td>
<td>0.15–0.3/36 months</td>
</tr>
<tr>
<td>Proportion of infections progressing directly to high-grade SIL (26, 28, 31)</td>
<td>0.1</td>
<td>0.05–0.5</td>
</tr>
<tr>
<td>Regression rate (age (years)) (low-grade SIL to HPV or Well) (27, 32–34)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>15–34</td>
<td>0.65/72 months</td>
<td>0.6–0.8/72 months</td>
</tr>
<tr>
<td>≥35</td>
<td>0.4/72 months</td>
<td>0.3–0.6/72 months</td>
</tr>
<tr>
<td>Proportion of low-grade SIL reverting to Well (27, 32–34)</td>
<td>0.9</td>
<td>0.5–1.0</td>
</tr>
<tr>
<td>Progression rate (age (years)) (low-grade SIL to high-grade SIL) (27, 32–34)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>15–34</td>
<td>0.1/72 months</td>
<td>0.1–0.3/72 months</td>
</tr>
<tr>
<td>≥35</td>
<td>0.35/72 months</td>
<td>0.3–0.572 months</td>
</tr>
<tr>
<td>Regression rate (high-grade SIL to low-grade SIL or Well) (27, 32–34)</td>
<td>0.35/72 months</td>
<td>0.3–0.5/72 months</td>
</tr>
<tr>
<td>Proportion of high-grade SIL reverting to Well (27, 32–34)</td>
<td>0.5</td>
<td>0–0.5</td>
</tr>
<tr>
<td>Progression rate (high-grade SIL to stage I cancer (8, 9, 35–38)</td>
<td>0.4/120 months</td>
<td>0.3–0.5/72 months</td>
</tr>
</tbody>
</table>

* Rates are converted to probabilities in the model.
† SIL, squamous intraepithelial lesion.

2-year cumulative incidence of high-grade SIL of 28 percent in women with HPV DNA; only 36 percent of these women had had a prior low-grade SIL smear. Consistent with our other estimates, our base-case estimates (table 2) were derived from the higher end of the reported ranges. Cases that progress directly to high-grade SIL are similar to the "rapidly progressive" cases used in other models (9). Again, we varied age-specific regression rates to produce an age-specific cancer incidence curve similar to that seen in unscreened populations.

Low-grade and high-grade squamous intraepithelial neoplasia. Determining transition probabilities from the literature that accurately reflect natural history is as difficult for SIL as it is for HPV infection. The difficulties in converting rates collected over varying, often-unspecified times and in heterogeneous populations are further magnified by differences in terminology. For example, many studies report transitions from HPV-associated cytologic changes to CIN I to CIN II to CIN III to carcinoma in situ, which may be difficult to translate into the Bethesda System (21) terminology.
of low-grade SIL and high-grade SIL. Our model assumes an age dependence in regression and progression rates (28, 32, 33). For our baseline case, we use the estimates of Syrjanen et al. (34), the largest cohort that reports results by using the Bethesda System. The length of time for high-grade SIL progression is more difficult to estimate from these data than are the other parameters; in this cohort, all patients who progressed to carcinoma in situ, according to an older classification system, were treated. Prior models have estimated the duration from severe dysplasia/carcinoma in situ to invasive cancer as 10–15 years. We used 12 years for the base case (table 2), since this interval resulted in an age-specific incidence of cervical cancer most consistent with observed data.

Natural history of invasive cancer. Almost no data exist for estimating the rates of progression from stage I through stage IV cervical carcinoma. There is also no way to determine the likelihood of developing symptoms. Since distribution of cases by stage in an unscreened population should be a function of both the progression rate and the likelihood of presentation with symptoms (since incident cases would be detected only upon presentation with symptoms), we adopted the approach taken by others (8, 9). We adjusted these estimates by varying the progression rates and the probability of presentation with symptoms across previously reported ranges so the proportion of cases represented by each stage was similar to that for cervical cancer patients who have never been screened (35–38). Our estimates are described in table 3.

Stage-specific survival. Survival probabilities at 1, 2, 3, 4, and 5 years postdiagnosis for each stage were obtained from the Patterns of Care Evaluation project of the American College of Surgeons (39, 40) (and A. Fremgen, American College of Surgeons, “personal communication,” 1997) (table 3). These values were chosen because they represent data from a wide range of facilities that treat women with cervical cancer and should be relatively representative of the range of current US practice. Five-year survival rates based on these data were as follows: stage I, 86.0 percent; stage II, 62.5 percent; stage III, 37.9 percent; and stage IV, 11.3 percent.

We assumed no cancer-related mortality after 5 years. Although the Patterns of Care Evaluation data show some deaths after 5 years for all stages, they are relatively rare compared with the first 5 years. Other models also have used 5-year survival. These data are disease specific; therefore, patients are also at risk for other causes of death during the 5-year postdiagnosis period.

Mortality from other causes. Mortality from causes other than cervical cancer was estimated by subtracting age-specific cervical cancer mortality rates from age-specific all-cause mortality rates by using US life tables from 1995 (41).

Hysterectomy for benign disease. We used age-specific hysterectomy rates obtained from the National Hospital Discharge Survey (42) and Maryland discharge data (43) to estimate age-specific hysterectomy rates.
rates. For two reasons, we did not correct these rates for hysterectomies performed because of cervical cancer. First, because the majority of surgically treated cervical cancer cases are radical hysterectomies, which have a separate International Classification of Diseases, Ninth Revision code, most hysterectomies for cervical cancer are not included in these data sets. Second, the proportion of nonradical hysterectomies performed for preinvasive diseases is relatively small: less than 2 percent of all cases performed over a 6-year period in North Carolina (E. R. Myers, Duke University Medical Center, unpublished data).

Sensitivity and specificity of cervical smears. We performed a meta-analysis of studies of conventional cervical smears, using colposcopy and histology as the reference standard (16). When we used a cytologic threshold of Atypical Squamous Cells of Uncertain Significance or higher and a histologic threshold of low-grade SIL or higher, we found a sensitivity of 51 percent and a specificity of 97 percent. These values were similar to those found in a previously published meta-analysis (44). We used these values to test the impact of screening at 1-, 2-, 3-, and 5-year intervals on the age-specific incidence of cervical cancer.

RESULTS

Age-specific prevalence of HPV and SIL

The age-specific prevalence of HPV infection in women whose cytology is normal, predicted by the model in which base-case estimates are used, is shown in figure 2, which also illustrates the predicted age-specific prevalence of low-grade and high-grade SIL lesions. The model predicts peak prevalences of HPV of 24.7 percent at age 21 years, low-grade SIL of 8.3 percent at age 28 years, and high-grade SIL of 2.6 percent at age 42 years.

Age-specific incidence of cervical cancer

Figure 3 shows the age-specific incidence of cervical cancer predicted by the base-case model parameters. The peak incidence is 81/100,000 at age 48 years. The predicted distribution of cases by stage was as follows: stage I, 46.4 percent; stage II, 27.0 percent, stage III, 18.1 percent; and stage IV, 8.5 percent.

Sensitivity analyses

We tested the impact of varying the age-specific incidence of HPV from one-half to twice the base-case estimates. As shown in figure 4, peak incidence and overall risk of cervical cancer varies with HPV incidence. Cancer incidence in younger women increases as HPV incidence increases, although the age of peak incidence does not change.

We also tested the impact of varying the prevalence of HPV and low-grade SIL at age 15 years on the subsequent incidence of cervical cancer (figure 5). We found that increasing the prevalence at younger ages without changing other parameters increases overall incidence and lowers the youngest ages at which cancer appears.

![Figure 2](https://academic.oup.com/aje/article-abstract/151/12/1158/55437/fig2)
Changing the age of peak HPV prevalence from 20 to 30 years changed the curve for cervical cancer incidence, but the peak age remained the same. However, delaying the age of peak HPV incidence and decreasing the annual probability that women would present with early-stage cancer did move the peak incidence to later ages (figure 6).

Accounting for hysterectomy incidence lowered the overall population risk of cervical cancer, especially at later ages. However, the estimated risk for women with a cervix is higher than that based on population-based estimates (23) (figure 7). We tested the impact of our natural history estimates on lifetime risk of cervical cancer in the absence of
screening. Table 4 presents the parameters, the input range for sensitivity analysis, and the resulting range of lifetime cervical cancer risk. On the basis of our sensitivity analysis of these parameters, the model suggests that cervical cancer risk is most related to HPV incidence, to the proportion of HPV infections that progress directly to high-grade SIL, and to low-grade SIL progression rates. Changes in these parameters result in two- to threefold differences in cervical cancer risk. Changes in low-grade SIL regression rates and in high-grade SIL progression and regression rates resulted in 50–75 percent differences in cancer risk. The proportion of low-grade SIL lesions that regressed directly to the Well state instead of to the HPV state, and the proportion of high-grade SIL lesions that regressed to Well instead of to low-grade SIL, had minimal impact on cervical cancer risk.

**Examples of model applications**

We estimated the lifetime risk of cervical cancer for women with no evidence of HPV DNA or SIL and with HPV, low-grade SIL, and high-grade SIL at various ages in the absence of further treatment. For a woman older than age 50 years who has no evidence of HPV infection, the risk of subsequent cervical cancer, even in the absence of screening, is less than 0.5 percent (figure 8).

We also tested the impact of screening at various intervals on the age-specific incidence of cancer (figure 9). With screening every 5 years, incidence increased markedly in younger women. As screening frequency increased, the proportion of cases in younger women also increased: with no screening, 47.7 percent of cases occurred in women younger than age 50 years, while 68.1 percent occurred in women younger than age 50 years who were screened every year.

**DISCUSSION**

We developed a Markov model that, when estimates for HPV incidence, regression, and progression as well
TABLE 4. Parameters, input range for sensitivity analysis, and range of lifetime cervical cancer risk*: Markov model

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Base case</th>
<th>Range</th>
<th>Lifetime risk (%) (base case, 3.67)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Relative risk of HPV† infection (age specific)</td>
<td>1.0</td>
<td>0.5–2.0</td>
<td>2.15–6.00</td>
</tr>
<tr>
<td>Proportion of HPV progressing directly to high-grade SIL†</td>
<td>0.1</td>
<td>0–0.3</td>
<td>2.84–5.35</td>
</tr>
<tr>
<td>Proportion of low-grade SIL regressing to Well instead of HPV</td>
<td>0.05</td>
<td>0–1</td>
<td>3.61–4.27</td>
</tr>
<tr>
<td>Proportion of high-grade SIL regressing to Well instead of low-grade SIL</td>
<td>0.01</td>
<td>0.1</td>
<td>3.47–3.67</td>
</tr>
<tr>
<td>HPV prevalence at age 15 years</td>
<td>0.1</td>
<td>0–0.15</td>
<td>3.57–3.72</td>
</tr>
<tr>
<td>HPV progression rate</td>
<td>0.2/36 months</td>
<td>0.15–0.3</td>
<td>2.5–4.7</td>
</tr>
<tr>
<td>Low-grade SIL progression rate‡</td>
<td>0.1–0.3/72 months</td>
<td>0.1–0.5</td>
<td>2.42–5.88</td>
</tr>
<tr>
<td>Low-grade SIL regression rate‡</td>
<td>0.4–0.65/120 months</td>
<td>0.3–0.8</td>
<td>2.92–4.83</td>
</tr>
<tr>
<td>High-grade SIL progression rate</td>
<td>0.35/72 months</td>
<td>0.3–0.5</td>
<td>2.91–4.1</td>
</tr>
<tr>
<td>High-grade SIL regression rate</td>
<td>0.4/120 months</td>
<td>0.3–0.5</td>
<td>2.98–3.8</td>
</tr>
</tbody>
</table>

* Uncorrected for hysterectomy.
† HPV, human papillomavirus; SIL, squamous intraepithelial lesion.
‡ Age dependent.

as for SIL regression and progression were used, resulted in a predicted age-specific incidence of cervical cancer similar to that seen in a number of unscreened populations (3, 45). The age-specific

FIGURE 8. Predicted lifetime risk of cervical cancer in the absence of treatment, by age (years), for women with (from top to bottom) no pathologic diagnosis (bottom curve), detectable human papillomavirus (HPV) (third curve), and biopsy-proven low-grade (second curve) and high-grade (top curve) squamous intraepithelial lesions (LSIL and HSIL, respectively); risks are corrected for hysterectomy.
incidence between ages 50 and 65 years with a more gradual decline. Their modeling suggests that some of this difference results from increased incidences in successive birth cohorts. The curves with peaks at earlier ages were observed primarily for western European countries between 1950 and 1975. The curves showing a later peak incidence are from either third-world countries or western countries in the 1930s and 1940s. Some of the difference in age-specific incidence may be due to differences in age at onset of sexual activity, number of partners, or overall prevalence of HPV in the sexually active population. Since incidence in unscreened populations also is a function of the likelihood of presenting with symptoms, some of the observed differences may be due to variations in access to care, or willingness to seek care, across time and place.

Our predicted distribution of cancer by stage also was similar to that reported in series of cases with no prior screening (35–38).

The incidence of HPV infection, the proportion of rapidly progressive infections, and low-grade SIL progression rates appear to have the largest impact on cervical cancer risk. This finding suggests the potential impact of primary prevention of HPV infection, by using either barrier methods of contraception, vaccination, or abstinence, on cervical cancer risk. Further refinement of the model will enable modeling of the effectiveness and cost-effectiveness of such strategies. The model also can be used to investigate the impact of testing for high-risk HPV types in screening strategies (49). It could even be adapted to model the impact of the probability of specific mutations in HPV-infected cells on cervical carcinogenesis.

Because multiple parameters can affect the predicted incidence of cervical cancer, similar results could be obtained by combining the estimates for the various natural history probabilities differently. Because our initial goal in constructing the model was to use it to analyze the cost-effectiveness of new technologies for improving the sensitivity of cervical smears (16), we chose estimates that resulted in a predicted age-specific incidence that would favor improved sensitivity (early age of peak incidence, relatively high progression rates). Given the range of reported estimates for natural history probabilities, our estimates clearly will not reflect the natural history of HPV infection in all populations. Our predicted prevalence patterns of HPV and SIL are similar to those reported in cross-sectional studies of average-risk populations (46, 47). The model does not predict a second peak in HPV in women later in life, after age 35 years, as was suggested in some studies (38, 50). Because cross-sectional data represent the prevalence in successive birth cohorts, it is likely that at least some of this second peak may be due to age differences in onset of sexual activity or other risk factors in different cohorts. We chose a cohort simulation for computational simplicity and speed. The impact of varying specific parameters in different cohorts on cross-sectional data could be tested easily by using our basic model.

Many of the parameters, especially those related to regression and progression, are reported as means. As Carson and DeMay (47) point out, the age-specific dis-
tribution of low-grade lesions does not appear to be
Gaussian, a finding that our model recreates. More
complete reporting of distributions would enable more
sophisticated modeling techniques that incorporate the
actual distribution of parameter estimates, in turn
allowing more precise estimates of the range of cervi-
cal cancer risk.

The model can be used to predict individual risk of
cervical cancer given a patient's age and histologic
diagnosis (figure 8). Varying the model to incorporate
the distribution of various input parameters could
enable the risk to be expressed as a point estimate with
confidence intervals. Incorporation of patient prefer-
ences for various treatment and follow-up options
could make the model useful for patient counseling.

Another application is in testing the impact of
screening and prevention strategies on cervical cancer
incidence. For example, the dramatic effect of decreas-
ing HPV incidence on cancer incidence (figure 4) sug-
gests the potential impact of effective HPV vaccines.
Assessing the effect of screening on age-specific inci-
dence (figure 9) is another example.

A third application is in exploring the effect of deriv-
ing incidence estimates from populations with varying
degrees of screening, as is common in the United States.
For example, the widely cited Eddy (9) Markov model
has served as the basis for several cost-effectiveness
analyses (9, 51). Although Eddy's model parameters
were adjusted to fit international data (52), the incidence
of invasive cervical cancer in an unscreened US popu-
lation was estimated by assuming that it would be three
times higher than that observed in a partially screened
population. However, this assumption does not account
for the fact that 30–50 percent of US cancer cases occur
in an unscreened population. Because the incidence of
cervical cancer in the United States reflects both
screened and unscreened populations as well as the
effect of different cohorts with varying exposure to
HPV, simply increasing the age-specific incidence by
threefold will overestimate the expected incidence in
unscreened patients at younger ages. If the distribution
of stages is not changed, then the ratio of incidence to
mortality will be overestimated; the ratio of early-stage
cases in the SEER registries is much higher in younger
women than in older women (66 percent localized in
women less than age 50 years compared with 37 percent
in women more than age 50 years) (2).

We were able to approximate the lifetime cervical
cancer risk of Eddy (2.5 percent) (9) by altering our
HPV incidence. However, the lifetime mortality risk
predicted by this model is 0.88 percent, substantially
lower than Eddy's estimate of 1.18 percent. We then
adjusted rates for progression between cancer stages
and symptoms to obtain similar incidence and mortal-
ity risks. By changing the progression rates to 90 per-
cent in 2.5 years for stage I to stage II, 75 percent in 1
year from stage II to stage III, and 100 percent in 1
year for stage III to stage IV and changing the proba-
bility of symptoms for stage III to 35 percent, we
obtained a lifetime risk of 2.52 percent and a mortality
risk of 1.14 percent. However, these progression rates
are inconsistent with those reported by Eddy.

Because detection of cervical cancer in younger
women included in the SEER data is more likely to be
due to screening and therefore occurs at both an earlier
age and an earlier phase of progression than in older
women, survival rates are likely to be higher than for
women who present with symptoms. The high inci-
dence-to-mortality ratio of the Eddy model (9) may be
secondary to extrapolations of distribution by stage in
unscreened populations to the SEER data for younger
women. Use of our model to examine the effects of
different screening intervals also supports this hypo-
thesis: as screening intervals decrease, the proportion
of early-stage disease increases, as does the proportion
of cases among younger women (figure 9), a finding that
has been reported in the British population (53). This
prediction of the model is also consistent with the find-
ing that in younger women, "rapid-onset" cervical can-
cer tends to be early-stage disease (54). In addition, we
have been able to recreate observed SEER incidence
and mortality data by modeling a cohort with varying
proportions of screening intervals, from no screening
over a lifetime to annual screening (16).

These comparisons illustrate the difficulty in esti-
mating the risk of cancer in unscreened populations
when most available data represent both screened and
unscreened populations. Previous models have used
estimates from case-control or cohort studies (33, 52).
However, the consistency of the shape of the curve for
age-specific incidence in unscreened women across
populations (1, 45) facilitates calibration of the model.
Other than the paper of Gustafsson and Adami (55),
we are unaware of another model that takes a similar
approach.

Similarly, relatively few published models of cervi-
cal cancer screening incorporate the HPV status of
normal women (49, 56, 57), and one includes the HPV
status of women infected with human immunodefi-
Ciency virus (58). Given the prevalence of HPV, the
growing insight into its molecular biology, and the
potential role of HPV testing in preventive strategies,
modeling cervical cancer prevention strategies in the
future may well require some method for incorporating
HPV status.

Obvious limitations inherent in any model are
uncertainty surrounding parameter estimates, assump-
tions that can be reasonably debated, and the effects of
changing epidemiologic parameters over time and space. In addition, the specific parameters of our model are based on US data. We used Federation Internationale de Obstetrique et Gynecologie staging of cervical cancer and calculated survival rates on the basis of published US data. Although use of these stages improves the clinical relevance of the model, stage-specific survival may well vary in other settings. Similarly, the effect of hysterectomy rates, a particularly important parameter in assessing the efficiency of screening strategies, may not be as important in other settings in which hysterectomy is not used as widely.

In summary, we have developed a model that synthesizes published data on HPV infection and cervical carcinogenesis and approximates reported patterns of age-specific incidence and prevalence. The model is designed to be updated easily as new evidence becomes available and enables modeling of both hypotheses about the biologic behavior of HPV-related disease and the potential impact of various strategies for preventing cervical cancer. Strategies that reduce HPV incidence can reduce cervical cancer incidence at least as much as strategies that improve the availability or sensitivity of cytologic screening. Given the importance of models in understanding the biology, epidemiology, and policy implications of HPV infection and cervical carcinogenesis, serious consideration should be given to development of a consensus model, or a series of models, for general use.

ACKNOWLEDGMENTS

This paper is based on work performed under contract (no. 290-97-0014) to the Agency for Health Care Policy and Research. The authors are solely responsible for its content. No agency endorsement of opinions or recommendations expressed herein should be stated or implied.

The authors thank Jane Kolimaga for overall project management and Jean Slutsky for project support.

REFERENCES


