Determinants of Human Papillomavirus Vaccine Initiation for female teens in the United States: A Classification and Regression Tree Analysis

Alexis R. Santos Lozada, MA
Diego Caraballo
José N. Caraballo, PhD

Introduction

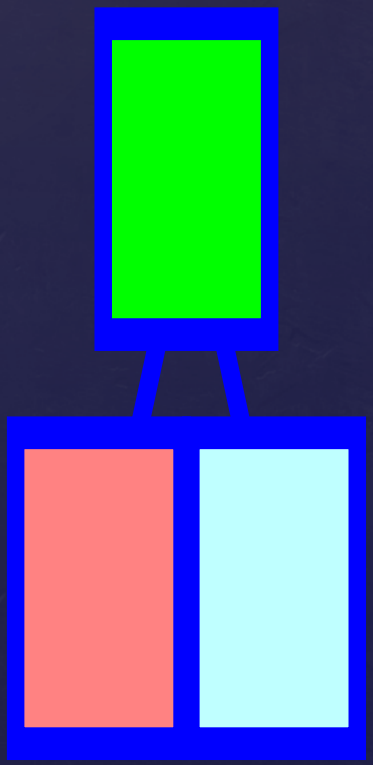
- Human Papillomavirus is one of the most common sexually transmitted infections in the United States.
- This infection has been catalogued as a requirement for the development of cervical cancer later in life.
- Since 2006 the USDA has approved a vaccine to prevent Human Papillomavirus Infection in females and males.
- The recommendation from the Advisory Committee on Immunization Practices (ACIP) is a three dose series before reaching adulthood.

Objective

- The purpose of this research is to study the determinants of Human Papillomavirus Vaccine Initiation in female teens in the United States using Classification and Regression Tree Methods.
- Initiation is a crucial issue to study as it may be influenced by several factors such as: race/ethnicity, culture, health insurance availability, poverty status, among others.

CART

- Classification and Regression Trees are employed in numerous ways in the United States.
- Examples:
  - UC – San Diego Medical Center
    - Classifying a patient as high risk or low risk
  - EPA
    - Measuring predicted values of water bodies based on its characteristics/pollution.



Data and Methods

- Data for this research come from the 2011 National Immunization Survey - Teen Sample.
  - This sample collects immunization information for a nationally representative sample of adolescents (13-17 years).
- Variable creation and descriptive statistics processes were completed using SAS.
- The method of analysis is Classification and Regression Trees. Classification Trees were generated using CART Software (Salford Systems).
- We employed a Gini Impurity Index for classification calculations and a minimum size of the terminal node of 15 cases.

Measurements

- Human Papillomavirus vaccine initiation is measured as a Yes or No.
- This analysis incorporates:
  - Race: Non-Hispanic White, Non-Hispanic Black, Hispanic, Non-Hispanic Other.
  - Health Insurance: Private Health Insurance, Medicaid, S-Chip, Other Health Insurance.
  - Poverty Status: Below poverty level, Above poverty under $75,000, Above poverty with income over $75,000.
  - Mother's marital status: Married, Non Married.
  - Teen's Age
  - Mother's Age and Education level.
  - Household knowledge about HPV and HPV Vaccine.

- We employed a learning-testing way of producing the classification trees in which half the dataset was used to create an initial tree and the other half was used to test the consistency of the results.

Original Dataset: n = 10,172

Learning Dataset

(n=5,077)

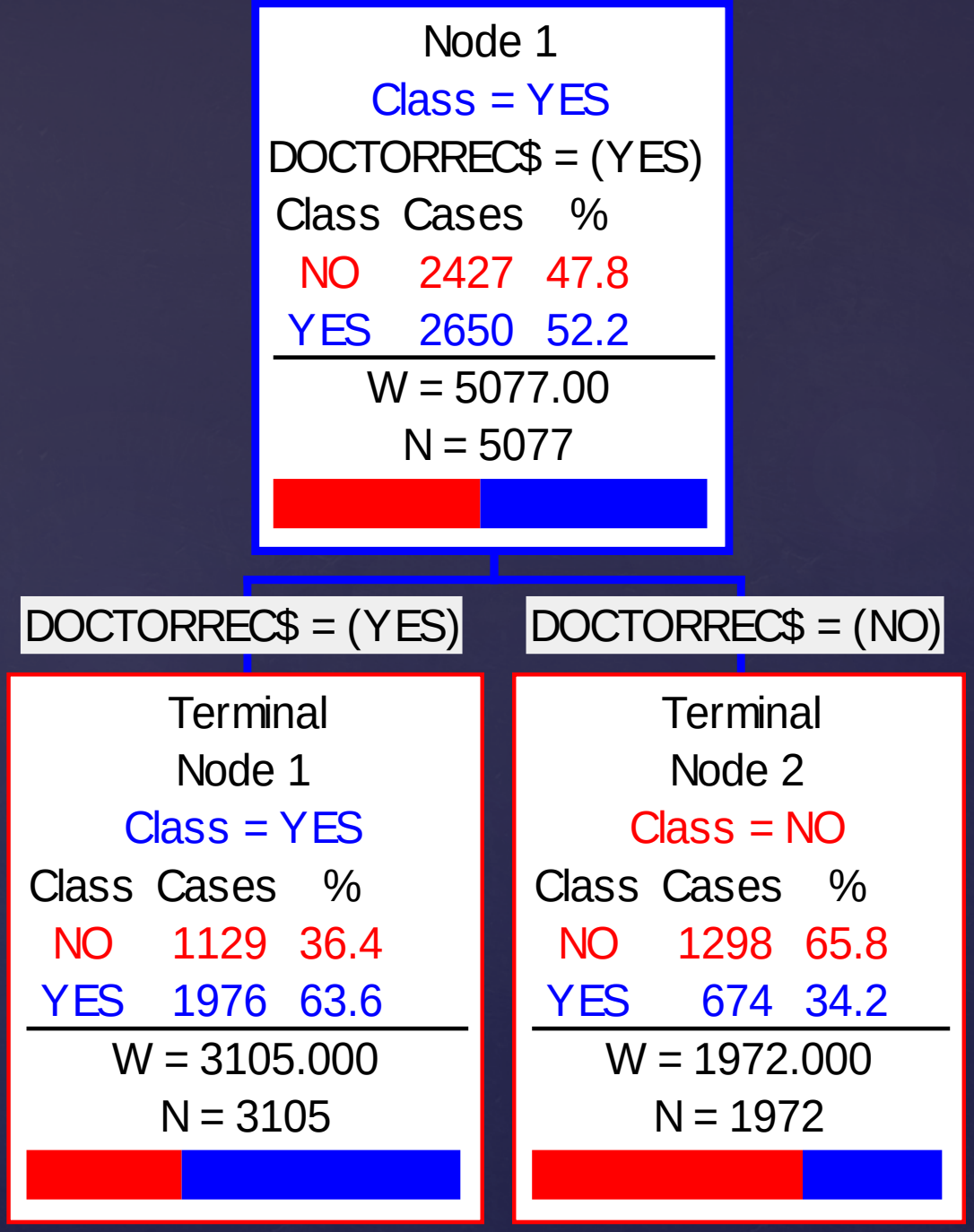
Testing Dataset

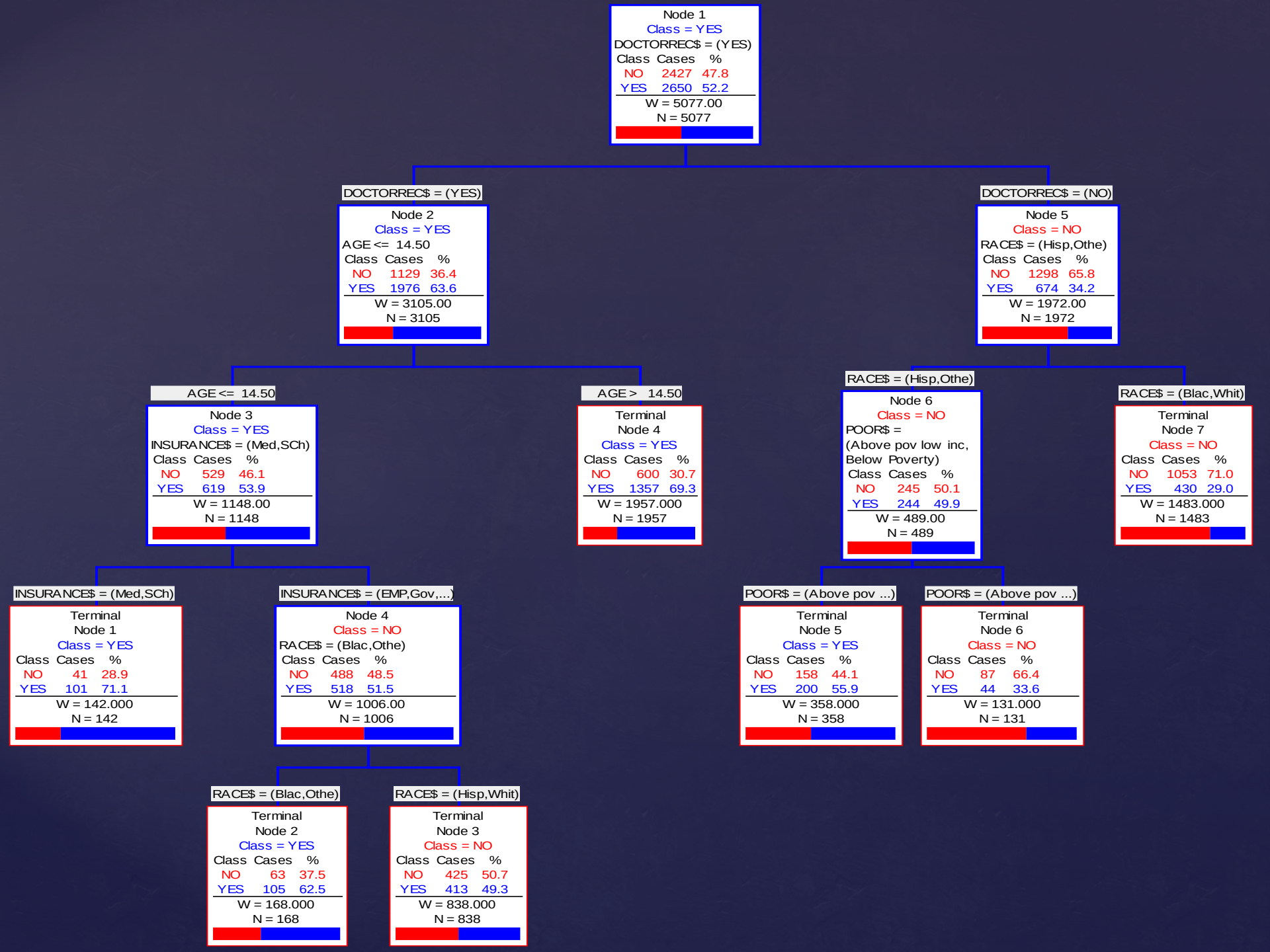
(n=5,095)

- Dataset splitting was done using a randomly assigned dummy identifier (0,1). Where cases with zero became part of the learning tree and cases with one were used to test the initial results.

Data Management

Results


Node 1
Class = YES
DOCTORREC$ = (YES)
Class Cases %
NO 2427 47.8
YES 2650 52.2
W = 5077.00
N = 5077
DOCTORREC$ = (YES)
DOCTORREC$ = (NO)
Terminal
Node 1
Class = YES
Class Cases %
NO 1129 36.4
YES 1976 63.6
W = 3105.000
N = 3105
Terminal
Node 2
Class = NO
Class Cases %
NO 1298 65.8
YES 674 34.2
W = 1972.000
N = 1972


Node 1
Class = YES
DOCTORREC$ = (YES)
Class Cases %
NO 2427 47.8
YES 2650 52.2
W = 5077.00
N = 5077
DOCTORREC$ = (YES)
Node 2
Class = YES
AGE <= 14.50
Class Cases %
NO 1129 36.4
YES 1976 63.6
W = 3105.00
N = 3105
DOCTORREC$ = (NO)
Node 5
Class = NO
RACE$ = (Hisp,Othe)
Class Cases %
NO 1298 65.8
YES 674 34.2
W = 1972.00
N = 1972
AGE <= 14.50
Node 3
Class = YES
INSURANCE$ = (Med,SCh)
Class Cases %
NO 529 46.1
YES 619 53.9
W = 1148.00
N = 1148
AGE > 14.50
Terminal
Node 4
Class = YES
Class Cases %
NO 600 30.7
YES 1357 69.3
W = 1957.000
N = 1957
RACE$ = (Hisp,Othe)
Node 6
Class = NO
POOR$ =
(Above pov low inc,
Below Poverty)
Class Cases %
NO 245 50.1
YES 244 49.9
W = 489.00
N = 489
RACE$ = (Blac,Whit)
Terminal
Node 7
Class = NO
Class Cases %
NO 1053 71.0
YES 430 29.0
W = 1483.000
N = 1483
INSURANCE$ = (Med,SCh)
Terminal
Node 1
Class = YES
Class Cases %
NO 41 28.9
YES 101 71.1
W = 142.000
N = 142
INSURANCE$ = (EMP,Gov,...)
Node 4
Class = NO
RACE$ = (Blac,Othe)
Class Cases %
NO 488 48.5
YES 518 51.5
W = 1006.00
N = 1006
POOR$ = (Above pov ...)
Terminal
Node 5
Class = YES
Class Cases %
NO 158 44.1
YES 200 55.9
W = 358.000
N = 358
POOR$ = (Above pov ...)
Terminal
Node 6
Class = NO
Class Cases %
NO 87 66.4
YES 44 33.6
W = 131.000
N = 131
RACE$ = (Blac,Othe)
Terminal
Node 2
Class = YES
Class Cases %
NO 63 37.5
YES 105 62.5
W = 168.000
N = 168
RACE$ = (Hisp,Whit)
Terminal
Node 3
Class = NO
Class Cases %
NO 425 50.7
YES 413 49.3
W = 838.000
N = 838

Conclusions

- Public policy initiatives aiming to deal with increasing the HPV vaccine initiation rate should target specific groups.
- Doctor's recommendation seems to be of outmost importance, physicians should continue to encourage their patients to vaccinate.
- Health Insurance coverage should also be targeted, as it seems those with public funded health insurance are more prone to have initiated the vaccine series.
- Specific initiatives to deal with the postponement of HPV initiation in terms of age, should also be targeted when dealing with promotion initiatives.

- Questions and comments are more than welcomed.

Thank you for your attention!