Longitudinal Studies Assessing the Population-Level Impact of Human Papillomavirus Vaccination on Reducing Cervical Premalignant Lesions

By

Jaimie Zhi Shing

Dissertation

Submitted to the Faculty of the

Graduate School of Vanderbilt University

in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

in

Epidemiology

May 14, 2021

Nashville, Tennessee

Approved:

Pamela C. Hull, Ph.D., M.A.

Staci L. Sudenga, Ph.D., M.P.H.

Marie R. Griffin, M.D., M.P.H.

Alicia Beeghly-Fadiel, Ph.D, M.P.H.

James C. Slaughter, Dr.P.H., M.S.

To the people without whom this journey would not have been possible, including

My incredible family,

Dr. Karen Z. Ching (Mommy), Dr. Jian Xing (Baba), and Dr. Elaine Z. Shing (Jie Jie),

For teaching me that anything is possible with perseverance and resilience

and

My amazing fiancé, Linh D. Nguyen,

For his unwavering love, support, and encouragement

ACKNOWLEDGEMENTS

I would first like to acknowledge the institutions that provided financial support for this dissertation, including the National Institutes of Health (TL1TR002244 and R01CA207401) and the Emerging Infections Cooperative Agreement from the Centers for Disease Control and Prevention (5U01C10003). I am indebted to the Division of TennCare of the Tennessee Department of Finance and Administration and the Human Papillomavirus Vaccine Impact Monitoring Project (HPV-IMPACT), who provided the data for this study. Of note, research statements in this dissertation are solely the responsibility of the author and do not necessarily represent the official views of the National Institutes of Health, the Centers for Disease Control and Prevention, and the Tennessee Department of Health.

I would like to thank each of member of my dissertation committee, including Drs. Pamela C. Hull, Staci L. Sudenga, Alicia Beeghly-Fadiel, Marie R. Griffin, and James C. Slaughter, for their guidance and support throughout the course of my doctoral training and dissertation research. Specifically, I would like to thank Drs. Pamela C. Hull and Staci L. Sudenga (committee co-chairs) for believing in me every step of the way, and Dr. Marie R. Griffin for her valued mentorship and content expertise. I would also like to thank Dr. Alicia Beeghly-Fadiel for her continued encouragement, and Dr. James (Chris) Slaughter for his statistical advice. Without the help and input from my committee, this work would not have been possible. I deeply value my entire committee for pushing me to think critically and for allowing me to grow professionally and personally.

I am grateful for Edward F. Mitchel, Jr., in the Department of Health Policy at Vanderbilt University Medical Center, for his integral role as the data programmer for each specific aim of

my dissertation. I would also like to thank Manideepthi Pemmaraju, in the Department of Health Policy at Vanderbilt University, for her friendship and hard work on the HPV-IMPACT team at Vanderbilt. Her knowledge and understanding of the HPV-IMPACT data and methods has helped me tremendously. Additionally, I would like to thank Rachel S. Chang, a medical student at Vanderbilt University School of Medicine, and Alyssa B. Rentuza, a Master of Public Health student at Vanderbilt University, for assisting in parts of this work. I would also like to acknowledge Dr. Ronald Alvarez in the Department of Obstetrics and Gynecology at Vanderbilt University Medical Center, for providing medical expertise regarding cervical premalignant lesions and administrative codes; Dr. Charreau Bell from Vanderbilt University's Data Science Institute, for providing methodologic and modeling advice for specific Aim 1; and Dr. Wen Wanqing from the Department of Medicine at Vanderbilt University, for providing initial program code and ensuring the validity of the modeling output for specific Aim 3.

Lastly, I would like to thank my family and friends for their unending love and support. Specifically, I want to acknowledge my parents, Drs. Karen Z. Ching and Jian Xing, for instilling a hard work ethic in me. Without their continued sacrifices and valuable life lessons, I would not be the person I am today. Additionally, I would like to acknowledge my big sister, Dr. Elaine Z. Shing, for always being there for me and setting a prime example of excellence and achievement. Next, I would like to thank my friends, including the ones I have met in Nashville, without whom I would be incredibly lonely, and my life would be less bright. Finally, I want to thank my fiancé, Linh D. Nguyen, for being my biggest cheerleader in both my life and career.

TABLE OF CONTENTS

LIST OF TABLES

Table Page

LIST OF FIGURES

Figure Page

LIST OF ABBREVIATIONS

2vHPV	Bivalent Human Papillomavirus Vaccine
4vHPV	Quadrivalent Human Papillomavirus Vaccine
9vHPV	Nonavalent Human Papillomavirus Vaccine
ACIP	Advisory Committee on Immunization Practices
ACOG	American College of Obstetricians and Gynecologists
ACS	American Cancer Society
AIS	Adenocarcinoma *in situ*
ASCUS	Atypical Squamous Cells of Undetermined Significance
CDC	Centers for Disease Control and Prevention
CI	Confidence Interval
CIN	Cervical Intraepithelial Neoplasia
CIN2	Cervical Intraepithelial Neoplasia Grade 2
CIN2/3	Features of both CIN2 and CIN3
CIN2+	Cervical premalignant lesions (e.g., CIN2, CIN2/3, CIN3, and AIS)
CIN3	Cervical Intraepithelial Neoplasia Grade 3
CPT	Current Procedural Terminology
DNA	Deoxyribonucleic Acid
FDA	Food and Drug Administration
HCPCS	Healthcare Common Procedure Coding System
HGSIL	High-Grade Squamous Intraepithelial Lesion
HPV	Human Papillomavirus

HPV-IMPACT	Human Papillomavirus Vaccine Impact Monitoring Project
ICD	International Classification of Diseases
ICD-9	International Classification of Diseases, 9th Revision
ICD-10	International Classification of Diseases, 10th Revision
ICD-CM	International Classification of Diseases, Clinical Modification
ICD-PCS	International Classification of Diseases, Procedure Coding System
LASSO	Least Absolute Shrinkage and Selection Operator
LGSIL	Low-Grade Squamous Intraepithelial Lesion
MSA	Metropolitan Statistical Area
NPV	Negative Predictive Value
Pap	Papanicolaou
PPV	Positive Predictive Value
RNA	Ribonucleic Acid
TennCare	Tennessee Medicaid
US	United States
USPSTF	United States Preventive Services Task Force

CHAPTER I

INTRODUCTION AND SPECIFIC AIMS

Overview

The human papillomavirus (HPV) is the most common sexually transmitted infection[1,2] and is associated with anogenital warts and cervical, anal, vaginal, penile, oropharyngeal, and vulvar cancers.[3,4] HPV infections contribute to considerable morbidity and costs, including healthcare expenditures, productivity loss, and premature death.[5–7] Since the HPV vaccine's introduction in 2006,[8] assessments of HPV-related adverse health outcomes and their secular trends have been vital in demonstrating the population impact of the HPV vaccine. The current nonavalent HPV vaccine was approved by the Food and Drug Administration in 2014,[9] and protects against HPV types 6, 11, 16, 18, 31, 33, 45, 52, and 58.[10–16] Prior research has attributed these nine HPV types to 90% of cervical cancer cases, suggesting the vaccine has the potential to prevent most cervical cancer cases worldwide.[10] However, in 2018, the estimated global age-standardized rate of incident cervical cancer cases was still 13.1 per 100,000 women.[11]

Because the latency period between an initial exposure to HPV and the development of cancer can be decades (15-20 years) and sometimes longer,[12–14] research aimed at assessing the impact of the HPV vaccine has focused on intermediate outcomes prior to cancer, such as cervical premalignant lesions, which typically develop within 1-3 years after infection.[15–17] Studies have shown notable decreases in the incidence of cervical premalignant lesions, including cervical intraepithelial neoplasia (CIN) grades 2 and 3 (CIN2 and CIN3), cervical lesions with features of both CIN2 and CIN3 (CIN2/3), and adenocarcinoma *in situ* (AIS)—

collectively, referred to as CIN2+, among age groups that may have likely benefited from the HPV vaccine; however, few of these studies have focused on populations with sub-optimal HPV vaccination coverage, such as the state of Tennessee.[18] Tennessee has consistently ranked among the lowest quantile for HPV vaccination; in 2019, the proportion of adolescents aged 13-17 years who initiated the vaccine (i.e., had at least one dose) and were up-to-date (i.e., had all recommended doses) were 61.9% and 43.0%, respectively, with large variation by urbanicity across the state.[18] Specifically, adolescent HPV vaccination proportions in 2019 among Tennessee residents living in metropolitan statistical area (MSA) with a principal city (i.e., urban cities) was 65.7% for initiation and 47.2% for those up-to-date, compared to just 52.5% (initiation) and 33.6% (up-to-date), respectively, among those living in non-MSAs (i.e., largely rural areas).[18] Understanding the impact of the HPV vaccine is critical for informing guidelines to increase vaccination coverage and decrease cervical cancer and CIN2+ incidence, particularly among states with sub-optimal vaccination and across populations with varying vaccination, such as MSAs versus non-MSAs. However, current methods for monitoring population-level CIN2+ in the United States (US) are labor-intensive and time-consuming, as CIN2+ case confirmation requires information from cervical biopsies, which are not included in most national cancer registries or surveillance systems.

Regular CIN2+ reporting in the US is only available among select populations through the state-based New Mexico HPV Pap Registry and the Human Papillomavirus Vaccine Impact Monitoring Project (HPV-IMPACT) in catchment areas of five US states.[19,20] The HPV-IMPACT monitoring project is nationally funded by the Centers for Disease Control and Prevention (CDC) and includes partnerships between the CDC, academic institutions, and Emerging Infections Programs at five state health departments in the following catchment areas:

1) Monroe County, New York; 2) New Haven County, Connecticut; 3) a subset of Washington and Multnomah Counties, Oregon; 4) a subset of Alameda County, California; and 5) Davidson County, Tennessee.[19] The goal of the HPV-IMPACT monitoring project is to improve surveillance of CIN2+ among females aged ≥18 years in the US and monitor the impact and effectiveness of the HPV vaccine on cervical premalignant lesions.[19] Enhanced surveillance, including testing for HPV genotypes on tissue specimens and collecting information on race/ethnicity, insurance status, cervical screening history, and vaccination history, are conducted for CIN2+ diagnoses among females aged 18-39 years in each catchment area.

Despite efforts to improve surveillance of CIN2+ in the US, populations and catchment areas without adequate population-based cervical biopsy data are unable to easily examine trends in CIN2+ to assess HPV vaccine impact. One potential solution is leveraging administrative data from healthcare or insurance databases, which includes rich information on patient procedures and diagnoses through billing claims. Administrative claims data may provide surrogate metrics, as these data include codes such as the International Classification of Diseases, Clinical Modification (ICD), Current Procedural Terminology (CPT), and Healthcare Common Procedure Coding System (HCPCS), which systematically classify health outcomes and patient services across various billing claims databases. However, claims-based models for estimating population-based CIN2+ incidence have not yet been validated for public health surveillance purposes.

Further, the recent ICD transition from the ninth (ICD-9) to tenth (ICD-10) revision in 2015[21] introduces challenges for assessing long-term population-based trends because ICD-10 codes may have different performance characteristics than ICD-9 codes. One US study that used claims data to examine trends in CIN2+ prevalence[22] did not assess trends past 2014 because

little is known on the interpretability of trends across both ICD-9 and ICD-10 eras. To expand CIN2+ surveillance and continue examining the impact of the HPV vaccine, methodologic insight regarding the validity of claims data for identifying CIN2+ pre-to-post ICD-10 transition is warranted. The ability to utilize claims data from the ICD-10 era will allow future studies to assess longer-term trends and include more recent data for continued monitoring of the HPV vaccine's impact on reducing CIN2+.

Specific Aims

Given the aforementioned research gaps, our overarching objective was to utilize administrative data from the Tennessee Medicaid (TennCare) program to retrospectively examine the impact of the HPV vaccine on reducing CIN2+ incidence from 2008 to 2018 among TennCare-enrolled women across the entire state of Tennessee, a population with sub-optimal vaccination coverage. To achieve this objective, we proposed the following three specific aims:

1. **Aim 1: To build and validate claims-based prediction models for identifying CIN2+ events in both ICD-9 and ICD-10 eras, and to compare the discriminative ability of models between ICD eras and by age group.** Pathology-confirmed CIN2+ events among TennCare-enrolled women in Davidson County, Tennessee, were identified by the HPV-IMPACT monitoring project, which served as our population-based gold standard events. Using TennCare billing claims data among women residing in Davidson County, Tennessee, we built and validated several models developed by various approaches, including 1) CIN2+ tissue diagnosis codes alone, 2) least absolute shrinkage and selection operator (LASSO), and 3) random forest classifiers. We assessed discrimination

and calibration and compared the performance of each model between ICD-9 and ICD-10 eras, and by age group.

2. **Aim 2: To examine the HPV vaccine's impact in a population with sub-optimal vaccination proportions by assessing trends in CIN2+ incidence from 2008 to 2018 among TennCare enrollees.** We identified age-group-specific (18-20, 21-24, 25-29, 30-34, and 35-39 years) annual CIN2+ incidence using the validated algorithm from Aim 1 among 1) TennCare-enrolled women aged 18-39 years and 2) those who were screened for cervical cancer to account for changes in screening patterns over time. Joinpoint regression was used to identify significant trend segments from 2008 to 2018, estimate annual percent changes for each trend segment, and estimate average annual percent changes from 2008 to 2018.
3. **Aim 3: To examine the HPV vaccine's impact in populations with varying vaccination proportions by assessing trends in CIN2+ incidence from 2008 to 2018 by urbanicity.** Similar to Aim 2, we identified age-group-specific annual CIN2+ incidence using the validated algorithm from Aim 1 among 1) TennCare-enrolled women aged 18-39 years and 2) those who were screened for cervical cancer to account for changes in screening patterns over time, stratifying by urbanicity (MSA versus non-MSA). Joinpoint regression was used to summarize trends, while age-period-cohort models were conducted to examine age, period, and cohort effects overall and by urbanicity.

CHAPTER II

BACKGROUND

Epidemiology and Burden of The Human Papillomavirus

The human papillomavirus (HPV) has over 200 identified genotypes and is the most common sexually transmitted infection in the United States (US).[1,2] Most HPV infections are transmitted sexually; however, non-sexual modes of transmission are also possible, including autoinoculation, direct contact with infected surfaces, and mother-to-child vertical transmission.[23] Prior to the introduction of the HPV vaccine, the prevalence of having any genital HPV infection, measured by HPV deoxyribonucleic acid (DNA) positivity, among US females aged 14-59 years was 43% in 2003-2006, with the highest prevalence in women aged 20-24 years (54%).[24] Additionally, before the availability of the HPV vaccine, the average probability of acquiring any type of HPV infection during one's lifetime was 91% and 85% for sexually active women and men, respectively, and 80% of these women and men would have acquired an HPV infection by age 45 years.[25] In a global-based systematic review of genital HPV-DNA prevalence, peak prevalence was consistently found to be among women aged 25 years and younger.[26] Such high prevalence of HPV infection is problematic because HPV infections are associated with several adverse health outcomes, including anogenital warts and cancers of the cervix, vulva, vagina, penis, anus, and oropharynx.[27]

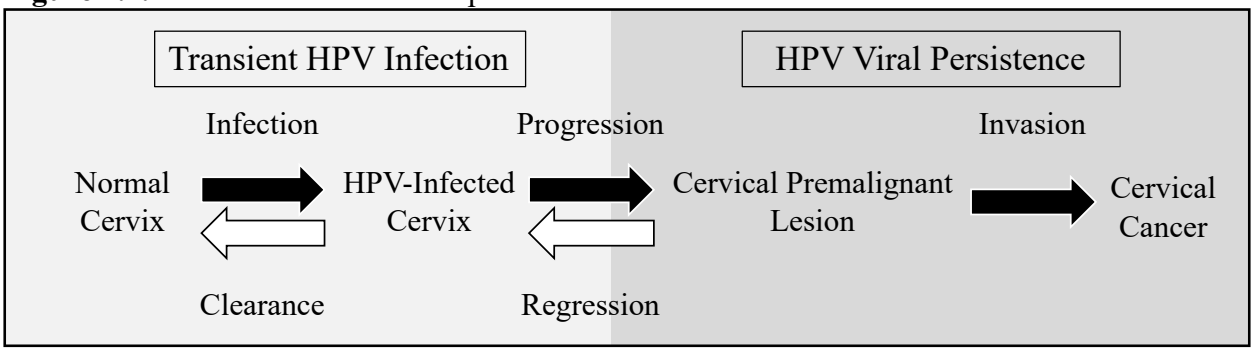
Although more than 90% of HPV infections are asymptomatic and clear within 6-24 months after infection,[2,28] HPV still contributes to considerable morbidity, mortality, and costs, including healthcare expenditures, productivity loss, and premature death.[5–7] Direct annual

medical costs for preventing and treating HPV-associated health outcomes in the US are estimated at $8 billion, of which $7 billion is toward cervical cancer screening, follow-up, and treatment.[29] Cervical cancer is also associated with other social and behavioral burdens, including significantly higher depression severity, lower quality of life, and more limitations in daily activities compared to women without cervical cancer.[7] Further, the 5-year relative cervical cancer survival rate in the US is approximately 66% with an annual mortality rate of 2.2 per 100,000 women.[30] Despite being vaccine-preventable, HPV-associated health outcomes are still prevalent in the US today, contributing to large direct and indirect costs that could be prevented through vaccination.

Pathogenesis of Cervical Premalignant Lesions and Cervical Cancer

The International Agency for Research on Cancer has identified twelve main high-risk, oncogenic HPV genotypes, including HPV 16, 18, 31, 33, 35, 39, 45, 51, 56, 58, and 59, with several others determined as possibly carcinogenic.[31] Of these, HPV types 16 and 18 alone attribute to approximately 70% of cervical cancer cases worldwide.[32] Most HPV infections are transient and typically clear or resolve spontaneously within two years.[2,28] However, 10% of cervical HPV infections persist for over two years[33]; persistent HPV infection is a major risk factor for the progression to a cervical premalignant lesion,[33,34] such as high-grade cervical intraepithelial neoplasia (CIN) grades 2 and 3 (CIN2 and CIN3), and adenocarcinoma *in situ* (AIS), collectively referred to as CIN2+, which can then develop into cervical cancer if left untreated (**Figure 2.1**). The strongest predictor of HPV persistence and progression to CIN2+ is HPV genotype, of which HPV 16 is most frequently detected in CIN, while HPV 18 is most prevalent in AIS.[34,35]

Figure 2.1.[a] Overview of the development of cervical cancer.



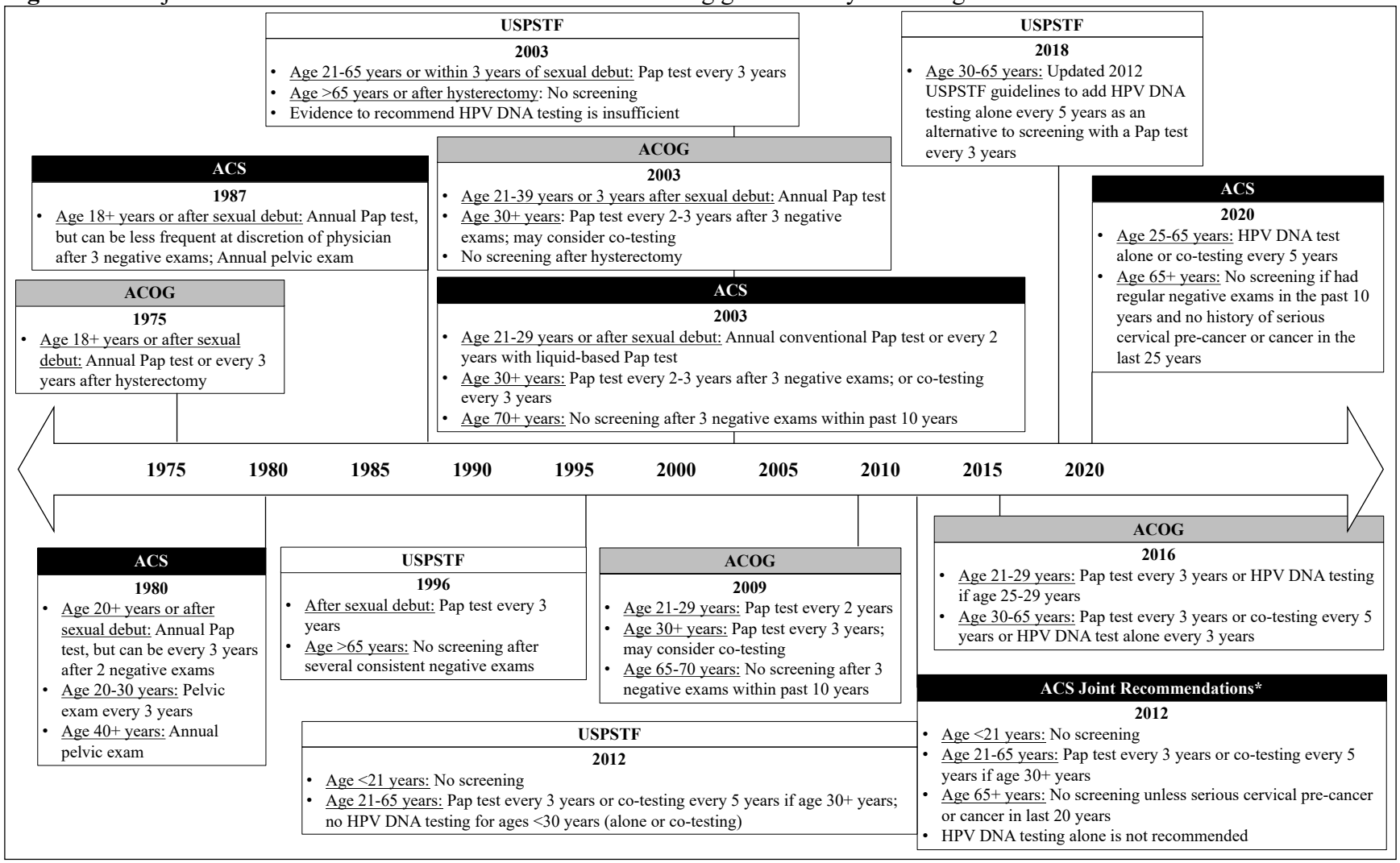
Abbreviations: HPV = Human Papillomavirus.
[a]Figure adapted from Schiffman *et al.* (2007). Lancet.[28]

CIN is identified by squamous cell abnormalities in the ectocervix (i.e., the surface of the cervix), while AIS is identified by glandular epithelium abnormalities in the endocervix (i.e., the cervical canal).[28] Cervical premalignant lesions generally develop within 1-3 years after infection[15–17] and are considered pre-cursors of cervical cancer; however, not all will progress to cancer due to regression. Of the premalignant lesions that do progress to invasive cervical cancer, the latency period from an initial HPV infection to the development of cancer, or even from premalignant lesion to cancer, can take decades (15-20 years) to develop.[12–14,36]

Cervical Cancer Prevention: Screening and The HPV Vaccine

To aid in the prevention of cervical premalignant lesions from progressing into cancer, several professional organizations in the US, including the American Cancer Society (ACS),[37] the United States Preventive Services Task Force (USPSTF),[38] the American College of Obstetricians and Gynecologists (ACOG),[39] and others, have released cervical cancer screening recommendations; however, these have varied in guidelines for age at screening initiation, frequency of screening, and circumstances for when to stop screening (**Figure 2.2**).

Figure 2.2. Major events in the timeline of cervical cancer screening guidelines by select organizations.



Abbreviations: ACOG = American College of Obstetricians and Gynecologists; ACS = American Cancer Society; USPSTF = United States Preventive Services Task Force.

*ACS recommendations in 2012 were made jointly with the American Society for Colposcopy and Cervical Pathology and the American Society for Clinical Pathology.

In the 1970s through the early 2000s, annual cytology, or Papanicolaou (Pap), tests were recommended after sexual debut or upon reaching a specific age (**Refer to Figure 2.2**). Some organizations recommended screening to begin by age 20 years, while others recommended by age 18 years or did not have a specific age guideline.[40–42] Frequency of screening also varied by organization; in 1975, ACOG recommended annual Pap tests or every 3 years after a hysterectomy[40]; in 1980, ACS recommended annual Pap tests or every 3 years after 2 negative exams[42]; and in 1996, USPSTF recommended Pap tests every 3 years for all women after sexual debut until age 65 years.[41] In 1996, USPSTF began recommending no cervical cancer screening for women aged 65 years and older due to the potential harms, such as the likelihood of false positives and the repercussions from invasive procedures, and because of the low yield of screening among older adults due to declining cervical premalignant lesions incidence after middle age.[41] In 2003, several organizations changed their recommendations to also stop screening at age 65 or 70 years, and to begin screening by age 21 years or after sexual debut[42–44]; during this time, screening was still recommended annually by ACS and ACOG; however, for those aged 30 and older, Pap tests were recommended every 2-3 years after 3 negative exams.[42,44]

Despite varying recommendations between different organizations early on, the guidelines became more homogeneous in 2012, when a joint recommendation was released by the American Cancer Society, the American Society for Colposcopy and Cervical Pathology, and the American Society for Clinical Pathology,[45] which were similar to the 2012 guidelines set forth by USPSTF and ACOG (**Refer to Figure 2.2**).[46] One of the major updates to the screening guidelines were that Pap tests were no longer recommended annually, but instead, recommended every 3 years for women aged 21-65 years, with the option of co-testing (i.e., Pap test with HPV

DNA test) every 5 years for women aged 30-65 years.[45,46] Prior to 2012, organizations were reluctant to recommend HPV DNA testing as an alternative for cervical cancer screening due to insufficient evidence to adequately examine the benefits and potential harms.[43] By 2012, several studies had indicated that HPV DNA testing was generally more sensitive and less specific at detecting CIN2 and CIN3 events than Pap tests, and thus may identify more false positives[46]; however, USPSTF concluded that having a longer screening interval of every 5 years for co-testing in women aged 30-65 years may reduce the opportunity for false-positives, but advised women choosing this screening method to be aware that persistent positive HPV results may result in increased surveillance and repeat testing.[46] Further, additional testing would be needed within the next 12 months for women receiving inconsistent results, such as a normal Pap test (i.e., negative cytology exam) and a positive HPV DNA test.[47] However, studies showed that this would only occur in 11% of women aged 30-34 years, and 3% of women aged 60-65 years.[48,49]

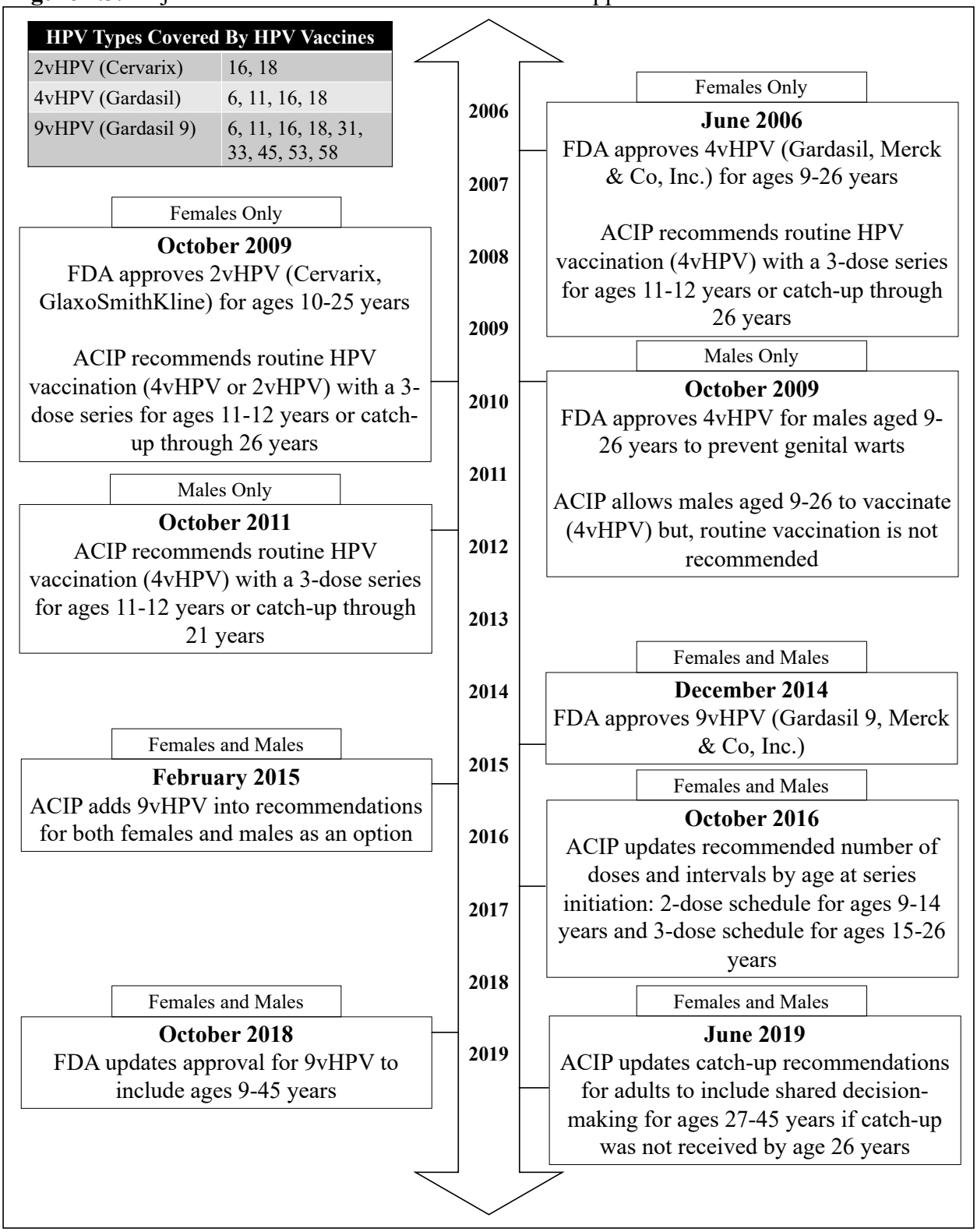
The updated recommendations in 2012 also emphasized the importance of not screening women younger than 21 years due to the potential harms for infections likely to spontaneous resolve and because of the rare occurrence of cervical cancer in women before age 21 years (**Refer to Figure 2.2**).[45,46] Specifically, the detection of abnormal cervical cells in young women may lead to more frequent and repeat testing and/or unnecessary invasive diagnostic procedures, such as a cervical biopsy or colposcopy, leading to potential vaginal bleeding, procedure-induced infections, pain, and anxiety.[46] Further, early intervention and treatment of cervical premalignant lesions, including cold-knife conizations and loop excisions, are associated with several adverse pregnancy outcomes, such as preterm birth, low birth weight, and perinatal death.[50] Through an in-depth decision analysis by the USPSTF, the harms of screening women before age 21 years were found to outweigh the benefits and were no longer recommended.[51]

In addition to cervical cancer screening, routine HPV vaccination also prevents cervical cancer. To protect against cervical cancer and many other HPV-associated health outcomes, the Food and Drug Administration (FDA) approved the licensure of the first available quadrivalent HPV vaccine (4vHPV, Gardasil) in 2006 for females aged 9-26 years, covering HPV 6, 11, 16, and 18 (**Figure 2.3**).[52] As mentioned, HPV 16 and 18 are the main high-risk, oncogenic genotypes responsible for 70% of cervical cancer cases.[32] The other two genotypes covered by the quadrivalent HPV vaccine, HPV 6 and 11, are considered low-risk to cause cervical cancer and are responsible for 90% of anogenital warts.[53] Along with the vaccine's approval in 2006 came recommendations by the Advisory Committee on Immunization Practices (ACIP) for routine HPV vaccination among female adolescents aged 11-12 years using a 3-dose series at intervals of 0, 1-2 months, and 6 months, with catch-up vaccination through age 26 years.[52]

In 2009, the FDA approved the bivalent HPV vaccine (2vHPV, Cervarix), covering HPV 16 and 18, for females only,[54] and also approved 4vHPV for males aged 9-26 years for genital wart prevention. However, at the time, ACIP did not recommend routine vaccination in males.[55] Following increased evidence and justification for the benefits of male vaccination, ACIP began recommending routine HPV vaccination among males aged 11-12 years using a 3-dose series in 2011, with catch-up vaccination through age 21 years.[56]

At the end of 2014, the FDA approved the most recent nonavalent HPV vaccine (9vHPV, Gardasil 9), protecting against HPV 6, 11, 16, 18, 31, 33, 45, 53, and 58 (**Figure 2.3**). These nine HPV types are responsible for 90% of cervical cancer cases worldwide.[10] Current ACIP recommendations include routine HPV vaccination for both females and males aged 11-12 years, with catch-up through age 26 years with the nonavalent HPV vaccine.[57,58] In 2016, the nonavalent HPV vaccine became the only HPV vaccine distributed in the US.[59] A 2-dose

Figure 2.3. Major events in the timeline of HPV vaccine approvals and recommendations.



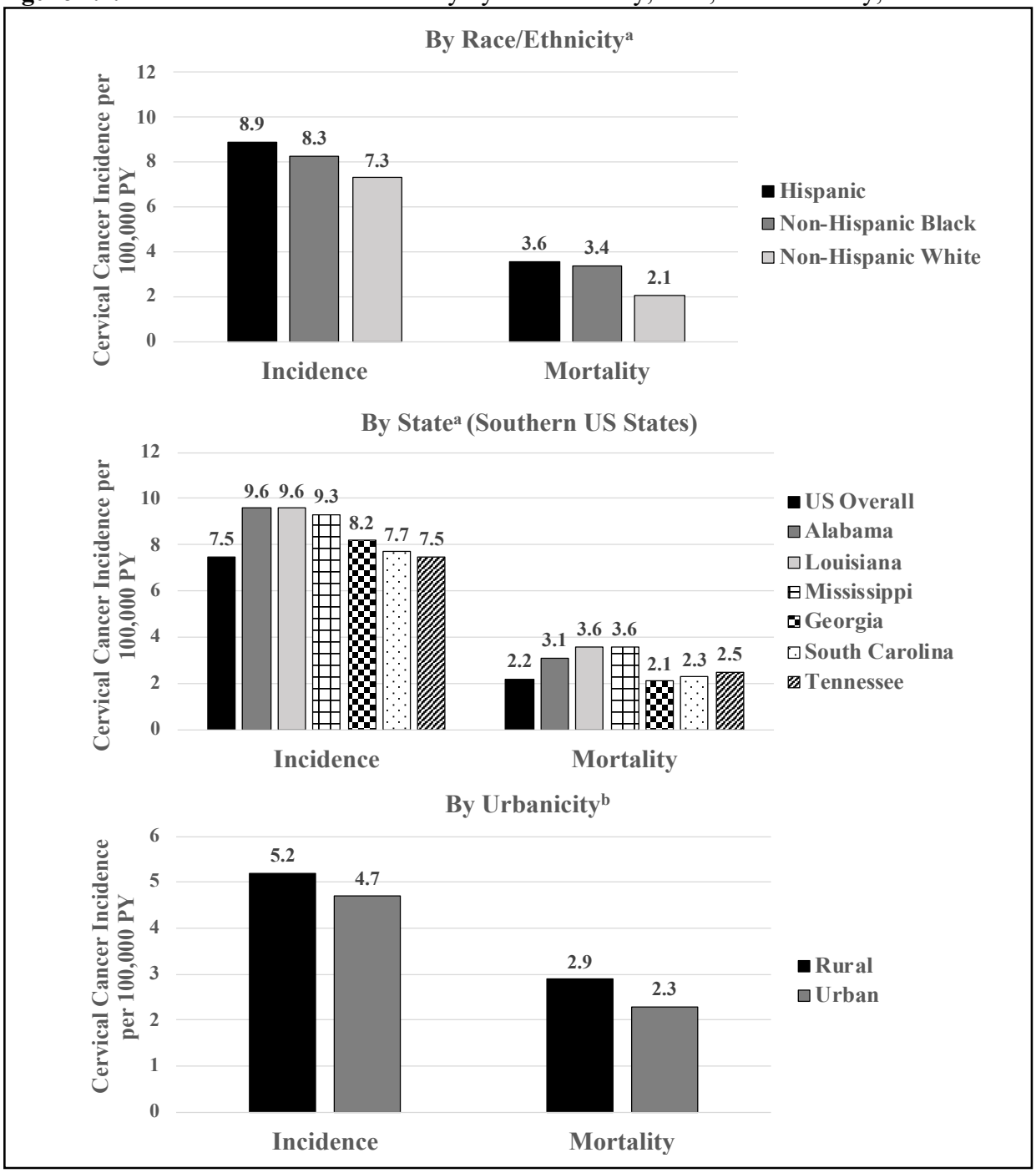
Abbreviations: 2vHPV = Bivalent HPV Vaccine; 4vHPV = Quadrivalent HPV Vaccine; 9vHPV = Nonavalent HPV Vaccine; ACIP = Advisory Committee on Immunization Practices; FDA = Food and Drug Administration; HPV = Human Papillomavirus.

schedule at intervals of 0 and 6-12 months is used for adolescents who initiate at age 9-14 years, while a 3-dose schedule at intervals of 0, 1-2 months, and 6 months is used for those who initiate at age 15-26 years.[57,58] For older adults, benefits of the HPV vaccine declines with increasing age because of the high likelihood of already being exposed to HPV (**Refer to Chapter II, Section: "Epidemiology and Burden of The Human Papillomavirus"**); thus, the vaccine is most beneficial for persons who have not yet engaged in sexual activity or young adults aged 26 years and younger. However, some older adults aged 27-45 years who have new sex partners may be at risk for acquiring a new HPV infection and can also be considered for HPV vaccination after shared decision making with their provider.[57]

Disparities in Cervical Cancer and HPV Vaccination

As a result of increased cervical cancer screening and improved management and treatment,[60] the age-adjusted cervical cancer incidence in the US has declined from 9.7 per 100,000 women in 1999 to 7.5 per 100,000 women in 2017.[61] However, despite declining trends overall, disparities in cervical cancer incidence and mortality exist between sociodemographic subgroups. Access to timely screening and quality of care differ by socioeconomic status, race/ethnicity, region, and urbanicity, all of which can impact stage at diagnosis and the resulting prognosis.[62] Specifically, women of Hispanic and non-Hispanic Black race/ethnicity, and those living in Southern US states or rural counties have higher incidence and mortality of cervical cancer compared to women of non-Hispanic White race/ethnicity and those living in a non-Southern US states or urban counties, respectively (**Figure 2.4**).[63–65]

Figure 2.4. Cancer incidence and mortality by race/ethnicity, state, and urbanicity, United States.



Abbreviations: US = United States.

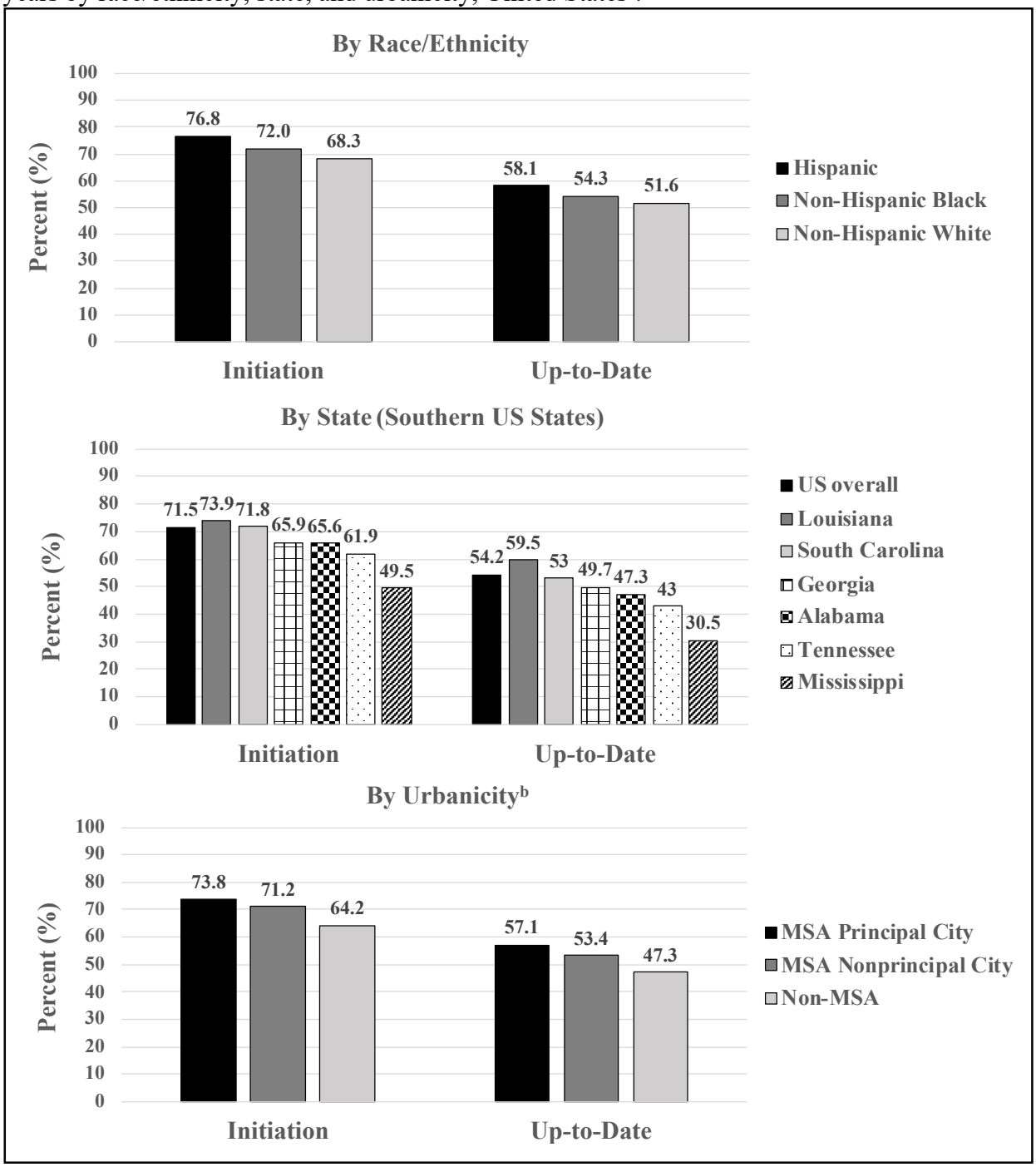
[a]Incidence and mortality rates by race/ethnicity and state are based on 2017 data reported by the Centers for Disease Control and Prevention.[61]

[b]Incidence rates by urbanicity are based on 2010-2014 data of localized cervical cancer cases reported by the National Cancer Institute's Surveillance, Epidemiology, and End Results Program[64]; Mortality rates by urbanicity are based on 2007 data from the national vital statistics mortality database.[65]

Likewise, racial, regional, and geographic disparities also exist for HPV vaccination (**Figure 2.5**). Adolescents of Hispanic and non-Hispanic Black race/ethnicity have historically had lower proportions of those who are up-to-date (i.e., had all recommended HPV vaccine doses) compared to their non-Hispanic White counterparts, despite having higher proportions of HPV vaccine initiation (i.e., had at least one dose).[18,66,67] Reasons for failure to complete the vaccine series among minority groups may be lack of parental knowledge about vaccine dosing schedules, lack of flexibility and convenience of the parent's and child's schedules for follow-up doses, and lack of clinic reminders and recall systems.[66] However, in recent years, HPV vaccine initiation and up-to-date coverage among Hispanics and non-Hispanic Blacks have surpassed those of non-Hispanic Whites.[18]

Regionally, HPV vaccination has generally been low in Southern US states compared to other states (**Figure 2.5**). In 2019, the overall HPV vaccine initiation and up-to-date proportions among adolescents aged 13-17 years was 71.5% and 54.2%, respectively, compared to 67.1% and 50.3%, respectively, for Region IV states (i.e., southeastern US states).[18] Of note, the state of Tennessee has consistently had less-than-optimal HPV vaccination proportions, with lower initiation (61.9%) and up-to-date (43.0%) proportions among adolescents aged 13-17 years compared to the national average.[18] Similar to the geographic disparities observed for cervical cancer, rural areas have also had poorer HPV vaccination coverage compared to urban areas. These geographic incongruences may be attributed to rural areas having more barriers to vaccination,[68] including lack of knowledge and awareness of HPV and its link to cancer,[69,70] more negative community messaging,[70] and more religious and cultural beliefs that may not support vaccination.[71,72]

Figure 2.5. HPV vaccination initiation and up-to-date coverage among adolescents aged 13-17 years by race/ethnicity, state, and urbanicity, United States[a].



Abbreviations: MSA = Metropolitan Statistical Area; HPV = Human Papillomavirus; US = United States.

[a]All data presented in the figure are based on 2019 initiation and up-to-date coverage from the National Immunization Survey-Teen reported by the Centers for Disease Control and Prevention.[18]

[b]MSAs represent urbanized areas, while non-MSAs represent largely rural areas, based on household county of residence and population counts determined by the US Census Bureau.

Summary of Population-Level Human Papillomavirus Vaccine Impact

One effective way of monitoring the population-level effects of the HPV vaccine is through surveillance studies assessing patterns and trends in HPV vaccination and HPV-associated health outcomes. Understanding the impact and effectiveness of the HPV vaccine is important for informing vaccination guidelines to aid in cancer prevention efforts. Individual-level observational studies comparing HPV-associated outcomes in vaccinated versus unvaccinated groups can demonstrate vaccine effectiveness, while population-level ecologic studies that assess aggregated trends in HPV-associated outcomes can demonstrate vaccine impact. The primary concern with ecologic studies is attributing changes in outcomes to a specific exposure when other exposures are fully or possibly responsible, which has been termed ecologic fallacy.[73] Mitigating this concern is important by taking into account other secular trends that might change outcomes—in this case, considering changes in screening and diagnosis for CIN2+, examining CIN2+ by groups that may have varying rates of disease detection (e.g., age group, urbanicity, income-level, race/ethnicity, etc.), and disentangling age, period, and cohort effects. Ultimately, ecologic studies examining HPV vaccine impact are vital for capturing both direct effects (i.e., vaccination) and indirect effects (i.e., herd effects) that are unable to be examined in effectiveness studies, which can be useful for assessing population-level impact.

In the US, ecologic studies examining HPV vaccine impact have demonstrated direct and indirect effects of the vaccine on reducing HPV-associated health outcomes, including HPV infection prevalence, anogenital warts, and cervical premalignant lesions (**Table 2.1**). Results from these studies have demonstrated significant declines in HPV-associated health outcomes among younger age groups that may have benefited from the vaccine's introduction.

Table 2.1. Summary of ecologic studies examining population-based HPV vaccine impact on HPV infection prevalence, anogenital warts, and cervical premalignant lesions.

Study	Population	Comparison Years	Main Results
		HPV Infection Prevalence	
Kahn *et al.* (2012)[74]	Females aged 13-26 years from primary care and sexually transmitted infection clinics in Cincinnati, Ohio	2009-2010 versus 2006-2007	• 58% decrease (31.7% to 13.4%) in 4vHPV-type prevalence from cervicovaginal swabs among females aged 13-26 years
Markowitz *et al.* (2013)[75]	Females aged 14-59 years from the National Health and Nutrition Examination Survey	2007-2010 versus 2003-2006	• 56% decrease (11.5% to 5.1%) in 4vHPV-type prevalence from cervicovaginal swabs among females aged 14-19 years; • No significant differences between time periods for other age groups
Dunne *et al.* (2015)[76]	Females aged 20-29 years screened for cervical cancer at Kaiser Permanente Northwest	2012-2013 versus 2007	• 42% decrease (10.6% to 6.2%) in 4vHPV-type prevalence from liquid cytology cervical specimens among females aged 20-29 years
Kahn *et al.* (2016)[77]	Females aged 13-26 years from primary care and sexually transmitted infection disease clinics in Cincinnati, Ohio	2013-2014 versus 2006-2007	• 75% decrease (34.8% to 8.7%) in 4vHPV-type prevalence from cervicovaginal swabs among females aged 13-26 years
Markowitz *et al.* (2016)[78]	Females aged 14-34 years from the National Health and Nutrition Examination Survey	2009-2012 versus 2003-2006	• 64% decrease (11.5% to 4.3%) in 4vHPV-type prevalence from cervicovaginal swabs among females aged 14-19 years; • 34% decrease (18.5% to 12.1%) among females aged 20-24 years; • No significant decreases for older age groups
Oliver *et al.* (2017)[79]	Females aged 14-34 years from the National Health and Nutrition Examination Survey	2011-2014 vs 2003-2006	• 71% decrease (11.5% to 3.3%) in 4vHPV-type prevalence from cervicovaginal swabs among females aged 14-19 years; • 61% decrease (18.5% to 7.2%) among females aged 20-24 years; • No significant decreases for older age groups
		Anogenital Warts	
Bauer *et al.* (2012)[80]	Females and males aged ≤10 years from claims data of	2010 versus 2007	• 34.8% decrease (0.9% to 0.1%) in genital warts among females aged 10-20 years;

	enrollees in the California Family Planning Access Care and Treatment program		• 10% decrease (1.0% to 0.9%) among females aged 21-25 years; • 18.6% decrease (2.7% to 2.2%) among males aged 10-21 years; • 11.2% decrease (5.1% to 4.5%) among males aged 21-25 years; • Increases or no significant trend for older age groups
Flagg *et al.* (2013)[81]	Females and males aged 10-39 years from the nationwide MarketScan Commercial Claims and Encounters Database of privately insured persons	Trends from 2003-2010	• 38% decrease (2.9 per 1000 PY in 2006 to 1.8 per 1000 PY in 2010) in anogenital wart prevalence among females aged 15-19 years; • 13% decrease (5.5 per 1000 PY in 2009 to 4.8 per 1000 PY in 2010) among females aged 20-24 years; • 9% decrease (4.1 per 1000 PY in 2009 to 3.7 per 1000 PY in 2010) among females aged 25-29 years; • 8% decrease (5.0 per 1000 PY in 2009 to 4.6 per 1000 PY in 2010) among males aged 20-24 years; • No significant decreases for older age groups
Perkins *et al.* (2015)[82]	Females and males aged 16-26 years from claims data of enrollees in an urban medical center and 6 community health centers in Boston, Massachusetts	2013 versus 2004	• 57% decrease (3.5% to 1.5%) in genital warts among females aged 16-26 years; • 19% decrease (3.6% to 2.9%) among males aged 16-26 years
Flagg and Torrone (2017)[83]	Females and males aged 15-39 years from the nationwide MarketScan Commercial Claims and Encounters Database of privately insured persons	Trends from 2008 to 2014	• 14% annual decrease (2.6 per 1000 PY in 2008 to 1.0 per 1000 PY in 2014) in anogenital wart prevalence among females aged 15-19 years; • 13% annual decrease (5.5 per 1000 PY in 2009 to 2.7 per 1000 PY in 2014) among females aged 20-24 years; • 6% annual decrease (4.1 per 1000 PY in 2009 to 2.9 per 1000 PY in 2014) among females aged 25-29 years; • 5% annual decrease from 2009 to 2014 (specific rates not presented in paper) among males aged 15-19 years; • 7% annual decrease from (5.0 per 1000 PY in 2009 to 3.6 per 1000 PY in 2014) among males aged 20-24 years; • Increases or no significant trend for older age groups

Shing *et al.* (2019)[84]	Females and males aged 15-39 years from claims data of enrollees in the Tennessee Medicaid Program	Trends from 2006-2014	• 11% annual decrease (3.1 per 1000 PY in 2006 to 1.3 per 1000 PY in 2014) in anogenital wart incidence among females aged 15-19 years; • 4% annual decrease (3.1 per 1000 PY in 2011 to 2.5 per 1000 PY in 2014) among females aged 20-24 years; • Increases or no significant trend for older age groups and males
Mann *et al.* (2019)[85]	Females and males of all ages from 27 sexually transmitted infection clinics across the United States	2016 versus 2010	• 13% annual decrease (2.3% to 0.9%) in anogenital wart prevalence among all females; • 8% annual decrease (7.3% to 4.4%) among males who have sex with females; • 11% annual decrease (6.2% to 2.9%) among males who have sex with males
Naleway *et al.* (2020)[86]	Females and males aged 11-39 years screened for cervical cancer at Kaiser Permanente Northwest	Post-vaccine era (2007-2016 for females; 2011-2016 for males) versus pre-vaccine era (2000-2006 for females; 2000-2010 for males)	• 67% decrease (44.2 per 10,000 PY to 14.6 per 10,000 PY) in anogenital wart incidence among females aged 15-19 years; • 48% decrease (60.2 per 10,000 PY to 31.5 per 10,000 PY) among females aged 20-24 years; • 45% decrease (11.9 per 10,000 PY to 6.5 per 10,000 PY) among males aged 15-19 years; • 31% decrease (53.8 per 10,000 PY to 36.9 per 10,000 PY) among males aged 20-24 years
Cervical Premalignant Lesions			
Niccolai *et al. (*2013)[87]	Females aged 18-39 years from a population-based surveillance site, New Haven, Connecticut	2011 versus 2008	• 18% decrease (834 per 100,000 PY to 688 per 100,000 PY) in CIN2+ incidence among females aged 21-24 years; • No significant decreases for older age groups
Hariri *et al.* (2015)[88]	Females aged 18-39 years from 4 population-based surveillance sites in California, Connecticut, New York, and Oregon	2012 versus 2008	• Significant decreases in CIN2+ incidence among females aged 18-20 years at all 4 sites; • Significant decreases among females aged 21-29 years in Connecticut and New York; • No significant decreases for older age groups
Flagg *et al.* (2016)[22]	Females aged 15-39 years from the nationwide MarketScan Commercial Claims and	Trends from 2007-2014	• 14% average annual decrease (14.8% in 2007 to 4.9% in 2014) in CIN2+ prevalence among females aged 15-19 years;

	Encounters Database of privately insured persons		• 8% average annual decrease (20.5% in 2007 to 11.3% in 2014) among females aged 20-25 years; • No significant decreases for older age groups
Benard *et al.* (2017)[89]	Females aged 15-29 years from a statewide surveillance system, New Mexico	Trends from 2007-2014	• 11% annual decrease (896.4 per 100,000 PY in 2007 to 414.9 per 100,000 PY in 2014) in CIN2 incidence among females aged 15-19 years; • 41% annual decrease (240.2 per 100,000 PY in 2007 to 0 per 100,000 PY in 2014) in CIN3 incidence among females aged 15-19 years; • 6% annual decrease (1027.7 per 100,000 PY in 2007 to 627.1 per 100,000 PY in 2014) in CIN2 incidence among females aged 20-24 years
Oakley *et al.* (2018)[90]	Females aged 18-39 years from a population-based surveillance site, Davidson County, Tennessee	Trends from 2008 to 2013	• 24% annual decrease (188.9 per 100,00 PY in 2008 to 58.7 per 100,000 PY in 2013) in CIN2+ incidence among females aged 18-20 years; • 10% annual decrease (495.6 per 100,000 PY in 2008 to 332.4 per 100,000 PY in 2013) among females aged 21-24 years; • No significant decreases for older age groups
Gargano *et al.* (2019)[91]	Females aged 18-39 years from 5 population-based surveillance sites in California, Connecticut, New York, Oregon, and Tennessee	2010-2011 and 2012-2013 versus 2008-2009	• Significant decreases in CIN2+ incidence among females aged 18-20 and 21-24 years at all 5 sites; • Varying trends among females aged 25-29, 30-34, and 35-39 years across study sites
McClung *et al.* (2019)[92]	Females aged ≥18 years from 5 population-based surveillance sites in California, Connecticut, New York, Oregon, and Tennessee	Trends from 2008 to 2016	• 38% average annual decrease (206 per 100,000 PY in 2008 to 12 per 100,000 PY in 2016) in CIN2+ incidence among females aged 18-19 years; • 15% average annual decrease (559 per 100,000 PY in 2008 to 151 per 100,000 PY in 2016) among females aged 20-24 years

Abbreviations: 4vHPV = Quadrivalent HPV Vaccine; CIN2 = Cervical Intraepithelial Neoplasia Grade 2; CIN2+ = Cervical Premalignant Lesions (e.g., cervical intraepithelial neoplasia grades 2 and 3 and adenocarcinoma *in situ*; CIN3 = Cervical Intraepithelial Neoplasia Grade 3; HPV = Human Papillomavirus; PY = Person-Years.

These studies have also shown increases or no significant trends among older age groups, suggesting the declines in younger ages may be attributed to the effects of vaccination. Some studies have also specifically examined trends in HPV vaccine type-specific HPV-infections[93] and cervical premalignant lesions[92,94] These studies have shown declines in vaccine type-specific HPV infections and cervical premalignant lesions, demonstrating HPV vaccine impact.[92–94] Further, one study showed declining trends in CIN2+ incidence among young women who tested positive for HPV 16 and 18 (types covered by the HPV vaccine) in both vaccinated and unvaccinated populations, suggesting both direct and indirect effects of the vaccine.[94]

Presently, significant declines in HPV-associated health outcomes have not been observed among older age groups due to lower vaccination rates in adults compared to adolescents, lower vaccine effectiveness in older persons due to the high likelihood of prior exposure to the virus, and background secular trends in unvaccinated adults. Future studies may begin to observe greater population impact of the vaccine as HPV vaccination rates continue to increase and younger cohorts age into older cohorts.

Because of the long latency period between an initial HPV infection to the development of cancer,[12–14,36] US studies observing meaningful vaccine impacts on reducing cervical cancer are only in the early stages, as the vaccine has only been available since 2006.[52] One recent US study that utilized the United States Cancer Statistics database, a population-based cancer registry, demonstrated declines in the incidence of cervical squamous cell carcinoma and adenocarcinoma from 1999 to 2017, with largest declines among women aged 15-20 years (12.7% average annual decline in cervical squamous cell carcinoma and 4.1% average annual decline in adenocarcinoma).[59] The authors noted that this age group is not typically screened for

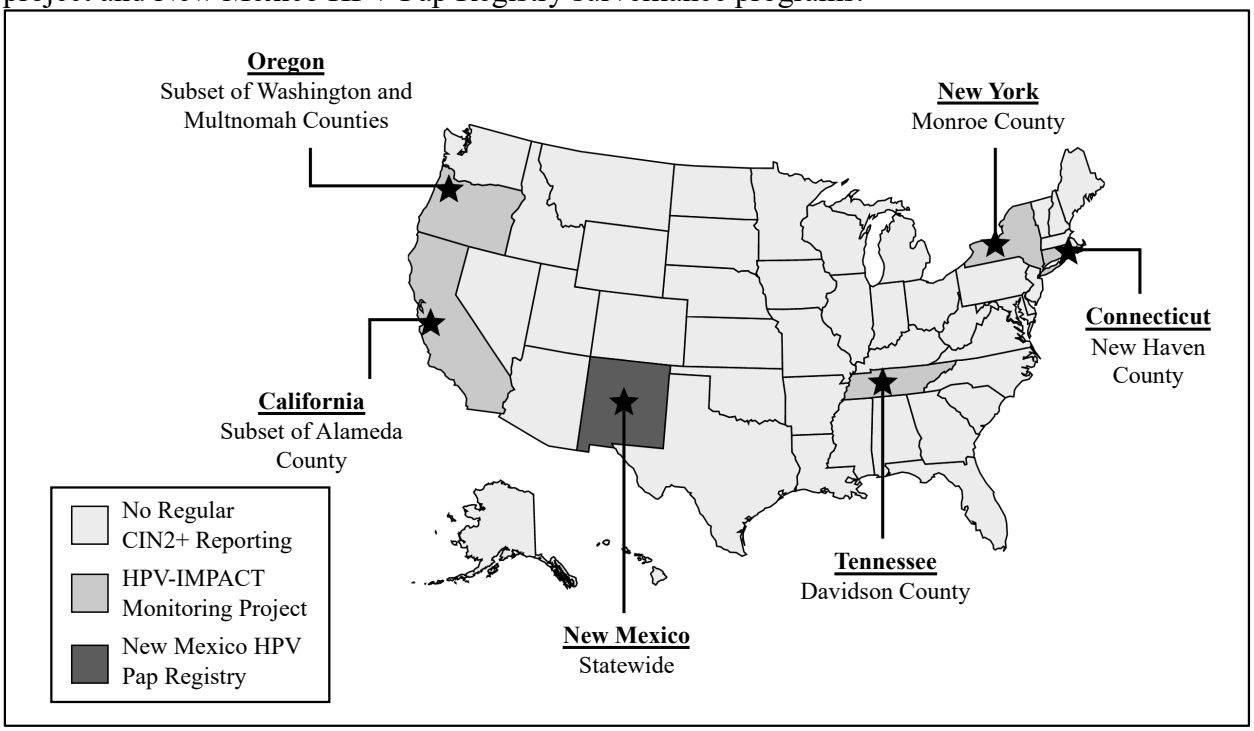
cervical cancer and adenocarcinoma is not easily detected with a Pap test, suggesting the results could indicate early evidence of HPV vaccine impact on reducing cervical cancer incidence.[59]

For now, a more efficient way to examine HPV vaccine impact is assessing intermediate outcomes to cancer, such as persistent HPV infections and CIN2+ (i.e., CIN2, CIN3, and adenocarcinoma *in situ*), which present decades earlier than cancer.[15–17] In the US, only a handful of populations and catchment areas have been able to examine the HPV vaccine's impact on reducing CIN2+ incidence using data from the New Mexico HPV Pap Registry and the Human Papillomavirus Vaccine Impact Monitoring Project (HPV-IMPACT).[19,20] Through these two surveillance programs, regular CIN2+ reporting is only available in six areas across the United States: 1) the state of New Mexico; 2) Monroe County, New York; 3) New Haven County, Connecticut; 4) a subset of Washington and Multnomah Counties, Oregon; 5) a subset of Alameda County, California; and 6) Davidson County, Tennessee (**Figure 2.6**).[19,20]

The New Mexico HPV Pap Registry was launched in 2006, becoming the first US program to capture statewide surveillance of cervical cancer screening and CIN through a fully electronic process.[20] Through this surveillance program, the state of New Mexico began requiring all reporting of cervical screening (Pap and HPV tests) and cervical diagnostic and treatment procedures (including cervical, vulvar, and vaginal pathology) to the New Mexico Notifiable Disease and Conditions. To date, the New Mexico HPV Pap Registry is the only surveillance system in the US to capture cervical screening and CIN since the FDA's initial approval of the HPV vaccine.

In 2008, the Centers for Disease Control and Prevention (CDC) initiated the HPV-IMPACT monitoring project, which consists of collaborations between the CDC, academic institutions, and Emerging Infections Programs at five state health departments in New York,

Figure 2.6. Catchment areas with regular CIN2+ reporting from the HPV-IMPACT monitoring project and New Mexico HPV Pap Registry surveillance programs.



Abbreviations: CIN2+ = Cervical premalignant lesions; HPV = Human Papillomavirus; HPV-IMPACT = Human Papillomavirus Vaccine Impact Monitoring Project; Pap = Papanicolaou.

Connecticut, Oregon, California, and Tennessee.[19] The goals of the HPV-IMPACT monitoring project are to assess HPV vaccine impact by monitoring trends in CIN2+ incidence and cervical screening utilization, describe the demographic and clinical characteristics, as well as the prevalence and distribution of HPV genotypes among women with CIN2+, and estimate HPV vaccine effectiveness for women with CIN2+.[19] In each of the participating populations, CIN2+ is a reportable condition using a standardized case definition across all sites, including CIN2, CIN2/3 (features of both CIN2 and CIN3), CIN3, and adenocarcinoma *in situ*. Enhanced reporting, including race/ethnicity, insurance status, cervical cancer screening history, and vaccination history is conducted for CIN2+ events among women aged 18-39 years and a tissue specimen is sent to CDC for HPV type testing.[19] Currently, HPV-IMPACT sites are working to

retrospectively identify cervical carcinomas from 2008 and onward to add to their former CIN2+ case definition.[19]

Challenges of Assessing Trends in Cervical Premalignant Lesion Incidence

Despite efforts to improve CIN2+ surveillance in the US, states without these population-based surveillance systems in place are unable to accurately monitor CIN2+ trends because CIN2+ tissue confirmation requires cervical biopsy data and surveillance of population-based cervical biopsies. In the US, this surveillance is only available through the New Mexico HPV Pap Registry and the HPV-IMPACT monitoring project. The typical screening process for a CIN2+ diagnosis begins with an abnormal result from a cytologic-based cervical screening test (i.e., Pap test) (**Table 2.2**), or a positive test result from a primary HPV DNA test (indicating the presence of a high-risk HPV type that is linked to cancer), or an abnormal test result from co-testing. Following an abnormal test result, patients are likely asked to return for a follow-up cervical diagnostic procedure, such as a colposcopy with a biopsy (most common next step) using an endocervical curettage, punch biopsy, or conization, to obtain tissue confirmation.[47,95] However, if the initial screening shows a high possibility of cancer, treatment might be needed right away.[47,95] Cervical biopsies following an initial abnormal result are important to check the cervical cells or tissues under a microscope and histologically confirm the final diagnosis because some cytologic screening results may be unclear. Thus, cervical biopsy data is vital to accurately monitor CIN2+ trends.

One potential resource to obtain CIN2+ data is through administrative billing claims databases; however, claims-based models for estimating population-based CIN2+ incidence have not yet been validated for public health surveillance purposes. Based on literature searches, one

Table 2.2. Possible abnormal Pap test results from cytology-based screening.

Test Result	Abbreviation	Explanation
Atypical Squamous Cells of Undetermined Significance	ASCUS	Some cells were detected that do not look completely normal; reasons may vary, including HPV infection, irritation, yeast infections, polyps, benign cysts, menopause, or changes in hormones during pregnancy.
Atypical Glandular Cells	AGC	Some abnormal glandular cells were detected.
Low-Grade Squamous Intraepithelial Lesions	LGSIL	Low-grade changes or mild dysplasia of cervical cells were detected; likely due to HPV infection; sometimes referred to as CIN1 but needs a biopsy for confirmation.
Atypical Squamous Cells, Cannot Exclude HGSIL	ASCH	Some abnormal squamous cells were detected that might be a high-grade squamous intraepithelial lesion, but diagnosis is unclear.
High-Grade Squamous Intraepithelial Lesions	HGSIL	High-grade changes or moderate to severe dysplasia of cervical cells were detected; most likely due to HPV infection; sometimes referred to as CIN2, CIN2/3, or CIN3 but needs a biopsy for confirmation; may turn into cervical cancer if left untreated.
Adenocarcinoma *in situ*	AIS	An advanced lesion or an area of abnormal growth in the glandular tissue of the cervix was detected; may turn into cervical cancer if left untreated.
Squamous Cell Carcinoma	SCC	Cervical cancer cells were detected.

Abbreviations: CIN1 = Cervical Intraepithelial Neoplasia Grade 1; CIN2 = Cervical Intraepithelial Neoplasia Grade 2; CIN2/3 = Features of both CIN2 and CIN3; CIN3 = Cervical Intraepithelial Neoplasia Grade 3; HPV = Human Papillomavirus; Pap = Papanicolaou.

study published in 2014 validated administrative algorithms (e.g., combinations of at least one or two diagnosis codes and one or two procedure codes), for identifying patients with high-grade squamous intraepithelial lesions, CIN2, CIN3, and cervical cancer all together as one outcome.[96] The algorithms were developed using International Classification of Diseases (ICD), 9th revision (ICD-9) claims codes among women aged 20-60 years who had outpatient codes for an abnormal Pap test from 2007-2009 in the Partners' Research Patient Data Registry from a Boston-based non-profit health organization (N = 24,426).[96] These algorithms were tested in the Harvard Pilgrim Health Care claims database among women aged 20-60 years and then validated using a linked electronic health record system.[96] Because the validation method used chart analyses among the algorithm-identified cases, only positive predictive value could be assessed with no sensitivity and specificity calculations; thus, many true events in the population could have been missed. Using positive predictive value alone to validate models may underestimate disease burden and is dependent on the population's disease prevalence.

One main challenge of using claims data to assess long-term CIN2+ trends in the US is determining how to handle the major coding transition from ICD-9 to the International Classification of Diseases, 10th revision (ICD-10) on October 1st, 2015. The new ICD-10 coding scheme has fundamental structural differences from ICD-9, with nineteen times as many procedural codes and five times as many diagnostic codes in ICD-10 than ICD-9.[21,97] Further, conversions between coding schemes are not always direct. For instance, a single ICD-9 code could be mapped to several ICD-10 codes or vice versa.

Only a few studies have evaluated the effect of the ICD-9 to ICD-10 coding transition on the validity and classification accuracy of claims data to identify health outcomes between the two ICD eras (**Table 2.3**).[98–102] Among these studies, conclusions have been mixed, with some

Table 2.3. Summary of US studies examining changes in the validity and classification of health outcomes from ICD-9 to ICD-10.

Study	Outcome	Conclusion
Slavova *et al.* (2018)[98]	Injury hospitalization outcomes (external cause of injury, injury intent, injury mechanism)	• Overall smooth transition in classification of external cause of injury from ICD-9 to ICD-10; • Significant changes in trends for some classifications of injury intent (unintentional and undetermined intent) and injury mechanism (poisoning, suffocation, struck by/against, transportation, and unspecified mechanism) from ICD-9 to ICD-10
Inscore *et al*. (2018)[99]	Acute injury among military personnel (e.g., traumatic brain injury, burns, factures, sprains, etc.)	• Overall similar broad classifications of acute injuries between ICD-9 to ICD-10 (injuries on the head and neck, spine and back, torso, etc.); • Differences in granular classifications (e.g., injury to sacrum/coccyx, pelvic organs, etc.) from ICD-9 to ICD-10
Panozzo *et al*. (2018)[100]	Acute myocardial infarction, angioedema, ischemic stroke, diabetes, hypertension	• Incidence and prevalence were similar for acute myocardial infarction and hypertension from ICD-9 to ICD-10; • Inconsistent trends for angioedema, ischemic stroke, and diabetes from ICD-9 to ICD-10
Salemi *et al*. (2019)[101]	Birth defects	• Most (33 of 46) birth defects had similar prevalence from ICD-9 to ICD-10; • Some (13 of 46) birth defects had significant changes in prevalence from ICD-9 to ICD-10, with 5 defects significantly decreasing and 8 defects significantly increasing immediately after the ICD-9 to ICD-10 transition
Sarayani *et al*. (2020)[102]	Pregnancy episodes	• Reasonable consistency and relatively stable trends in the identification of pregnancy episodes from ICD-9 to ICD-10

Abbreviations: ICD-9 = International Classification of Diseases, 9th revision; ICD-10 = International Classification of Diseases, 10th revision; US = United States.

health outcomes having consistent classifications between ICD-9 and ICD-10 eras (e.g., external cause of injury, broad definitions of acute injuries, acute myocardial infarction, hypertension, most birth defects, and pregnancy episodes), and other health outcomes having significant changes immediately after the coding transition (e.g., injury intent, injury mechanism, granular definitions of acute injuries, angioedema, ischemic stroke, diabetes, and some birth defects).

Because the ICD-9 to ICD-10 transition has impacted classification accuracy differently depending on the health outcome, studies specifically focused on examining the validity of ICD-9 and ICD-10 codes for identifying CIN2+ events are important for improving future long-term CIN2+ surveillance. To our knowledge, no studies have compared the discriminative ability between ICD-9 and ICD-10 codes for identifying CIN2+ events. Without understanding the impact of the coding transition on classifying CIN2+ events, whether observed changes in CIN2+ incidence in future epidemiologic studies examining long-term CIN2+ trends are confounded by changes in coding schemes is unclear.

Leveraging Administrative Billing Claims Data

Despite the challenge of addressing the period discontinuity between the ICD-9 and ICD-10 eras, claims data may still be leveraged to provide adequate surrogate metrics for assessing CIN2+ trends in future surveillance studies. Administrative claims data are collected by organizations, such as healthcare and insurance systems, for record-keeping on billing, procedures, registration information, and more. For these reasons, claims data are rich sources of information on large groups of people under a common system. The first ICD coding system, known as the International List of Causes of Death, was developed in 1893 so countries could

share mortality data with each other.[103] Since then, several revisions have been implemented and present-day codes not only represent causes of death, but also medical diagnoses and procedures.

Presently, several administrative coding sets are widely used in the US, including the ICD Clinical Modification (ICD-CM), ICD Procedure Coding System (ICD-PCS), Healthcare Common Procedure Coding System (HCPCS), and Current Procedure Terminology (CPT). ICD-CM codes are developed and maintained by the CDC and is used by all US providers to report medical diagnoses.[104] ICD-PCS codes are developed and maintained by the Centers for Medicare and Medicaid Services and is used to report inpatient procedures in the US.[104] HCPCS codes are divided into two levels, both of which are used to report medical procedures. Level I HCPCS codes are identical to CPT codes, which are developed and maintained by the American Medical Association; these codes are used by hospital providers for ambulatory and outpatient procedures only.[104] Level II HCPCS codes are developed by the Centers for Medicare and Medicaid Services and are used by hospitals, physicians, and other health care professionals who bill to Medicare and Medicaid to report procedures that are not covered under CPT/Level I HCPCS and other services and equipment, such as drugs, prosthetics, medical devices, etc.[104]

Due to the standardized nature of diagnostic and procedural codes used for health care billing claims, administrative databases are useful tools for epidemiologic studies to systematically classify health outcomes and patient services across insurance databases. Therefore, with proper validation of claims data for identifying CIN2+ events in both ICD-9 and ICD-10 eras, future surveillance studies examining HPV vaccine impact on reducing cervical premalignant lesions in areas without access to population-based cervical biopsy data is possible.

CHAPTER III

IMPROVING SURVEILLANCE OF CERVICAL PREMALIGNANT LESIONS WITH ADMINISTRATIVE DATA: ASSESSING THE VALIDITY OF CLAIMS-BASED PREDICTION MODELS IN ICD-9 AND ICD-10 ERAS*

*Portions of this chapter have been published in Shing *et al*., Improving cervical precancer surveillance: Validity of claims-based prediction models in ICD-9 and ICD-10 eras, The Journal of the National Cancer Institute Cancer Spectrum, 2020, Volume 5, Issue 1[105]; permission to reproduce these portions have been granted by Oxford University Press and my co-authors, Marie R. Griffin, Linh D. Nguyen, James C. Slaughter, Edward F. Mitchel, Manideepthi Pemmaraju, Alyssa B. Rentuza, and Pamela C. Hull.

Abstract

Capturing cervical premalignant lesions, including cervical intraepithelial neoplasia grades 2, 3, and adenocarcinoma *in situ* (CIN2+) requires cervical biopsy information not included in most US cancer registries. Billing codes could provide surrogate metrics; however, the 2015 transition in International Classification of Diseases, ninth (ICD-9) to tenth (ICD-10) revision disrupts trends. We built, validated, and compared claims-based prediction models to identify CIN2+ events in both ICD eras. A database of Davidson County, Tennessee, pathology-confirmed CIN2+ events from the Human Papillomavirus Vaccine Impact Monitoring Project (HPV-IMPACT) provided gold standard events. Using Tennessee Medicaid, 2008-2017 cervical diagnostic procedures (N = 8,549) among Davidson County women aged 18-39 years were

randomly split (60/40 training/testing). Relevant diagnosis, procedure, and screening codes were used to build models from: 1) CIN2+ tissue diagnosis codes alone, 2) least absolute shrinkage and selection operator (LASSO), and 3) random forest classifiers. Model-classified index events were counted to estimate incident events. From 2008 to 2017, HPV-IMPACT confirmed 983 incident CIN2+ events among Tennessee Medicaid-enrolled, Davidson County women. Claims-based models identified 1,007 (LASSO), 1,245 (CIN2+ tissue diagnosis codes alone), and 957 (random forest) incident events. LASSO performed well in ICD-9 and ICD-10 eras: 77.3% (95% Confidence Interval [CI] = 72.5%-81.5%) versus 81.1% (95% CI = 71.5%-88.6%) sensitivity, 93.0% (95% CI = 91.9%-94.0%) versus 90.2% (95% CI = 87.2%-92.7%) specificity, 61.3% (95% CI = 56.6%-65.8%) versus 60.3% (95% CI = 51.0%-69.1%) positive predictive value, 96.6% (95% CI = 95.8%-97.3%) versus 96.3% (95% CI = 94.1%-97.8% negative predictive value, 91.0% (95% CI = 89.9%-92.1%) versus 88.8% (95% CI = 85.9%-91.2%) accuracy, 85.1% (95% CI = 82.9%-87.4%) versus 85.6% (95% CI = 81.4%-89.9%) C-indices, respectively; performance did not statistically significantly differ between eras (95% CIs all overlapped). Results confirmed model utility with good performance across both ICD eras for CIN2+ surveillance. Validated claims-based models may be used in future CIN2+ trend analyses to estimate HPV vaccine impact where population-based biopsies are unavailable.

Introduction

The human papillomavirus (HPV) vaccine's impact on cervical premalignant lesions, including cervical intraepithelial neoplasia (CIN) grades 2 and 3 and adenocarcinoma *in situ* (together referred to as CIN2+) may be observed sooner than the vaccine's impact on cervical cancer.[12] The HPV vaccine can prevent nearly 80% of CIN2+,[92] and preventing cervical premalignant lesions will ultimately prevent cervical cancer and its associated premature mortality.[28] Additionally, premalignant lesions are associated with considerable preventable morbidity and costs.[6,106,107] In the United States (US), 196,000 CIN2+ events were diagnosed in 2016.[92] Despite declines from 216,000 CIN2+ events in 2008,[92] the HPV vaccine is not yet reaching its full potential. HPV vaccination lags behind other recommended adolescent vaccines, including tetanus, diphtheria, and pertussis, and meningococcal conjugate vaccines, and there is substantial variation in vaccination rates across states.[18] Monitoring trends in CIN2+ is critical for evaluating the impact of HPV vaccination over time and targeting vaccine promotion and cervical cancer screening efforts.

Examining CIN2+ in the US is challenging. CIN2+ diagnosis confirmation requires cervical biopsies, which are not included in most US cancer registries or surveillance systems. Several states have monitored CIN2+ rates through the state-based Pap registry in New Mexico and the population-based Human Papillomavirus Vaccine Impact Monitoring Project (HPV-IMPACT) in five states.[87–92,94] However, the vast majority of states do not have such surveillance capacity, so it is not possible to examine national CIN2+ trends or variation across states.

A potential solution is leveraging administrative data using International Classification of Diseases, Clinical Modification (ICD), Current Procedural Terminology (CPT), and Healthcare Common Procedure Coding System (HCPCS) codes, which systematically classify diseases and

patient procedures, providing surrogate metrics. However, the transition from the ninth (ICD-9) to tenth (ICD-10) coding revision in 2015[21] disrupts trends since ICD-9 and ICD-10 codes differ in structure.[21,97] Compared to ICD-9, ICD-10 codes have more detail about laterality, severity, and complexity of health conditions, allowing for increased specificity and accuracy.[97]

Despite the need to expand options for the surveillance of CIN2+ incidence, limited information is available on the validity of claims data for identifying incident CIN2+ events between ICD-9 and ICD-10 eras. To our knowledge, no studies have validated claims-based CIN2+ models that can detect trends in CIN2+ across both ICD eras. While such models are not intended to provide the highest accuracy that would be needed for clinical decision-making, they would be useful for detecting trends in public health surveillance. To address this gap, we aimed to build and validate claims-based models identifying CIN2+ events in ICD-9 and ICD-10 eras as a method to estimate the number of CIN2+ events in the population, and we compared three model building approaches to identify an optimal model. In addition, to provide insight into unifying period continuity across ICD-9 and ICD-10 eras for future trend analyses of HPV vaccine impact, we compared model performance between the two ICD eras.

Methods

Study Population

Billing codes from the Tennessee Medicaid program (TennCare) identified women with cervical diagnostic procedural encounters from 2008 to 2017 who were TennCare-enrolled at the time of procedure (**Table 3.1**). We included women aged 18-39 years residing in Davidson

Table 3.1. Administrative codes and groupings used for study population inclusion criteria and potential predictors of CIN2+ event status.

Coding System	Code	Code Description
Cervical Diagnostic Procedures		
Cervical Diagnostic Procedure Codes		
CPT	57420-57421	Colposcopy of the entire vagina with and without biopsy
	57450	Colposcopy of the cervix including upper/adjacent vagina; with loop electrode biopsy of the cervix
	57452	Colposcopy of the cervix including upper/adjacent vagina
	57454	Colposcopy of the cervix, with biopsy of the cervix and endocervical curettage
	57455	Colposcopy of the cervix, with biopsy
	57456	Colposcopy of the cervix, with endocervical curettage
	57460	Colposcopy of the cervix, with loop electrode biopsy
	57461	Colposcopy of the cervix, with loop electrode conization
	57500	Cervical biopsy, single or multiple, or local excision of lesion, with or without fulguration
	57505	Endocervical curettage
	57520	Cervical conization
	57522	Loop excision
	58110	Endometrial biopsy
Cervical Screening Tests		
Human Papillomavirus Screening Test Codes		
ICD-9	V73.81	Encounter for screening for HPV
ICD-10	Z11.51	Encounter for screening for HPV
Pap Smear/Test Codes		
ICD-9	V72.31	Routine gynecological examination with or without Papanicolaou cervical smear
	V72.32	Pap smear to confirm findings of recent normal smear following initial abnormal smear
	V76.2	Screening for malignant neoplasm of cervix (Pap smear) outside of a routine gynecological examination
	V76.47	Vaginal pap smear for confirmation of recent normal following initial abnormal smear done
	795.06	Papanicolaou smear of cervix with cytologic evidence of malignancy
	91.46	Microscopic examination of specimen from female genital tract, cell block and Papanicolaou smear

ICD-10	Z01.411, Z01.419	Routine gynecological examination with or without Papanicolaou cervical smear
	Z01.42	Pap smear to confirm findings of recent normal smear following initial abnormal smear
	Z12.4	Screening for malignant neoplasm of cervix (Pap smear) outside of a routine gynecological examination
	Z12.72	Vaginal pap smear for confirmation of recent normal following initial abnormal smear done
	R87.614	Papanicolaou smear of cervix with cytologic evidence of malignancy
CPT	88141-88145, 88147-88148, 88150-88158, 88164-88167, 88174-88175	Cytology testing of the vagina and/or cervix
HCPCS	P3000-P3001, G0101, G0123-G0124, G0141, G0143-G0145, G0147-G0148, Q0091	Cytology testing of the vagina and/or cervix
Human Papillomavirus DNA Test Codes		
ICD-9	795.05, 795.09	Infectious agent detection by nucleic acid (DNA or RNA); papillomavirus, human, direct/amplified, quantification/probe technique
ICD-10	R87.10, R87.820	Infectious agent detection by nucleic acid (DNA or RNA); papillomavirus, human, direct/amplified, quantification/probe technique
CPT	87620-87622, 87623-87625	Infectious agent detection by nucleic acid (DNA or RNA); papillomavirus, human, direct/amplified, quantification/probe technique
Cervical-Related Diagnoses		
CIN2+ Tissue Diagnosis Codes		
ICD-9	233.1	Cervical intraepithelial neoplasia III
	622.12	Cervical intraepithelial neoplasia II
ICD-10	D06.0, D06.1, D06.7, D06.9, N87.2	Cervical intraepithelial neoplasia III
	N87.1	Cervical intraepithelial neoplasia II

Non-Specific CIN Tissue Diagnosis Codes		
ICD-9	622.10	Dysplasia of cervix, unspecified
ICD-10	N87.9	Dysplasia of cervix, unspecified
High-Grade Cervical Intraepithelial Lesion Cytology Diagnosis Codes		
ICD-9	795.04	High grade squamous intraepithelial lesion on cytologic smear of cervix
ICD-10	R87.613	High grade squamous intraepithelial lesion on cytologic smear of cervix
CIN1 Tissue Diagnosis Codes		
ICD-9	622.11	Cervical intraepithelial neoplasia I
ICD-10	N87.1	Cervical intraepithelial neoplasia I
Low-Grade Cervical Intraepithelial Lesion Cytology Diagnosis Codes		
ICD-9	795.03	Low grade squamous intraepithelial lesion on cytologic smear of cervix
ICD-10	R87.612	Low grade squamous intraepithelial lesion on cytologic smear of cervix
Atypical Squamous Cells of Undetermined Significance Diagnosis Codes		
ICD-9	795.01	Atypical squamous cells of undetermined significance on Pap test
	795.02	Atypical squamous cells cannot exclude high grade squamous intraepithelial lesion on Pap test/cytologic smear of cervix
ICD-10	R87.610	Atypical squamous cells of undetermined significance on Pap test
	R87.611	Atypical squamous cells cannot exclude high grade squamous intraepithelial lesion on Pap test/cytologic smear of cervix
Cervical Treatment Procedures		
Cervical Treatment Procedure Codes		
CPT	57511	Cryotherapy of Cervix
	57510	Electro or thermal cautery of cervix
	57513	Laser ablation
	57530–57531	Trachelectomy or Cervicectomy
	57540, 57545, 57550, 57555, 57556	Excision of cervical stump
	57520	Cervical conization
	57522	Loop excision

Cervical or Vaginal Biopsy Codes		
CPT	57421, 57450,	Colposcopy of the entire vagina with biopsy
	57454, 57455,	Colposcopy of the cervix including upper/adjacent vagina; with loop electrode biopsy of the
	57460,	cervix
	57500,	Colposcopy of the cervix, with biopsy of the cervix and endocervical curettage
	58110	Colposcopy of the cervix, with biopsy
		Colposcopy of the cervix, with loop electrode biopsy
		Cervical biopsy, single or multiple, or local excision of lesion, with or without fulguration
		Endometrial biopsy

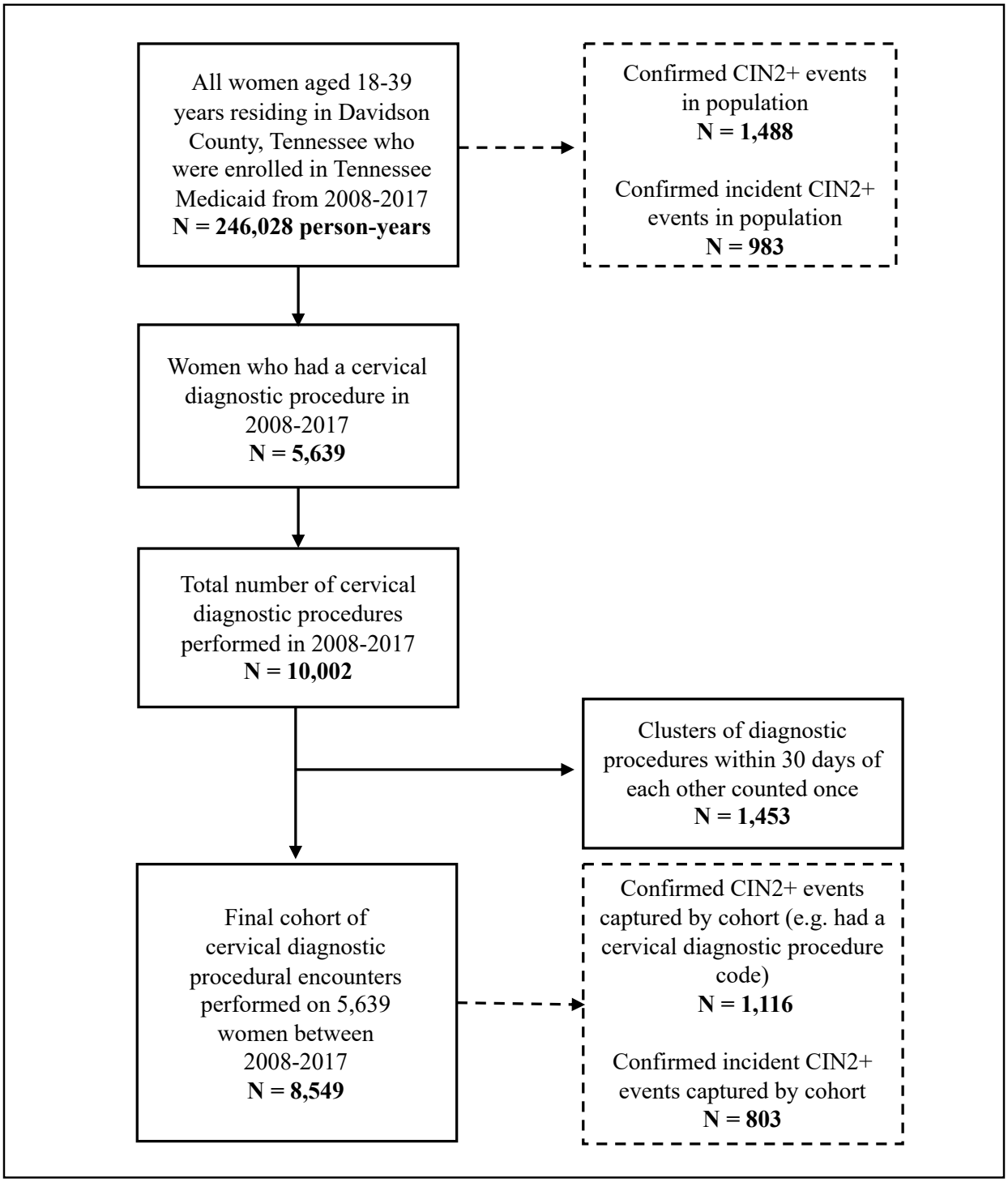
Abbreviations: CIN = Cervical Intraepithelial Lesion; CPT = Current Procedural Terminology; DNA = Deoxyribonucleic Acid; HCPCS = Healthcare Common Procedure Coding System; HPV = Human Papillomavirus; ICD = International Classification of Diseases; Pap = Papanicolaou; RNA = Ribonucleic Acid.

County, Tennessee, because our gold standard dataset for validation had the same age and geographic inclusion (**Figure 3.1**). Although young women aged 18-21 years are no longer recommended for cervical cancer screening, prior to 2012, screening was still recommended for women after sexual debut if this occurred before age 21 years (**Refer to Chapter II, Section: "Cervical Cancer Prevention: Screening and The HPV Vaccine")**; therefore, examining CIN2+ incidence in younger ages could still be beneficial. We counted encounters rather than women to account for women with multiple encounters. Procedures within 30 days of each other were considered clusters of associated procedures and counted as one encounter; 1,453 clusters were identified among 10,002 total procedures in 2008-2017 (final sample = 8,549 encounters). This research was approved by the Division of TennCare and deemed public health surveillance, thereby exempt by the Tennessee Department of Health and Vanderbilt University Institutional Review Boards.

Gold Standard

Biopsy-confirmed CIN2+ events, including CIN2, CIN3, and adenocarcinoma *in situ*, in Davidson County, Tennessee were collected and validated by the HPV-IMPACT team at Vanderbilt University Medical Center as part of the HPV-IMPACT monitoring project,[19] a nationally funded program consisting of partnerships between the Centers for Disease Control and Prevention, academic institutions, and Emerging Infections Programs in five state health departments.[19] Since 2008, HPV-IMPACT has conducted enhanced surveillance on CIN2+ events among women aged 18-39 years in select catchment areas across the United States, including Davidson County, Tennessee, making CIN2+ a reportable disease in Tennessee.

Figure 3.1. Flow diagram to capture cohort of cervical diagnostic procedural encounters from 2008 to 2017 among TennCare-enrolled women aged 18-39 years residing in Davidson County, Tennessee.



Abbreviations: CIN = Cervical Intraepithelial Lesion; TennCare = Tennessee Medicaid.

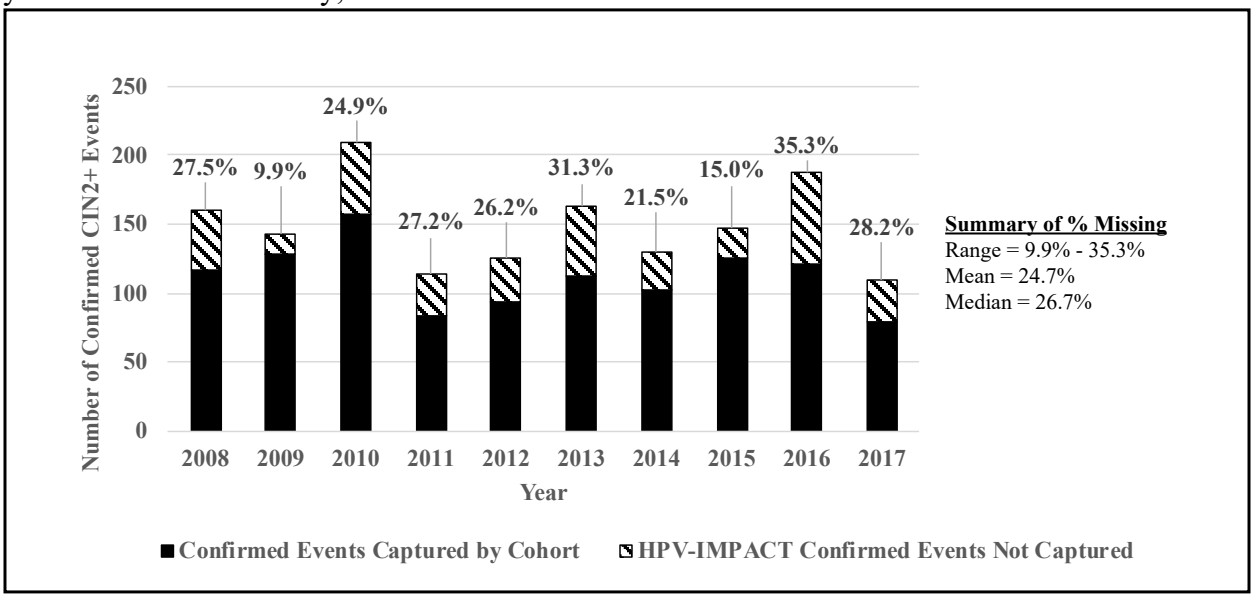
The HPV-IMPACT team receives reports from pathology laboratories serving Davidson County, and reviews charts of women with pathologically-confirmed CIN2+ to assure these women were Davidson County residents at the time of biopsy and that the biopsy reflected an incident event. Records of women with cervical biopsies identified through administrative databases, including TennCare, the Hospital Discharge Data System, and Ambulatory Surgery Treatment Center, are also audited to assure all CIN2+ events are captured in the HPV-IMPACT surveillance. The Hospital Discharge Data System includes data on hospital-based inpatient and outpatient surgical procedures, while the Ambulatory Surgery Treatment Center includes data on non-hospital outpatient procedures. Among the HPV-IMPACT confirmed incident CIN2+ events of women enrolled in TennCare at the time of their diagnosis, the TennCare audits identified an additional 4-20% incident CIN2+ events annually from 2008 to 2017, which were then added to the final number of gold standard confirmed events (**Appendix Table A1**). Altogether, from 2008 to 2017, HPV-IMPACT identified a total of 1,488 CIN2+ events among TennCare-enrolled women aged 18-39 years residing in Davidson County, of which 983 were incident events (**Refer to Figure 3.1**).

In our analytic sample, encounters were considered confirmed events if the diagnostic procedure was from a woman with an HPV-IMPACT confirmed CIN2+ event. We found the interval between these women's diagnostic procedure dates and their closest HPV-IMPACT confirmed diagnosis date ranged from 0 to 3,131 days (median = 28 days). Therefore, we used pre-determined conservative parameters to associate diagnoses with their most probable corresponding procedures. Encounters were only considered confirmed events if the confirmed diagnosis date was within +/-60 days of procedure date or within 60 days before the first diagnostic procedure in a cluster (procedures within 30 days of each other) and 60 days after the

last diagnostic procedure in the cluster. Given the inclusion criteria and pre-specified parameters, 1,116 confirmed events were captured in our final sample of cervical diagnostic procedures among TennCare-enrolled women aged 18-39 years residing in Davidson County, of which 803 were incident events (**Refer to Figure 3.1**).

In breaking down the HPV-IMPACT confirmed events captured by year, we discovered 10%-35% of HPV-IMPACT confirmed CIN2+ events were missed by our inclusion criteria (i.e., CPT codes for a cervical diagnostic procedure among TennCare-enrolled women) annually from 2008 to 2017 (**Figure 3.2**). We performed Pearson's chi-squared trend tests for proportions and linear regression to determine whether the proportion of HPV-IMPACT confirmed CIN2+ events missed by our inclusion criteria was associated with year. To diagnose reasons for missingness,

Figure 3.2. Annual percent of HPV-IMPACT confirmed CIN2+ events missing from cohort of cervical diagnostic procedures from 2008 to 2017 among TennCare-enrolled women aged 18-39 years in Davidson County, Tennessee.



Abbreviations: CIN = Cervical Intraepithelial Lesion; TennCare = Tennessee Medicaid.

we abstracted a subset of missed gold standard events from one health center for which we had access to medical charts (n = 68) and examined each woman's medical procedures and insurance status.

Predictors

From each diagnostic procedure date, we used the same interval parameters for determining confirmed event status to search for presence of ICD-9, ICD-10, CPT, and HCPCS codes relating to a 1) CIN2+ tissue diagnosis, 2) non-specific CIN tissue diagnosis, 3) high-grade squamous intraepithelial lesion (HGSIL) cytology diagnosis, 4) CIN1 tissue diagnosis, 5) low-grade squamous intraepithelial lesion (LGSIL) cytology diagnosis, 6) atypical squamous cells of undetermined significance (ASCUS) diagnosis, 7) HPV screening test, 8) Papanicolaou (Pap) smear/test, 9) HPV deoxyribonucleic acid (DNA) test, 10) cervical treatment procedure, and 11) cervical or vaginal biopsy. We consulted an expert panel to determine appropriate groupings for each predictor using either a single code or combination of codes (**Refer to Table 3.1**).

Model Building

The data were randomly split into 60% training and 40% testing sets by era. To assess ICD-10 coding implementation lag for the transition cut-off (October 1, 2015), we examined crossover usage of ICD-9 codes after September 30, 2015, and ICD-10 codes before October 1, 2015. Only 16 of 1,444 (1.11%) encounters used an ICD-9 code after September 30, 2015, and 5 of 7,105 (0.07%) encounters used an ICD-10 code before October 1, 2015. Because crossover was minimal, we retained the original cut-off; ICD-9 and ICD-10 eras consisted of encounters

during January 1, 2008-September 30, 2015, and October 1, 2015-December 31, 2017, respectively.

To determine which method provides an optimal model, we built models identifying CIN2+ events using three distinct algorithms:

1. A pre-specified set of CIN2+ tissue diagnosis codes alone, a method used by a prior claims-based study[22] that classified CIN2+ event status using ICD-9 codes for a specific CIN2+ tissue diagnosis—622.12 (CIN2) and 233.1 (CIN3). To identify CIN2+ events in the ICD-10 era, we mapped the ICD-9 codes used by Flagg *et al*. to corresponding ICD-10 codes (N87.1, N87.2, D06.0, D06.1, D06.7, and D06.9).
2. Least Absolute Shrinkage and Selection Operator (LASSO) using logistic regression, a machine learning method to build a parsimonious model when there are correlated predictors by simultaneously conducting variable selection and regularization to enhance prediction ability.[108] We assessed the distribution of predicted probabilities between gold standard events versus non-events and the classification performance at cut-off values of 0.3, 0.5, and 0.6 to determine an appropriate cut-off value for our model. As noted elsewhere,[109] an adequately performing model should have minimal overlap of predicted probabilities between gold standard events and non-events.
3. Random forest classifiers, a machine learning method that creates and averages several bootstrapped decision trees with various predictors and cut-off values to reduce overfitting and improve accuracy.[110] We conducted parameter tuning for random forest algorithms using a randomized search method, which tested various combinations of number of trees, maximum predictor selection methods, maximum tree depths, minimum number of samples

for a split, and minimum number of samples in a leaf node. The final random forest parameters were selected based on the highest mean validation score.

Models derived from the LASSO and random forest algorithms were built using ICD-9 and ICD-10 training sets combined, creating a uniform model across eras. Correlation matrices were built in R (R core team, Vienna, Austria) to confirm selected predictors were not highly correlated. LASSO was trained in Stata 16 (StataCorp, College Station, TX). Random forest was trained in Python 3.7.4 using the RandomForestClassifiers function from the scikit-learn package.

Model Comparison and Validation

We examined bivariate associations of demographic and coding characteristics of cervical diagnostic procedures between ICD eras using two-sided Pearson's chi-squared tests. We also assessed bivariate associations between coding characteristics and CIN2+ status (confirmed CIN2+ event versus non-event) among the overall sample and by ICD era using two-sided Pearson's chi-squared tests. To compare concordance between each model building methodology, we calculated percent agreement and Cohen's kappa statistics. For bivariate and concordance tests, p-values less than 0.05 were considered statistically significant.

Confusion matrices determined true positives, false positives, false negatives, and true negatives. To assess apparent validity, defined as model performance among samples used to develop the models,[111] we examined discrimination and calibration of LASSO and random forest models among training sets, by era. Discrimination was assessed using six performance measures: 1) sensitivity, 2) specificity, 3) positive predictive value (PPV), 4) negative predictive value (NPV), 5) accuracy, and 6) C-index. Calibration was assessed using calibration plots.

Apparent validity was not assessed for CIN2+ tissue diagnosis codes alone because this method was not trained. We also examined discrimination and calibration in testing sets, by era. Because CIN2+ diagnosis trends differ across ages, we also assessed model performance by age group (18-24, 25-29, 30-39 years) and ICD era among testing sets. We considered assessing model performance in women aged 18-20 years to capture differences in screening recommendation among this age group; however, the sample size for the numerator (i.e., CIN2+ events) was too small to conduct such analyses. Therefore, women aged 18-24 years were combined to allow for more power in our age-group-stratified analyses.

Binomial 95% confidence intervals (CIs) were calculated for all six performance measures using the Clopper-Pearson Exact method to test for statistically significant differences in model performance between the ICD eras. We assessed generalizability by comparing performance in training and testing sets by model and era; comparisons were considered statistically significant if 95% CIs did not overlap. Lastly, to determine an optimal model for public health surveillance (i.e., appropriate for examining trends over time), we counted annual index events classified by each model and compared model-identified index events with HPV-IMPACT's confirmed annual number of incident CIN2+ events in the population.

Sensitivity Analyses

The study[22] on which we based our first model (CIN2+ tissue diagnosis codes alone) was restricted to women who were screened for cervical cancer. Therefore, we replicated our methods among a cohort of women with cervical screening tests to assess differences in administrative code patterns and model performance between women with cervical diagnostic procedures (our study population) versus those with cervical screening tests (Flagg *et al*.'s[22]

study population). Of note, twenty-five screening codes from our identified list were not used by Flagg *et al*.,[22] including ICD-9 Codes: V73.81, V72.31, 795.06, 795.05, 795.09, 91.46; ICD-10 Codes: Z11.51, Z01.411, Z01.419, Z01.42, Z12.4, Z12.72, R87.6141, R87.10, R87.820; and CPT Codes: 88144, 88145, 88151, 88155-88158, 87623-87625. For comprehensiveness rather than pure replication, we used our longer list of screening codes to create a similar cohort of screened women that could potentially capture missed events. Subsequently, we built and validated three claims-based models built by CIN2+ tissue diagnosis codes alone, LASSO, and random forest classifiers.

For each cervical screening test encounter, we searched within +/- 365 days of the screening date or within 365 days before the first screening test in a cluster (i.e., tests within 365 days of each other) and 365 days after the last screening test in the cluster for the presence of administrative codes relating to the following predictors: a 1) CIN2+ tissue diagnosis, 2) non-specific CIN tissue diagnosis, 3) HGSIL cytology diagnosis, 4) CIN1 tissue diagnosis, 5) LGSIL cytology diagnosis, 6) atypical squamous cells of undetermined significance diagnosis, 7) cervical treatment procedure, 8) specific biopsy procedure, and 9) cervical diagnostic procedure. To identify screening test encounters that resulted in a confirmed CIN2+ event, we determined whether the encounter was performed on a woman with an HPV-IMPACT confirmed CIN2+ event. Screening tests were only considered confirmed events if the confirmed diagnosis date was within 365 days after the screening date or within 365 days before the first screening test in a cluster and 365 days after the last screening test in the cluster.

Results

Characteristics of Cervical Diagnostic Procedures

We identified 5,639 TennCare-enrolled women aged 18-39 years residing in Davidson County, Tennessee, with a total of 8,549 (ICD-9 = 7,105; ICD-10 = 1,444) cervical diagnostic procedures from 2008 to 2017 (**Table 3.2**). In the ICD-9 era, 885 of 7,105 (12.5%) confirmed CIN2+ events occurred among women with cervical diagnostic procedures compared to 231 of 1,444 (16.0%) in the ICD-10 era ($p < 0.001$). The Pearson's chi-squared trend test for proportions found that the annual proportion of HPV-IMPACT confirmed CIN2+ events missed by our inclusion criteria increased by 0.8% per year ($p = 0.04$); however, after regressing the proportion of events missed on year, the linear trend was no longer significant ($p = 0.40$). When we examined patient medical charts of women with confirmed CIN2+ events who were missed by our inclusion criteria (n = 68), we found 37 (54%) had non-TennCare insurance at the time of their cervical diagnostic procedure, causing them not to be captured by our cohort inclusion criteria. However, these women were insured by TennCare at the time of their diagnosis and thus, captured by HPV-IMPACT as a TennCare-enrolled event and therefore, counted as a confirmed CIN2+ event in the population. The other 31 (45%) events (e.g., those that had TennCare at the time of their procedure but were not captured in our sample) did not have any of the billing codes listed in our inclusion criteria or any codes indicating a cervical procedure.

Compared to the ICD-9 era, a greater proportion of women who had cervical diagnostic procedures in the ICD-10 era were aged 30-39 years (ICD-9 = 27.6% versus ICD-10 = 46.8%) and other/unknown race/ethnicity (ICD-9 = 38.8% versus ICD-10 = 47.8%) ($p < 0.001$) (**Table 3.2**). The proportion of administrative codes used in the ICD-9 era versus ICD-10 era statistically

Table 3.2. Characteristics of cervical diagnostic procedures (N = 8,549) among 5,639 TennCare-enrolled women aged 18-39 years in Davidson County, Tennessee by ICD era[a].

Characteristic	ICD-9 Era N = 7,105 n (column %)	ICD-10 Era N = 1,444 n (column %)	P-Value
Confirmed CIN2+ Event			<0.001*
Yes	885 (12.5)	231 (16.0)	
No	6,220 (87.5)	1,213 (84.0)	
Age Group, years			<0.001*
18-24	3,062 (43.1)	261 (18.1)	
25-29	2,081 (29.3)	507 (35.1)	
30-39	1,962 (27.6)	676 (46.8)	
Race/Ethnicity			<0.001*
NH White	2,011 (28.3)	341 (23.6)	
NH Black	2,184 (30.7)	382 (26.5)	
NH Other/Unknown	2,755 (38.8)	690 (47.8)	
Hispanic	155 (2.2)	31 (2.2)	
CIN2+[b] Tissue Diagnosis Code			<0.001*
Yes	1,508 (21.2)	381 (26.4)	
No	5,597 (78.8)	1,063 (73.6)	
Non-Specific CIN Tissue Diagnosis Code			<0.001*
Yes	808 (11.4)	119 (8.2)	
No	6,297 (88.6)	1,325 (91.8)	
High-Grade Squamous Intraepithelial Lesion Cytologic Diagnosis Code			0.204
Yes	845 (11.9)	189 (13.1)	
No	6,260 (88.1)	1,255 (86.9)	
CIN1 Tissue Diagnosis Code			0.951
Yes	1,831 (25.8)	371 (25.7)	
No	5,274 (74.2)	1,073 (74.3)	
Low-Grade Squamous Intraepithelial Lesion Cytologic Diagnosis Code			0.436
Yes	2,492 (35.1)	491 (34.0)	
No	4,613 (64.9)	953 (66.0)	
Atypical Squamous Cells of Undetermined Significance Diagnosis Code			0.031*
Yes	2,480 (34.9)	547 (37.9)	
No	4,625 (65.1)	897 (62.1)	
Human Papillomavirus Screening Test Code			<0.001*
Yes	173 (2.4)	167 (11.6)	
No	6,932 (97.6)	1,277 (88.4)	
Pap Smear/Test Code			0.017*
Yes	4,987 (70.2)	1,059 (73.3)	

No	2,118 (29.8)	385 (26.7)	
Human Papillomavirus DNA Test Code			0.112
Yes	3,434 (48.3)	731 (50.6)	
No	3,671 (51.7)	713 (49.4)	
Cervical Treatment Procedure Code			0.652
Yes	469 (6.6)	100 (6.9)	
No	6,636 (93.4)	1,344 (93.1)	
Cervical or Vaginal Biopsy Code			<0.001*
Yes	3,140 (44.2)	735 (50.9)	
No	3,965 (55.8)	709 (49.1)	

Abbreviations: CIN = Cervical Intraepithelial Lesion; DNA = Deoxyribonucleic Acid; NH = Non-Hispanic; ICD = International Classification of Diseases, Clinical Modification; TennCare = Tennessee Medicaid.

[a]The ICD-9 era includes procedures from January 1, 2008, through September 30, 2015; the ICD-10 era includes procedures from October 1, 2015, through December 31, 2017.

*Asterisks denote $p < 0.05$.

significantly differed for the following coding groupings: CIN2+ tissue diagnosis, non-specific CIN tissue diagnosis, ASCUS diagnosis, HPV screening test, Pap smear/test, and cervical or vaginal biopsy ($p < 0.05$).

Overall and in both ICD eras, administrative codes associated with cervical diagnostic procedures of women with confirmed CIN2+ events versus non-events were presence of CIN2+ tissue diagnosis, non-specific CIN tissue diagnosis, HGSIL cytologic diagnosis, or cervical or vaginal biopsy codes, and absence of LGSIL cytologic diagnosis or cervical treatment procedure codes (**Table 3.3**; $p < 0.05$). In the ICD-9 era only, coding patterns of cervical diagnostic procedures of women with confirmed CIN2+ events versus non-events were absence of codes for an ASCUS diagnosis, a Pap smear/test, and an HPV DNA test.

Model Building Results

Models were trained using 60% of the total 8,549 encounters, resulting in 5,129 encounters (ICD-9 = 4,263; ICD-10 = 866). Among the training set (N = 5,129), LASSO selected all code groupings as strong independent predictors of CIN2+ events; the strongest individual predictor was having a code for a CIN2+ tissue diagnosis (**Table 3.4,** beta coefficient = 5.34). Other positive predictors included codes for a non-specific CIN diagnosis, HGSIL cytologic diagnosis, LGSIL cytologic diagnosis, ASCUS diagnosis, cervical treatment procedure, or cervical or vaginal biopsy. Negative predictors included codes for a CIN1 tissue diagnosis, HPV screening test, Pap smear/test, and HPV DNA test. Individual predictors were not highly correlated with one another; all correlation coefficients were between -0.2 (LGSIL and ASCUS) and 0.5 (Pap smear/test and HPV DNA test) (**Figure 3.3**).

Table 3.3. Coding characteristics of confirmed CIN2+ events versus non-events among cervical diagnostic procedures of TennCare-enrolled women aged 18-39 years in Davidson County, Tennessee by ICD era[a].

	Overall N = 8,549			ICD-9 Era N = 7,105			ICD-10 Era N = 1,444		
Code Grouping	Confirmed CIN2+ Event n (Col %)	Non-Events n (Col %)	P-Value	Confirmed CIN2+ Event n (Col %)	Non-Events n (Col %)	P-Value	Confirmed CIN2+ Event n (Col %)	Non-Events n (Col %)	P-Value
CIN2+ Tissue Diagnosis			<0.001*			<0.001*			<0.001*
Yes	1,079 (96.7)	810 (10.9)		859 (97.1)	649 (10.4)		220 (95.2)	161 (13.3)	
No	37 (3.3)	6,623 (89.1)		26 (2.9)	5,571 (89.6)		11 (4.8)	1,052 (86.7)	
Non-Specific CIN Tissue Diagnosis			<0.001*			<0.001*			<0.001*
Yes	311 (27.9)	616 (8.3)		254 (28.7)	554 (8.9)		57 (24.7)	62 (5.1)	
No	805 (72.1)	6,817 (91.7)		631 (71.3)	5,666 (91.1)		174 (75.3)	1,151 (94.9)	
High-Grade Squamous Intraepithelial Lesion Cytologic Diagnosis			<0.001*			<0.001*			<0.001*
Yes	413 (37.0)	621 (8.4)		331 (37.4)	514 (8.3)		82 (35.5)	107 (8.8)	
No	703 (63.0)	6,812 (91.7)		554 (62.6)	5,706 (91.7)		149 (64.5)	1,106 (91.2)	
CIN1 Tissue Diagnosis			0.636			0.456			0.663
Yes	281 (25.2)	1,921 (25.8)		219 (24.8)	1,612 (25.9)		62 (26.8)	309 (25.5)	
No	835 (74.8)	5,512 (74.2)		666 (75.3)	4,608 (74.1)		169 (73.2)	904 (74.5)	
Low-Grade Squamous Intraepithelial Lesion Cytologic Diagnosis			<0.001*			<0.001*			<0.001*
Yes	251 (22.5)	2,732 (36.8)		198 (22.4)	2,294 (36.9)		53 (22.9)	438 (36.1)	
No	865 (77.5)	4,701 (63.2)		687 (77.6)	3,926 (63.1)		178 (77.1)	775 (63.9)	
Atypical Squamous Cells of Undetermined Significance Diagnosis			<0.001*			<0.001*			0.089
Yes	265 (23.8)	2,762 (37.2)		189 (21.4)	2,291 (36.8)		76 (32.9)	471 (38.8)	
No	851 (76.3)	4,671 (62.8)		696 (78.6)	3,929 (63.2)		155 (67.1)	742 (61.2)	
Human Papillomavirus Screening Test			0.667			0.916			0.700
Yes	47 (4.2)	293 (3.94)		22 (2.5)	151 (2.4)		25 (10.8)	142 (11.7)	
No	1,069 (95.8)	7,140 (96.1)		863 (97.5)	6,069 (97.6)		206 (89.2)	1,071 (88.3)	
Pap Smear/Test			<0.001*			<0.001*			0.127
Yes	676 (60.6)	5,370 (72.3)		516 (58.3)	4,471 (71.9)		160 (69.3)	899 (74.1)	
No	440 (39.4)	2,063 (27.8)		369 (41.7)	1,749 (28.1)		71 (30.7)	314 (25.9)	

Human Papillomavirus DNA Test			<0.001*			<0.001*			0.086
Yes	422 (37.8)	3,743 (50.4)		317 (35.8)	3,117 (50.1)		105 (45.5)	626 (51.6)	
No	694 (62.2)	3,690 (49.6)		568 (64.2)	3,103 (49.9)		126 (54.6)	587 (48.4)	
Cervical Treatment Procedure			<0.001*			<0.001*			<0.001*
Yes	275 (24.6)	294 (4.0)		222 (25.1)	247 (4.0)		53 (22.9)	47 (3.9)	
No	841 (75.4)	7,139 (96.0)		663 (74.9)	5,973 (96.0)		178 (77.1)	1,166 (96.1)	
Cervical or Vaginal Biopsy			<0.001*			<0.001*			0.038*
Yes	612 (54.8)	3,263 (43.9)		480 (54.2)	2,660 (42.8)		132 (57.1)	603 (49.7)	
No	504 (45.2)	4,170 (56.1)		405 (45.8)	3,560 (57.2)		99 (42.9)	610 (50.3)	

Abbreviations: CIN = Cervical Intraepithelial Lesion; Col = Column; DNA = Deoxyribonucleic Acid; ICD = International Classification of Diseases, Clinical Modification; TennCare = Tennessee Medicaid.

[a]The ICD-9 era includes procedures from January 1, 2008, through September 30, 2015; the ICD-10 era includes procedures from October 1, 2015, through December 31, 2017.

*Asterisks denote $p < 0.05$

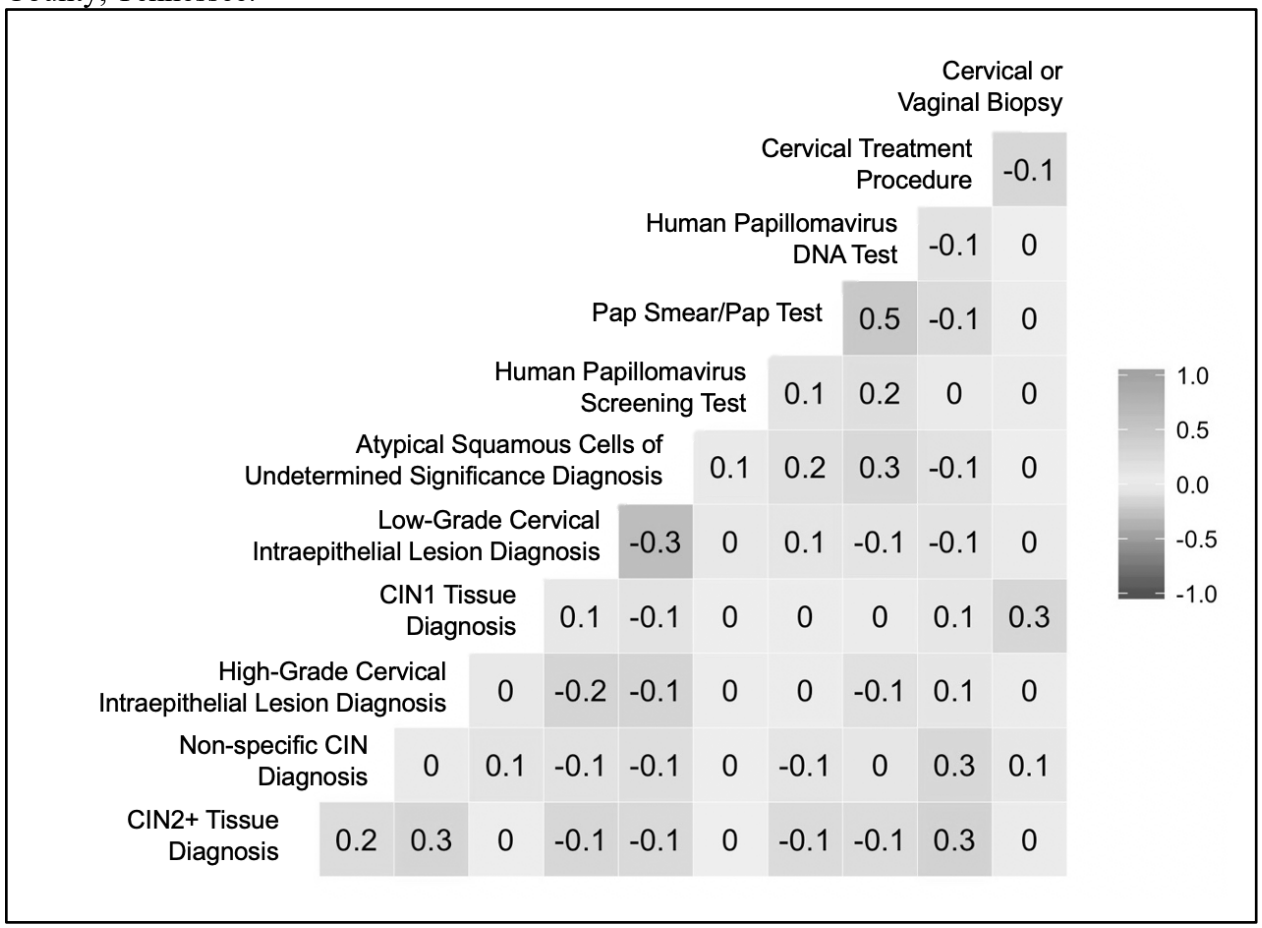
Table 3.4. Beta coefficients and predictor importance scores of LASSO and random forest models[a] to classify CIN2+ event status in the training set (N = 5,129) of cervical diagnostic procedures among TennCare-enrolled women aged 18-39 years in Davidson County, Tennessee.

Predictors	LASSO Beta Coefficients	Random Forest Predictor Importance Scores
Constant	-5.915605	—
CIN2+ Tissue Diagnosis	5.341873	0.695894
Cervical Treatment Procedure	0.9440706	0.089150
Cervical or Vaginal Biopsy	0.9414902	0.032999
High-Grade Squamous Intraepithelial Lesion Diagnosis	0.9338596	0.095700
Non-Specific CIN Diagnosis	0.3964537	0.028032
Low-Grade Squamous Intraepithelial Lesion Diagnosis	0.3541705	0.010605
Atypical Squamous Cells of Undetermined Significance Diagnosis	0.2838765	0.010486
CIN1 Tissue Diagnosis	-0.2115674	0.015590
Human Papillomavirus DNA Test	-0.2082338	0.008846
Pap Smear/Test	-0.1695168	0.011962
Human Papillomavirus Screening Test	-0.0893877	0.000737

Abbreviations: CIN = Cervical Intraepithelial Lesion; DNA = Deoxyribonucleic Acid; ICD = International Classification of Diseases, Clinical Modification; LASSO = Least Absolute Shrinkage and Selection Operator; TennCare = Tennessee Medicaid.

[a]Models were built using training sets of both ICD-9 and ICD-10 eras combined.

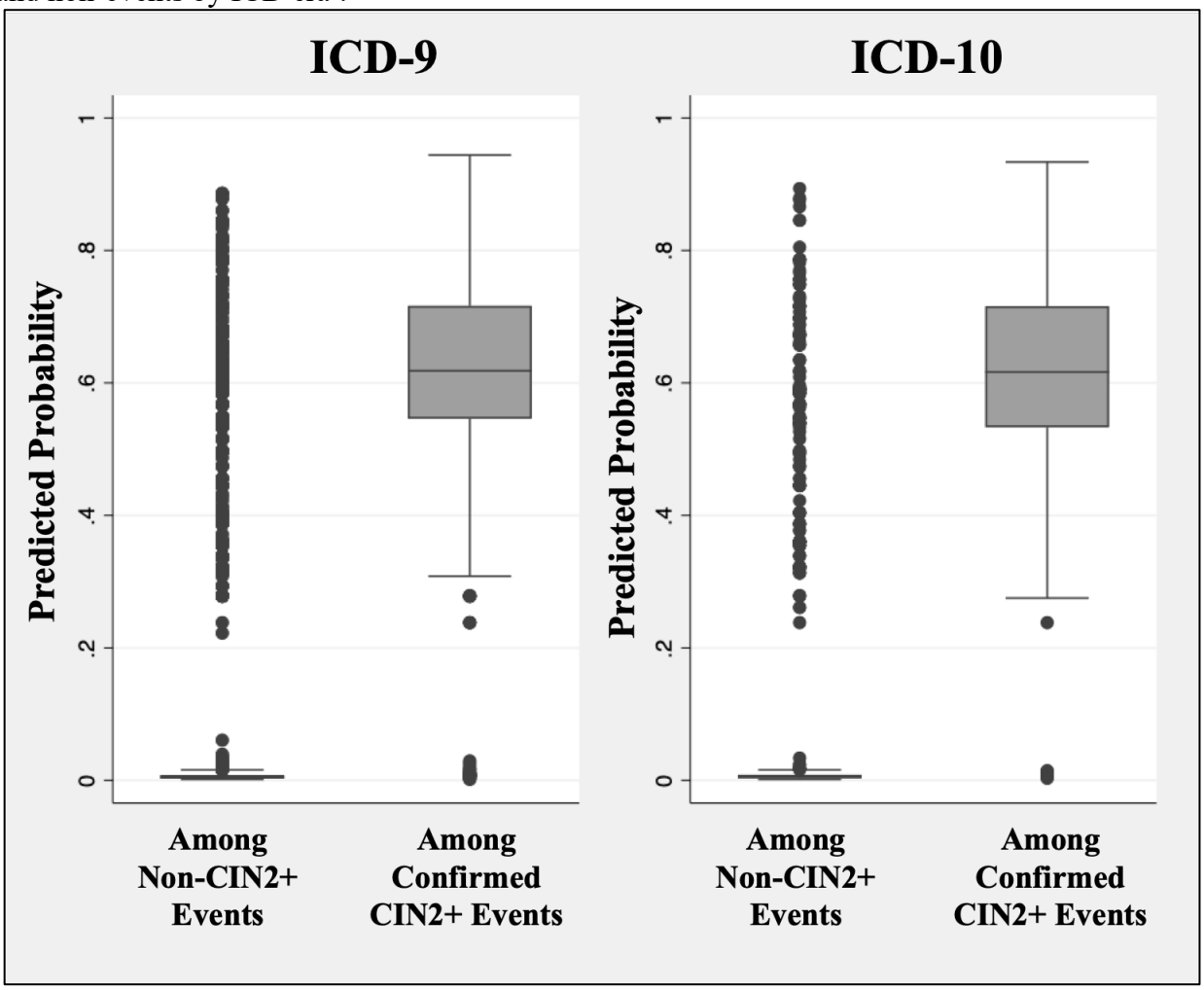
Figure 3.3. Correlation matrix of predictors selected in the model built by LASSO among cervical diagnostic procedures of TennCare-enrolled women aged 18-39 years in Davidson County, Tennessee.



Abbreviations: CIN = Cervical Intraepithelial Lesion; DNA = Deoxyribonucleic Acid; LASSO = Least Absolute Shrinkage and Selection Operator; TennCare = Tennessee Medicaid.

When assessing an appropriate LASSO cut-off value, the predicted probabilities among the HPV-IMPACT confirmed CIN2+ events had both a mean and median of 0.6 compared to a mean of 0.1 and median of 0.01 among non-events (**Figure 3.4**). Boxplots of predicted probabilities between confirmed CIN2+ events versus non-events among each ICD era showed

Figure 3.4. Boxplot of predicted probabilities among HPV-IMPACT confirmed CIN2+ events and non-events by ICD era[a].



Abbreviations: CIN = Cervical Intraepithelial Lesion; HPV-IMPACT = Human Papillomavirus Vaccine Impact Monitoring Project; ICD = International Classification of Diseases, Clinical Modification.

[a]The ICD-9 era includes procedures from January 1, 2008, through September 30, 2015; the ICD-10 era includes procedures from October 1, 2015, through December 31, 2017.

minimum predicted probabilities among confirmed CIN2+ events were approximately 0.3 for both ICD- 9 and ICD-10 ears, with some outliers below 0.3. After comparing classification results and performance metrics of the LASSO algorithm between cut-off values of 0.3, 0.5, and 0.6, we determined that the cut-off value of 0.5 was appropriate because many false positives would be introduced with a 0.3 cut-off, while the 0.6 cut-off would underestimate the true number of confirmed events.

Optimal parameters for the random forest model included 23 trees, an automatic maximum predictor selection method, 36 maximum tree depth, 5 minimum samples for a split, and 8 minimum samples in a leaf node (**Table 3.5**). In the random forest model, having a CIN2+ tissue diagnosis code was the strongest predictor of CIN2+ event status (**Refer to Table 3.4,** importance score = 0.70).

Table 3.5. Randomized search results[a] of random forest algorithms to classify CIN2+ event status in the training set (N = 5,129) of cervical diagnostic procedures among TennCare-enrolled women aged 18-39 years in Davidson County, Tennessee.

Model Rank	Number of Trees	Maximum Predictors Selection Method	Maximum Tree Depth	Minimum Number of Samples for a Split	Minimum Number of Samples in a Leaf Node	Mean Validation Score ± Standard Deviation
1	23	Automatic	36	5	8	0.919669 ± 0.011364
2	177	Automatic	100	5	4	0.919281 ± 0.011434
3	177	Square Root	43	10	1	0.919280 ± 0.010409
4	45	Automatic	57	10	8	0.919084 ± 0.013300
5	67	Square Root	71	5	6	0.918694 ± 0.012235

Abbreviations: CIN = Cervical Intraepithelial Lesion; TennCare = Tennessee Medicaid.
[a]Only the best five performing models are reported.

Model Performance

The testing set included 40% of the total 8,549 encounters, resulting in 3,420 encounters (ICD-9 = 2,842; ICD-10 = 578). Concordance between all models was high, with percent agreements ranging from 92%-98% (Kappa range = 0.74-0.91) (**Figure 3.5**). The highest concordance was between models built by LASSO and random forest for both ICD eras, with percent agreements of 98% (Kappa = 0.90) and 97% (Kappa = 0.91) in the ICD-9 and ICD-10 era, respectively.

Among the ICD-9 testing set (N = 2,842), 356 encounters were confirmed CIN2+ events (**Figure 3.6**). In the ICD-9 era, CIN2+ tissue diagnosis codes alone classified 624 encounters as CIN2+ events, of which 342 were correctly classified. Among the ICD-10 testing set (N = 578), 90 encounters were confirmed events. In the ICD-10 era, CIN2+ tissue diagnosis codes alone classified 160 cervical diagnostic procedures as CIN2+ events, of which 88 were correctly classified. CIN2+ tissue diagnosis codes alone performed similarly between ICD-9 and ICD-10 eras: 96.1% (95% CI = 93.5%-97.8%) versus 97.8% (95% CI = 92.2%-99.7%) sensitivity, 88.7% (95% CI = 87.3%-89.9%) versus 85.3% (95% CI = 81.8%-88.3%) specificity, 54.8% (95% CI = 50.8%-58.8%) versus 55.0% (95% CI = 46.9%-62.9%) PPV, 99.4% (95% CI = 98.9%-99.7%) versus 99.5% (95% CI = 98.3%-99.9%) NPV, 89.6% (95% CI = 88.4%-90.7%) versus 87.2% (95% CI = 84.2%-89.8%) accuracy, and C-indices of 92.4% (95% CI = 91.2%-93.6%) versus 91.5% (95% CI = 89.3%-93.7%), respectively (**Table 3.6**, 95% CIs overlapped).

Performance between training and testing sets for the model developed by LASSO was similar in both eras (**Table 3.6**, 95% CIs overlapped). All LASSO performance measures in the testing set were similar between ICD-9 and ICD-10 eras: 77.3% (95% CI = 72.5%-81.5%) versus 81.1% (95% CI = 71.5%-88.6%) sensitivity, 93.0% (95% CI = 91.9%-94.0%) versus 90.2%

Figure 3.5. Concordance between model building methodologies in the testing set (N = 3,420) of cervical diagnostic procedures among TennCare-enrolled women aged 18-39 years in Davidson County, Tennessee by ICD era[a].

Concordance Between CIN2+ Tissue Diagnosis Codes Alone and LASSO

ICD-9 Era

					Agreement	Kappa	P-Value
					93.8%	0.80	<0.001
		LASSO Classifications					
CIN2+ Tissue Diagnosis Codes Alone Classifications		CIN2+ Event	Non-Event	Total			
	CIN2+ Event	449	175	624			
	Non-Event	0	2,218	2,218			
	Total	449	2,393	2,842			

ICD-10 Era

					Agreement	Kappa	P-Value
					93.3%	0.82	<0.001
		LASSO Classifications					
CIN2+ Tissue Diagnosis Codes Alone Classifications		CIN2+ Event	Non-Event	Total			
	CIN2+ Event	121	39	160			
	Non-Event	0	418	418			
	Total	121	457	578			

Concordance Between LASSO and Random Forest

ICD-9 Era

					Agreement	Kappa	P-Value
					97.5%	0.90	<0.001
		Random Forest Classifications					
LASSO Classifications		CIN2+ Event	Non-Event	Total			
	CIN2+ Event	391	58	449			
	Non-Event	12	2,381	2,393			
	Total	403	2,439	2,842			

ICD-10 Era

					Agreement	Kappa	P-Value
					97.1%	0.91	<0.001
		Random Forest Classifications					
LASSO Classifications		CIN2+ Event	Non-Event	Total			
	CIN2+ Event	109	12	121			
	Non-Event	5	452	457			
	Total	114	464	578			

Concordance Between Random Forest and CIN2+ Tissue Diagnosis Codes Alone

ICD-9 Era

					Agreement	Kappa	P-Value
					92.2%	0.74	<0.001
		CIN2+ Tissue Diagnosis Codes Alone Classifications					
Random Forest Classifications		CIN2+ Event	Non-Event	Total			
	CIN2+ Event	403	0	403			
	Non-Event	221	2,218	2,439			
	Total	624	2,218	2,842			

ICD-10 Era

					Agreement	Kappa	P-Value
					92.0%	0.78	<0.001
		CIN2+ Tissue Diagnosis Codes Alone Classifications					
Random Forest Classifications		CIN2+ Event	Non-Event	Total			
	CIN2+ Event	114	0	114			
	Non-Event	46	418	464			
	Total	160	418	578			

Abbreviations: CIN = Cervical Intraepithelial Lesion; ICD = International Classification of Diseases, Clinical Modification; LASSO = Least Absolute Shrinkage and Selection Operator; TennCare = Tennessee Medicaid.

[a]The ICD-9 era includes procedures from January 1, 2008, through September 30, 2015; the ICD-10 era includes procedures from October 1, 2015, through December 31, 2017.

Figure 3.6. Confusion matrices of claims-based models to classify CIN2+ event status in the testing set (N = 3,420) of cervical diagnostic procedures among TennCare-enrolled women aged 18-39 years in Davidson County, Tennessee by ICD era[a].

Legend

	Confirmed Events (n)	Confirmed Non-Events (n)	Total Classified (N)
Classified Events (n)	True Positives	False Positives	Total Classified Events
Classified Non-Events (n)	False Negatives	True Negatives	Total Classified Non-Events
Total Gold Standard (N)	Total Confirmed Events	Total Confirmed Non-Events	Total Sample Size

CIN2+ Tissue Diagnosis Codes Alone

ICD-9 Era

	Confirmed Events (n)	Confirmed Non-Events (n)	Total Classified (N)
Classified Events (n)	342	282	624
Classified Non-Events (n)	14	2,204	2,218
Total Gold Standard (N)	356	2,486	2,842

ICD-10 Era

	Confirmed Events (n)	Confirmed Non-Events (n)	Total Classified (N)
Classified Events (n)	88	72	160
Classified Non-Events (n)	2	416	418
Total Gold Standard (N)	90	488	578

LASSO

ICD-9 Era

	Confirmed Events (n)	Confirmed Non-Events (n)	Total Classified (N)
Classified Events (n)	275	174	449
Classified Non-Events (n)	81	2,312	2,393
Total Gold Standard (N)	356	2,486	2,842

ICD-10 Era

	Confirmed Events (n)	Confirmed Non-Events (n)	Total Classified (N)
Classified Events (n)	73	48	121
Classified Non-Events (n)	17	440	457
Total Gold Standard (N)	90	488	578

Random Forest

ICD-9 Era

	Confirmed Events (n)	Confirmed Non-Events (n)	Total Classified (N)
Classified Events (n)	250	153	403
Classified Non-Events (n)	106	2,333	2,439
Total Gold Standard (N)	356	2,486	2,842

ICD-10 Era

	Confirmed Events (n)	Confirmed Non-Events (n)	Total Classified (N)
Classified Events (n)	68	46	114
Classified Non-Events (n)	22	442	464
Total Gold Standard (N)	90	488	578

Abbreviations: CIN = Cervical Intraepithelial Lesion; ICD = International Classification of Diseases, Clinical Modification; LASSO = Least Absolute Shrinkage and Selection Operator; TennCare = Tennessee Medicaid.

[a]The ICD-9 era includes procedures from January 1, 2008, through September 30, 2015; the ICD-10 era includes procedures from October 1, 2015, through December 31, 2017.

Table 3.6. Performance of prediction models to classify CIN2+ event status among cervical diagnostic procedures of TennCare-enrolled women aged 18-39 years in Davidson County, Tennessee by ICD era[a].

Performance Measure	CIN2+ Tissue Diagnosis Codes Alone		LASSO				Random Forest			
	ICD-9 (N = 7,105)	ICD-10 (N = 1,444)	ICD-9 (N = 7,105)		ICD-10 (N = 1,444)		ICD-9 (N = 7,105)		ICD-10 (N = 1,444)	
	Testing Set (n = 2,842)	Testing Set (n = 578)	Training Set (n = 4,263)	Testing Set (n = 2,842)	Training Set (n = 866)	Testing Set (n = 578)	Training Set (n = 4,263)	Testing Set (n = 2,842)	Training Set (n =866)	Testing Set (n = 578)
Sensitivity, % (95% CI)	96.1 (93.5, 97.8)	97.8 (92.2, 99.7)	82.0 (78.5, 85.2)	77.3 (72.5, 81.5)	75.2 (67.2, 82.1)	81.1 (71.5, 88.6)	79.8[b] (76.1, 83.1)	70.2[b] (65.2, 74.9)	75.9 (68.0, 82.7)	75.6 (65.4, 84.0)
Specificity, % (95% CI)	88.7 (87.3, 89.9)	85.3 (81.8, 88.3)	94.2 (93.4, 94.9)	93.0 (91.9, 94.0)	93.1 (91.0, 94.8)	90.2 (87.2, 92.7)	95.0 (94.3, 95.7)	93.8 (92.8, 94.8)	93.5 (91.5, 95.2)	90.6 (87.6, 93.0)
PPV, % (95% CI)	54.8 (50.8, 58.8)	55.0 (46.9, 62.9)	66.8 (63.0, 70.4)	61.3 (56.6, 65.8)	68.0 (60.0, 75.2)	60.3 (51.0, 69.1)	69.5 (65.7, 73.2)	62.0 (57.1, 66.8)	69.5 (61.6, 76.6)	59.6 (50.1, 68.7)
NPV, % (95% CI)	99.4 (98.9, 99.7)	99.5 (98.3, 99.9)	97.4 (96.8, 97.9)	96.6 (95.8, 97.3)	95.1 (93.2, 96.5)	96.3 (94.1, 97.8)	97.1[b] (96.5, 97.6)	95.7[b] (94.8, 96.4)	95.2 (93.4, 96.7)	95.3 (92.9, 97.0)
Accuracy, % (95% CI)	89.6 (88.4, 90.7)	87.2 (84.2, 89.8)	92.7 (91.9, 93.5)	91.0 (89.9, 92.1)	90.2 (88.0, 92.1)	88.8 (85.9, 91.2)	93.2[b] (92.4, 93.9)	90.9[b] (89.8, 91.9)	90.6 (88.5, 92.5)	88.2 (85.3, 90.7)
C-Index, % (95% CI)	92.4 (91.2, 93.6)	91.5 (89.3, 93.7)	88.1 (86.5, 89.8)	85.1 (82.9, 87.4)	84.1 (80.5, 87.8)	85.6 (81.4, 89.9)	87.4[b] (85.7, 89.2)	82.0[b] (79.6, 84.5)	84.7 (81.1, 88.4)	83.1 (78.4, 87.7)

Abbreviations: CI = Confidence Interval; CIN = Cervical Intraepithelial Lesion; ICD = International Classification of Diseases, Clinical Modification; LASSO = Least Absolute Shrinkage and Selection Operator; NPV = Negative Predictive Value; PPV = Positive Predictive Value; TennCare = Tennessee Medicaid.

[a]The ICD-9 era includes procedures from January 1, 2008, through September 30, 2015; the ICD-10 era includes procedures from October 1, 2015, through December 31, 2017.

[b]Performance between the training and testing sets are statistically significantly different (confidence intervals do not overlap with each other).

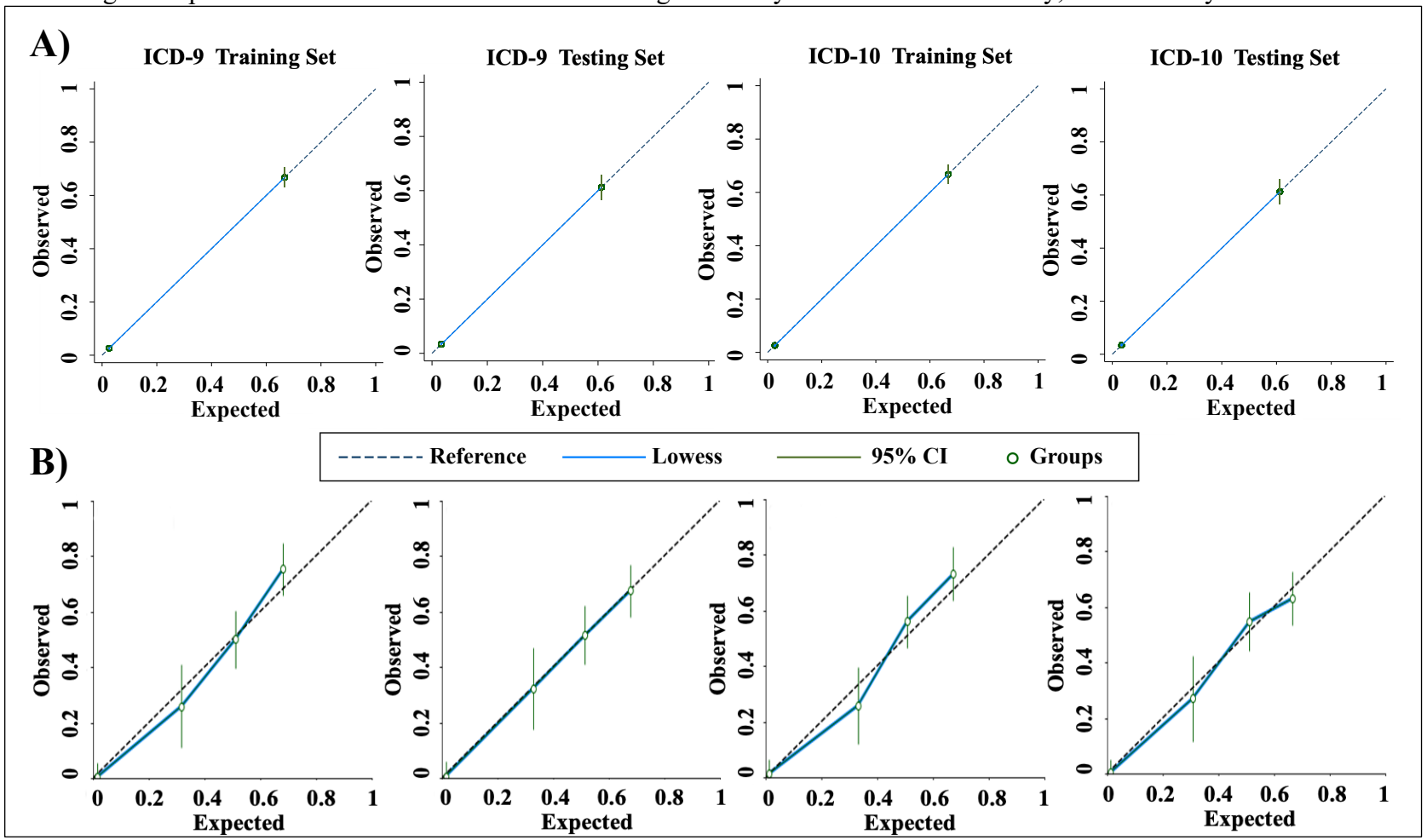
(95% CI = 87.2%-92.7%) specificity, 61.3% (95% CI = 56.6%-65.8%) versus 60.3% (95% CI = 51.0%-69.1%) PPV, 96.6% (95% CI = 95.8%-97.3%) versus 96.3% (95% CI = 94.1%-97.8%) NPV, 91.0% (95% CI = 89.9%-92.1%) versus 88.8% (95% CI = 85.9%-91.2%), accuracy, and C-indices of 85.1% (95% CI = 82.9%-87.4%) versus 85.6% (95% CI = 81.4%-89.9%), respectively (95% CIs overlapped). LASSO was well calibrated in both ICD eras and testing and training sets; expected and observed probabilities were similar (**Figure 3.7**).

Performance of the model developed by random forest in testing sets of ICD-9 and ICD-10 eras was similar: 70.2% (95% CI = 65.2%-74.9%) versus 75.6% (95% CI = 65.4%-84.0%) sensitivity, 93.8% (95% CI = 92.8%-94.8%), versus 90.6% (95% CI = 87.6%-93.0%) specificity, 62.0% (95% CI = 57.1%-66.8%) versus 59.6% (95% CI=50.1%-68.7%) PPV, 95.7% (95% CI = 94.8%-96.4%) versus 95.3% (95% CI = 92.9%-97.0%) NPV, 90.9% (95% CI = 89.8%-91.9%) versus 88.2% (95% CI = 85.3%-90.7%) accuracy, and C-indices of 82.0% (95% CI = 79.6%-84.5%) versus 83.1% (95% CI = 78.4%-87.7%), respectively (**Refer to Table 3.6**, 95% CIs overlapped). However, this model was not generalizable in the ICD-9 era, with statistically significant differences in sensitivity, NPV, accuracy, and C-index between training and testing sets (95% CIs did not overlap). Random forest was well-calibrated for both eras and testing and training sets (**Figure 3.7**).

Model Performance by Age Group

Performance of CIN2+ tissue diagnosis codes alone was similar between ICD-9 and ICD-10 eras across all age groups, except for C-index among ages 25-29 years (**Table 3.7**). C-indices statistically significantly differed between ages 18-24 (94.6%; 95% CI = 93.4%-95.8%) versus

Figure 3.7. Calibration plots of models built by A) LASSO and B) random forest algorithms to classify CIN2+ event status among cervical diagnostic procedures of TennCare-enrolled women aged 18-39 years in Davidson County, Tennessee by ICD era[a].



Abbreviations: CI = Confidence Interval; ICD = International Classification of Diseases, Clinical Modification; LASSO = Least Absolute Shrinkage and Selection Operator.

[a]The ICD-9 era includes procedures from January 1, 2008 through September 30, 2015; the ICD-10 era includes procedures from October 1, 2015 through December 31, 2017.

Table 3.7. Performance of claims-based models to classify CIN2+ event status by age group in the testing set (N = 3,420) of cervical diagnostic procedures among TennCare-enrolled women aged 18-39 years in Davidson County, Tennessee by ICD era[a].

Performance Measure	Ages 18-24 Years (N = 1,349)		Ages 25-29 Years (N =1,033)		Ages 30-39 Years (N = 1,038)	
	ICD-9 (n = 1,248)	ICD-10 (n = 101)	ICD-9 (n = 835)	ICD-10 (n = 198)	ICD-9 (n = 759)	ICD-10 (n = 279)
CIN2+ Tissue Diagnosis Codes Alone						
Sensitivity, % (95% CI)	99.2 (95.5, 99.8)	91.7 (61.5, 99.8)	93.0 (86.6, 96.9)	100.0 (91.4, 100.0)	95.8 (90.5, 98.6)	97.3 (85.8, 99.9)
Specificity, % (95% CI)	90.0 (88.1, 91.7)	89.9 (81.7, 95.3)	87.2 (84.6, 89.6)	80.9 (73.9, 86.7)	88.0 (85.2, 90.4)	86.4 (81.4, 90.4)
PPV, % (95% CI)	51.7 (45.1, 58.3)	55.0 (31.5, 76.9)	53.3 (46.3, 60.6)	57.7 (45.4, 69.4)	59.9 (52.6, 66.9)	52.2 (39.8, 64.4)
NPV, % (95% CI)	99.9 (99.5, 100.0)	98.8 (93.3, 100.0)	98.7 (97.5, 99.5)	100.0 (97.1, 100.0)	99.1 (98.0, 99.7)	99. 5 (97.4, 100.0)
Accuracy, % (95% CI)	90.9 (89.1, 92.4)	90.1 (82.5, 95.2)	88.0 (85.6, 90.2)	84.9 (79.1, 89.5)	89.2 (86.8, 91.3)	87.8 (83.4, 91.4)
C-Index, % (95% CI)	94.6 (93.4, 95.8)[b]	90.8 (82.0, 99.5)	98.7 (97.5, 99.5)[b,c]	90.5 (87.4, 93.5)[c]	91.9 (89.7, 94.1)[b]	91.8 (88.4, 95.3)
LASSO						
Sensitivity, % (95% CI)	81.1 (73.1, 87.7)	75.0 (42.8, 94.5)	74.6 (65.6, 82.3)	80.5 (65.1, 91.2)	75.8 (67.2, 83.2)	83.8 (68.0, 93.8)
Specificity, % (95% CI)	93.7 (92.1, 95.0)	94. 4 (87.4, 98.2)	92.0 (89.7, 93.8)	87.9 (81.7, 92.6)	93.0 (90.7, 94.8)	90.1 (85.6, 93.5)
PPV, % (95% CI)	58.2 (50.4, 65.7)	64.3 (35.1, 87.2)	59.4 (50.9, 67.6)	63.5 (49.0, 76.4)	66.9 (58.3, 74.7)	56.4 (42.3, 69.7)
NPV, % (95% CI)	97.9 (96.8, 98.6)	96.6 (90.3, 99.3)	95.8 (94.0, 97.2)	94.5 (89.5, 97.6)	95.3 (93.4, 96.9)	97.3 (94.3, 99.0)
Accuracy, % (95% CI)	92.5 (90.1, 93.9)	92.1 (85.0, 96.5)	89.6 (87.3, 91.6)	86.4 (80.8, 90.8)	90.3 (87.9, 92.3)	89.3 (85.0, 92.6)
C-Index, % (95% CI)	87.4 (83.9, 91.0)	84.7 (71.7, 97.7)	83.3 (79.1, 87.4)	84.2 (77.5, 90.9)	84.4 (80.4, 88.4)	86.9 (80.6, 93.2)
Random Forest						
Sensitivity, % (95% CI)	77.0 (68.6, 84.2)	50.0 (21.1, 78.9)	64.9 (55.4, 73.6)	73.2 (57.1, 85.8)	68.3 (59.2, 76.5)	86.5 (71.2, 95.5)
Specificity, % (95% CI)	94.7 (93.2, 95.9)	94.4 (87.4, 98.2)	92.4 (90.2, 94.2)	89.8 (84.0, 94.1)	94.1 (91.9, 95.8)	89.7 (85.1, 93.2)
PPV, % (95% CI)	61.0 (52.9, 68.8)	54.5 (23.4, 83.3)	57.4 (48.4, 66.0)	65.2 (49.8, 78.6)	68.3 (59.2, 76.5)	56.1 (42.4, 69.3)
NPV, % (95% CI)	97.4 (96.3, 98.3)[b]	93.3 (86.1, 97.5)	94.3 (92.4, 95.9)[b]	92.8 (87.4, 96.3)	94.1 (91.9, 95.8)	97.7 (94.8, 99.3)
Accuracy, % (95% CI)	93.0 (91.4, 94.3)	89.1 (81.4, 94.4)	88.6 (86.3, 90.7)	86.4 (80.8, 90.8)	90.0 (87.6, 92.0)	89.3 (85.0, 92.6)
C-Index, % (95% CI)	85.9 (82.1, 89.7)	72.2 (57.2, 87.2)	78.6 (74.1, 83.2)	81.5 (74.2, 88.8)	81.2 (76.9, 85.5)	88.1 (82.2, 94.0)

Abbreviations: CI = Confidence Interval; CIN = Cervical Intraepithelial Lesion; ICD = International Classification of Diseases, Clinical Modification; LASSO = Least Absolute Shrinkage and Selection Operator; NPV = Negative Predictive Value; PPV = Positive Predictive Value; TennCare = Tennessee Medicaid.

[a]The ICD-9 era includes procedures from January 1, 2008, through September 30, 2015; the ICD-10 era includes procedures from October 1, 2015, through December 31, 2017.

[b]Performance between age groups, either ages 18-24 versus 25-29 years, ages 18-24 versus 30-39 years, or ages 25-29 versus 35-39 years are statistically significantly different (confidence intervals do not overlap with each other).

[c]Performance between the ICD-9 and ICD-10 eras within each age group are statistically significantly different (confidence intervals do not overlap with each other).

25-29 (98.7%; 95% CI = 97.5%-99.5%) years and 25-29 (98.7%; 95% CI = 97.5%-99.5%) versus 30-39 (91.9%; 95% CI = 89.7%-94.1%) years in the ICD-9 era for CIN2+ tissue diagnosis codes alone (95% CIs did not overlap). Models performed similarly between ICD-9 and ICD-10 eras across all age groups for LASSO and random forest. When comparing between age groups across ICD eras, all measures were similar for LASSO. For random forest, NPV statistically significantly differed between ages 18-24 (97.4%; 95% CI = 96.3%-98.3%) versus 25-29 (94.2%; 95% CI = 92.4%-95.9%) years in the ICD-9 era (95% CIs did not overlap).

Model Estimation of Incident CIN2+ Events

From 2008 to 2017, HPV-IMPACT identified 983 confirmed incident CIN2+ events among TennCare-enrolled women aged 18-39 years residing in Davidson County (**Table 3.8 and Figure 3.8**). When counting model-identified incident events compared to HPV-IMPACT's confirmed incident events, all claims-based models showed declining trends in CIN2+ incidence from 2008 tp 2017, with some yearly classification variation. LASSO (n = 1,007) and random forest (n = 957) more closely captured the true number of population HPV-IMPACT incident events (n = 983) compared to CIN2+ tissue diagnosis codes alone (n = 1,245).

Sensitivity Analysis: Results Among Cervical Screening Tests

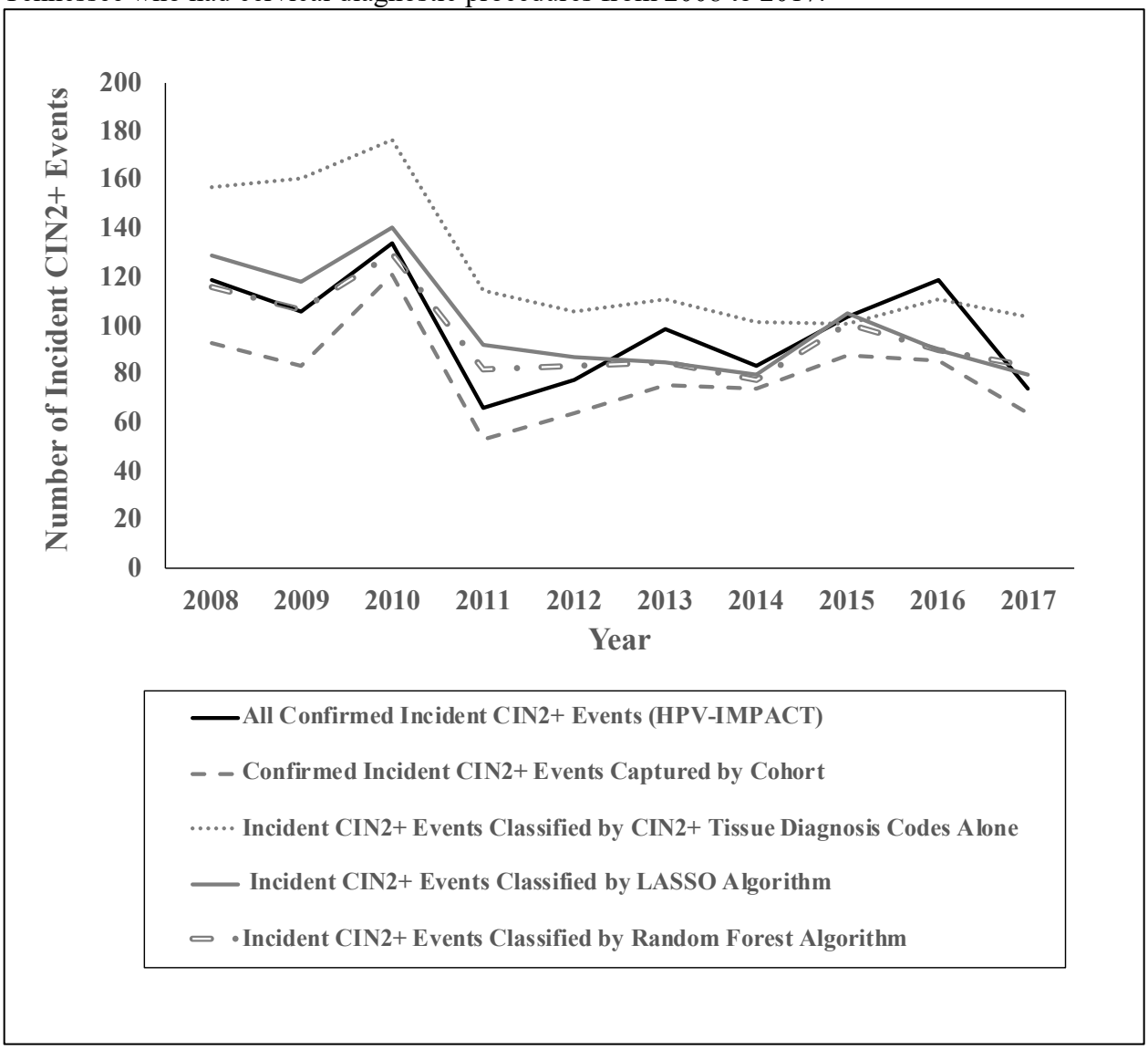
To compare administrative code patterns and model performance between women with cervical diagnostic procedures versus cervical screening tests, we replicated all prior methods among a cohort of cervical screening tests. We identified 42,324 TennCare-enrolled women aged 18-39 years residing in Davidson County, Tennessee who had a total of 88,765 (ICD-9 = 69,792, ICD-10 = 18,973) cervical screening tests from 2008 to 2017 (**Appendix Figure A1**).

Table 3.8. Annual number of incident CIN2+ events identified by claims-based models among TennCare-enrolled women aged 18-39 years residing in Davidson County, Tennessee who had cervical diagnostic procedures from 2008 to 2017.

Year	All Confirmed Incident CIN2+ Events in Population Identified by HPV-IMPACT	Confirmed Incident CIN2+ Events Captured by Cohort	Model		
			CIN2+ Tissue Diagnosis Codes Alone	LASSO	Random Forest
2008	119	93	157	129	116
2009	106	84	161	118	107
2010	134	121	177	141	130
2011	66	53	115	92	82
2012	78	64	106	87	84
2013	99	76	111	85	85
2014	84	74	102	80	78
2015	104	88	101	105	101
2016	119	86	111	90	90
2017	74	64	104	80	84
Total	983	803	1245	1007	957

Abbreviations: CIN = Cervical Intraepithelial Lesion; HPV-IMPACT = Human Papillomavirus Vaccine Impact Monitoring Project; LASSO = Least Absolute Shrinkage and Selection Operator; TennCare = Tennessee Medicaid.

Figure 3.8. Trends in the annual number of incident[a] CIN2+ events identified by claims-based models among TennCare-enrolled women aged 18-39 years residing in Davidson County, Tennessee who had cervical diagnostic procedures from 2008 to 2017.

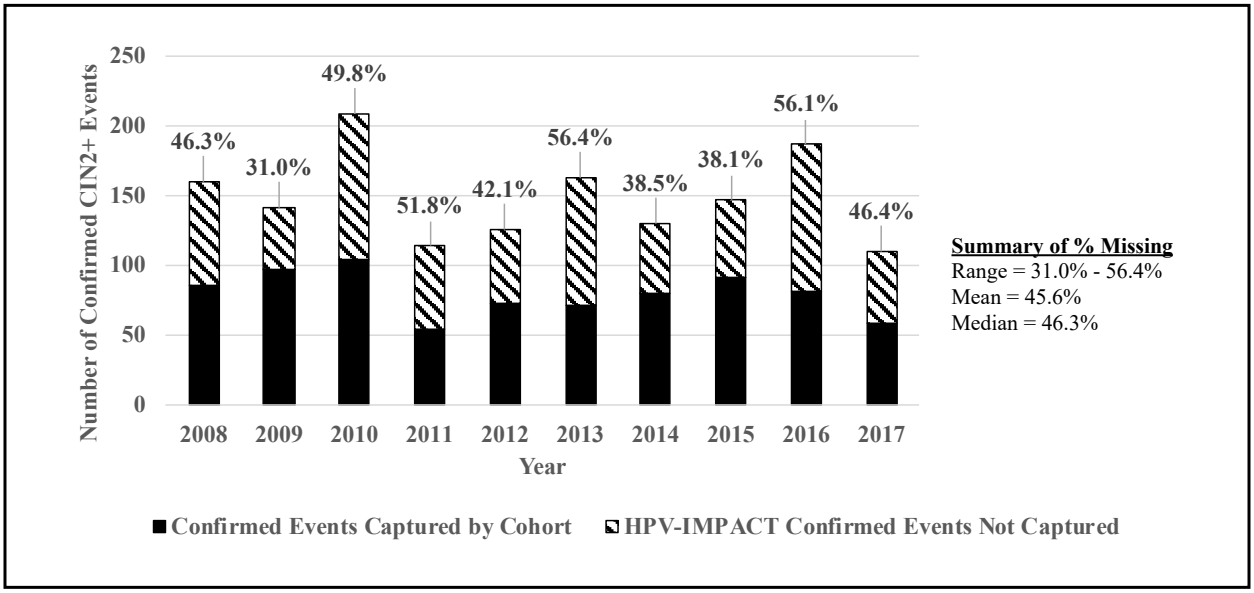


Abbreviations: CIN = Cervical Intraepithelial Lesion; HPV-IMPACT = Human Papillomavirus Vaccine Impact Monitoring Project; LASSO = Least Absolute Shrinkage and Selection Operator; TennCare = Tennessee Medicaid.

[a]Incident events are determined by applying each model to the cohort of cervical diagnostic procedures and counting index events classified by each model.

When limiting the cohort to women who had qualifying billing codes for a cervical screening test, only 800 (54%) of the 1,488 gold standard, HPV-IMAPCT confirmed CIN2+ events from the target population (i.e., TennCare-enrolled women aged 18-39 years residing in Davidson County) were captured in the cohort. The percent of HPV-IMPACT confirmed CIN2+ events missed ranged from 31% to 56% annually from 2008 to 2017 (**Figure 3.9**), compared to just 10%-35% missed events among cervical diagnostic procedures (**Refer to Figure 3.2**).

Figure 3.9. Annual percent of HPV-IMPACT confirmed CIN2+ events missing from cohort of cervical screening tests from 2008 to 2017 among TennCare-enrolled women aged 18-39 years in Davidson County, Tennessee.



Abbreviations: CIN = Cervical Intraepithelial Lesion; HPV-IMPACT = Human Papillomavirus Vaccine Impact Monitoring Project; TennCare = Tennessee Medicaid.

Similar to characteristics we observed among cervical diagnostic procedures, compared to the ICD-9 era, a greater proportion of women who had cervical screening tests in the ICD-10 era were aged 30-39 years (ICD-9 = 34.8% versus ICD-10 = 44.7%) and other/unknown race/ethnicity (ICD-9 = 41.6% versus ICD-10 = 49.2%) (**Appendix Table A2**, $p < 0.001$). The proportion of administrative codes used in the ICD-9 era versus ICD-10 era statistically

significantly differed for all code groupings ($p < 0.05$). For cervical screening tests overall and in both ICD eras, coding patterns were also statistically significantly different between confirmed CIN2+ events and non-events for all code groupings (**Appendix Table A3**, $p < 0.001$).

Among cervical screening tests, cervical or vaginal biopsy codes were highly correlated with cervical diagnostic procedure and CIN1 tissue diagnosis codes, with correlation coefficients of 0.7 and 0.6, respectively (**Appendix Figure A2**); therefore, we removed cervical or vaginal biopsy codes when building the model using LASSO among cervical screening tests. The optimized parameters for the model developed by the random forest algorithm among cervical screening tests were the same as those found among cervical diagnostic procedures: 23 trees, an automatic maximum predictor selection method, 36 maximum tree depth, 5 minimum samples for a split, and 8 minimum samples in a leaf node (**Appendix Table A4**). Contrary to the final model built by LASSO among cervical diagnostic procedures, which selected all code groupings as important predictors of CIN2+ status (**Refer to Table 3.4**), the LASSO model among cervical screening tests only selected the following predictors: CIN2+ tissue diagnosis, non-specific CIN diagnosis, HGSIL diagnosis, LGSIL diagnosis, cervical treatment procedure, and cervical diagnostic procedure (**Appendix Table A5**). All of the code groupings selected by LASSO among cervical screening tests were positive predictors of CIN2+ status; no negative predictors were selected. Both final models built by LASSO and random forest among cervical screening tests determined having codes for a CIN2+ tissue diagnosis was the strongest predictor of CIN2+ status (LASSO beta coefficient = 6.02; random forest importance score = 0.67).

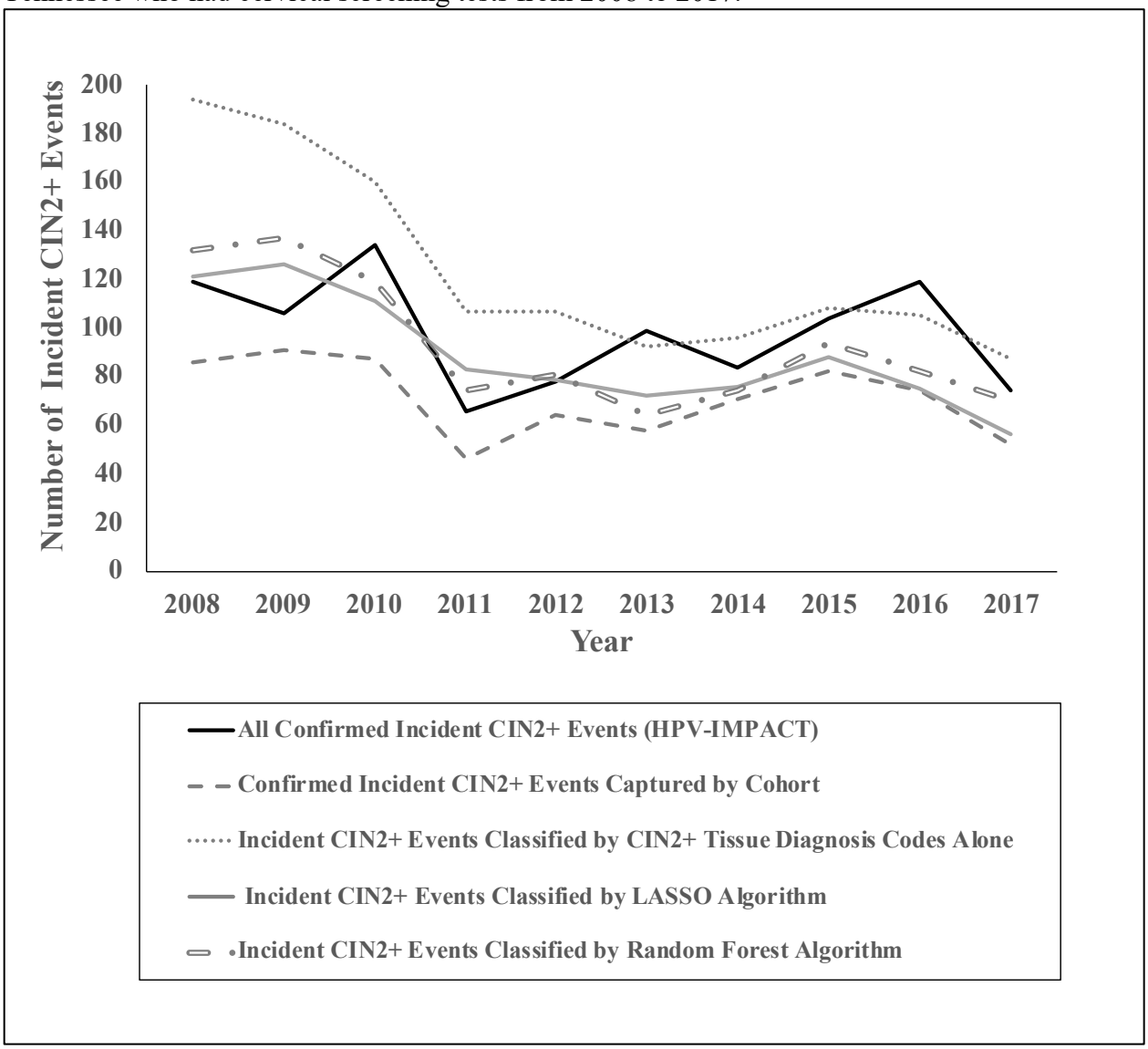
Among the ICD-9 testing set (N = 69,792), 641 cervical screening test encounters were confirmed CIN2+ events (**Appendix Figure A3**). In the ICD-9 era, CIN2+ tissue diagnosis codes alone classified 1,281 encounters as CIN2+ events, of which 628 were correctly classified.

Among the ICD-10 testing set (N = 18,973), 159 cervical screening test encounters were confirmed events. In the ICD-10 era, CIN2+ tissue diagnosis codes alone classified 285 encounters as CIN2+ events, of which 155 were correctly classified. Among cervical screening tests, CIN2+ tissue diagnosis codes alone had similar discriminative performance between ICD-9 and ICD-10 eras in terms of sensitivity, PPV, NPV, accuracy, and C-index (95% CIs all overlapped), but statistically significantly differed in specificity between eras (ICD-9 = 99.1%; 95% CI = 98.9%-99.2% versus ICD-10 = 99.5%; 95% CI = 99.3%-99.6%) (**Appendix Table A6**).

For the model built by LASSO among cervical screening tests, discriminative performance statistically significantly differed in specificity and accuracy between ICD eras; accuracy also differed between training and testing sets for the ICD-10 era (95% CIs did not overlap). For the model built by random forest among cervical screening tests, specificity statistically significantly differed between ICD-9 (99.5%; 95% CI = 99.5%-99.6%) and ICD-10 (99.8%; 95% CI = 99.6%-99.9%) eras.

When counting model-classified index events among cervical screening tests and comparing to HPV-IMPACT confirmed incident events (n = 983), models built by LASSO (n = 887) and random forest (n = 927) estimated incident events more closely than CIN2+ tissue codes alone (n = 1240) (**Figure 3.10**). Overall trends of all models followed to gold standard declining trend from 2008 to 2017; however, CIN2+ tissue diagnosis codes alone consistently and substantially overestimated incident CIN2+ events until 2013.

Figure 3.10. Trends in the annual number of incident[a] CIN2+ events identified by claims-based models among TennCare-enrolled women aged 18-39 years residing in Davidson County, Tennessee who had cervical screening tests from 2008 to 2017.



Abbreviations: CIN = Cervical Intraepithelial Lesion; HPV-IMPACT = Human Papillomavirus Vaccine Impact Monitoring Project; LASSO = Least Absolute Shrinkage and Selection Operator; TennCare = Tennessee Medicaid.

[a]Incident events are determined by applying each model to the cohort of cervical diagnostic procedures and counting index events classified by each model.

Discussion

We validated claims-based models for estimating the number of CIN2+ events in ICD-9 and ICD-10 eras, which performed well and are optimized for public health surveillance and trend analyses. Among women with cervical diagnostic procedures, the LASSO model most closely identified the population's confirmed incident CIN2+ events, with no statistically significant differences in performance between ICD eras. Because LASSO and random forest performed comparably well, model averaging could be an acceptable method; however, LASSO was more internally generalizable, with no statistically significant differences in performance between testing and training sets in both eras. Further, LASSO may be easier to understand and replicate within other databases compared to random forest because LASSO is a linear model.

When comparing coding patterns and claims-based models between cervical diagnostic procedures and cervical screening tests, models among cervical screening tests had higher specificity and accuracy; however, the models among cervical screening tests were less internally generalizable and demonstrated statistically significant differences between ICD-9 and ICD-10 eras for some performance measures. Further, when limiting the population to women with screening tests, nearly half of all CIN2+ events in the population were not captured, compared to just a quarter not captured among cervical diagnostic procedures.

When stratifying by age group, LASSO model performance among cervical diagnostic procedures was similar across age groups and eras. In our study population, the distribution of age groups differed across eras, which may be explained by epidemiologic shifts in disease occurrence and detection from changes in cervical screening guidelines and the impact of HPV vaccination over time. In 2012, updated guidelines recommended against cervical screening for women aged <21 years,[45] contributing to decreases in young women receiving Pap smears and

cervical diagnostic procedures. Further, cervical biopsy data from the HPV-IMPACT monitoring project demonstrated declines in CIN2+ incidence among younger ages (18-24 years) who may have benefited from the HPV vaccine, and increasing trends among older ages (30-39 years) from 2008 to 2015.[91] Due to differences in screening and CIN2+ trends across ages over time, changes in characteristics between eras seems reasonable.

Although using CIN2+ tissue diagnosis codes alone to identify CIN2+ events is intuitive, this approach had relatively low specificity, resulting in over-estimated event classifications. We also observed over-estimated incident CIN2+ event estimations after applying this approach to cervical diagnostic procedures and counting classified index events. When aiming for an accurate estimation of "true" population disease rates, specificity should be optimized to "rule-in" identified events. Although random forest had the highest specificity, this model was not as generalizable as LASSO. At the same time, from a data science perspective, it is important not to build a perfectly accurate model since this might mean the model is over-trained and is merely memorizing the training set's data patterns, which would limit the external generalizability of the model.

One study[96] validated claims-based algorithms identifying high-grade cervical dysplasia, including HGSIL, CIN2, CIN3, and cervical cancer; however, the study was published prior to ICD-10; therefore, it only included ICD-9 codes, which is not useful for assessing trends past 2015. Additionally, events were identified based on sets of rules (e.g., at least one-two diagnosis codes and one-two procedure codes), then confirmed using chart analyses from a linked electronic health system. Thus, the algorithm measured positive predictive value and not sensitivity; many true events may have been missed by the inclusion criteria. The study was

restricted to women with abnormal Pap test codes, which we found to exclude nearly half of true events within our study population.

Claims data comes with limitations. We were unable to test model performance by demographic sectors besides age group, such as race/ethnicity, because 40% of our sample self-reported their race/ethnicity as other/unknown, with increases in this classification over time due to increasing proportions of enrollees not identifying in a single racial group.[84,112] Specifically in TennCare, multi-race is not collected and beginning in 2018, race/ethnicity was an optional field, resulting in higher proportions of enrollees with unknown race. Further, annual reports by the Substance Abuse and Mental Health Services Administration, which utilized Medicaid enrollment demographics to characterize race/ethnicity of persons in Tennessee served by the State Mental Health Authority, also showed increasing proportions of unknown race/ethnicity even before TennCare allowed for the reporting of race to be optional.[113–117] Another limitation was that because our sample was limited to women with qualifying diagnostic procedural codes, only 75% of HPV-IMPACT's confirmed events were captured. Missingness may be because some women were retroactively enrolled in TennCare around the time of diagnosis and procedure codes were not captured, codes may have been non-specific, or procedures were not billed. Whether these issues would apply to other insurance databases is unknown. We were unable to validate outside of Davidson County, Tennessee; thus, model performance in populations with different CIN2+ prevalence or demographics, such as in non-Medicaid populations, may differ, and should be examined in future studies.

Our study had notable strengths. We utilized gold standard data from HPV-IMPACT, which underwent extensive audits to ensure high-quality data. Because these data were population-based, we could build models optimized for surveillance and trend analyses,

prioritizing the estimation of population CIN2+ incidence. Additionally, we built models using machine learning methods, gaining valuable information from each method. To our knowledge, our study is the first to validate claims-based models for identifying incident CIN2+ events in ICD-9 and ICD-10 eras. Although we did not have access to external data to validate our models in outside populations, we were still able to demonstrate good internal generalizability between training and testing sets from our own sample. A potential next step could be an external validation of our models in another HPV-IMPACT partnering site to test how well these models perform in a different population.

Examining CIN2+ incident trends after 2015 is valuable for evaluating the HPV vaccine's impact on reducing cervical premalignant lesions. These ecologic analyses are important because the vaccine has both direct effects (on vaccinated persons) and indirect effects (on those exposed to vaccinated persons) that are not captured in traditional vaccine effectiveness analyses. Since the vaccine's introduction in 2006,[8] assessing US trends in CIN2+ incidence has been limited to populations with adequate cervical biopsy data.[87–92,94] Claims-based studies without access to population-based cervical biopsies are limited by the 2015 ICD-10 transition.[22] Our study bridges these gaps by developing a simple model that may be uniformly applied to ICD-9 and ICD-10 eras with similar performance to assess more recent CIN2+ trends. This study expands options for CIN2+ surveillance by providing an alternate metric for identifying CIN2+ events in populations where cervical biopsy data are unavailable.

CHAPTER IV

EVIDENCE OF HUMAN PAPILLOMAVIRUS VACCINE IMPACT ON REDUCING CERVICAL PREMALIGNANT LESION INCIDENCE, TENNESSEE MEDICAID 2008-2018

Abstract

Demonstrating human papillomavirus (HPV) vaccine impact is critical for informing guidelines to increase vaccination rates and decrease HPV-related outcomes, particularly in states with sub-optimal vaccination rates, such as Tennessee. We examined HPV vaccine impact among Tennessee Medicaid (TennCare) enrollees by assessing trends in the incidence of cervical premalignant lesions among 1) all women aged 18-39 and 2) the subset of those women who were screened for cervical cancer. We used a validated claims-based model to identify incident cervical premalignant lesions, including cervical intraepithelial neoplasm grades 2 or 3 or adenocarcinoma *in situ* (together referred to as CIN2+) events and calculated annual age-group-specific incidence rates from TennCare billing data, 2008-2018. Significant trends, annual percent changes, and average annual percent changes (AAPC) were determined by Joinpoint regression. From 2008 to 2018, CIN2+ incidence significantly declined in women aged 18-20 years [AAPC = -31.9; 95% confidence interval (CI) = -38.6, 24.6], 21-24 years (AAPC = -12.9; 95% CI = -22.3, -2.4), and 25-29 years (AAPC = -6.4; 95% CI= -8.1, -4.6). Among screened women, CIN2+ incidence significantly declined for ages 18-20 years (AAPC = -20.3; 95% I = -25.3, -15.0), 21-24 years (AAPC = -10.2; 95% CI = -12.6, -7.8), and 25-29 years (AAPC = -2.6; 95% CI = -3.9, -1.2). No significant declines were observed among older age groups (30-34 and

35-39 years). Results from this ecologic study show reduced CIN2+incidence in age groups most likely to have benefited from the HPV vaccine. Declines among young, screened women indicate that these declines were not entirely due to decreases in screening. Evidence of HPV vaccine impact in populations with low vaccination rates, such as Tennessee, is promising.

Introduction

Cervical cancers and cervical premalignant lesions (e.g., high-grade cervical lesions) are preventable outcomes that are associated with considerable costs, such as direct medical expenses, premature death, and loss of productivity.[5–7] An estimated 90% of cervical cancer cases and 76% of high-grade cervical lesions are attributable to the nine genotypes covered by the nonavalent human papillomavirus (HPV) vaccine.[9,10,92] Despite the Advisory Committee on Immunization Practices' recommendations for routine adolescent HPV vaccination,[57] HPV vaccination coverage lags behind other recommended adolescent vaccines in the United States (US).[18] In 2019, the proportion of US adolescents aged 13-17 years who had initiated (had at least one dose) and were up-to-date (had all recommended doses) with the HPV vaccine were 72% and 54%, respectively, with large variation across states, ranging from 49% to 92% for initiation and 31% to 79% for those up-to-date.[18]

Tennessee has consistently ranked among the lowest quartile for HPV vaccination, with initiation and up-to-date proportions of 62% and 43%, respectively, among adolescents aged 13-17 years in 2019.[18,118] Understanding the impact of the HPV vaccine is critical for informing guidelines to increase vaccination coverage and decrease cervical cancer and premalignant lesion incidence, particularly among states with sub-optimal vaccination, such as Tennessee.

While the latency period between an initial HPV exposure and the development of cervical cancer can be up to 20 years, cervical premalignant lesions, including cervical intraepithelial neoplasia (CIN) grades 2 and 3 and adenocarcinoma *in situ* (collectively referred to as CIN2+), may be detected within a few years of infection.[12–14] Thus, CIN2+ has frequently been used to monitor HPV vaccine impact as an intermediate outcome for cervical cancer.[88–91,119] Studies have shown notable decreases in CIN2+ incidence among age groups that may have

likely benefited from the HPV vaccine[88–91,119]; however, few of these studies have focused on populations with sub-optimal HPV vaccination rates.

To our knowledge, only one study has assessed the HPV vaccine's impact on reducing CIN2+ incidence in a low vaccine coverage population, which used data from the New Mexico HPV Pap registry.[89] Other ecologic studies documenting the vaccine's impact on CIN2+ have been conducted by the Human Papillomavirus Vaccine Impact Monitoring Project (HPV-IMPACT), a surveillance program funded by the Centers for Disease Control and Prevention in five catchment areas across the US[19]; however, these studies are limited by the areas that have adequate population-based cervical biopsy data from partnering institutions within the program and do not specifically focus on assessing vaccine impact in low-coverage populations.

A surrogate metric for capturing CIN2+ events in populations without population-based cervical biopsy data may be insurance billing claims codes. Only one US study has used claims data to examine CIN2+ trends from 2007-2014.[22] The use of more recent claims data to assess trends has been limited by the transition from the International Classification of Diseases 9th revision (ICD-9) to 10th revision (ICD-10) codes in 2015.[21] In Chapter III (Aim 1), we validated a claims-based model for capturing CIN2+ events in both ICD eras.[105] In this present study, we aimed to examine the HPV vaccine's impact in a population with sub-optimal vaccination rates by assessing trends in CIN2+ incidence from 2008 through 2018 among 1) women aged 18-39 years enrolled in the Tennessee Medicaid (TennCare) program and 2) the subset of those women who were screened for cervical cancer to account for changes in screening patterns over time.

Methods

Study Population

We used TennCare claims data to identify women aged 18-39 years with at least one year of consecutive TennCare enrollment from 2008 to 2018. To account for changes in cervical cancer screening trends over time, we also identified the annual subpopulation of women who were screened for cervical cancer, defined by having at least one billing code for:

- an HPV screening examination (ICD-9 code V73.81 or ICD-10 code Z11.51); or
- a Papanicolaou (Pap) smear/test (ICD-9 codes V72.31, V72.32, V76.2, V76.47, 795.06, 91.46, or ICD-10 codes Z01.411, Z01.419, Z01.42, Z12.4, Z12.72, R87.614, or CPT codes 88141-88145, 88147-88148, 88150-88158, 88164-88167, 88174-88175, or HCPCS codes P3000-P3001, G0101, G0123-G0124, G0141, G0143-G0145, G0147-G0148, Q0091); or
- an HPV DNA test (ICD-9 codes 795.05, 795.09, or ICD-10 codes R87.10, R87.820, or CPT codes 87620-87622, 87623-87625).

This study was considered to be public health surveillance and was thus exempt by Institutional Review Boards at Vanderbilt University and the Tennessee Department of Health. This research activity was reviewed and approved by the Tennessee Department of Finance and Administration Division of TennCare.

Definition of Incident CIN2+ Events

Using our previously validated claims-based model,[105] (**Refer to Chapter III**) we identified incident CIN2+ events, including CIN2, CIN3, and adenocarcinoma *in situ*, from 2008

to 2018, using data from 2007 to 2018. Briefly, this model was built by least absolute shrinkage and selection operator logistic regression and uses a linear combination of diagnosis, screening, and treatment codes to calculate prediction scores for CIN2+ event status among women with cervical diagnostic procedures (**Table 4.1**). When computing prediction scores, the beta coefficient for the constant was forced into each calculation; therefore, no billing codes were needed in order to include the constant. Positive beta coefficients indicate predictors that add to the prediction score (or likelihood) of being a CIN2+ event, while negative beta coefficients indicate predictors that subtract to the prediction score of being a CIN2+ event. Using the prediction scores, we generated predicted probabilities using the following equation: $\frac{1}{1+\exp(-prediction\ score)}$. Finally, CIN2+ events were determined by identifying all predicted probabilities ≥0.5.

For model-identified CIN2+ events, the corresponding diagnostic procedure date was used as a proxy for diagnosis date because not all events had a specific ICD diagnosis code for CIN2+. Events were only counted among women with at least one year of consecutive TennCare enrollment from their diagnostic procedure date. For the subpopulation of women screened for cervical cancer, model-identified CIN2+ events were only counted if the screening date was within one year prior to the diagnostic procedure date. We defined incident events as those among women who were event-free for at least one year prior to their diagnostic procedure date. Additional information regarding model building and validation is described elsewhere.[105]

Statistical Analyses

Assuming that the occurrence of events, additions, and losses was homogenously distributed, we estimated annual person-time from 2008 to 2018 by counting the number of

Table 4.1. Administrative billing codes and beta (β) coefficients used to determine prediction scores[a] for CIN2+ status.

Predictor[b]	Billing Codes	β
1. Constant	–	-5.915605
2. CIN2+ Tissue Diagnosis	233.1, 622.12, D06.0, D06.1, D06.7, D06.9, N87.2, N87.1	5.341873
3. Cervical Treatment Procedure	57511, 57510, 57513, 57530–57531, 57540, 57545, 57550, 57555, 57556, 57520, 57522	0.9440706
4. Cervical or Vaginal Biopsy	57421, 57450, 57454, 57455, 57460, 57500, 58110	0.9414902
5. High-Grade Cervical Intraepithelial Lesion Diagnosis	795.04, R87.613	0.9338596
6. Non-Specific CIN Diagnosis	622.10, N87.9	0.3964537
7. Low-Grade Cervical Intraepithelial Lesion Diagnosis	795.03, R87.612	0.3541705
8. Atypical Squamous Cells of Undetermined Significance Diagnosis	795.01 795.02, R87.610, R87.611	0.2838765
9. Human Papillomavirus DNA Test	795.05, 795.09, R87.10, R87.820, 87620-87622, 87623-87625	0.2082338
10. Human Papillomavirus Screening Examination	V73.81, Z11.51	-0.0893877
11. Papanicolaou Smear/Test	V72.31, V72.32, V76.2, V76.47, 795.06, 91.46, Z01.411, Z01.419, Z01.42, Z12.4, Z12.72, R87.614, 88141-88145, 88147-88148, 88150-88158, 88164-88167, 88174-88175, P3000-P3001, G0101, G0123-G0124, G0141, G0143-G0145, G0147-G0148, Q0091	-0.1695168
12. CIN1 Tissue Diagnosis	622.11, N87.1	-0.2115674

Abbreviations: CIN = Cervical Intraepithelial Neoplasia; DNA = Deoxyribonucleic Acid.

[a]Prediction scores to determine CIN2+ status were calculated by the following equation:

β1 + (β2*Predictor2) + (β3*Predictor3) + (β4*Predictor4) + (β5*Predictor5) + (β6*Predictor6) + (β7*Predictor7) + (β8*Predictor8) + (β9*Predictor9) + (β10*Predictor10) + (β11*Predictor11) + (β12*Predictor12).

[b]Predictors were coded as 1 if any of the corresponding billing codes were identified within 60 days from a woman's cervical diagnostic procedure date; if a woman had a cluster of cervical diagnostic procedures within 30 days of each other, predictor billing codes were searched within -60 days to +60 days from the earliest and latest date, respectively, in the cluster of procedures; predictors were coded as 0 if the search criteria were not met.

women enrolled in TennCare on July 1st of each year with at least one year of consecutive enrollment, stratified by age group (18-20, 21-24, 25-29, and 30-39 years). To estimate annual person-time for the subpopulation of screened women, we counted women enrolled in TennCare on July 1st of each year with at least one year of consecutive enrollment who were screened for cervical cancer within one year prior to the current year (e.g., person-time estimation for 2008 included women screened between July 1st, 2007 to July 1st, 2008), stratified by age group.

Annual incidence rates were calculated by dividing the total number of women meeting the incident CIN2+ case definition by the total estimated person-years for each year and age group, and then multiplying by 100,000 to express annual CIN2+ incidence per 100,000 person-years. We used Joinpoint Desktop Software version 4.5.0.1 (National Cancer Institute, Bethesda, MD)[120] to identify significant trends between 2008 and 2018, determined by the best fit log-linear model with the fewest inflection years among each age group. We estimated annual percent changes (APCs, i.e., beta coefficients for each trend) and average annual percent changes (AAPCs, i.e., weighted averages of APCs before and after the detected inflection year) using permutation tests with Poisson variance. With a threshold two-sided α of 0.05, 95% confidence intervals (CI) that excluded 0 were considered statistically significant.

Results

Descriptive Characteristics

A total of 549,671 TennCare-enrolled women aged 18-39 years with 2.3 million person-years of data over an 11-year study period (2008-2018) were evaluated (**Table 4.2**). Over one half were aged 25-34 years. Approximately one-third (34.1%) of the study population

Table 4.2. Characteristics of TennCare-enrolled women aged 18-39 years with at least 1 year of consecutive enrollment from 2008 to 2018 and among those screened for cervical cancer by age group.

Characteristic	Year											
	Overall %	2008 %	2009 %	2010 %	2011 %	2012 %	2013 %	2014 %	2015 %	2016 %	2017 %	2018 %
	All Women											
Total PY	2,332,477	164,695	166,528	179,207	185,209	185,571	184,990	204,357	263,169	300,837	249,254	248,660
Age Group, Years												
18-20	16.7	18.1	19.1	19.1	18.4	17.0	16.3	16.2	15.9	15.3	15.5	14.9
21-24	18.2	20.7	19.3	19.7	18.7	18.1	17.8	18.0	18.9	19.1	15.5	15.5
25-29	24.1	23.6	23.6	23.4	24.0	24.5	24.7	24.7	24.3	24.1	23.4	24.1
30-34	22.2	19.5	20.2	20.6	21.7	22.7	23.2	22.9	22.3	22.3	23.8	23.8
35-39	18.9	18.1	17.9	17.2	17.1	17.7	18.1	18.3	18.6	19.1	21.8	21.7
Race/ethnicity												
White	34.1	41.1	39.8	39.4	37.8	36.0	33.7	31.2	29.6	28.9	31.2	33.9
Black	17.5	23.7	22.9	21.7	20.6	19.7	18.0	15.8	14.5	13.9	14.0	14.9
Hispanic	0.7	0.7	0.7	0.8	0.8	0.8	0.7	0.7	0.6	0.6	0.7	0.8
Other/Unknown	47.7	34.4	36.6	38.2	40.8	43.5	47.5	52.3	55.2	56.6	54.2	50.4
Urbanicity[a]												
MSA	72.9	72.8	72.9	72.9	73.2	73.4	73.2	73.0	72.9	72.9	72.4	72.1
Non-MSA	27.1	27.1	26.9	27.0	26.7	26.5	26.7	27.0	27.1	27.1	27.5	27.9
	Women Screened for Cervical Cancer											
Total PY	759,269	67,322	70,633	74,885	72,018	67,947	62,683	62,908	72,928	79,327	65,500	63,118
Age Group, Years												
18-20	13.8	20.4	20.5	20.1	17.6	14.3	12.5	10.9	9.9	8.8	8.5	7.3
21-24	22.2	24.8	23.1	23.7	23.1	22.7	22.2	21.9	22.6	22.5	18.8	18.3
25-29	26.9	25.1	25.1	24.6	25.9	27.2	27.8	28.2	28.1	28.1	27.9	29.0
30-34	21.4	16.6	18.0	18.6	20.3	21.8	22.7	23.1	22.6	23.2	24.6	25.1
35-39	15.6	13.1	13.3	13.0	13.0	14.0	14.8	16.0	16.8	17.4	20.2	20.3
Race												
White	31.3	37.1	35.8	35.5	33.4	31.3	29.0	27.5	27.5	27.2	28.7	30.9
Black	19.9	23.8	23.2	22.2	21.8	21.1	20.0	18.5	17.0	16.9	16.8	16.9
Hispanic	0.7	0.6	0.6	0.7	0.7	0.7	0.6	0.6	0.7	0.6	0.7	0.8

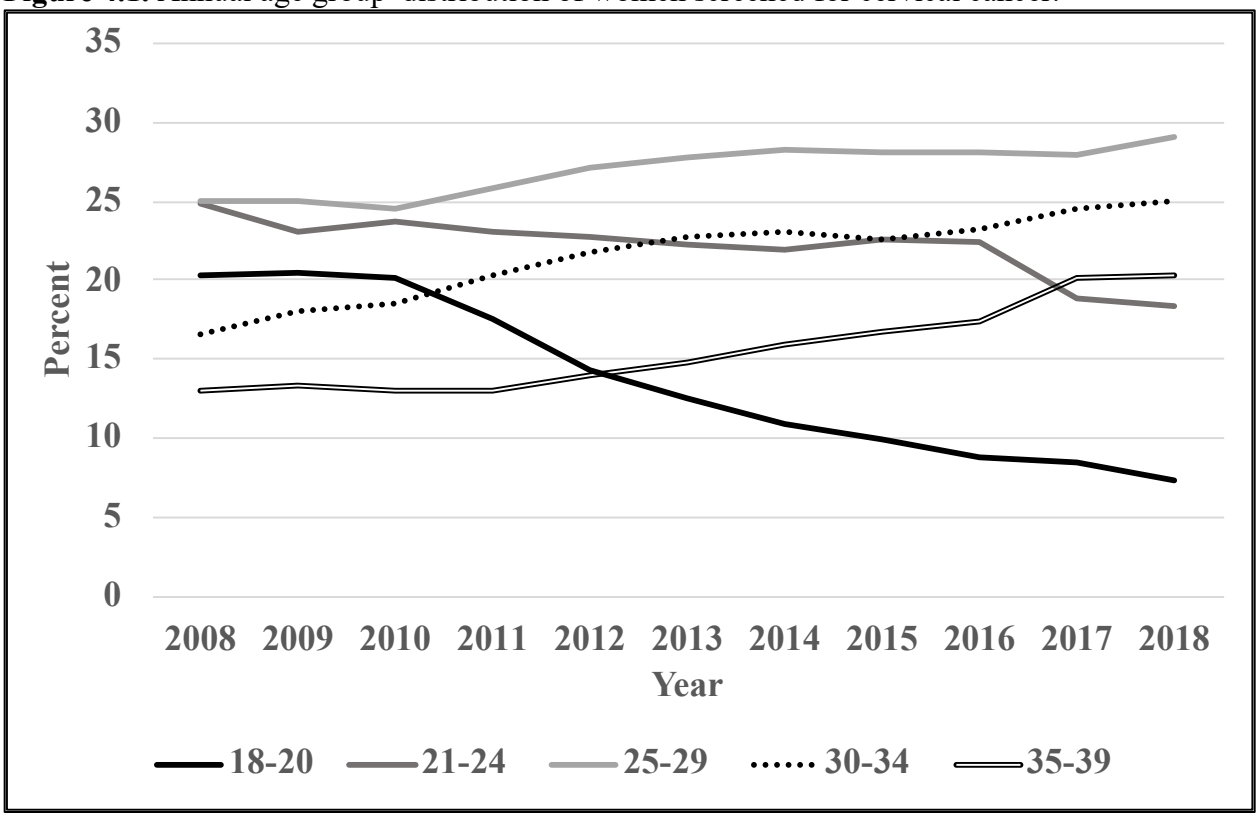
Other/Unknown	48.2	38.5	38.5	41.6	44.1	47.0	50.4	53.4	54.8	55.3	53.8	51.4
Urbanicity[a]												
MSA	74.2	73.4	73.4	73.6	74.0	74.7	75.0	74.9	74.5	74.3	74.2	73.7
Non-MSA	25.8	26.5	26.6	26.4	26.0	25.3	25.0	25.1	25.5	25.6	25.8	26.3

Abbreviations: MSA = Metropolitan Statistical Area; TennCare = Tennessee Medicaid; TN = Tennessee; PY = Person-Years.
[a]Urbanicity was categorized by county of residence using MSA definitions and boundaries set by the US Census Bureau, which classifies MSAs as counties associated with at least one urbanized area that has a population of at least 50,000 persons.

self-identified as White, 17.5% as Black, and 0.7% as Hispanic; the other/unknown race/ethnicity category progressively increased from 34.4% in 2008 to 50.4% in 2018. Most women (72.9%) lived in a metropolitan statistical area (MSA). Age group and urbanicity distributions remained stable across calendar years.

Among women screened for cervical cancer, the urbanicity distribution remained stable, while age group and racial/ethnic distributions changed over time (**Refer to Table 4.2**). Specifically, the proportion of young women screened decreased from 20.4% in 2008 to 7.3% in 2018 for those 18-20 years, and 24.8% in 2008 to 18.3% in 2018 for those 21-24 years (**Figure 4.1**). Conversely, the proportion of older women increased from 16.6% in 2008 to 25.1% in 2018

Figure 4.1. Annual age group[a] distribution of women screened for cervical cancer.



[a]Age groups are expressed in years.

for those 30-34 years, and 13.1% in 2008 to 20.3% in 2018 for those 34-39 years. With respect to race, the other/unknown racial/ethnic group increased from 38.5% in 2008 to 51.4% in 2018.

Trends Among All Women

Among all TennCare-enrolled women from 2008 to 2018, CIN2+ incidence was highest for those aged 21-24 years (646.4/100,000 person-years) and 25-29 years (666.1/100,000 person-years) (**Table 4.3**)**.** Across the 11-year study period, the steepest declines in CIN2+ incidence were observed among the youngest age group, 18-20 years from 720.1/100,000 person-years in 2008 to 24.2/100,000 person-years in 2018 (**Table 4.3, Table 4.4, Figure 4.2**). Significant declines in CIN2+ incidence were also observed among women aged 21-24 years, from 1193.6/100,000 person-years in 2008 to 293.7/100,000 person-years in 2018 (AAPC = -12.9; 95% CI = -22.3, -2.4). Women aged 25-29 years experienced a smaller, yet significant decline in incidence from 845.7/100,000 person-years in 2008 to 526.6/100,000 person-years in 2018 (AAPC = -6.4; 95% CI = -8.1, -4.6). Among older age groups (30-34 years and 35-39 years), trends in CIN2+ incidence from 2008 to 2018 were not significant.

Annual percent changes for trend segments were determined after identifying none, one, or two Joinpoint-detected inflections **(Table 4.4, Figure 4.2**). Among women aged 18-20 years, declines in CIN2+ incidence were not significant until 2010 (APC [2010-2018] = -36.2; 95% CI = -43.8, -27.5). Among women aged 21-24 years, CIN2+ incidence significantly declined from 2012-2016 (APC = -23.6; 95% CI = -41.3, -0.6), but was stable from 2016 to 2018. Among women aged 30-34 years, CIN2+ incidence initially non-significantly increased from 2008 to 2010, then significantly declined from 2010 to 2016 (APC = -7.9; 95% CI = -10.6, -5.3), followed by a stable trend from 2016 to 2018. Similarly, among women aged 35-39 years,

Table 4.3. Age-group-specific annual CIN2+ incidence[a] among TennCare-enrolled women aged 18-39 years and those screened for cervical cancer, 2008-2018.

	Year											
Age Group, Years	Overall IR	2008 IR	2009 IR	2010 IR	2011 IR	2012 IR	2013 IR	2014 IR	2015 IR	2016 IR	2017 IR	2018 IR
All Women												
18-20	244.6	720.1	693.1	557.7	317.3	294.6	146.3	93.6	59.7	13.0	20.7	24.2
21-24	646.4	1193.6	1020.3	1118.2	939.3	835.0	721.1	444.1	396.3	302.4	310.4	293.7
25-29	666.1	845.7	913.9	880.7	754.8	791.9	752.1	569.5	540.3	556.7	493.6	526.6
30-34	463.1	487.9	598.7	611.9	539.6	518.3	494.2	432.1	411.1	361.9	411.3	409.1
35-39	313.9	334.9	402.9	380.2	378.1	332.0	308.4	270.0	273.8	273.5	287.5	306.1
Women Screened for Cervical Cancer												
18-20	877.1	1485.5	1495.9	1268.5	787.7	924.5	562.4	406.6	333.9	71.5	144.0	173.5
21-24	1521.0	2230.1	1909.1	2173.9	1855.1	1726.3	1625.3	1126.9	1110.2	879.4	812.0	864.6
25-29	1716.4	1868.4	1955.1	1930.4	1718.6	1833.2	1877.7	1496.1	1572.1	1649.0	1448.9	1580.5
30-34	1374.8	1318.6	1493.1	1501.1	1405.0	1406.1	1398.0	1309.5	1412.5	1201.7	1373.7	1353.2
35-39	1076.1	1067.9	1196.3	1137.4	1138.7	1071.7	956.6	985.5	1026.2	1055.2	1033.9	1175.9

Abbreviations: CIN = Cervical Intraepithelial Lesion; IR= Incidence Rate; TennCare = Tennessee Medicaid.

[a]Incidence rates are expressed per 100,000 person-years.

Table 4.4. Age-group-specific trends in CIN2+ incidence among TennCare-enrolled women aged 18-39 years and those who were screened for cervical cancer, 2008-2018.

Age Group, Years	Average Annual Percent Change[a]			Annual Percent Change[b]			
	Time Period	AAPC	95% CI	Inflection Year	Time Period	APC	95% CI
All Women							
18-20	2008-2018	-31.9	-38.6, -24.6*	2010	2008-2010	-12.0	-40.2, 29.6
					2010-2018	-36.2	-43.8, -27.5*
21-24	2008-2018	-12.9	-22.3, -2.4*	2012	2008-2012	-7.4	-18.5, 5.3
				2016	2012-2016	-23.6	-41.3, -0.6*
					2016-2018	0.2	-51.1, 105.4
25-29	2008-2018	-6.4	-8.1, -4.6*	--	--	--	--
30-34	2008-2018	-2.0	-4.4, 0.6	2010	2008-2010	9.8	-4.9, 26.7
				2016	2010-2016	-7.9	-10.6, -5.3*
					2016-2018	5.9	-6.3, 19.6
35-39	2008-2018	-1.6	-3.7, 0.6	2010	2008-2010	8.0	-3.9, 21.4
				2014	2010-2014	-9.8	-14.9, -4.4*
					2014-2018	2.6	-0.6, 5.6
Women Screened for Cervical Cancer							
18-20	2008-2018	-20.3	-25.3, -15.0*	--	--	--	--
21-24	2008-2018	-10.2	-12.6, -7.8*	--	--	--	--
25-29	2008-2018	-2.6	-3.9, -1.2*	--	--	--	--
30-34	2008-2018	-1.0	-2.2, 0.3	--	--	--	--
35-39	2008-2018	-0.4	-1.9, 1.1	--	--	--	--

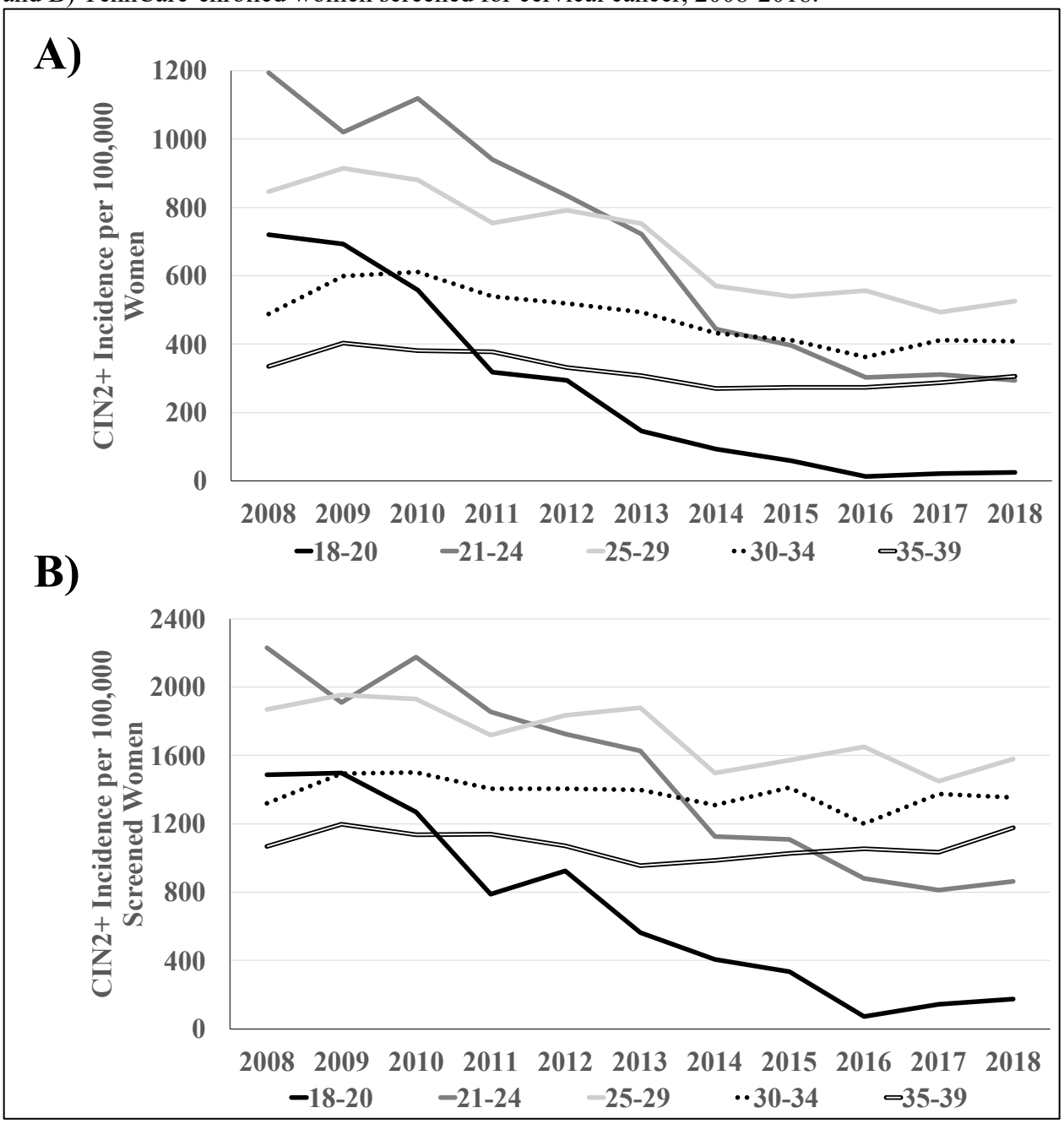
Abbreviations: AAPC= Average annual percent change; APC= Annual Percent Change; CIN = Cervical Intraepithelial Lesion; TennCare = Tennessee Medicaid.

[a]Average annual percent changes are weighted averages of the annual percent changes of all time periods or the average annual percent changes across the entire 11-year study period.

[b]Annual percent changes were determined by the β-coefficient of the best fit log-linear model using a permutation test and Poisson variance; if no inflection year was detected, then only the AAPC is reported because the AAPC is equal to the APC.

*Astericks indicate statistical significance ($p < 0.05$).

Figure 4.2. Age-group-specific[a] annual CIN2+ incidence among A) TennCare-enrolled women and B) TennCare-enrolled women screened for cervical cancer, 2008-2018.



Abbreviations: CIN = Cervical Intraepithelial Neoplasia; TennCare = Tennessee Medicaid.
[a]Age groups are expressed in years.

CIN2+ incidence non-significantly increased from 2008 to 2010, then significantly decreased (APC = -9.8; 95% CI = -14.9, -4.4) from 2010 to 2014, followed by a stable trend from 2014 to 2018.

Trends Among Screened Women

Similar to all TennCare-enrolled women, CIN2+ incidence from 2008 to 2018 among women screened for cervical cancer was highest for those aged 21-24 years (1521.0/100,000 person years) and 25-29 years (1716.4/100,000 person-years) **(Refer to Table 4.3).** CIN2+ trends across the 11-year study period among screened women mirrored that of all women, with the youngest age group showing the steepest declines from 1485.5/100,000 person-years in 2008 to 173.5/100,000 person-years in 2018 (AAPC = -20.3, 95% CI = -25.3, -15.0) (**Refer to Table 4.3, Table 4.4, Figure 4.1**). Significant declines were also observed in women aged 21-24 years, from 2230.1/100,000 person-years in 2008 to 864.6/100,000 person-years in 2018 (APC = -10.2; 95% CI = -12.6, -7.8), and in women aged 25-29 years, from 1868.4/100,000 person-years in 2008 to 1580.5/100,000 person-years in 2018 (APC = -2.6; 95% CI = -3.9, -1.2). CIN2+ incidence among older age groups (30-34 years and 36-39 years) were not significant. Among screened women, no significant trend shifts were detected for any age group.

Discussion

Among a population with sub-optimal HPV vaccination coverage (62% initiation and 43% up-to-date in 2019),[18,118] we observed a reduction in the incidence of cervical premalignant lesions (CIN2+) associated with HPV vaccine introduction. Most notable declines were in young women aged 18-20 years, an age group most likely to have benefited from the HPV vaccine's

approval in 2006 and the Advisory Committee on Immunization Practices' recommendations for routine vaccination among adolescents aged 11-12 years.[52] Declines were also observed among women aged 21-24 and 25-29 years, while stable trends were detected among older women in age groups that were less likely to have benefited from the HPV vaccine (30-34 and 35-39 years). After restricting to women who were screened for cervical cancer, declines were still observed in young women (aged 18-20, 21-24, and 25-29 years), suggesting that these declines were not simply due to decreases in screening.

Cervical cancer screening patterns have changed over time, contributing to changes in CIN2+ detection. Specifically, in our study population, we observed decreases in the age group distribution of screened women who were aged 18-20 years from 20.4% in 2008 to 7.3% in 2018. These declines were expected given changes in cervical screening guidelines over time. Prior to 2009, cervical cancer screening was recommended for women after sexual debut, regardless of age; however, in 2009, the American College of Obstetricians and Gynecologists recommended screening to begin at age 21 years (**Refer to Chapter II, Section: "Cervical Cancer Prevention: Screening and The HPV Vaccine"**).[121] However, because HPV infections are common and typically clear spontaneously within 24 months of infection,[2,28] guidelines were updated to recommend against screening for women younger than 21 years to protect adolescents and young women from unnecessary invasive gynecologic procedures that could put them at risk for cervical damages.[45] Therefore, declines in screening among women younger than 21 years are reasonable and should be considered when interpreting general CIN2+ trends. Conversely, the proportion of older women (age 30-34 and 35-39 years), increased over time, most notably from 13.1% in 2008 to 20.3% in 2018 for those aged 34-39 years. Increased prevalence of

screening among older women is likely due to increased adoption of cervical cancer screening guidelines.

To account for changes in screening patterns over time, we examined age-group-specific CIN2+ incidence among a subpopulation of women who were screened for cervical cancer. Among young, screened women aged 18-20, 21-24, and 25-29 years, declining trends in CIN2+ incidence were statistically significant, yet less pronounced, compared to declines among all women; this is likely because trends among all women are due to both declines in screening and HPV vaccine impact, while trends among screened women removes the confounding effect of declines in screening.

Declines in CIN2+ incidence were expected among younger women, particularly because of increasing HPV vaccination trends in Tennessee. In 2019, HPV vaccine initiation among Tennessee females aged 13-17 years was 70%, rising from just 30% in 2008.[18] Conversely, stable trends in CIN2+ incidence were expected among older women because of vaccine ineligibility at the time of the vaccine's first approval in 2006,[52] and because of low HPV vaccination coverage among even age-eligible adults.[122] It was only in 2019 that the Advisory Committee on Immunization Practices recommended that patients aged 27-45 years who are at-risk for a new HPV infection may consider getting the vaccine through shared decision-making with their provider.[57] Despite this recommendation, most insurance companies do not currently cover the vaccine for persons older than 26 years of age.[123] Therefore, continued monitoring of the vaccine's impact on HPV-related health outcomes is warranted in the future.

Our results corroborate findings of other ecologic studies demonstrating HPV vaccine impact on reducing cervical premalignant lesions among population-based surveillance sites across the US, which have also reported significant declines in CIN2+ incidence among young

women and no significant declines among older women.[87–92] Compared to populations with cervical biopsy surveillance data, we found similar declining trends in CIN2+ incidence using claims-based models. The most recent analysis by the HPV-IMPACT monitoring project[92] reported average annual decreases in CIN2+ incidence of 38% and 15% among women aged 18-19 and 20-24 years, respectively, from 2008 to 2016, compared to average annual decreases of 32% and 13% among women aged 18-20 and 21-24 years, respectively, from 2008 to 2018 in our study.

Data from the HPV-IMPACT monitoring project also showed declines in CIN2+ incidence among women aged 25-29 from 2008 to 2016; however, these declines were not significant.[92] With the inclusion of more recent data through 2018, we observed significant declines in CIN2+ incidence among women aged 25-29 years in our study population. As younger cohorts begin to age and vaccination rates continue to rise, increased evidence of the HPV vaccine's population impact may continue to become more apparent, as demonstrated by the findings from our study.

Our study has a few limitations. We did not take into account individual-level vaccination data because we aimed to capture both direct effects (from vaccination) and indirect effects (from herd effects) of the vaccine, which would only be observable through an ecologic perspective. Additionally, there are no current cervical screening guidelines based on vaccination status, as all age-eligible women are still encouraged to be screened. Thus, examining population-based CIN2+ trends is still useful regardless of vaccination status. Given that this was an ecologic analysis, we cannot definitively conclude that CIN2+ incident trends were directly due to vaccination; however, our temporal findings and correlative results from different age groups support our conclusion of the impact of the HPV vaccine on the population. Our study

only included women enrolled in the Tennessee Medicaid program so, our results may not be generalizable to populations with different socio-demographics or populations in other regions, such as non-Medicaid populations. Because of eligibility for the Vaccines for Children program among low-income individuals, adolescents insured by Medicaid have historically had higher vaccination coverage compared to the general public[124]; thus, evidence of HPV vaccine impact on reducing CIN2+ incidence may be more prominent among the Medicaid population compared to the general population.

Our study has several strengths. To our knowledge, this is the first claims-based study in the US to examine trends in CIN2+ incidence across both ICD-9 and ICD-10 coding eras (2008-2018) using a validated model. Prior claims-based studies have been unable to assess trends past 2015[22] because the transition from ICD-9 to ICD-10 occurred on October 1st, 2015, and the discriminative ability of ICD-9 and ICD-10 codes to identify CIN2+ events was unknown. Since then, we validated claims-based models to classify CIN2+ events in both ICD-9 and ICD-10 eras (**Refer to Chapter III**), reporting no significant differences in discriminative performance; therefore, we were able to include more recent CIN2+ data in our analyses to expand upon prior research and CIN2+ surveillance. Moreover, utilizing claims-based data is an efficient and timely method to monitor HPV vaccination impact through capturing frequency of clinical events such as CIN2+. This study is also the first to examine CIN2+ incident trends in a US population outside of the catchment areas with population-based cervical biopsy data which include the New Mexico HPV Pap registry and the HPV-IMPACT monitoring project, where the most recent CIN2+ trends have only been analyzed through 2016.

In summary, both HPV vaccination and cervical cancer screening are methods to decrease incidence of cervical precancers. While cervical screening adoption is increasing, HPV

vaccination coverage remains low in certain states, including Tennessee. However, our results suggest evidence of HPV vaccine impact in a population with low vaccination coverage, such as Tennessee.

CHAPTER V

AGE, PERIOD, AND COHORT EFFECTS ON TRENDS IN CERVICAL PREMALIGNANT LESIONS IN URBAN AND RURAL AREAS, TENNESSEE MEDICAID 2008-2018

Abstract

Disparities in human papillomavirus (HPV) vaccination exist between urban (metropolitan statistical areas [MSAs]) and rural (non-MSAs) regions; thus, vaccine impact may differ by urbanicity. We examined trends in cervical premalignant lesions, including age, period, and birth cohort effects, in MSAs and non-MSAs among all Tennessee Medicaid (TennCare)-enrolled women aged 18-39 years and the subset of women who were screened for cervical cancer. Using TennCare claims data, we identified annual incidence of age-group-specific (18-20, 21-24, 25-29, 30-34, and 35-39 years) cervical premalignant lesions, including cervical intraepithelial neoplasia grades 2 or 3 or adenocarcinoma *in situ* (together referred to as CIN2+) from 2008 to 2018. Joinpoint regression was used to identify trends over time. Age-period-cohort Poisson regression models were used to evaluate age, period, and cohort effects. All analyses were performed with stratification by urbanicity (MSA vs non-MSA), determined by county of residence, for all women and for the annual subset with procedure codes indicating cervical cancer screening. From 2008 to 2018, a total of 11,243 incident CIN2+ events (7,956 in MSAs; 3,287 in non-MSAs) were identified among TennCare-enrolled women aged 18-39 years. CIN2+ incident trends were similar between women in MSAs and non-MSAs, with the largest declines among women aged 18-20 (MSA average annual percent change [AAPC] = -30.4, 95%

confidence interval [CI] = -35.4, -25.0; non-MSA AAPC = -30.9, 95% CI = -36.8, -24.5) and 21-24 (MSA AAPC = -14.8, 95% CI = -18.1, -11.3; non-MSA AAPC = -15.1, 95% CI = -17.9, -12.2) years. Trends were generally stable among older women. Trends were largely driven by age and cohort effects. Patterns were consistent among screened women. Significant declines in CIN2+ incidence were observed regardless of urbanicity, suggesting HPV vaccine impact in populations with varying vaccination rates.

Introduction

The current nonavalent human papillomavirus (HPV) vaccine can prevent up to 90% of cervical cancer cases.[10] Despite being vaccine-preventable, cervical cancer remains the fourth most common incident cancer in women worldwide, contributing to over 300,000 cervical cancer-related deaths annually.[11] Since 2006, when the HPV vaccine was first introduced in the United States (US),[52] studies have demonstrated reductions in surrogates for cervical cancer, such as cervical premalignant lesions, including cervical intraepithelial lesions grades 2 and 3, and adenocarcinoma *in situ* (together referred to as CIN2+).[87–92,94] Evaluating the HPV vaccine's impact on CIN2+ as an intermediate outcome to cervical cancer is more efficient than evaluating the vaccine's impact on cancer, which can take decades to develop compared to just a few years for CIN2+.[28,125]

While studies have documented overall declines in HPV-related adverse health outcomes among younger age groups who were most likely to have benefited from the introduction of the vaccine, disparities in HPV vaccination exist between urban and rural geographical regions,[126–131] raising concern that the vaccine may impact these populations disproportionately. Specifically, HPV vaccine initiation among adolescents aged 13-17 years in urban areas, known as metropolitan statistical areas (MSAs), ranged from 49% to 74% from 2013 to 2019, compared to 37% to 64% in rural areas (non-MSAs) in the US.[18,130] Despite increasing adolescent HPV vaccination within urban and rural areas over time, annual vaccination coverage is significantly lower in rural areas compared to urban areas.[130]

Similar geographic disparities have also been demonstrated among adults, with a 42% lower odds of HPV vaccine initiation among adults aged 18-26 years in rural areas compared to urban areas across 8 US states.[131] These geographic differences may be attributed to rural areas

having more barriers to vaccination,[68] including lack of health care access,[129] lack of knowledge and awareness of HPV and its link to cancer,[69,70] increased negative community messaging regarding the vaccine,[70] and religious and cultural beliefs that may not support vaccination.[71,72] Given these large geographic disparities in both adolescent and adult HPV vaccination, examining whether urbanicity has modified the vaccine's impact on reducing HPV-related outcomes is important for informing national HPV vaccination guidelines and public health interventions to improve vaccination rates.

A few studies have examined trends in HPV-associated health outcomes by urbanicity[84,87,132,133]; of these, most have focused on the vaccine's impact on reducing anogenital warts.[84,132,133] Only one study has assessed trends in CIN2+ incidence by urbanicity, reporting significant declines in CIN2+ incidence from 2008 to 2011 among women aged 21-24 years in both urban and rural counties in Connecticut.[87] However, this study did not examine CIN2+ trends by urbanicity for other age groups and did not control for possible changes in cervical cancer screening over time. To better understand the HPV vaccine's impact on CIN2+ by urbanicity, we examined secular trends in CIN2+ incidence, including age, period, and birth cohort effects, from 2008 through 2018 in urban and rural areas in Tennessee among 1) women aged 18-39 years enrolled in the Tennessee Medicaid program and 2) a subset of those women who were screened for cervical cancer to control for changes in screening rates over time.

Methods

Study Population

We used data from the Tennessee Medicaid (TennCare) program to identify women aged 18-39 years who were enrolled in TennCare from 2008 to 2018. For the subset of screened women, we used billing codes for an HPV screening examination, Papanicolaou (Pap) test, and HPV DNA test to identify TennCare-enrolled women who were screened for cervical cancer at least once during any given year (**Refer to Chapter IV, Section: "Methods," Subsection: "Study Population"**). To be able to examine CIN2+ trends by urbanicity, women with missing data on residence were excluded.

Urbanicity was categorized by county of residence using the MSA definitions and boundaries set by the US Census Bureau. Urban areas, or metropolitan statistical areas (MSAs), were Tennessee counties with at least one area with a population of at least 50,000 persons, while all other counties were considered rural areas (non-MSAs).[134] This study was considered public health surveillance and not human research by the Institutional Review Boards at Vanderbilt University and the Tennessee Department of Health. This research activity was reviewed and approved by the Tennessee Department of Finance and Administration Division of TennCare.

Definition of Incident CIN2+ Events

Methods for identifying CIN2+ events, including CIN2, CIN3, and adenocarcinoma *in situ*, were conducted in the same manner as in Aim 2 (**Refer to Chapter IV, Section: "Methods," Subsection: Definition of Incident CIN2+ Events**). Briefly, we used our validated claims-based model (**Refer to Chapter III**)[105] to search for women with cervical diagnostic procedures who were consecutively enrolled in TennCare for at least one year from their diagnostic procedure date. Billing codes were searched for a relevant diagnosis (CIN2+ tissue,

non-specific CIN, high-grade cervical intraepithelial lesion, CIN1, low-grade cervical intraepithelial lesion, and atypical squamous cells of undetermined significance), screening (HPV screening examination, Pap test, and HPV DNA test), treatment procedure, and cervical or vaginal biopsy to calculate CIN2+ prediction scores (**Refer to Table 4.1**). Prediction scores were then transformed into predicted probabilities, and CIN2+ events were determined by probabilities of at least 0.5. More detailed information on the model building and validation is presented elsewhere.[105] The date of the first high probability cervical diagnostic procedure was considered the CIN2+ incident diagnosis date.

For the subset of women who were screened for cervical cancer, CIN2+ events were counted if the screening date was within one year of the diagnostic procedure date. Incident CIN2+ events were model-identified events that did not have another CIN2+ event for at least one year prior to the diagnostic procedure date. Incident CIN2+ events in MSAs and non-MSAs were only counted if the woman resided in an MSA or non-MSA county, respectively, on the date of their cervical diagnostic procedure.

Denominator and Rates

Annual person-years for each age group (18-20, 21-24, 25-29, 30-34, and 35-39 years) was estimated by counting the total number of women who were enrolled in TennCare on July 1st of each year with at least one year of consecutive enrollment. For example, total person-years for 2008 comprised of TennCare-enrolled women who were continuously enrolled between July 1st, 2007 and July 1st, 2008. Screened person-years included the subset of total women who had a least one cervical cancer screening code during the year prior to July 1 of each year. Only women residing in an MSA or non-MSA county on July 1st of each year were counted towards

the person-time estimation for MSA and non-MSA populations, respectively. Annual CIN2+ incidence rates per 100,000 person-years were calculated by dividing the total number of women meeting the incident CIN2+ event definition by the estimated person-time for each year and age group among all TennCare-enrolled women and those residing in MSA and non-MSA counties, and then multiplying by 100,000.

Joinpoint Trend Analyses

We identified CIN2+ incident trends and significant changes in trends (i.e., changes in slope) over time using the Joinpoint Desktop Software version 4.5.01 (National Cancer Institute, Bethesda, MD), which calculated average percent changes (APCs, beta coefficients for each trend) and average annual percent changes (AAPCs, weighted averages of APCs) from 2008 to 2018, by urbanicity.[120] Using grid search and permutation tests, we allowed for a maximum of two joinpoints detected per model with uncorrelated errors to determine the best fit log-linear models. Using a two-sided α threshold of 0.05, 95% confidence intervals (CI) that excluded 0 were considered statistically significant. Because Joinpoint analyses were performed on TennCare-enrolled women overall (not stratified by urbanicity) in Aim 2 (**Refer to Chapter IV**), in this chapter, only Joinpoint results stratified by urbanicity are discussed; however, the results are still presented in the tables and figures to be able to easily compare the combined results with the urbanicity-stratified results.

Age-Period-Cohort Analyses

Age (A), period (P), and birth cohort (C) effects were evaluated using the Clayton and Schiffler modeling approach for age-period-cohort analyses. [135,136] The model building process

begins with an age model, then adds a "drift" parameter (i.e., sum of the linear period and cohort effects).[135,136] Derivatives of the drift parameter are estimated and regressed on period and cohort to estimate their effects on trends. [135,136] We used this approach to derive the following submodels: 1) age, 2) age-drift, 3) age-cohort, 4) age-period, and 5) age-period-cohort.

The general multiplicative formula for the age-period-cohort models were based on Poisson regression to derive incidence rates ($\log(\lambda\ A, P)$) at age A in a period P for persons in birth cohort C: $\log(\lambda\ A, P)) = f(A) + g(P) + h(C)$, where A, P, and C represent the mean age, period, and birth cohort for the observational units, respectively, and *f*, *g*, and *h*, represent the functions for each effect.[137] Synthetic birth cohort groups were calculated by subtracting the midpoint of each age group (18-20, 21-24, 25-29, 30-34, and 35-39 years) from each one-year period (2008, 2009, 2010, 2011, 2012, 2013, 2014, 2015, 2016, 2017, and 2018).

Because of linear dependency between age, period, and cohort effects (C=P-A), the simultaneous linear effects of all three effects cannot be estimated; therefore, any parameterization of the age-period-cohort model includes two fixed levels and one slope among the three functions.[135,136] We parameterized our models based on the maximum likelihood of the age-period-cohort model, considering age effects as incidence rates for the reference period (2008) and period effects as rate ratios relative to the reference period (2008).[138] Cohort effects were constrained to be 0 on average with 0 slope.[138] We estimated annual percent changes (EAPC), or the overall linear trends, from the net drift in the age-drift models.

Model goodness-of-fit was examined using residual deviance statistics. Using the Clayton and Schiffler approach,[135,136] model fit was assessed for each submodel, comparing each iterative model to the primary model of age alone by sequentially adding cohort and period effects to determine whether these added parameters significantly improved model fit. Then,

model fit was deductively assessed by iteratively removing parameters and testing whether this significantly deteriorated model fit. We tested for significant differences in residual deviance of each pairwise comparison using chi-squared tests. All age-period-cohort analyses were conducted using the *apc.fit* function from the Epi package in R (version 3.6.2).[137,138] P-values <0.05 were considered statistically significant.

Results

Age-Specific Trends in CIN2+ Incidence

Between 2008 and 2018, we identified 7,956 incident CIN2+ events in MSA counties, compared to 3,287 incident CIN2+ events in non-MSA counties (total = 11,243) (**Table 5.1**). Of the total number of events, 10,540 (94%) women (7,470 MSA; 3,070 non-MSA) had a cervical screening code identified in the year prior to their incident event.

Among women residing in MSAs, CIN2+ incidence significantly declined from 2008 to 2018 for those aged 18-20 years (AAPC = -30.4; 95% CI = -35.4, -25.0), 21-24 years (AAPC = -14.8; 95% CI = -18.1, -11.3), and 35-39 years (AAPC = -3.9; 95% CI = -5.8, -1.9) (**Table 5.2, Figure 5.1**). However, after restricting to screened women, declines in CIN2+ were only observed for the youngest three age groups. Several Joinpoint-detected inflections (e.g., time points where there are significant changes in slopes across time periods) were identified (**Table 5.3**). Among women residing in MSAs, inflections were only observed for those aged 30-34 years, with significant increases in CIN2+ incidence from 2008 to 2010 (APC=13.3; 95% CI=1.5, 26.4), followed by significant decreases from 2010-2016 (APC=-8.0; 95% CI=-9.9, -6.1). This pattern was mirrored, yet less pronounced and non-significant, among screened

Table 5.1. Annual age-group-specific CIN2+ incidence per 100,000 person-years among all women enrolled in Tennessee Medicaid and those screened for cervical cancer, by urbanicity.

	Overall					MSA[a] Residence					Non-MSA[a] Residence				
Age (yrs)	18-20	21-24	25-29	30-34	35-39	18-20	21-24	25-29	30-34	35-39	18-20	21-24	25-29	30-34	35-39
								All Women							
Total PY	388284	423390	560644	518451	440074	279140	308959	412173	380147	318804	109144	114431	148471	138304	121270
Events	952	2756	3747	2403	1385	640	1909	2664	1713	1030	312	847	1083	690	355
Year															
2008	720.5	1197.9	849.1	488.3	335.2	720.5	1098.7	771.7	462.4	352.2	720.5	1471.7	1076.1	557.1	292.7
2009	693.7	1031.3	917.4	596.3	403.2	669.3	946.9	869.1	529.2	420.0	755.5	1267.8	1056.5	779.7	360.1
2010	557.9	1136.2	888.6	612.1	383.6	477.8	1113.3	827.0	584.6	362.9	758.7	1200.4	1061.4	689.1	436.0
2011	317.4	945.8	755.5	547.3	384.6	289.1	887.2	706.2	542.1	435.4	390.8	1110.6	895.0	561.9	251.2
2012	294.8	835.6	788.3	518.6	332.1	284.0	808.0	779.5	521.1	340.0	323.4	913.9	813.5	511.4	311.3
2013	149.6	733.7	755.2	494.4	308.5	133.7	692.3	683.4	468.2	323.2	190.6	851.0	959.4	568.1	270.1
2014	96.7	444.4	568.1	428.1	270.2	92.7	441.5	574.8	422.4	295.5	106.7	452.3	549.0	443.8	203.7
2015	59.8	398.6	542.3	411.5	276.0	49.9	369.2	513.7	397.5	275.4	84.7	476.1	622.0	450.1	277.4
2016	13.0	304.2	557.2	362.3	273.7	12.0	276.6	566.1	349.2	293.4	15.5	376.1	532.6	398.3	220.1
2017	20.7	308.0	502.6	415.1	287.6	14.4	318.4	518.8	407.6	291.9	37.0	281.0	460.1	435.4	276.2
2018	24.3	293.9	527.0	406.2	304.4	26.3	259.2	534.5	422.5	281.1	19.0	380.1	507.5	363.3	366.6
								Women Screened for Cervical Cancer							
Total PY	104774	168751	204585	162692	118383	75181	123842	153229	122514	88209	29593	44909	51356	40178	30174
Events	921	2581	3522	2239	1277	616	1798	2521	1591	944	305	783	1001	648	333
Year															
2008	1485.6	2236.2	1874.6	1319.0	1068.2	1497.6	2081.3	1702.6	1221.0	1080.7	1456.2	2660.4	2411.1	1601.1	1035.2
2009	1496.2	1927.8	1961.1	1485.7	1196.7	1451.8	1787.5	1867.3	1317.5	1206.8	1603.8	2319.1	2239.1	1979.6	1170.1
2010	1268.5	2208.0	1946.9	1501.1	1147.7	1106.1	2170.3	1823.2	1377.9	1033.4	1663.3	2313.6	2315.5	1872.1	1462.7
2011	787.8	1861.3	1719.2	1425.5	1159.9	696.1	1734.7	1569.5	1373.9	1281.5	1015.4	2216.6	2169.7	1583.3	815.7
2012	924.8	1726.3	1822.8	1406.1	1071.9	892.1	1695.8	1758.4	1385.4	1060.7	1011.2	1814.1	2023.1	1470.2	1104.3
2013	575.2	1639.8	1883.4	1398.1	956.6	506.9	1547.3	1683.0	1277.9	955.4	760.8	1914.8	2497.7	1787.3	959.9
2014	421.2	1127.1	1490.4	1302.9	985.6	401.7	1118.5	1495.3	1223.7	1054.9	472.2	1151.0	1475.1	1562.5	782.5
2015	333.9	1110.4	1581.9	1412.8	1034.5	288.5	999.2	1474.9	1300.1	1016.9	452.7	1409.1	1899.6	1770.4	1088.4
2016	71.5	885.1	1649.2	1207.2	1055.2	58.5	822.7	1626.2	1137.2	1079.2	107.4	1046.9	1718.0	1418.0	980.7
2017	144.0	812.0	1476.4	1386.5	1034.0	99.9	835.8	1500.3	1298.8	1016.1	258.1	748.3	1409.6	1651.7	1089.7
2018	173.5	864.7	1580.8	1340.7	1168.1	178.8	739.8	1627.1	1393.0	1041.0	159.1	1193.5	1453.7	1191.6	1561.0

Abbreviations: CIN = Cervical Intraepithelial Lesion; IR = Incidence Rate; MSA = Metropolitan Statistical Area; PY = Person-Years.
[a]Urbanicity was categorized by county of residence using MSA definitions and boundaries set by the US Census Bureau, which classifies MSAs as counties associated with at least one urbanized area that has a population of at least 50,000 persons.
[b]Incidence rates are expressed per 100,000 person-years.

Table 5.2. Average annual percent changes in age-group-specific CIN2+ incidence among all women enrolled in Tennessee Medicaid and those screened for cervical cancer, by urbanicity, 2008-2018.

	Overall		MSA[a] Residence		Non-MSA[a] Residence	
Age, Years	AAPC[b]	95% CI	AAPC[b]	95% CI	AAPC[b]	95% CI
All Women						
18-20	-31.8*	-38.4, -24.5	-30.4*	-35.4, -25.0	-30.9*	-36.8, -24.5
21-24	-14.4*	-19.6, -8.8	-14.8*	-18.1, -11.3	-15.1*	-17.9, -12.2
25-29	-6.4*	-8.0, -4.6	-5.3	-7.1, -3.6	-8.8*	-11.3, -6.3
30-34	-2.0	-4.4, 0.6	-0.8	-2.6, 1.1	-6.2*	-8.5, -3.8
35-39	-1.6	-3.9, 0.8	-3.9*	-5.8, -1.9	-1.5	-6.1, 3.2
Women Screened for Cervical Cancer						
18-20	-20.2*	-25.2, -14.9	-21.1*	-26.1, -15.8	-19.8*	-26.5, -12.4
21-24	-10.3*	-12.7, -7.8	-10.4*	-13.2, -7.6	-10.0*	-12.7, -7.1
25-29	-2.5*	-3.9, -1.2	-2.6*	-2.9, -0.2	-4.9	-7.5, -2.3
30-34	-1.0	-2.2, 0.3	1.3	-2.4, 5.2	-2.5	-5.0, 0.1
35-39	-0.5	-2.0, 1.0	-1.1	-2.7, 0.5	1.1	-3.4, 5.8

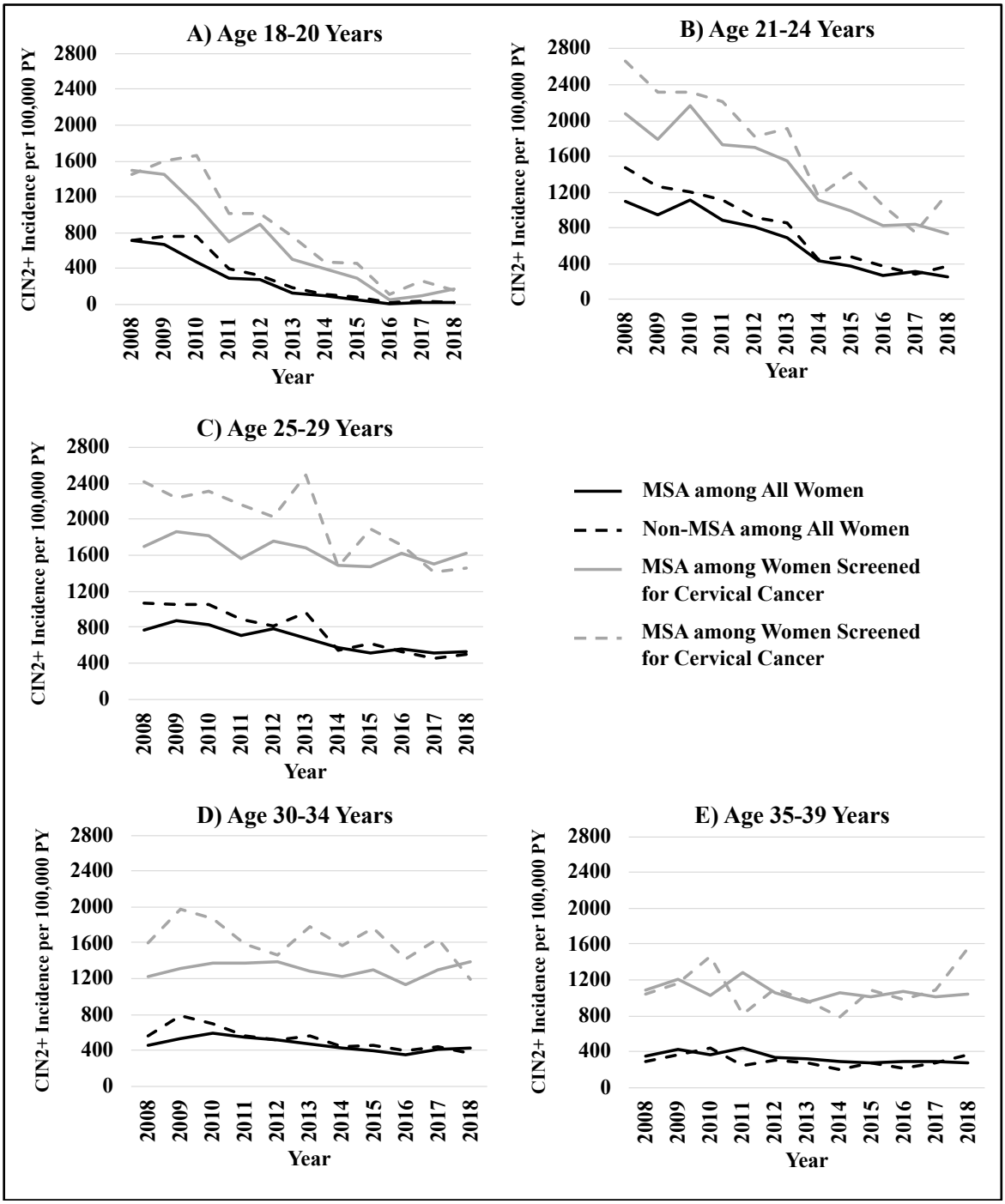
Abbreviations: AAPC= Average annual percent change; CIN = Cervical Intraepithelial Lesion; MSA = Metropolitan Statistical Area.

[a]Urbanicity was categorized by county of residence using MSA definitions and boundaries set by the US Census Bureau, which classifies MSAs as counties associated with at least one urbanized area that has a population of at least 50,000 persons.

[b]Average annual percent changes are weighted averages of annual percent changes from 2008 to 2018.

*Astericks indicate statistical significance ($p < 0.05$).

Figure 5.1. Annual CIN2+ incidence per 100,000 person-years among all women enrolled in Tennessee Medicaid and those screened for cervical cancer who resided in MSAs[a] versus non-MSAs[a] aged A) 18-20 years, B) 21-24 years, C) 25-29 years, D) 30-34 years, and E) 35-39 years, 2008-2018.



Abbreviations: CIN = Cervical Intraepithelial Lesion; MSA = Metropolitan Statistical Area; PY = Person-Years

[a]Urbanicity was categorized by county of residence using MSA definitions and boundaries set by the US Census Bureau, which classifies MSAs as counties associated with at least one urbanized area that has a population of at least 50,000 persons.

Table 5.3. Annual percent changes in age-group-specific CIN2+ incidence among all women enrolled in Tennessee Medicaid and those screened for cervical cancer, by urbanicity, 2008-2018.

Age, Years	Overall				MSA[a] Residence				Non-MSA[a] Residence			
	Inflection Year	Time Period	APC[b]	95% CI	Inflection Year	Time Period	APC[b]	95% CI	Inflection Year	Time Period	APC[b]	95% CI
						All Women						
18-20	2010	2008-2010	-12.0	-40.2, 29.6	--	2008-2018	-30.4*	-35.4, -25.0	2010	2008-2010	0.2	-29.4, 42.4
		2010-2018	-36.0*	-43.7, -23.4						2010-2018	-37.0*	-43.4, -29.9
21-24	2010	2008-2010	-0.7	-29.2, 39.2	--	2008-2018	-14.8*	-18.1, -11.3	--	2008-2018	-15.1*	-17.9, -12.2
		2010-2018	-17.5*	-21.6, -13.2								
25-29	--	2008-2018	-6.4*	-8.0, -4.6	--	2008-2018	-5.3*	-7.1, -3.6	--	2008-2018	-8.8*	-11.3, -6.3
30-34	2010	2008-2010	10.0	-4.8, 27.1	2010	2008-2010	13.3*	1.5, 26.4	--	2008-2018	-6.2*	-8.5, -3.8
	2016	2010-2016	-8.0*	-10.6, -5.4	2016	2010-2016	-8.0*	-9.9, -6.1				
		2016-2018	5.7	-6.5, 19.6		2016-2018	9.0	-0.2, 19.2				
35-39	2010	2008-2010	8.6	-4.7, 23.7	--	2008-2018	-3.9*	-5.8, -1.9	--	2008-2018	-1.5	-6.1, 3.2
	2014	2010-2014	-10.0*	-15.6, -3.9								
		2014-2018	2.4	-1.1, 6.1								
						Women Screened for Cervical Cancer						
18-20	--	2008-2018	-20.2*	-25.2, -14.9	--	2008-2018	-21.1*	-26.1, -15.8	2010	2008-2010	4.5	-26.2, 47.8
										2010-2018	-24.9*	-32.5, -16.5
21-24	--	2008-2018	-10.3*	-12.7, -7.8	--	2008-2018	-10.4*	-13.2, -7.6	--	2008-2018	-10.0	-12.7, -7.1
25-29	--	2008-2018	-2.5*	-3.9, -1.2	--	2008-2018	-2.6*	-2.9, -0.2	--	2008-2018	-4.9*	-7.5, -2.3
30-34	--	2008-2018	-1.0	-2.2, 0.3	2010	2008-2010	7.5	-13.3, 33.3	--	2008-2018	-2.5	-5.0, 0.1
					2016	2010-2016	-2.8	-6.7, 1.2				
						2016-2018	8.4	-9.4, 29.7				
35-39	--	2008-2018	-0.5	-2.0, 1.0	--	2008-2016	-1.1	-2.7, 0.5	--	2008-2018	1.1	-3.4, 5.8

Abbreviations: APC= Annual Percent Change; CIN = Cervical Intraepithelial Lesion; MSA = Metropolitan Statistical Area.

[a]Urbanicity was categorized by county of residence using MSA definitions and boundaries set by the US Census Bureau, which classifies MSAs as counties associated with at least one urbanized area that has a population of at least 50,000 persons.

[b]Annual percent changes were determined by the β-coefficient of the best fit log-linear model using a permutation test and Poisson variance for each time period detected by Joinpoint.

*Astericks indicate statistical significance ($p < 0.05$).

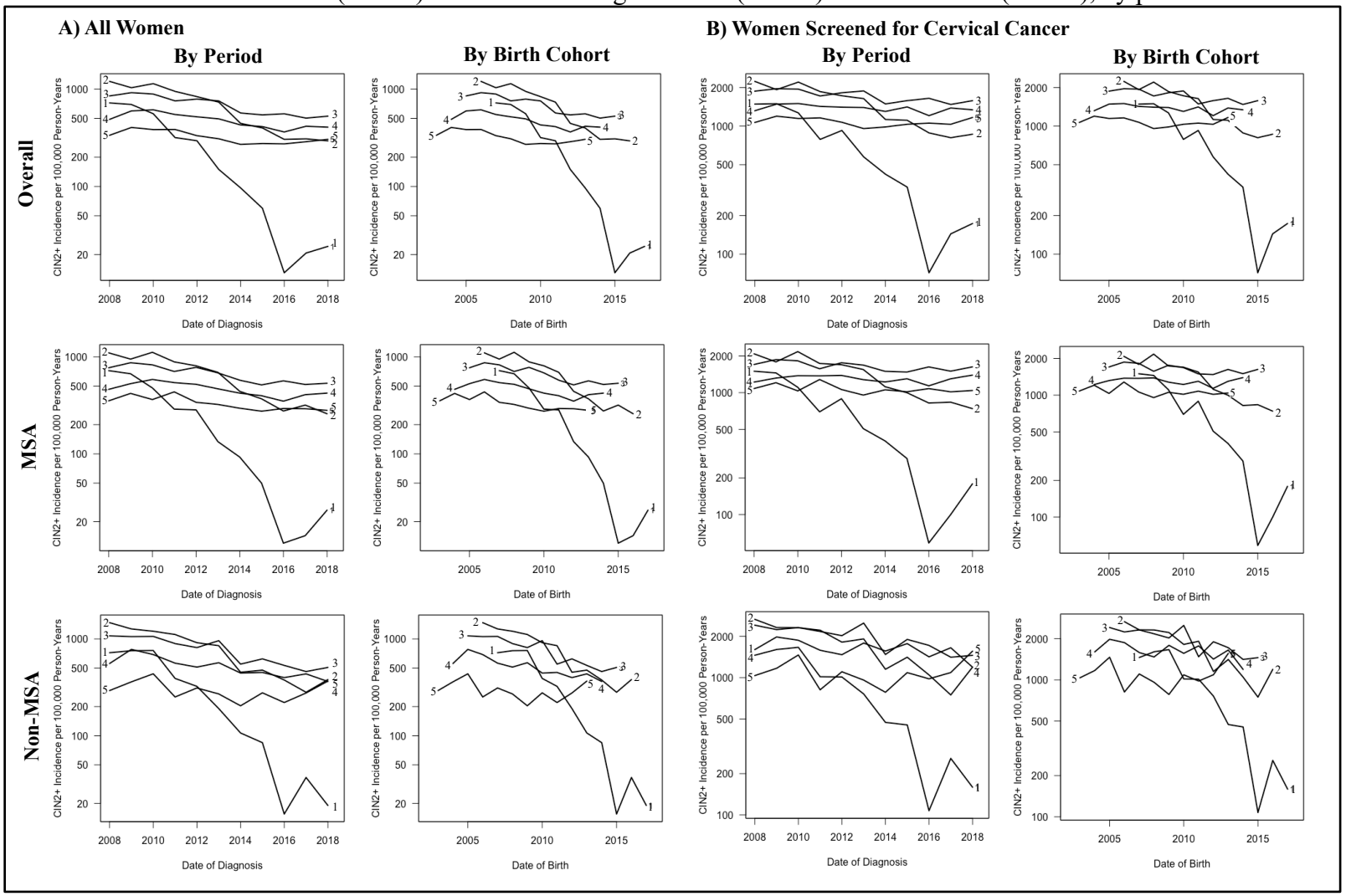
women aged 30-34 years who resided in MSAs.

Among women residing in non-MSAs, CIN2+ incidence significantly declined for those aged 18-20 years (AAPC = -30.9; 95% CI = -36.8, -24.5), 21-24 years (AAPC = -17.9, 95% CI = -17.9, -12.2), 25-29 years (AAPC = -8.8; 95% CI = -11.3, -6.3), and 30-34 years (AAPC =-6.2; 95% CI =-8.5, -3.8) (**Refer to Table 5.2, Figure 5.1**). After restricting to women who were screened for cervical cancer, significant declines were only observed for the two youngest age groups. Among women residing in non-MSAs, an inflection was only observed among those aged 18-20 years, with stable trends from 2008 to 2010, followed by significant declines in CIN2+ incidence from 2010 to 2018 (APC = -37.0; 95% CI = -43.4, -29.9) (**Refer to Table 5.3**). This pattern was similar to that of screened women aged 18-20 years who resided in non-MSAs.

Descriptive Age, Period, Cohort Effects

Among TennCare-enrolled women overall (both MSA and non-MSA combined), age-group-specific CIN2+ incidence rates varied by period and birth cohort (**Figure 5.2**). At the start of the study (2008), CIN2+ incidence was highest among younger age groups (18-20, 21-24, 25-29 years) compared to older age groups (30-34 and 35-39 years); by 2018, the two youngest age groups (18-20 and 21-24 years) had the lowest CIN2+ incidence rates (24.3/100,000 person-years and 293.9/100,000 person-years, respectively), with drastically lower rates among women aged 18-20 years. Age-group specific CIN2+ incidence rates by birth cohort demonstrated that within the same age group, young women (18-20, 21-24, and 25-29 years) who were born later had lower CIN2+ incidence rates. Among older women (30-34 and 35-39 years) in the same age group, CIN2+ incidence rates were stable regardless of date of birth. These age-group-specific patterns by period and birth cohort were also observed among TennCare-enrolled women who

Figure 5.2. Age-group-specific[a] CIN2+ incidence per 100,000 person-years among A) all women enrolled in TennCare and B) those screened for cervical cancer overall (Row 1) and those residing in MSAs (Row 2) and non-MSAs (Row 3), by period and birth cohort.



Abbreviations: CIN = Cervical Intraepithelial Lesion; MSA = Metropolitan Statistical Area; TennCare.= Tennessee Medicaid.
[a]Age-group-specific lines in each graph are interpreted as follows: 1 (age 18-20 years), 2 (age 21-24 years), 3 (age 25-29 years), 4 (age 30-34 years), and 5 (age 35-39 years).

were screened for cervical cancer, women residing in MSAs, and screened women residing in MSAs.

Women residing in non-MSAs showed some similarities and differences in rates by period and birth cohort compared to women residing in MSAs (**Figure 5.2**). Patterns were similar in that young women aged 18-20, 21-24, and 25-29 years had higher CIN2+ incidence rates in 2008 compared to older women, with the highest rates among women aged 21-24 years (1471.7/100,000 person-years). The most drastic changes in CIN2+ rates were in the youngest age group (18-20 years = 720.5/100,000 person-years in 2008 to 19.0/100,000 person-years in 2018). Rates of decline varied by age group but by 2018, incidence rates were similar between ages 21-24 (380.1/100,000 person-years), 30-34 (363.3/100,000 person-years), and 35-39 years (366.6/100,000 person-years). Within the same age group, women residing in non-MSAs who were born later had lower CIN2+ incidence rates for all age groups, except for age 35-39 years. This pattern was also observed in screened women residing in non-MSAs.

Age, Period, Cohort Effects in Regression Models

Among all TennCare-enrolled women, age-period-cohort Poisson regression models indicated decreasing CIN2+ incidence from 2008 to 2018 (overall EAPC = 0.90%/year; MSA EAPC = 0.90%/year; non-MSA EAPC = 0.89%/year) (**Table 5.4**). Significant improvements in model fit were found when adding drift (i.e., the overall linear trend in CIN2+ incidence), period, and cohort effects ($p < 0.001$). The best fitting model included all three effects (age-period-cohort), indicated by the lowest residual deviance (residual deviance = 814.3). Age-period-cohort models also had the best fit among women residing in MSAs (residual deviance = 662.6) and non-MSAs (residual deviance = 410.9). Model comparisons demonstrated notably larger cohort

Table 5.4. Age-period-cohort models for CIN2+ incidence among all women enrolled in Tennessee Medicaid and those screened for cervical cancer, by urbanicity, 2008-2018.

	Goodness-of-Fit			Model Comparison					
	Residual df	Residual Deviance	P-value	Comparison	Interpretation	Change in df	Change in Deviance	P-value	EAPC (95% CI)
All Women									
Overall									0.90 (0.89, 0.90)
1. Age	238	2806.86	--			--	--	--	
2. Age-Drift	237	1668.58	<0.001	2 versus 1	Trend (drift)	1	1138.28	<0.001	
3. Age-Cohort	234	909.39	<0.001	3 versus 2	Nonlinear cohort effect	3	759.19	<0.001	
4. Age-Period	234	1598.01	<0.001	4 versus 2	Nonlinear period effect	3	70.57	<0.001	
5. Age-Period-Cohort	231	814.30	<0.001	5 versus 3	Period effect adjusted for cohort	3	95.09	<0.001	
				5 versus 4	Cohort effect adjusted for period	3	783.71	<0.001	
MSA Residence									0.90 (0.90, 0.91)
1. Age	238	2006.75	--			--	--	--	
2. Age-Drift	237	1295.52	<0.001	2 versus 1	Trend (drift)	1	711.23	<0.001	
3. Age-Cohort	234	733.25	<0.001	3 versus 2	Nonlinear cohort effect	3	562.27	<0.001	
4. Age-Period	234	1244.25	<0.001	4 versus 2	Nonlinear period effect	3	51.27	<0.001	
5. Age-Period-Cohort	231	662.64	<0.001	5 versus 3	Period effect adjusted for cohort	3	70.60	<0.001	
				5 versus 4	Cohort effect adjusted for period	3	581.60	<0.001	
Non-MSA Residence									0.89 (0.88, 0.90)
1. Age	238	1080.50	--			--	--	--	
2. Age-Drift	237	639.10	<0.001	2 versus 1	Trend (drift)	1	441.40	<0.001	
3. Age-Cohort	234	435.57	<0.001	3 versus 2	Nonlinear cohort effect	3	203.53	<0.001	
4. Age-Period	234	619.29	<0.001	4 versus 2	Nonlinear period effect	3	19.81	<0.001	
5. Age-Period-Cohort	231	410.92	<0.001	5 versus 3	Period effect adjusted for cohort	3	24.66	<0.001	
				5 versus 4	Cohort effect adjusted for period	3	208.37	<0.001	
Women Screened for Cervical Cancer									
Overall									0.95 (0.95, 0.96)
1. Age	238	1104.09	--			--	--	--	
2. Age-Drift	237	832.40	<0.001	2 versus 1	Trend (drift)	1	271.69	<0.001	
3. Age-Cohort	234	491.25	<0.001	3 versus 2	Nonlinear cohort effect	3	341.15	<0.001	
4. Age-Period	234	814.39	<0.001	4 versus 2	Nonlinear period effect	3	18.01	<0.001	
5. Age-Period-Cohort	231	461.38	<0.001	5 versus 3	Period effect adjusted for cohort	3	29.87	<0.001	

				5 versus 4	Cohort effect adjusted for period	3	353.01	<0.001	
MSA Residence									0.96 (0.95, 0.96)
1. Age	238	891.14	--			--	--	--	
2. Age-Drift	237	729.89	<0.001	2 versus 1	Trend (drift)	1	161.25	<0.001	
3. Age-Cohort	234	463.58	<0.001	3 versus 2	Nonlinear cohort effect	3	266.30	<0.001	
4. Age-Period	234	714.29	<0.001	4 versus 2	Nonlinear period effect	3	15.60	0.001	
5. Age-Period-Cohort	231	436.61	<0.001	5 versus 3	Period effect adjusted for cohort	3	26.97	<0.001	
				5 versus 4	Cohort effect adjusted for period	3	277.68	<0.001	
Non-MSA Residence									0.94 (0.93, 0.96)
1. Age	238	499.55	--			--	--	--	
2. Age-Drift	237	383.80	<0.001	2 versus 1	Trend (drift)	1	1115.75	<0.001	
3. Age-Cohort	234	301.08	0.002	3 versus 2	Nonlinear cohort effect	3	82.72	<0.001	
4. Age-Period	234	380.60	<0.001	4 versus 2	Nonlinear period effect	3	3.20	0.362	
5. Age-Period-Cohort	231	296.86	0.002	5 versus 3	Period effect adjusted for cohort	3	4.21	0.239	
				5 versus 4	Cohort effect adjusted for period	3	83.73	<0.001	

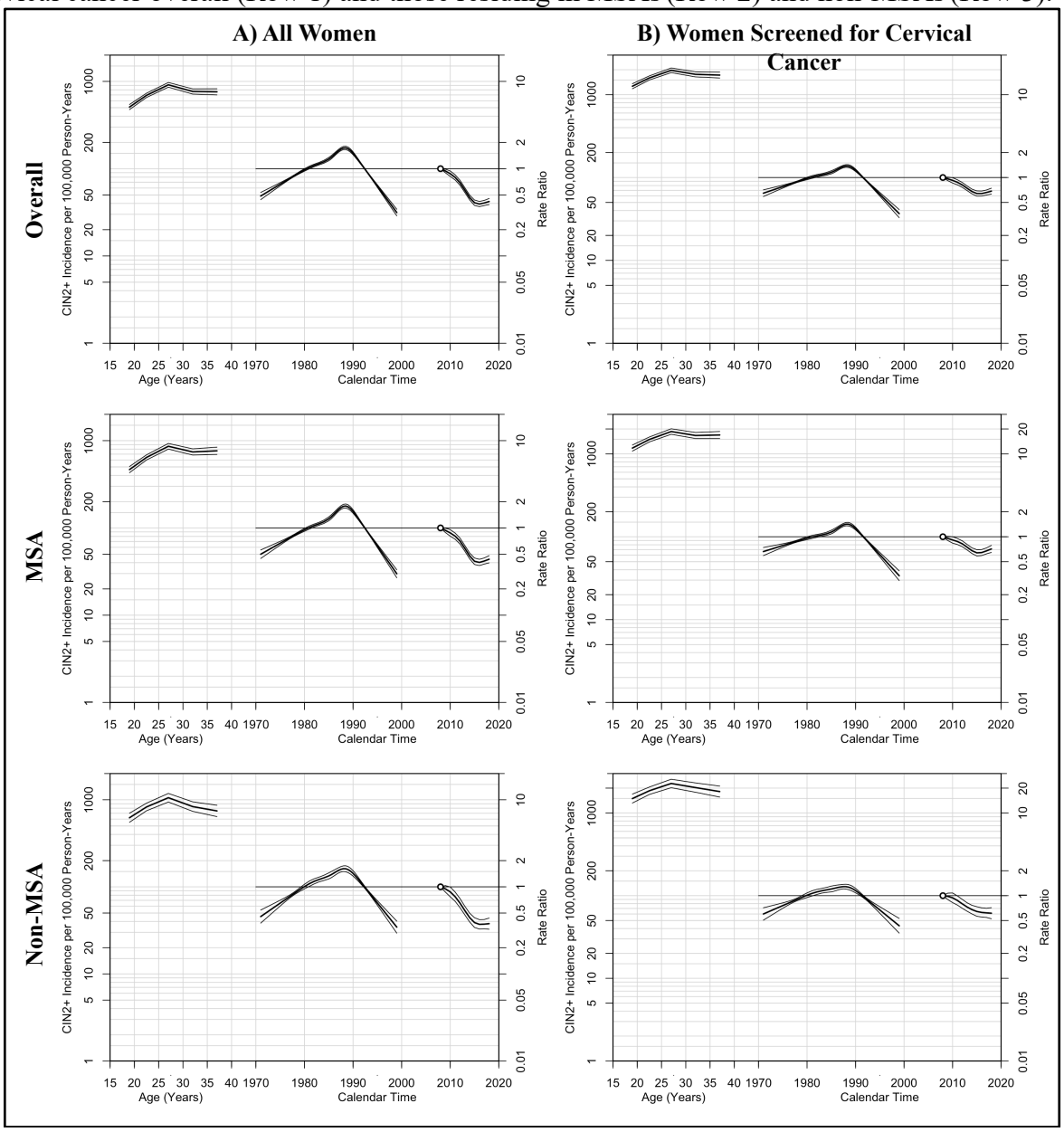
Abbreviations: CI = Confidence Interval; CIN = Cervical Intraepithelial Lesion; df = Degrees of Freedom; EAPC = Estimated Annual Percent Change; MSA = Metropolitan Statistical Area.

effects than period effects in women overall, those residing in MSAs, and those residing in non-MSAs (change in deviance for nonlinear cohort effects versus nonlinear period effects = 759.2 versus 70.6 [overall], 562.3 versus 51.2 [MSA], 203.5 versus 19.8 [non-MSA]).

Among all three groups of screened women (overall, those residing in MSA, and those residing in non-MSAs), age-cohort and age-period-cohort models had similar goodness-of-fit, with slightly better fit for age-period-cohort models (residual deviance for age-cohort versus age-period-cohort = 491.3 versus 461.4 [overall], 463.6 versus 436.6 [MSA], 301.1 versus 296.9 [non-MSA] (**Refer to Table 5.4**). Among screened women overall and those residing in MSAs, cohort effects were larger than period effects (change in deviance for nonlinear cohort effects versus nonlinear period effects = 341.2 versus 18.0 [overall], 266.3 versus 15.6 [MSA]. Among screened women residing in non-MSAs, nonlinear period effects and period effects adjusted for cohort effects were not significant ($p > 0.05$); however, the drift, nonlinear cohort effects, and cohort effects adjusted for period effects were significant ($p < 0.001$).

Age effects for women overall, those residing in MSAs, and in non-MSAs showed increasing CIN2+ incidence with increasing age among younger women until a peak of around age 27 years, followed by plateauing or decreasing CIN2+ incidence with increasing age among older women (**Figure 5.3**). Cohort effects demonstrated that women born between 1970 to 1988 experienced higher CIN2+ incidence with later years of birth, while women born after 1988 experienced lower CIN2+ incidence with later years of birth. In all women overall, those residing in MSAs, and in non-MSAs, period effects demonstrated decreasing CIN2+ incidence from 2009 to the mid 2010's, and then a mild increasing curvature in the late 2010's. For screened women, CIN2+ incidence had similar, yet less steep and prominent, period effect patterns compared to all women.

Figure 5.3. Age, cohort, and period effects[a] among A) all women and B) women screened for cervical cancer overall (Row 1) and those residing in MSAs (Row 2) and non-MSAs (Row 3).



Abbreviations: CIN = Cervical Intraepithelial Lesion; MSA = Metropolitan Statistical Area.
[a]Each plot's horizontal axis is divided into two parts: age, ranging from 15-40 years (left) and calendar time, ranging from 1970-2020 (right). Each plot contains two vertical axes: CIN2+ incidence per 100,000 person-years (left) and rate ratios (right), and three sets of curves: **age effects**, interpretable as cross-sectional CIN2+ incidence rates per 100,000 women at risk for the reference period, 2008, adjusted for cohort effects, with corresponding 95% confidence intervals (left), **cohort effects**, constrained to be 0 on average with 0 slope, interpretable as rate ratios relative to the age-period predictions (i.e., residual rate ratios) with corresponding 95% confidence intervals (middle), and **period effects**, interpretable as rate ratios relative to the reference period, 2008 (indicated by the hollow circle), with corresponding 95% confidence intervals (right).

Discussion

We examined secular trends in CIN2+ incidence, taking into account age, period, and cohort effects, among TennCare-enrolled women from 2008 to 2018 by urbanicity. In both MSAs and non-MSAs, our results demonstrated declining trends in CIN2+ incidence among women aged 18-39 years from 2008 onward, with the most drastic declines among young women aged 18-20 years and 21-24 years. Declines were likely because of the HPV vaccine's introduction in 2006 and the Advisory Committee on Immunization Practices' recommendations for adolescent HPV vaccination, as well as changes in cervical cancer screening recommendations and aggressiveness of approach.[45,52,121] Although patterns and rates of decline in CIN2+ incidence were similar between women residing in MSAs and non-MSAs, significant declines were delayed until 2010 for women residing in non-MSAs, unlike in MSAs, which began in 2008. After restricting our analyses to women screened for cervical cancer to control for the confounding effects of changing screening rates over time, HPV vaccine impact was still evident, regardless of urbanicity.

Our age-period-cohort analyses indicated that trends in CIN2+ incidence were largely driven by age and cohort effects, even after adjusting for period effects. Young women in more recent generations had lower rates of CIN2+ compared to young women born earlier. These effects are likely attributable to generational differences in vaccine eligibility, vaccination behaviors, and screening recommendations. When the Food and Drug Administration approved the first quadrivalent HPV vaccine in 2006 for females aged 9-26 years,[52] older women were ineligible for the vaccine. Even among age-eligible adults aged 18-26 years, HPV vaccination coverage in the US has historically been low, ranging from 22.1%-39.9% (initiation) and 13.8%-21.5% (completion) from 2013-2018.[122] Additionally, HPV vaccination in women aged over 26

years is often not covered by insurance and is ineffective among those who have already been infected with vaccine genotypes, creating both financial and biologic barriers in preventing CIN2+ in this age group. Further, in 2009, guidelines by the American College of Obstetricians and Gynecologists were updated to recommend against screening for women younger than 21 years to protect adolescents and young women from unnecessary invasive gynecologic procedures that could put them at risk for cervical damages.[121] In 2012, consensus screening guidelines were released by several major organizations, including the American Cancer Society and the United States Preventive Services Task Force, to also recommend against screening for women younger than 21 years.[45,46] These changes in guidelines contributed to decreases in CIN2+ detection among younger women. Further, guidelines for the aggressiveness of approach, such as frequency of screening, were also changed, contributing to less frequent screening and fewer colposcopies and biopsies to detect CIN2+ in screened women.[45,46]

In looking at urbanicity-stratified CIN2+ incidence by age group over time, we found similar patterns and evidence of HPV vaccine impact on reducing CIN2+ incidence in both MSAs and non-MSAs, despite varying HPV vaccination coverage by urbanicity. In Tennessee from 2016 to 2019, HPV vaccine initiation and completion in MSAs ranged from 66 to 69%and 46-47%, respectively, compared to 46-53% and 25-34% respectively, in non-MSAs.[18] A prior study in Connecticut also reported significant declines in CIN2+ incidence among young women in both urban and rural counties.[87] This suggests that while HPV vaccination coverage rates are lower in rural communities than in urban communities, CIN2+ still significantly declined across urban and rural settings. Additionally, despite varying HPV vaccination rates in urban and rural areas, a global-based meta-analysis reported similar genital HPV infection prevalence in urban (10%) and rural (11%) areas after the introduction of the HPV vaccine,[139] suggesting that HPV

infection rates are comparable regardless of urbanicity. This finding is corroborated by our prior work among TennCare-enrolled women, showing similar age-group-specific anogenital wart incidence, an HPV-associated outcome, by urbanicity.[84]

Among women who were screened for cervical cancer, HPV vaccine impact was still evident in MSAs and non-MSAs, with similar declining CIN2+ incidence between young, screened women (aged 18-20 and 21-24 years) residing in MSAs and non-MSAs. However, in MSAs, significant declines were observed in screened women aged 25-29 years, while declines in non-MSAs for this age group were not significant. This may be due to improved accessibility of HPV vaccination in urban centers upon first release. Further, age, period, and cohort effects were all significant among screened women residing in MSAs, while period effects were not significant for those residing in non-MSAs, indicating that CIN2+ incident trends in non-MSAs were mostly driven by age and cohort effects. Because our sample size for women residing in MSAs was roughly double that of non-MSAs, we cannot rule out the possibility that differences between MSAs and non-MSAs were also due to differences in power and sample size.

Our study has limitations. The study represents a unique population of Tennessee Medicaid women; thus, results may not be generalizable to other geographical regions or to populations of higher socioeconomic status. Additionally, our results did not consider race/ethnicity or income-level, which are both associated with CIN2+.[140] Specifically, women of Black race and those with higher levels of poverty have been shown to have higher CIN2+ rates[140]; thus, these factors may impact CIN2+ trends by urbanicity. Furthermore, because this is an ecologic study, we were unable to examine individual-level vaccination data, but instead, were able to assess direct and indirect effects of the HPV vaccine.

Our study has notable strengths. This is the first study to describe CIN2+ incident trends by urbanicity using a validated claims-based model, demonstrating the applicability of utilizing claims data for CIN2+ surveillance research. Examining population-based CIN2+ trends in the US is costly and limited to populations with adequate surveillance of cervical biopsies through the New Mexico HPV Pap registry[89] and the HPV Vaccine Impact Monitoring Project.[19] Utilizing claims data is a more efficient way to monitor HPV vaccine impact, and we were able to leverage TennCare claims data to examine vaccine impact on reducing CIN2+ incidence among Tennessee Medicaid enrollees, regardless of urbanicity. Further, this is the first US study to examine secular time trends in CIN2+ incidence using age-period-cohort models. Prior studies examining HPV vaccine impact on CIN2+ incidence have focused on evaluating overall linear trends using Joinpoint or pre-to-post vaccine era CIN2+ incidence using incidence rate ratios.[87–92,94] We expand upon these prior studies by attempting to disentangle the age, period, and cohort effects on CIN2+ trends using age-period-cohort models. However, due to the linear dependency of all three effects, the magnitude of each effect cannot be entirely isolated. Our study also has a large sample size, increasing the power of our study and reinforces the validity of our findings. Additionally, we are able to utilize our data from screened women to control for changes in cervical cancer screening patterns.

In summary, we demonstrated significant declines in CIN2+ incidence in both MSAs and non-MSAs among TennCare-enrolled women, particularly in younger women who likely could have benefited from the HPV vaccine through direct or indirect effects. CIN2+ trends were mostly driven by age and cohort effects, but effects in non-MSAs were delayed compared to MSAs, suggesting an impact of lower vaccination rates, such as delayed increases in vaccination in non-MSAs.

CHAPTER VI

SUMMARY AND FUTURE DIRECTIONS

HPV infections are associated with several adverse health outcomes, including anogenital warts and cancers of the cervix, anus, vagina, penis, oropharynx, and vulva,[3,4] most of which could be prevented through vaccination. Since the introduction of the HPV vaccine in 2006,[52] studies have shown significant declines in HPV-associated outcomes, including HPV infections,[74–79] anogenital warts,[80–86] and cervical premalignant lesions,[22,87–92] among young women who likely benefited from either direct or indirect effects of the HPV vaccine (e.g., getting vaccinated or being exposed to vaccinated persons). Due to the considerable costs, morbidity, and mortality associated with cancer,[5–7] the primary long-term goal of the HPV vaccine is reducing cancer incidence; however, observing the HPV vaccine's impact on cancer is not yet possible as the latency period between an initial HPV infection to the development of cancer can take several decades.[12–14,36] Therefore, examining trends in intermediate endpoints to cancer, including cervical premalignant lesions, such as cervical intraepithelial neoplasia grades 2 and 3 and adenocarcinoma *in situ* (CIN2+) is more feasible and can be detected earlier than the HPV vaccine's impact on cancer.

We built and validated claims-based models to identify cervical premalignant lesions, such as cervical intraepithelial neoplasia grades 2 and 3 and adenocarcinoma *in situ* (CIN2+), to be used for public health surveillance and to identify population-based CIN2+ trends (Chapter III). Using our validated model, we demonstrated population-level impact of the human papillomavirus (HPV) vaccine on reducing CIN2+ incidence in a population with low

vaccination coverage by observing significant declines in CIN2+ incidence among young Tennessee Medicaid (TennCare)-enrolled women who likely benefited from the HPV vaccine coupled with stable trends among older women (Chapter IV). Trends were largely driven by age and cohort effects in that younger women who were born in more recent cohorts had lower CIN2+ incidence compared to older cohorts, likely due to generational differences in vaccine availability and eligibility, as well as changes in cervical cancer screening recommendations (Chapter V). Further, declining trends in CIN2+ incidence and age and cohort effects were similar in populations with known varying vaccination coverage, such as urban and rural areas, suggesting evidence of HPV vaccine impact, regardless of the urbanicity of the geographic residence (Chapter V). However, we did observe some delays in declining CIN2+ rates in non-MSAs compared to MSAs, likely due to the geographic disparities between the two groups. Despite significant declines in CIN2+ regardless of urbanicity, identifying and reducing barriers to vaccination in rural communities should still remain a priority to further improve vaccine impact on reducing CIN2+ incidence.

Our results provide evidence of reductions in population-based CIN2+ incidence and demonstrate the applicability of utilizing administrative claims data to monitor HPV vaccine impact on CIN2+. Currently, monitoring CIN2+ incidence in the US is labor-intensive and costly because CIN2+ diagnoses require tissue confirmation from cervical biopsies. Few states have access to population-based or statewide surveillance of cervical biopsy data from the New Mexico HPV Pap Registry and the Human Papillomavirus Vaccine Impact Monitoring Project (HPV-IMPACT) across five states.[19,20] A more efficient and cost-effective solution is utilizing administrative claims databases, which systematically captures patient procedures and diagnoses across an entire network. Because of the uncertainty around the validity of claims codes to

identify CIN2+ events, prior claims-based studies examining CIN2+ trends are limited. Further, because of the major administrative coding transition from the International Classification of Diseases, Clinical Modification (ICD) 9th (ICD-9) to 10th (ICD-10) revision, no claims-based studies have been able to assess trends past 2015, when the ICD-9-to-ICD-10 transition occurred.

Our study addresses the gaps by validating a simple linear claims-based model for future CIN2+ trend studies to utilize across ICD eras, as we observed no significant differences in the model's ability to discriminate CIN2+ event status between the ICD-9 and ICD-10 era. Using the validated model, we were able to include more recent data to examine trends in CIN2+ incidence between 2008 and 2018 among TennCare-enrolled women, which has not yet been done in the US. Because billing claims codes include rich information on patient procedures, we were also able to feasibly examine CIN2+ incident trends among the subset of TennCare-enrolled women who were screened for cervical cancer to account for changes in screening patterns over time.

When comparing results from our claims-based study to other studies with population-based cervical biopsy data, our results showed similar patterns (i.e., significant declines in CIN2+ among young women and stable trends among older women) and similar rates of decline among young women, which reinforces the validity and applicability of claims-based studies as alternates to biopsy-based studies for examining population-based trends in CIN2+ incidence. Our study expands upon other studies demonstrating population-level HPV vaccine impact on CIN2+ incidence by specifically examining secular trends, including age, period, and cohort effects, and examining these secular CIN2+ trends by urbanicity. Other studies have only discussed the overarching CIN2+ trends over time without explaining whether the trends are attributed to either age, period, or cohort effects. Using advanced age-period-cohort Poisson regression models, we were able to determine that trends in CIN2+ incidence over time are

largely driven by age and cohort effects, which further strengthens the notion that the reductions in CIN2+ incidence are likely attributed in part by the introduction of the HPV vaccine and changes in cervical screening recommendations over time.

Several limitations of our study should be noted. The study population was limited to Tennessee Medicaid enrollees; thus, the results may not be generalizable to populations with different socio-demographics, such as those with higher income and greater access to health care; however, we were able to build the claims-based models and examine CIN2+ trends among a large retrospective cohort of all TennCare-enrolled women aged 18-39 years to improve power and validity. Further, although we restricted the trend analyses to women who were screened for cervical cancer to remove the confounding effects of changing screening patterns over time, the possibility of residual confounding (e.g., race/ethnicity, income, access to care, education, etc.) cannot be ruled out. We were, however, able to stratify the analyses by urbanicity (Chapter V) to address one of the factors that may influence CIN2+ rates. We were unable to examine trends by other covariates, such as race/ethnicity, due to the ambiguity of the race/ethnicity variable coded in the TennCare database (i.e., pre-combined other and unknown race/ethnicity into a single category).

Our study had several notable strengths. We had the unique opportunity to leverage data from one of the only population-based cervical biopsy surveillance systems in the US (HPV-IMPACT) to build and validate claims-based models for identifying CIN2+ events (Chapter III). The gold standard data from HPV-IMPACT were carefully audited to ensure high quality and comprehensiveness of the data. Further, because the gold standard data were population-based, we were able to optimize the models for public health surveillance to identify trends over time. This is also the first study to be able to include CIN2+ data from the ICD-10 era (2015 and

onward) to examine CIN2+ incident trends in a population outside of the catchment areas with cervical biopsy surveillance data (Chapters IV and V).

In conclusion, our study demonstrates the utility and feasibility of claims data for future trend studies of CIN2+ incidence among populations without access to population-based cervical biopsy data. Our study was unique in that we documented HPV vaccine impact on reducing CIN2+ incidence in several ways, including among a population with low vaccination coverage to examine whether vaccine impact is reaching low coverage populations, among women screened for cervical cancer to account for screening changes over time, and among women residing in urban and rural areas to compare vaccine impact in groups with varying vaccination coverage. The three specific aims in the overarching study all provide valuable additions to the HPV vaccine impact literature and can be utilized in future studies to continue monitoring HPV vaccine impact on reducing CIN2+ incidence in the US.

REFERENCES

1. Satterwhite CL, Torrone E, Meites E, et al. Sexually transmitted infections among US women and men: Prevalence and incidence estimates, 2008. *Sex Transm Dis*. 2013;40(3):187-193. doi:10.1097/OLQ.0b013e318286bb53

2. de Sanjosé S, Brotons M, Pavón MA. The natural history of human papillomavirus infection. *Best Pract Res Clin Obstet Gynaecol*. 2018;47:2-13. doi:10.1016/j.bpobgyn.2017.08.015

3. Braaten KP, Laufer MR. Human papillomavirus (HPV), HPV-related disease, and the HPV vaccine. *Rev Obstet Gynecol*. 2008;1(1):2-10.

4. Saraiya M, Unger ER, Thompson TD, et al. US assessment of HPV types in cancers: Implications for current and 9-valent HPV vaccines. *J Natl Cancer Inst*. 2015;107(6):djv086. doi:10.1093/jnci/djv086

5. Nwankwo C, Corman SL, Shah R, Kwon Y. HSR19-102: Direct and indirect economic burden of cervical cancer (CxCa) in the United States in 2015: A mixed-methods analysis. *Journal of the National Comprehensive Cancer Network*. 2019;17(3.5):HSR19-102-HSR19-102. doi:10.6004/jnccn.2018.7182

6. Östensson E, Silfverschiöld M, Greiff L, et al. The economic burden of human papillomavirus-related precancers and cancers in Sweden. *PLOS ONE*. 2017;12(6):e0179520. doi:10.1371/journal.pone.0179520

7. Shah R, Nwankwo C, Kwon Y, Corman SL. Economic and humanistic burden of cervical cancer in the United States: Results from a nationally representative survey. *Journal of Women's Health*. 2020;29(6):799-805. doi:10.1089/jwh.2019.7858

8. Markowitz LE, Dunne E, Saraiya M, Lawson H, Chesson H, Unger E. Quadrivalent human papillomavirus vaccine: Recommendations of the Advisory Committee on Immunization Practices (ACIP). *MMWR Morb Mortal Wkly Rep*. 2007;56(2):1-24.

9. Petrosky E, Bocchini JA, Hariri S, et al. Use of 9-valent human papillomavirus (HPV) vaccine: updated HPV vaccination recommendations of the advisory committee on immunization practices. *MMWR Morb Mortal Wkly Rep*. 2015;64(11):300-304.

10. de Sanjosé S, Serrano B, Tous S, et al. Burden of human papillomavirus (HPV)-related cancers attributable to HPVs 6/11/16/18/31/33/45/52 and 58. *JNCI Cancer Spectr*. 2018;2(4). doi:10.1093/jncics/pky045

11. Arbyn M, Weiderpass E, Bruni L, et al. Estimates of incidence and mortality of cervical cancer in 2018: a worldwide analysis. *The Lancet Global Health*. 2020;8(2):e191-e203. doi:10.1016/S2214-109X(19)30482-6

12. Ylitalo N, Josefsson A, Melbye M, et al. A prospective study showing long-term infection with human papillomavirus 16 before the development of cervical carcinoma in situ. *Cancer Res*. 2000;60(21):6027-6032.

13. Meijer CJLM, Snijders PJF, Brule A. Screening for cervical cancer: Should we test for infection with high-risk HPV? *CMAJ*. 2000;163(5):535-538.

14. Watson RA. Human papillomavirus: Confronting the epidemic—A urologist's perspective. *Rev Urol*. 2005;7(3):135-144.

15. Winer RL, Kiviat NB, Hughes JP, et al. Development and duration of human papillomavirus lesions, after initial infection. *J Infect Dis*. 2005;191(5):731-738. doi:10.1086/427557

16. Woodman CB, Collins S, Winter H, et al. Natural history of cervical human papillomavirus infection in young women: A longitudinal cohort study. *Lancet*. 2001;357(9271):1831-1836. doi:10.1016/S0140-6736(00)04956-4

17. Baseman JG, Koutsky LA. The epidemiology of human papillomavirus infections. *J Clin Virol*. 2005;32 Suppl 1:S16-24. doi:10.1016/j.jcv.2004.12.008

18. Elam-Evans LD. National, Regional, State, and Selected Local Area Vaccination Coverage Among Adolescents Aged 13–17 Years — United States, 2019. *MMWR Morb Mortal Wkly Rep*. 2020;69. doi:10.15585/mmwr.mm6933a1

19. Centers for Disease Control and Prevention. Human papillomavirus vaccine impact monitoring project (HPV-IMPACT). Published February 21, 2018. Accessed February 9, 2019. https://www.cdc.gov/ncird/surveillance/hpvimpact/overview.html

20. The University of New Mexico. NMHPVPR-The New Mexico HPV Pap Registry. Accessed November 9, 2020. https://hpvprevention.unm.edu/nmhpvpr/

21. Cartwright DJ. ICD-9-CM to ICD-10-CM Codes: What? Why? How? *Adv Wound Care (New Rochelle)*. 2013;2(10):588-592. doi:10.1089/wound.2013.0478

22. Flagg EW, Torrone EA, Weinstock H. Ecological association of human papillomavirus vaccination with cervical dysplasia prevalence in the United States, 2007-2014. *Am J Public Health*. 2016;106(12):2211-2218. doi:10.2105/AJPH.2016.303472

23. Sabeena S, Bhat P, Kamath V, Arunkumar G. Possible non-sexual modes of transmission of human papilloma virus. *Journal of Obstetrics and Gynaecology Research*. 2017;43(3):429-435. doi:10.1111/jog.13248

24. Hariri S, Unger ER, Sternberg M, et al. Prevalence of genital human papillomavirus among females in the United States, the National Health And Nutrition Examination Survey, 2003-2006. *J Infect Dis*. 2011;204(4):566-573. doi:10.1093/infdis/jir341

25. Chesson HW, Dunne EF, Hariri S, Markowitz LE. The estimated lifetime probability of acquiring human papillomavirus in the United States. *Sex Transm Dis*. 2014;41(11):660-664. doi:10.1097/OLQ.0000000000000193

26. Smith JS, Melendy A, Rana RK, Pimenta JM. Age-Specific Prevalence of Infection with human papillomavirus in females: A global review. *Journal of Adolescent Health*. 2008;43(4):S5.e1-S5.e62. doi:10.1016/j.jadohealth.2008.07.009

27. Saraiya M, Unger ER, Thompson TD, et al. US assessment of HPV types in cancers: implications for current and 9-valent HPV vaccines. *J Natl Cancer Inst*. 2015;107(6):djv086. doi:10.1093/jnci/djv086

28. Schiffman M, Castle PE, Jeronimo J, Rodriguez AC, Wacholder S. Human papillomavirus and cervical cancer. *Lancet*. 2007;370(9590):890-907. doi:10.1016/S0140-6736(07)61416-0

29. Chesson HW, Ekwueme DU, Saraiya M, Watson M, Lowy DR, Markowitz LE. Estimates of the annual direct medical costs of the prevention and treatment of disease associated with human papillomavirus in the United States. *Vaccine*. 2012;30(42):6016-6019. doi:10.1016/j.vaccine.2012.07.056

30. National Cancer Institute. Cancer of the cervix uteri - cancer stat facts. SEER. Accessed October 29, 2020. https://seer.cancer.gov/statfacts/html/cervix.html

31. IARC Working Group on the Evaluation of Carcinogenic Risks to Humans. *Biological Agents*. International Agency for Research on Cancer; 2012.

32. de Sanjose S, Quint WG, Alemany L, et al. Human papillomavirus genotype attribution in invasive cervical cancer: A retrospective cross-sectional worldwide study. *Lancet Oncol*. 2010;11(11):1048-1056. doi:10.1016/S1470-2045(10)70230-8

33. Koshiol J, Lindsay L, Pimenta JM, Poole C, Jenkins D, Smith JS. Persistent human papillomavirus infection and cervical neoplasia: A systematic review and meta-analysis. *Am J Epidemiol*. 2008;168(2):123-137. doi:10.1093/aje/kwn036

34. Wheeler CM. The natural history of cervical human papillomavirus infections and cervical cancer: Gaps in knowledge and future horizons. *Obstet Gynecol Clin North Am*. 2013;40(2):165-176. doi:10.1016/j.ogc.2013.02.004

35. Khan MJ, Castle PE, Lorincz AT, et al. The elevated 10-year risk of cervical precancer and cancer in women with human papillomavirus (HPV) type 16 or 18 and the possible utility of type-specific HPV testing in clinical practice. *J Natl Cancer Inst*. 2005;97(14):1072-1079. doi:10.1093/jnci/dji187

36. Vink MA, Bogaards JA, van Kemenade FJ, de Melker HE, Meijer CJLM, Berkhof J. Clinical progression of high-grade cervical intraepithelial neoplasia: Estimating the time to preclinical cervical cancer from doubly censored national registry data. *Am J Epidemiol*. 2013;178(7):1161-1169. doi:10.1093/aje/kwt077

37. Fontham ETH, Wolf AMD, Church TR, et al. Cervical cancer screening for individuals at average risk: 2020 guideline update from the American Cancer Society. *CA: A Cancer Journal for Clinicians*. 2020;70(5):321-346. doi:10.3322/caac.21628

38. US Preventive Services Task Force, Curry SJ, Krist AH, et al. Screening for cervical cancer: US Preventive Services Task Force recommendation statement. *JAMA*. 2018;320(7):674-686. doi:10.1001/jama.2018.10897

39. American College of Obstetricians and Gynecologists (ACOG). Practice bulletin no. 157: Cervical cancer screening and prevention. *Obstet Gynecol*. 2016;127(1):e1-e20. doi:10.1097/AOG.0000000000001263

40. American College of Obstetricians and Gynecologists. The frequency with which a cervical-vaginal cytology examination should be performed in gynecologic practice. *ACOG Technical Bulletin*. 1975;(29).

41. United States Preventive Services Task Force (USPSTF). *US Preventive Services Task Force Guides to Clinical Preventive Services*. 2nd ed. Agency for Healthcare Research and Quality (US); 1996.

42. American Cancer Society. History of ACS recommendations for the early detection of cancer in people without symptoms. Accessed November 11, 2019. https://www.cancer.org/health-care-professionals/american-cancer-society-prevention-early-detection-guidelines/overview/chronological-history-of-acs-recommendations.html

43. United States Preventive Services Task Force (USPSTF). Screening for cervical cancer: Recommendations and Rationale. *AFP*. 2003;67(8):1759.

44. Committee on Practice Bulletins—Gynecology. ACOG practice bulletin number 45, August 2003: Committee on Practice Bulletins-Gynecology. Cervical cytology screening. *Obstetrics & Gynecology*. 2003;102(2):417-427. doi:10.1016/S0029-7844(03)00745-2

45. Saslow D, Solomon D, Lawson HW, et al. American Cancer Society, American Society for Colposcopy and Cervical Pathology, and American Society for Clinical Pathology Screening Guidelines for the Prevention and Early Detection of Cervical Cancer. *CA Cancer J Clin*. 2012;62(3):147-172. doi:10.3322/caac.21139

46. Moyer VA. Screening for cervical cancer: U.S. Preventive Services Task Force recommendation statement. *Ann Intern Med*. 2012;156(12):880-891, W312. doi:10.7326/0003-4819-156-12-201206190-00424

47. National Cancer Institute. Understanding cervical changes: A health guide for women. Published May 2017. Accessed February 8, 2021. https://www.cancer.gov/publications/patient-education/understanding-cervical-changes

48. Datta SD, Koutsky LA, Ratelle S, et al. Human papillomavirus infection and cervical cytology in women screened for cervical cancer in the United States, 2003-2005. *Ann Intern Med*. 2008;148(7):493-500. doi:10.7326/0003-4819-148-7-200804010-00004

49. Castle PE, Fetterman B, (ASCP) S, et al. Five-year experience of human papillomavirus DNA and papanicolaou test cotesting. *Obstet Gynecol*. 2009;113(3):595-600. doi:10.1097/AOG.0b013e3181996ffa

50. Vesco KK, Whitlock EP, Eder M, et al. *Screening for Cervical Cancer: A Systematic Evidence Review for the U.S. Preventive Services Task Force*. Agency for Healthcare Research and Quality (US); 2011. Accessed February 7, 2021. http://www.ncbi.nlm.nih.gov/books/NBK66099/

51. Kulasingam SL, Havrilesky L, Ghebre R, Myers ER. *Screening for Cervical Cancer: A Decision Analysis for the U.S. Preventive Services Task Force*. Agency for Healthcare Research and Quality (US); 2011. Accessed February 7, 2021. http://www.ncbi.nlm.nih.gov/books/NBK92546/

52. Centers for Disease Control and Prevention. Quadrivalent human papillomavirus vaccine: Recommendations of the Advisory Committee on Immunization Practices (ACIP). *MMWR Morb Mortal Wkly Rep*. 2007;56(02):1-24.

53. Yanofsky VR, Patel RV, Goldenberg G. Genital warts: A comprehensive review. *J Clin Aesthet Dermatol*. 2012;5(6):25-36.

54. Centers for Disease Control and Prevention. FDA licensure of bivalent human papillomavirus vaccine (HPV2, Cervarix) for use in females and updated HPV vaccination recommendations from the Advisory Committee on Immunization Practices (ACIP). *MMWR Morbidity and mortality weekly report*. 2010;59(20):626.

55. Centers for Disease Control and Prevention (CDC). FDA licensure of quadrivalent human papillomavirus vaccine (HPV4, Gardasil) for use in males and guidance from the Advisory Committee on Immunization Practices (ACIP). *MMWR Morb Mortal Wkly Rep*. 2010;59(20):630-632.

56. Centers for Disease Control and Prevention. Recommendations on the use of quadrivalent human papillomavirus vaccine in males--Advisory Committee on Immunization Practices (ACIP), 2011. *MMWR Morb Mortal Wkly Rep*. 2011;60(50):1705-1708.

57. Meites E, Szilagyi PG, Chesson HW, Unger ER, Romero JR, Markowitz LE. Human papillomavirus vaccination for adults: Updated recommendations of the Advisory Committee on Immunization Practices. *MMWR Morb Mortal Wkly Rep*. 2019;68(32):698-702. doi:10.15585/mmwr.mm6832a3

58. Meites E. Use of a 2-dose schedule for human papillomavirus vaccination — Updated recommendations of the Advisory Committee on Immunization Practices. *MMWR Morb Mortal Wkly Rep*. 2016;65. doi:10.15585/mmwr.mm6549a5

59. Mix JM, Dyne EAV, Saraiya M, Hallowell BD, Thomas CC. Assessing impact of HPV vaccination on cervical cancer incidence among women aged 15–29 Years in the United States, 1999–2017: An ecologic study. *Cancer Epidemiol Biomarkers Prev*. 2021;30(1):30-37. doi:10.1158/1055-9965.EPI-20-0846

60. Fuentes A. Advancements in cervical cancer prevention and management of persistent, recurrent, and metastatic disease: 2016 update. *American Journal of Hematology / Oncology®*. 2017;12(12). Accessed November 7, 2020. https://www.gotoper.com/publications/ajho/2017/2016dec/advancements-in-cervical-cancer-prevention-and-management-of-persistent-recurrent-and-metastatic-disease-2016-update

61. U.S. Cancer Statistics Working Group. U.S. Cancer Statistics Data Visualizations Tool, based on 2019 submission data (1999-2017): U.S. Department of Health and Human Services, Centers for Disease Control and Prevention and National Cancer Institute. Published June 2020. Accessed November 6, 2020. www.cdc.gov/cancer/dataviz

62. Landy R, Pesola F, Castañón A, Sasieni P. Impact of cervical screening on cervical cancer mortality: estimation using stage-specific results from a nested case-control study. *Br J Cancer*. 2016;115(9):1140-1146. doi:10.1038/bjc.2016.290

63. Yoo W, Kim S, Huh WK, et al. Recent trends in racial and regional disparities in cervical cancer incidence and mortality in United States. *PLoS One*. 2017;12(2). doi:10.1371/journal.pone.0172548

64. Yu L. Rural–urban and racial/ethnic disparities in invasive cervical cancer incidence in the United States, 2010–2014. *Prev Chronic Dis*. 2019;16. doi:10.5888/pcd16.180447

65. Singh GK. Rural–urban trends and patterns in cervical cancer mortality, incidence, stage, and survival in the United States, 1950–2008. *J Community Health*. 2012;37(1):217-223. doi:10.1007/s10900-011-9439-6

66. Jeudin P, Liveright E, Del Carmen MG, Perkins RB. Race, ethnicity, and income factors impacting human papillomavirus vaccination rates. *Clin Ther*. 2014;36(1):24-37. doi:10.1016/j.clinthera.2013.11.001

67. Hirth J. Disparities in HPV vaccination rates and HPV prevalence in the United States: A review of the literature. *Human Vaccines & Immunotherapeutics*. 2019;15(1):146-155. doi:10.1080/21645515.2018.1512453

68. Vanderpool RC, Stradtman LR, Brandt HM. Policy opportunities to increase HPV vaccination in rural communities. *Hum Vaccin Immunother*. 2019;15(7-8):1527-1532. doi:10.1080/21645515.2018.1553475

69. Mohammed KA, Subramaniam DS, Geneus CJ, et al. Rural-urban differences in human papillomavirus knowledge and awareness among US adults. *Prev Med*. 2018;109:39-43. doi:10.1016/j.ypmed.2018.01.016

70. Blake KD, Ottenbacher AJ, Finney Rutten LJ, et al. Predictors of human papillomavirus awareness and knowledge in 2013: Gaps and opportunities for targeted communication strategies. *Am J Prev Med*. 2015;48(4):402-410. doi:10.1016/j.amepre.2014.10.024

71. Dilley SE, Peral S, Straughn JM, Scarinci IC. The challenge of HPV vaccination uptake and opportunities for solutions: Lessons learned from Alabama. *Prev Med*. 2018;113:124-131. doi:10.1016/j.ypmed.2018.05.021

72. Vanderpool RC, Dressler EVM, Stradtman LR, Crosby RA. Fatalistic beliefs and completion of the HPV vaccination series among a sample of young Appalachian Kentucky women. *J Rural Health*. 2015;31(2):199-205. doi:10.1111/jrh.12102

73. Cohen HW. Limitations of an ecological study: A review. *Am J Hypertens*. 2005;18(6):750-750. doi:10.1016/j.amjhyper.2005.04.002

74. Kahn JA, Brown DR, Ding L, et al. Vaccine-type human papillomavirus and evidence of herd protection after vaccine introduction. *Pediatrics*. 2012;130(2):e249-256. doi:10.1542/peds.2011-3587

75. Markowitz LE, Hariri S, Lin C, et al. Reduction in human papillomavirus (HPV) prevalence among young women following HPV vaccine introduction in the United States, National Health and Nutrition Examination Surveys, 2003-2010. *J Infect Dis*. 2013;208(3):385-393. doi:10.1093/infdis/jit192

76. Dunne EF, Naleway A, Smith N, et al. Reduction in human papillomavirus vaccine type prevalence among young women screened for cervical cancer in an integrated US healthcare delivery system in 2007 and 2012-2013. *J Infect Dis*. 2015;212(12):1970-1975. doi:10.1093/infdis/jiv342

77. Kahn JA, Widdice LE, Ding L, et al. Substantial decline in vaccine-type human papillomavirus (HPV) among vaccinated young women during the first 8 years after HPV vaccine introduction in a community. *Clin Infect Dis*. 2016;63(10):1281-1287. doi:10.1093/cid/ciw533

78. Markowitz LE, Liu G, Hariri S, Steinau M, Dunne EF, Unger ER. Prevalence of HPV after introduction of the vaccination program in the United States. *Pediatrics*. 2016;137(3):e20151968. doi:10.1542/peds.2015-1968

79. Oliver SE, Unger ER, Lewis R, et al. Prevalence of human papillomavirus among females after vaccine introduction—National Health and Nutrition Examination Survey, United States, 2003–2014. *J Infect Dis*. 2017;216(5):594-603. doi:10.1093/infdis/jix244

80. Bauer HM, Wright G, Chow J. Evidence of human papillomavirus vaccine effectiveness in reducing genital warts: An analysis of California public family planning administrative claims data, 2007-2010. *Am J Public Health*. 2012;102(5):833-835. doi:10.2105/AJPH.2011.300465

81. Flagg EW, Schwartz R, Weinstock H. Prevalence of anogenital warts among participants in private health plans in the United States, 2003-2010: Potential impact of human papillomavirus vaccination. *Am J Public Health*. 2013;103(8):1428-1435. doi:10.2105/AJPH.2012.301182

82. Perkins RB, Legler A, Hanchate A. Trends in male and female genital warts among adolescents in a safety-net health care system 2004–2013: Correlation with introduction of female and male human papillomavirus vaccination. *Sexually Transmitted Diseases*. 2015;42(12):665-668. doi:10.1097/OLQ.0000000000000369

83. Flagg EW, Torrone EA. Declines in anogenital warts among age groups most likely to be impacted by human papillomavirus vaccination, United States, 2006-2014. *Am J Public Health*. Published online November 21, 2017:e1-e8. doi:10.2105/AJPH.2017.304119

84. Shing JZ, Hull PC, Zhu Y, et al. Trends in anogenital wart incidence among Tennessee Medicaid enrollees, 2006–2014: The impact of human papillomavirus vaccination. *Papillomavirus Res*. Published online April 10, 2019. doi:10.1016/j.pvr.2019.04.007

85. Mann LM, Llata E, Flagg EW, et al. Trends in the prevalence of anogenital warts among patients at sexually transmitted disease clinics-sexually transmitted disease surveillance network, United States, 2010-2016. *J Infect Dis*. 2019;219(9):1389-1397. doi:10.1093/infdis/jiy684

86. Naleway AL, Crane B, Smith N, Francisco M, Weinmann S, Markowitz LE. Temporal trends in the incidence of anogenital warts: Impact of human papillomavirus vaccination. *Sexually Transmitted Diseases*. 2020;47(3):179-186. doi:10.1097/OLQ.0000000000001103

87. Niccolai LM, Julian PJ, Meek JI, McBride V, Hadler JL, Sosa LE. Declining rates of high-grade cervical lesions in young women in Connecticut, 2008-2011. *Cancer Epidemiol Biomarkers Prev*. 2013;22(8):1446-1450. doi:10.1158/1055-9965.EPI-13-0272

88. Hariri S, Johnson ML, Bennett NM, et al. Population-based trends in high-grade cervical lesions in the early human papillomavirus vaccine era in the United States. *Cancer*. 2015;121(16):2775-2781. doi:10.1002/cncr.29266

89. Benard VB, Castle PE, Jenison SA, et al. Population-based incidence rates of cervical intraepithelial neoplasia in the human papillomavirus vaccine era. *JAMA Oncol*. 2017;3(6):833-837. doi:10.1001/jamaoncol.2016.3609

90. Oakley F, Desouki MM, Pemmaraju M, et al. Trends in high-grade cervical cancer precursors in the human papillomavirus vaccine era. *Am J Prev Med*. 2018;55(1):19-25. doi:10.1016/j.amepre.2018.03.015

91. Gargano JW, Park IU, Griffin MR, et al. Trends in high-grade cervical lesions and cervical cancer screening in five states, 2008-2015. *Clin Infect Dis*. Published online April 8, 2019. doi:10.1093/cid/ciy707

92. McClung NM, Gargano JW, Park IU, et al. Estimated number of cases of high-grade cervical lesions diagnosed among women — United States, 2008 and 2016. *MMWR Morb Mortal Wkly Rep*. 2019;68(15):337-343. doi:10.15585/mmwr.mm6815a1

93. Dickson EL, Vogel RI, Luo X, Downs LS. Recent trends in type-specific HPV infection rates in the United States. *Epidemiol Infect*. 2015;143(5):1042-1047. doi:10.1017/S0950268814001538

94. McClung NM, Gargano JW, Bennett NM, et al. Trends in human papillomavirus vaccine types 16 and 18 in cervical precancers, 2008–2014. *Cancer Epidemiol Biomarkers Prev*. 2019;28(3):602-609. doi:10.1158/1055-9965.EPI-18-0885

95. Mayo Clinic. HPV test - Mayo Clinic. Published May 22, 2020. Accessed February 8, 2021. https://www.mayoclinic.org/tests-procedures/hpv-test/about/pac-20394355

96. Kim SC, Gillet VG, Feldman S, et al. Validation of claims-based algorithms for identification of high-grade cervical dysplasia and cervical cancer. *Pharmacoepidemiol Drug Saf*. 2013;22(11). doi:10.1002/pds.3520

97. Centers for Disease Control and Prevention. ICD - ICD-10-CM - International classification of diseases, (ICD-10-CM/PCS transition. Published March 1, 2019. Accessed March 18, 2019. https://www.cdc.gov/nchs/icd/icd10cm_pcs_background.htm

98. Slavova S, Costich JF, Luu H, et al. Interrupted time series design to evaluate the effect of the ICD-9-CM to ICD-10-CM coding transition on injury hospitalization trends. *Inj Epidemiol*. 2018;5(1):36. doi:10.1186/s40621-018-0165-8

99. Inscore MC, Gonzales KR, Rennix CP, Jones BH. The effect of transitioning to ICD-10-CM on acute injury surveillance of active duty service members. *Inj Epidemiol*. 2018;5(1):32. doi:10.1186/s40621-018-0162-y

100. Panozzo CA, Woodworth TS, Welch EC, et al. Early impact of the ICD-10-CM transition on selected health outcomes in 13 electronic health care databases in the United States. *Pharmacoepidemiol Drug Saf*. 2018;27(8):839-847. doi:10.1002/pds.4563

101. Salemi JL, Tanner JP, Kirby RS, Cragan JD. The impact of the ICD-9-CM to ICD-10-CM transition on the prevalence of birth defects among infant hospitalizations in the United States. *Birth Defects Res*. 2019;111(18):1365-1379. doi:10.1002/bdr2.1578

102. Sarayani A, Wang X, Thai TN, Albogami Y, Jeon N, Winterstein AG. Impact of the transition from ICD-9-CM to ICD-10-CM on the identification of pregnancy episodes in US health insurance claims data. *Clin Epidemiol*. 2020;12:1129-1138. doi:10.2147/CLEP.S269400

103. World Health Organization. Classification of Diseases (ICD). Accessed November 16, 2020. https://www.who.int/standards/classifications/classification-of-diseases

104. Centers for Medicare and Medicaid Services. ICD-10-CM, ICD-10-PCS, CPT, and HCPCS code sets. Published online September 2020. https://www.cms.gov/outreach-and-education/medicare-learning-network-mln/mlnproducts/downloads/icd9-10cm-icd10pcs-cpt-hcpcs-code-sets-educational-tool-icn900943.pdf

105. Shing JZ, Griffin MR, Nguyen LD, et al. Improving cervical precancer surveillance: Validity of claims-based prediction models in ICD-9 and ICD-10 eras. *JNCI Cancer Spectrum*. 2020;(pkaa112). doi:10.1093/jncics/pkaa112

106. Henk HJ, Insinga RP, Singhal PK, Darkow T. Incidence and costs of cervical intraepithelial neoplasia in a US commercially insured population. *J Low Genit Tract Dis*. 2010;14(1):29-36. doi:10.1097/LGT.0b013e3181ac05e9

107. Novaes HMD, Itria A, Silva GA e, Sartori AMC, Rama CH, Soárez PC de. Annual national direct and indirect cost estimates of the prevention and treatment of cervical cancer in Brazil. *Clinics (Sao Paulo)*. 2015;70(4):289-295. doi:10.6061/clinics/2015(04)12

108. Tibshirani R. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*. 1996;58(1):267-288. doi:10.1111/j.2517-6161.1996.tb02080.x

109. Steyerberg EW, Vickers AJ, Cook NR, et al. Assessing the performance of prediction models: a framework for some traditional and novel measures. *Epidemiology*. 2010;21(1):128-138. doi:10.1097/EDE.0b013e3181c30fb2

110. Tin Kam Ho. Random decision forests. In: *Proceedings of 3rd International Conference on Document Analysis and Recognition*. Vol 1. ; 1995:278-282 vol.1. doi:10.1109/ICDAR.1995.598994

111. Steyerberg EW, Bleeker SE, Moll HA, Grobbee DE, Moons KGM. Internal and external validation of predictive models: A simulation study of bias and precision in small samples. *Journal of Clinical Epidemiology*. 2003;56(5):441-447. doi:10.1016/S0895-4356(03)00047-7

112. Guerino P, James C. *Detailed Race, Ethnicity, and Language Preference in the Health Insurance Marketplaces, 2017*. Centers for Medicare and Medicaid Services, Office of Minority Health; 2017.

113. Substance Abuse and Mental Health Services Administration. *Tennessee 2014 Mental Health National Outcome Measures (NOMS): SAMHSA Uniform Reporting System*.; 2014. Accessed February 9, 2021. https://www.samhsa.gov/data/sites/default/files/URSTables2014/Tennessee.pdf

114. Substance Abuse and Mental Health Services Administration. *Tennessee 2015 Mental Health National Outcome Measures (NOMS): SAMHSA Uniform Reporting System*.; 2015. Accessed February 9, 2021. https://www.samhsa.gov/data/sites/default/files/Tennessee.pdf

115. Substance Abuse and Mental Health Services Administration. *Tennessee 2016 Mental Health National Outcome Measures (NOMS): SAMHSA Uniform Reporting System*.; 2016. Accessed February 9, 2021. https://www.samhsa.gov/data/sites/default/files/Tennessee-2016.pdf

116. Substance Abuse and Mental Health Services Administration. *Tennessee 2017 Mental Health National Outcome Measures (NOMS): SAMHSA Uniform Reporting System*.; 2017. Accessed February 9, 2021. https://www.samhsa.gov/data/sites/default/files/cbhsq-reports/Tennessee-2017.pdf

117. Substance Abuse and Mental Health Services Administration. *Tennessee 2018 Mental Health National Outcome Measures (NOMS): SAMHSA Uniform Reporting System*.; 2018. Accessed February 9, 2021. https://www.samhsa.gov/data/sites/default/files/cbhsq-reports/Tennessee-2018.pdf

118. Walker TY. National, Regional, State, and Selected Local Area Vaccination Coverage Among Adolescents Aged 13–17 Years — United States, 2018. *MMWR Morb Mortal Wkly Rep*. 2019;68. doi:10.15585/mmwr.mm6833a2

119. Markowitz LE, Gee J, Chesson H, Stokley S. Ten Years of Human Papillomavirus Vaccination in the United States. *Academic Pediatrics*. 2018;18(2):S3-S10. doi:10.1016/j.acap.2017.09.014

120. Average Annual Percent Change (AAPC) — Joinpoint help system 4.5.0.1. Accessed December 4, 2017. https://surveillance.cancer.gov/help/joinpoint/setting-parameters/method-and-parameters-tab/average-annual-percent-change-aapc

121. Committee on Practice Bulletins—Gynecology. ACOG committee opinion No. 431: routine pelvic examination and cervical cytology screening. *Obstet Gynecol*. 2009;113(5):1190-1193. doi:10.1097/AOG.0b013e3181a6d022

122. Boersma P, Black LI. Human papillomavirus vaccination among adults aged 18-26, 2013-2018. *NCHS Data Brief*. 2020;(354):1-8.

123. Withers R. People ages 27 to 45 can now get the HPV vaccine. But will insurance cover it? Slate Magazine. Published October 9, 2018. Accessed December 3, 2020. https://slate.com/technology/2018/10/hpv-vaccine-insurance-coverage-age-27-45.html

124. Lindley MC, Smith PJ, Rodewald LE. Vaccination coverage among U.S. adolescents aged 13–17 years eligible for the Vaccines for Children Program, 2009. *Public Health Rep*. 2011;126(Suppl 2):124-134.

125. Wheeler CM. Natural history of human papillomavirus infections, cytologic and histologic abnormalities, and cancer. *Obstet Gynecol Clin North Am*. 2008;35(4):519-536; vii. doi:10.1016/j.ogc.2008.09.006

126. Curtis CR, Dorell C, Yankey D, et al. National human papillomavirus vaccination coverage among adolescents aged 13-17 years-National Immunization Survey--teen, United States, 2011. *MMWR supplements*. 2014;63(2):61-70.

127. Henry KA, Stroup AM, Warner EL, Kepka D. Geographic factors and human papillomavirus (HPV) vaccination initiation among adolescent girls in the United States.

Cancer Epidemiol Biomarkers Prev. 2016;25(2):309-317. doi:10.1158/1055-9965.EPI-15-0658

128. Henry KA, Swiecki-Sikora AL, Stroup AM, Warner EL, Kepka D. Area-based socioeconomic factors and Human Papillomavirus (HPV) vaccination among teen boys in the United States. *BMC Public Health*. 2017;18(1):19. doi:10.1186/s12889-017-4567-2

129. Monnat SM, Rhubart DC, Wallington SF. Differences in human papillomavirus vaccination among adolescent girls in metropolitan versus non-metropolitan areas: Considering the moderating Roles of maternal socioeconomic status and health care access. *Matern Child Health J*. 2016;20(2):315-325. doi:10.1007/s10995-015-1831-x

130. Walker TY, Elam-Evans LD, Williams CL, et al. Trends in human papillomavirus (HPV) vaccination initiation among adolescents aged 13–17 by metropolitan statistical area (MSA) status, National Immunization Survey – Teen, 2013 – 2017. *Human Vaccines & Immunotherapeutics*. 2020;16(3):554-561. doi:10.1080/21645515.2019.1671765

131. Lee M, Gerend MA, Boakye EA. Rural–Urban differences in human papillomavirus vaccination among young adults in 8 U.S. states. *American Journal of Preventive Medicine*. 2020;0(0). doi:10.1016/j.amepre.2020.07.023

132. Smith MA, Liu B, McIntyre P, Menzies R, Dey A, Canfell K. Trends in genital warts by socioeconomic status after the introduction of the national HPV vaccination program in Australia: analysis of national hospital data. *BMC Infect Dis*. 2016;16. doi:10.1186/s12879-016-1347-z

133. Thompson LH, Nugent Z, Blanchard JF, Ens C, Yu BN. Increasing incidence of anogenital warts with an urban–rural divide among males in Manitoba, Canada, 1990–2011. *BMC Public Health*. 2016;16. doi:10.1186/s12889-016-2885-4

134. US Census Bureau Geography. 2010 geographic terms and concepts - Core based statistical areas and related statistical areas. Accessed December 3, 2017. https://www.census.gov/geo/reference/gtc/gtc_cbsa.html

135. Clayton D, Schifflers E. Models for temporal variation in cancer rates. I: Age-period and age-cohort models. *Stat Med*. 1987;6(4):449-467. doi:10.1002/sim.4780060405

136. Clayton D, Schifflers E. Models for temporal variation in cancer rates. II: Age-period-cohort models. *Stat Med*. 1987;6(4):469-481. doi:10.1002/sim.4780060406

137. Carstensen B. Age–period–cohort models for the Lexis diagram. *Statistics in Medicine*. 2007;26(15):3018-3045. doi:10.1002/sim.2764

138. Carstensen B, Plummer M, Laara E, Hills M. *Epi: Statistical Analysis in Epidemiology*.; 2020. Accessed December 12, 2020. https://CRAN.R-project.org/package=Epi

139. Sabeena S, Bhat PV, Kamath V, et al. Community-based prevalence of genital human papilloma virus (HPV) Infection: A systematic review and meta-analysis. *Asian Pac J Cancer Prev*. 2017;18(1):145-154. doi:10.22034/APJCP.2017.18.1.145

140. Niccolai LM, Julian PJ, Bilinski A, et al. Geographic poverty and racial/ethnic disparities in cervical cancer precursor rates in connecticut, 2008–2009. *Am J Public Health*. 2013;103(1):156-163. doi:10.2105/AJPH.2011.300447

APPENDIX TABLE OF CONTENTS

Page

Appendix A – Chapter III Appendix

APPENDIX A – CHAPTER III APPENDIX

Appendix Table A1. Additional confirmed incident CIN2+[a] events identified by TennCare audits.

Year	Total Gold Standard Confirmed Incident CIN2+ Events	New Incident CIN2+ Events Added from Audit n (%[b])
2008	119	--
2009	106	--
2010	134	19[c] (5.6)
2011	66	11 (20.0)
2012	78	3 (4.0)
2013	99	4 (4.2)
2014	84	4 (5.0)
2015	104	8 (8.3)
2016	119	7 (6.3)
2017	74	3 (4.2)

CIN = Cervical Intraepithelial Lesion; TennCare = Tennessee Medicaid.

[a]CIN2+ includes CIN2, CIN3, and adenocarcinoma *in situ*.

[b]The annual percent of new incident events added from the TennCare audit is calculated by the following equation:

((Total Gold Standard Confirmed Incident CIN2+ Events / (Total Gold Standard Confirmed Incident CIN2+ Events – New Incident CIN2+ Events Added from Audit)) – 1) x 100.

[c]The reported 2010 TennCare audit represents new events identified from 2008 to 2010 combined.

Appendix Table A2. Characteristics of cervical screening tests (N = 88,765) among 42,324 TennCare-enrolled women aged 18-39 years in Davidson County, Tennessee by ICD era[a].

Characteristic	ICD-9 Era N = 69,792 n (column %)	ICD-10 Era N = 18,973 n (column %)	P-Value
Confirmed CIN2+ Event			0.299
Yes	641 (0.9)	159 (0.8)	
No	69,151 (99.1)	18,814 (99.2)	
Age Group, years			<0.001*
18-24	26,052 (37.3)	5,136 (27.1)	
25-29	19,459 (27.9)	5,363 (28.3)	
30-39	24,281 (34.8)	8,474 (44.7)	
Race/Ethnicity			<0.001*
NH White	15,310 (21.9)	3,588 (18.9)	
NH Black	23,724 (34.0)	5,647 (29.8)	
NH Other/Unknown	29,003 (41.6)	9,338 (49.2)	
Hispanic	1,755 (2.5)	400 (2.1)	
CIN2+[b] Tissue Diagnosis Code			0.002*
Yes	1,281 (1.8)	285 (1.5)	
No	68,511 (98.2)	18,688 (98.5)	
Non-Specific CIN Tissue Diagnosis Code			<0.001*
Yes	937 (1.3)	136 (0.7)	
No	68,855 (98.7)	18,837 (99.3)	
High-Grade Squamous Intraepithelial Lesion Cytologic Diagnosis Code			<0.001*
Yes	892 (1.3)	175 (0.9)	
No	68,900 (98.7)	18,798 (99.1)	
CIN1 Tissue Diagnosis Code			<0.001*
Yes	1,987 (2.9)	337 (1.8)	
No	67,805 (97.2)	18,636 (98.2)	
Low-Grade Squamous Intraepithelial Lesion Cytologic Diagnosis Code			<0.001*
Yes	4,178 (6.0)	825 (4.4)	
No	65,614 (94.0)	18,148 (95.7)	
Atypical Squamous Cells of Undetermined Significance Diagnosis Code			<0.001*
Yes	6,654 (9.5)	1,260 (6.6)	
No	63,138 (90.5)	17,713 (93.4)	
Human Papillomavirus Screening Test Code			<0.001*
Yes	1,040 (1.5)	609 (3.2)	
No	68,752 (98.5)	18,364 (96.8)	

Pap Smear/Test Code			<0.001*
Yes	62,242 (89.2)	16,328 (86.2)	
No	7,550 (10.8)	2,645 (13.9)	
Human Papillomavirus DNA Test Code			<0.001*
Yes	6,510 (9.3)	2,036 (10.7)	
No	63,282 (90.7)	16,937 (89.3)	
Cervical Treatment Procedure Code			<0.001*
Yes	185 (0.3)	19 (0.1)	
No	69,607 (99.7)	18,954 (99.9)	
Cervical or Vaginal Biopsy Code			<0.001*
Yes	3,150 (4.5)	671 (3.5)	
No	66,642 (95.5)	18,302 (96.5)	
Had a Cervical Diagnostic Procedure Code			<0.001*
Yes	5,579 (8.0)	1,053 (5.6)	
No	64,213 (92.0)	17,920 (94.5)	

Abbreviations: CIN = Cervical Intraepithelial Lesion; DNA = Deoxyribonucleic Acid; NH = Non-Hispanic; ICD = International Classification of Diseases, Clinical Modification; TennCare = Tennessee Medicaid.

[a]The ICD-9 era includes procedures from January 1, 2008, through September 30, 2015; the ICD-10 era includes procedures from October 1, 2015, through December 31, 2017.

*Asterisks denote $p < 0.05$.

Appendix Table A3. Coding characteristics of confirmed CIN2+ events versus non-events among cervical screening tests of TennCare-enrolled women aged 18-39 years in Davidson County, Tennessee by ICD era[a].

	Overall N = 88,765			ICD-9 Era N = 69,792			ICD-10 Era N = 18,973		
Code Grouping	Confirmed CIN2+ Event n (Col %)	Non-Events n (Col %)	P-Value	Confirmed CIN2+ Event n (Col %)	Non-Events n (Col %)	P-Value	Confirmed CIN2+ Event n (Col %)	Non-Events n (Col %)	P-Value
CIN2+ Tissue Diagnosis			<0.001*			<0.001*			<0.001*
Yes	783 (97.9)	783 (0.9)		628 (98.0)	653 (0.9)		155 (97.5)	130 (0.7)	
No	17 (2.1)	87,182 (99.1)		13 (2.0)	68,498 (99.1)		4 (2.5)	18,684 (99.3)	
Non-Specific CIN Tissue Diagnosis			<0.001*			<0.001*			<0.001*
Yes	255 (31.9)	818 (0.9)		213 (33.2)	724 (1.1)		42 (26.4)	94 (0.5)	
No	545 (68.1)	87,147 (99.1)		428 (66.8)	68,427 (99.0)		117 (73.6)	18,720 (99.5)	
High-Grade Squamous Intraepithelial Lesion Cytologic Diagnosis			<0.001*			<0.001*			<0.001*
Yes	365 (45.6)	702 (0.8)		299 (46.7)	593 (0.9)		66 (41.5)	109 (0.6)	
No	435 (54.4)	87,263 (99.2)		342 (53.4)	68,558 (99.1)		93 (58.5)	18,705 (99.4)	
CIN1 Tissue Diagnosis			<0.001*			<0.001*			<0.001*
Yes	277 (34.6)	2,047 (2.3)		226 (35.3)	1,761 (2.6)		51 (32.1)	286 (1.5)	
No	523 (65.4)	85,918 (97.7)		415 (64.7)	64,390 (97.5)		108 (67.9)	18,528 (98.5)	
Low-Grade Squamous Intraepithelial Lesion Cytologic Diagnosis			<0.001*			<0.001*			<0.001*
Yes	301 (37.6)	4,702 (5.4)		249 (38.9)	3,929 (5.7)		52 (32.7)	773 (4.1)	
No	499 (62.4)	83,263 (94.7)		392 (61.2)	65,222 (94.3)		107 (67.3)	18,041 (95.9)	
Atypical Squamous Cells of Undetermined Significance Diagnosis			<0.001*			<0.001*			<0.001*
Yes	324 (40.5)	7,590 (8.6)		256 (39.9)	6,398 (9.3)		68 (42.8)	1,192 (6.3)	
No	476 (59.5)	80,375 (91.4)		385 (60.1)	62,753 (90.8)		91 (57.2)	17,622 (93.7)	
Cervical Treatment Procedure			<0.001*			<0.001*			<0.001*
Yes	47 (5.9)	157 (0.2)		41 (6.4)	144 (0.2)		6 (3.8)	13 (0.1)	
No	753 (94.1)	87,808 (99.8)		600 (93.6)	69,007 (99.8)		153 (96.2)	18,801 (99.9)	
Cervical or Vaginal Biopsy			<0.001*			<0.001*			<0.001*
Yes	692 (86.5)	3,129 (3.6)		558 (87.1)	2,592 (3.8)		134 (84.3)	537 (2.9)	

No	108 (13.4)	84,836 (96.4)		83 (13.0)	66,559 (96.3)		25 (15.7)	18,277 (97.2)	
Cervical Diagnostic Procedure			<0.001*			<0.001*			<0.001*
Yes	785 (98.1)	5,847 (6.7)		628 (98.0)	4,951 (7.2)		157 (98.7)	896 (4.8)	
No	15 (1.9)	82,118 (93.4)		13 (2.0)	64,200 (92.8)		2 (1.3)	17,918 (95.2)	

Abbreviations: CIN = Cervical Intraepithelial Lesion; Col = Column; DNA = Deoxyribonucleic Acid; ICD = International Classification of Diseases, Clinical Modification; TennCare = Tennessee Medicaid.

[a]The ICD-9 era includes procedures from January 1, 2008, through September 30, 2015; the ICD-10 era includes procedures from October 1, 2015, through December 31, 2017.

*Asterisks denote $p < 0.05$.

Appendix Table A4. Randomized search results[a] of random forest algorithms to classify CIN2+ event status in the training set of cervical screening tests (N = 53,259) among TennCare-enrolled women aged 18-39 years in Davidson County, Tennessee.

Model Rank	Number of Trees	Maximum Predictors Selection Method	Maximum Tree Depth	Minimum Number of Samples for a Split	Minimum Number of Samples in a Leaf Node	Mean Validation Score ± Standard Deviation
1	23	Automatic	36	5	8	0.994386 ± 0.000343
2	89	Square Root	78	10	10	0.994311 ± 0.000401
3	111	Square Root	57	10	10	0.994255 ± 0.000338
4	200	Square Root	36	5	8	0.994217 ± 0.000401
5	45	Automatic	57	10	8	0.994217 ± 0.000439

Abbreviations: CIN = Cervical Intraepithelial Lesion; TennCare = Tennessee Medicaid.
[a]Only the best five performing models are reported.

Appendix Table A5. Beta coefficients and predictor importance scores of LASSO and random forest models[a] to classify CIN2+ event status in the training set (N = 53,259) of cervical screening tests among TennCare-enrolled women aged 18-39 years in Davidson County, Tennessee.

Predictors	LASSO Beta Coefficients	Random Forest Predictor Importance Scores
Constant	-9.529154	—
CIN2+ Tissue Diagnosis	6.028418	0.666390
Non-Specific CIN Diagnosis	0.4756844	0.027637
High-Grade Squamous Intraepithelial Lesion Diagnosis	0.7419008	0.061153
CIN1 Tissue Diagnosis	—	0.017307
Low-Grade Squamous Intraepithelial Lesion Diagnosis	0.2370433	0.010115
Atypical Squamous Cells of Undetermined Significance Diagnosis	—	0.006796
Cervical Treatment Procedure	0.2136401	0.001213
Cervical or Vaginal Biopsy	—[b]	0.126696
Cervical Diagnostic Procedure	3.363726	0.082693

Abbreviations: CIN = Cervical Intraepithelial Lesion; ICD = International Classification of Diseases, Clinical Modification; LASSO = Least Absolute Shrinkage and Selection Operator; TennCare = Tennessee Medicaid.

[a]Models were built using training sets of both ICD-9 and ICD-10 eras combined.

[b]Among cervical screening tests, the predictor Cervical or Vaginal Biopsy was highly correlated with Cervical Diagnostic Procedure and CIN1 Tissue Diagnosis; thus, we removed Cervical or Vaginal Biopsy when building the LASSO algorithm among cervical screening tests.

Appendix Table A6. Performance of prediction models to classify CIN2+ event status among cervical screening tests of TennCare-enrolled women aged 18-39 years in Davidson County, Tennessee by ICD era[a].

Performance Measure	CIN2+ Tissue Diagnosis Code Alone		LASSO				Random Forest			
	ICD-9 (N = 69,792)	ICD-10 (N = 18,973)	ICD-9 (N = 69,792)		ICD-10 (N = 18,973)		ICD-9 (N = 69,792)		ICD-10 (N = 18,973)	
	Testing Set (n = 27,917)	Testing Set (n = 7,589)	Training Set (n = 41,875)	Testing Set (n = 27,917)	Training Set (n = 11,384)	Testing Set (n = 7,589)	Training Set (n = 41,875)	Testing Set (n = 27,917)	Training Set (n = 11,384)	Testing Set (n = 7,589)
Sensitivity, % (95% CI)	99.2 (97.0, 99.9)	95.1 (86.3, 98.9)	81.7 (77.6, 85.4)	82.6 (77.2, 87.2)	72.5 (62.5, 81.0)	75.4 (62.7, 85.5)	86.9 (83.2, 90.0)	87.7 (82.8, 91.6)	85.7 (77.2, 92.0)	80.3 (68.2, 89.4)
Specificity, % (95% CI)	99.1[b] (98.9, 99.2)	99.5[b] (99.3, 99.6)	99.5 (99.5, 99.6)	99.5[b] (99.4, 99.6)	99.6 (99.5, 99.7)	99.8[b] (99.7, 99.9)	99.6 (99.5, 99.6)	99.5[b] (99.5, 99.6)	99.6 (99.5, 99.7)	99.8[b] (99.6, 99.9)
PPV, % (95% CI)	47.6 (43.1, 52.1)	58.6 (48.2, 68.4)	63.3 (59.0, 67.4)	58.7 (53.2, 64.1)	62.3 (52.7, 71.2)	76.7 (64.0, 86.6)	66.7 (62.5, 70.7)	61.8 (56.4, 67.0)	65.6 (56.7, 73.8)	74.2 (62.0, 84.2)
NPV, % (95% CI)	100.0 (100.0, 100.0)	100.0 (99.9, 100.0)	99.8 (99.8, 99.9)	99.9 (99.8, 99.9)	99.8 (99.7, 99.8)	99.8 (99.7, 99.9)	99.9 (99.8, 99.9)	99.9 (99.8, 99.9)	99.9 (99.8, 99.9)	99.8 (99.7, 99.9)
Accuracy, % (95% CI)	99.1 (99.0, 99.2)	99.4 (99.2, 99.6)	99.4 (99.3, 99.4)	99.4[b] (99.3, 99.5)	99.4[c] (99.2, 99.5)	99.6 [b,c] (99.7, 99.9)	99.5 (99.4, 99.5)	99.4 (99.3, 99.5)	99.5 (99.3, 99.6)	99.6 (99.5, 99.7)
C-Index, % (95% CI)	99.1 (98.5, 99.7)	97.3 (94.5, 100.0)	90.6 (88.8, 92.5)	91.1 (88.6, 93.5)	86.0 (81.6, 90.5)	87.6 (82.2, 93.1)	93.2 (91.6, 95.7)	93.6 (89.2, 96.1)	92.7 (89.2, 96.1)	90.1 (85.0, 95.1)

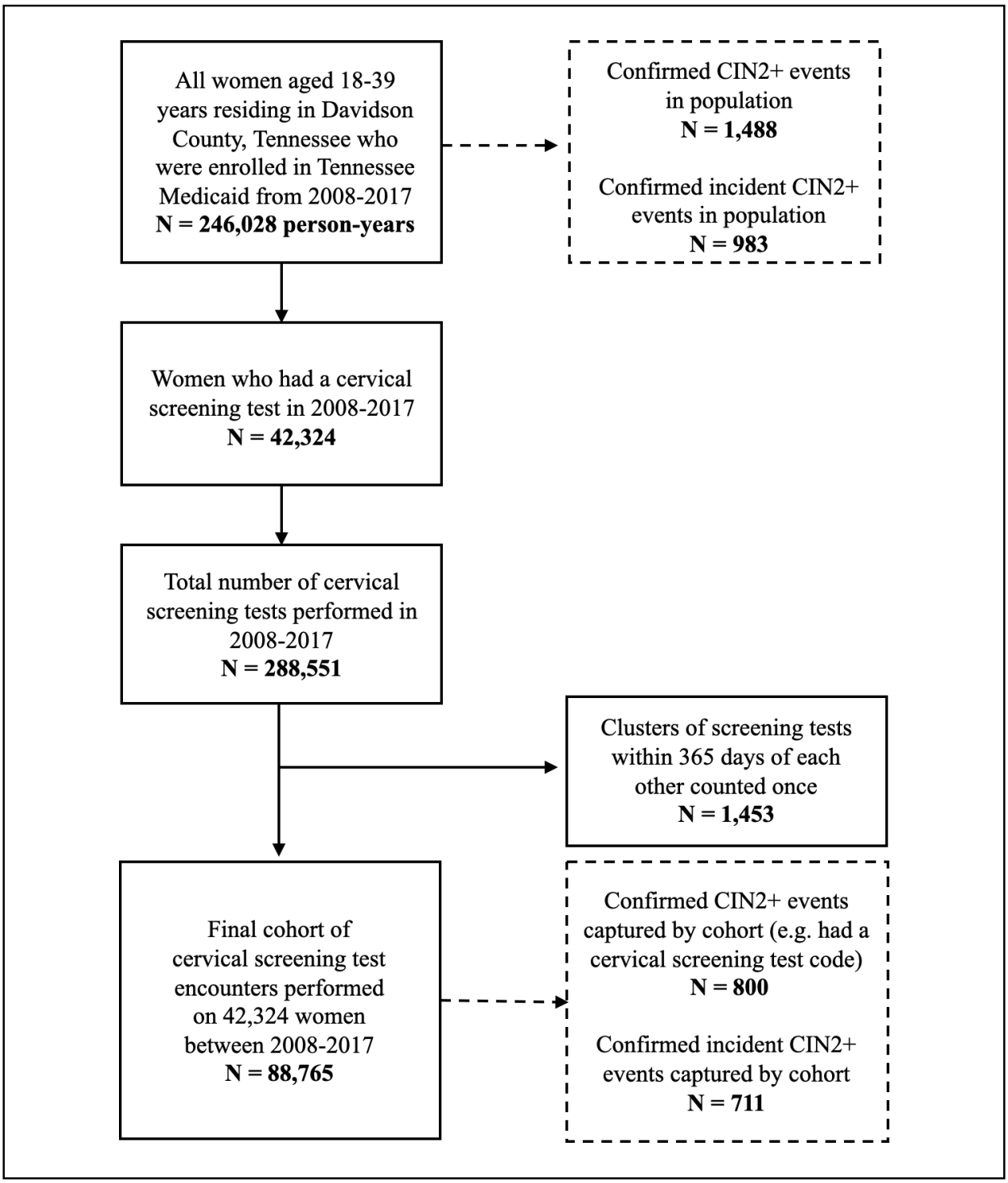
Abbreviations: CI = Confidence Interval; CIN = Cervical Intraepithelial Lesion; ICD = International Classification of Diseases, Clinical Modification; LASSO = Least Absolute Shrinkage and Selection Operator; NPV = Negative Predictive Value; PPV = Positive Predictive Value; TennCare = Tennessee Medicaid.

[a]The ICD-9 era includes procedures from January 1, 2008, through September 30, 2015; the ICD-10 era includes procedures from October 1, 2015, through December 31, 2017.

[b]Performance in the testing sets among the ICD-9 and ICD-10 eras within each model type are statistically significantly different (i.e., confidence intervals do not overlap with each other).

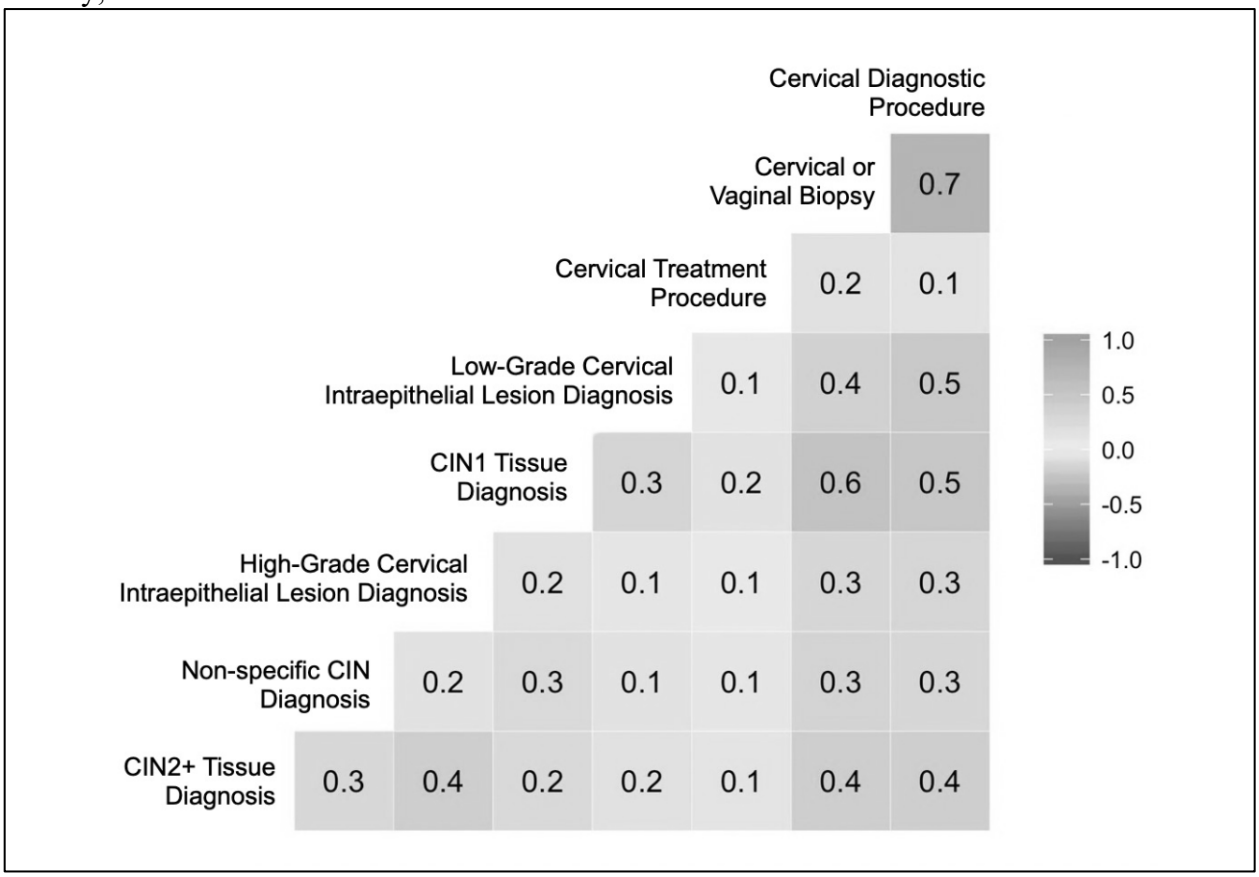
[c]Performance between the training and tests sets within each model type statistically significantly different (confidence intervals do not overlap with each other).

Appendix Figure A1. Flow diagram to capture cohort of cervical screening test encounters from 2008 to 2017 among TennCare-enrolled women aged 18-39 years residing in Davidson County, Tennessee.



Abbreviations: CIN = Cervical Intraepithelial Lesion; TennCare = Tennessee Medicaid.

Appendix Figure A2. Correlation matrix of predictors selected in the model built by LASSO among cervical screening tests of TennCare-enrolled women aged 18-39 years in Davidson County, Tennessee.



Abbreviations: CIN = Cervical Intraepithelial Lesion; LASSO = Least Absolute Shrinkage and Selection Operator; TennCare = Tennessee Medicaid.

Appendix Figure A3. Confusion matrices of claims-based models to classify CIN2+ event status in the testing set (N = 35,506) of cervical screening tests among TennCare-enrolled women aged 18-39 years in Davidson County, Tennessee by ICD era[a].

Legend

	Confirmed Events (n)	Confirmed Non-Events (n)	Total Classified (N)
Classified Events (n)	True Positives	False Positives	Total Classified Events
Classified Non-Events (n)	False Negatives	True Negatives	Total Classified Non-Events
Total Gold Standard (N)	Total Confirmed Events	Total Confirmed Non-Events	Total Sample Size

CIN2+ Tissue Diagnosis Codes Alone

ICD-9 Era

	Confirmed Events (n)	Confirmed Non-Events (n)	Total Classified (N)
Classified Events (n)	628	653	1,281
Classified Non-Events (n)	13	68,498	68,511
Total Gold Standard (N)	641	69,151	69,792

ICD-10 Era

	Confirmed Events (n)	Confirmed Non-Events (n)	Total Classified (N)
Classified Events (n)	155	130	285
Classified Non-Events (n)	4	18,684	18.688
Total Gold Standard (N)	159	18,814	18,973

LASSO

ICD-9 Era

	Confirmed Events (n)	Confirmed Non-Events (n)	Total Classified (N)
Classified Events (n)	159	78	237
Classified Non-Events (n)	77	27,603	27,680
Total Gold Standard (N)	236	27,681	27,917

ICD-10 Era

	Confirmed Events (n)	Confirmed Non-Events (n)	Total Classified (N)
Classified Events (n)	38	7	45
Classified Non-Events (n)	23	7,521	7,544
Total Gold Standard (N)	61	7,528	7,589

Random Forest

ICD-9 Era

	Confirmed Events (n)	Confirmed Non-Events (n)	Total Classified (N)
Classified Events (n)	207	128	335
Classified Non-Events (n)	29	27,553	27,582
Total Gold Standard (N)	236	27,681	27,917

ICD-10 Era

	Confirmed Events (n)	Confirmed Non-Events (n)	Total Classified (N)
Classified Events (n)	49	17	66
Classified Non-Events (n)	12	7,511	7,523
Total Gold Standard (N)	61	7,528	7,589

Abbreviations: CIN = Cervical Intraepithelial Lesion; ICD = International Classification of Diseases, Clinical Modification; LASSO = Least Absolute Shrinkage and Selection Operator; TennCare = Tennessee Medicaid.

[a]The ICD-9 era includes procedures from January 1, 2008, through September 30, 2015; the ICD-10 era includes procedures from October 1, 2015, through December 31, 2017.