Effect of COVID-19 burden on changes in HPV vaccination coverage among females between 2019 and 2022

Rayleen Lewis

This uses MS Word as output format. [See here](https://quarto.org/docs/output-formats/ms-word.html) for more information. You can switch to other formats, like html or pdf. See [the Quarto documentation](https://quarto.org/) for other formats.

**Authors**

* Rayleen Lewis (ORCID: 0000-0000-1234-5678)

**Author affiliations**

1. College of Public Health, University of Georgia, Athens, GA, USA.

Corresponding author: rml20258@uga.edu

Disclaimer: The opinions expressed in this article are the author’s own and don’t reflect their employer or University.

# 1. Summary/Abstract

*Write a summary of your project.*

# 2. Introduction

## 2.1 General Background Information

*Provide enough background on your topic that others can understand the why and how of your analysis*

## 2.2 Description of data and data source

Information on the dat sources can be found in the Data Acquisition and Data Importing and Cleaning sections below. Briefly, all datasets will be reduced to include only countries with vaccination coverage estimates in both 2019 and 2022. Average coverage will be compared in 2022 to 2019. Multivariable linear regression modeling the difference in coverage between the two years as the outcome will be conducting using measures of COVID-19 burden, HPV vaccination program characteristics, country income information, and country-level demographics. Measure(s) of COVID-19 burden will be considered the explanatory variable of interest. Other characteristics will be considered for adjustment.

## 2.3 Questions/Hypotheses to be addressed

The objective of this work is to determine if HPV vaccination coverage changed between 2019 and 2022 among countries offering HPV vaccination as of 2019, and if so, are these changes correlated with COVID-19 burden in 2020-2021. I hypothesize that HPV vaccination coverage was lower in 2022 than 2019 among countries with high burden of COVID-19 in 2020-2021 and comparable in countries with a low COVID-19 burden.

# 3. Methods

Briefly, all datasets will be reduced to include only countries with vaccination coverage estimates in both 2019 and 2022. Average coverage will be compared in 2022 to 2019. Multivariable linear regression modeling the difference in coverage between the two years as the outcome will be conducting using measures of COVID-19 burden, HPV vaccination program characteristics, country income information, and country-level demographics. Measure(s) of COVID-19 burden will be considered the explanatory variable of interest. Other characteristics will be considered for adjustment.

## 3.1 Schematic of workflow

Sometimes you might want to show a schematic diagram/figure that was not created with code (if you can do it with code, do it). [Figure 1](#fig-schematic) is an example of some - completely random/unrelated - schematic that was generated with Biorender. We store those figures in the assets folder.

|  |
| --- |
| Figure 1: A figure that is manually generated and shows some overview/schematic. This has nothing to do with the data, it’s just a random one from one of our projects I found and placed here. |

## 3.2 Data aquisition

Data on HPV vaccination coverage and vaccination program details are available at the country level through the World Health Organization’s (WHO) [HPV Dashboard](https://www.who.int/teams/immunization-vaccines-and-biologicals/diseases/human-papillomavirus-vaccines-(HPV)/hpv-clearing-house/hpv-dashboard). Briefly, coverage data reflect administrative and official HPV vaccination coverage reported annually through the WHO/UNICEF Joint Reporting Form on Immunization. HPV Dashboard include country income level (low, lower middle, upper middle, high), whether there was a national HPV vaccine schedule, year of vaccine introduction, primary delivery strategy (school-based, facility based, varies by region, mixed), number of doses in routine schedule, whether the program is gender neutral (target males and females), WHO region, and year-specific coverage from 2010 to 2023.

Country-specific weekly COVID-19 case and death counts are available from starting in 1/4/2020 through the WHO [COVID-19 dashboard](https://data.who.int/dashboards/covid19/data?n=o). In August 2023, WHO stopped requiring weekly reporting, although some countries continued to voluntarily report COVID-19 data.

Additional country-level covariates were collected from other publicly available data sources. Information on country income level by year is available through the [World Bank](https://datahelpdesk.worldbank.org/knowledgebase/articles/906519-world-bank-country-and-lending-groups#:~:text=For%20the%20current%202025%20fiscal,those%20with%20a%20GNI%20per). Country-level demographics including life expectancy of females at birth and population size were available from the [United Nations](https://population.un.org/wpp/downloads?folder=Standard%20Projections&group=CSV%20format)

All data sources are publicly available for download and were downloaded on January 26-27, 2025.

## 3.3 Data import and cleaning

*Write code that reads in the file and cleans it so it’s ready for analysis. Since this will be fairly long code for most datasets, it might be a good idea to have it in one or several R scripts. If that is the case, explain here briefly what kind of cleaning/processing you do, and provide more details and well documented code somewhere (e.g. as supplement in a paper). All materials, including files that contain code, should be commented well so everyone can follow along.* Each data source must be imported. Data sources can be merged based on country. Each dataset will be restricted to countries with vaccination coverage estimates in 2019 and 2022.

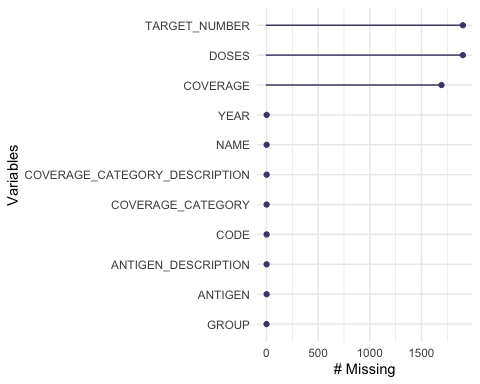
#### 3.3.0.1 HPV coverage and vaccination program data

There are two data files to be imported for the HPV vaccination data, one file contains yearly coverage estimates and the other contains information about the HPV vaccination program.

##### 3.3.0.1.1 HPV vaccination coverage

A total of 196 countries are included in the data set. Coverage estimates are available from 2010 to 2023, but most countries are missing on some years of data. Data manipulation will be needed to restrict to the years of interest and convert data form long to wide.

#Importing yearly HPV vaccination coverage  
hpv\_coverage\_raw <- read.xlsx(here("data","raw-data","Human Papillomavirus (HPV) vaccination coverage 2025-21-01 11-57 UTC.xlsx"), sheet = "Sheet1", colNames = TRUE)  
  
#There are missing data, data will need to be converted from long to wide format to get a better sense of complete data  
gg\_miss\_var(hpv\_coverage\_raw)



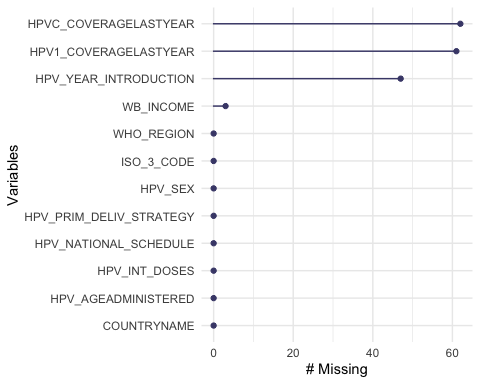
#Gives total number of countries in the dataset  
table(hpv\_coverage\_raw$YEAR)

2010 2011 2012 2013 2014 2015 2016 2017 2018 2019 2020 2021 2022 2023   
 196 196 196 196 196 196 196 196 196 196 196 196 196 195

##### 3.3.0.1.2 HPV vaccination program characteristics

A total of 196 countries are included in the data set. About one third of countries are missing year of introduction, and a few are missing country income. Other characteristics (e.g., delivery strategy) appear complete.

#Importing HPV vaccination program characteristics  
hpv\_program\_raw <- read.csv(here("data","raw-data","who-dashboard-1dosecoverage-1-27-25.csv"), header = TRUE)  
  
#There are missing data, year of introduction is missing for about a third of countries  
gg\_miss\_var(hpv\_program\_raw)



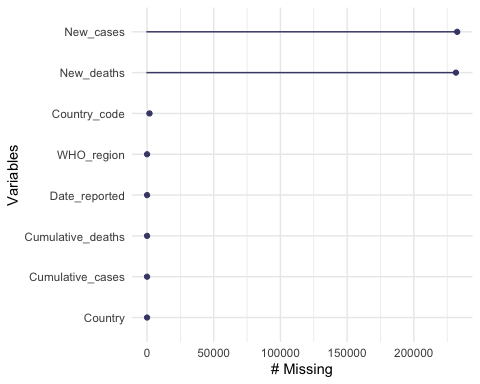
#Gives total number of countries in the dataset  
nrow(hpv\_program\_raw)

[1] 196

#### 3.3.0.2 COVID19 data

This data source includes weekly COVD case count data on 239 countries. The data set includes 440640 rows, but about half have missing data on new case counts. Completeness of data will have to be reviewed once the data are restricted to countries with vaccination coverage estimates in the years of interest.

#Importing HPV vaccination program characteristics  
covid\_raw <- read.csv(here("data","raw-data","WHO-COVID-19-global-daily-data.csv"), header = TRUE)  
  
#There are missing data, year of introduction is missing for about a third of countries  
gg\_miss\_var(covid\_raw)



#Gives total number of countries in the dataset  
nrow(covid\_raw)

[1] 440640

#Getting the number of countires - 239  
countries <- table(covid\_raw$Country\_code)  
dim(countries)

[1] 239

#### 3.3.0.3 United Nations data

There are 238 countries with yearly data from 1950 to 2023. Data appear mostly complete (ISO2 Alpha-code is missing for some rows but this is missing for all World and Region specific rows since this is a code to indicate country). Data will be restricted to the two years of interest during the next part of the project.

u\_n\_raw <- read.xlsx(here("data","raw-data","WPP2024\_GEN\_F01\_DEMOGRAPHIC\_INDICATORS\_COMPACT.xlsx"), sheet = "Estimates")  
header\_row <- 10  
column\_names\_un <- as.character(unlist(u\_n\_raw[header\_row, ]))  
  
# Read the data again, skipping the rows before the header  
u\_n <- read.xlsx(here("data","raw-data","WPP2024\_GEN\_F01\_DEMOGRAPHIC\_INDICATORS\_COMPACT.xlsx"), sheet = "Estimates", startRow = 18, colNames = FALSE)  
colnames(u\_n) <- column\_names\_un  
  
#Getting number of countries  
un\_countries <- table(u\_n$`ISO3 Alpha-code`)  
dim(un\_countries)

[1] 238

#238 countries  
  
#Getting number of years  
table(u\_n$Year)

1950 1951 1952 1953 1954 1955 1956 1957 1958 1959 1960 1961 1962 1963 1964   
 5 297 297 297 297 297 297 297 297 297 297 297 297 297 297 297   
1965 1966 1967 1968 1969 1970 1971 1972 1973 1974 1975 1976 1977 1978 1979 1980   
 297 297 297 297 297 297 297 297 297 297 297 297 297 297 297 297   
1981 1982 1983 1984 1985 1986 1987 1988 1989 1990 1991 1992 1993 1994 1995 1996   
 297 297 297 297 297 297 297 297 297 297 297 297 297 297 297 297   
1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012   
 297 297 297 297 297 297 297 297 297 297 297 297 297 297 297 297   
2013 2014 2015 2016 2017 2018 2019 2020 2021 2022 2023   
 297 297 297 297 297 297 297 297 297 297 297

#Year ranges from 1950 to 2023  
  
#There are missing data, year of introduction is missing for about a third of countries  
gg\_miss\_var(u\_n)



#### 3.3.0.4 World Bank data

There are 218 countries with yearly data from 1987 to 2023. Some data are missing, but this is mostly complete. Data will be restricted to the two years of interest during the next part of the project.

# A tibble: 6 × 39  
 country\_code country `1987` `1988` `1989` `1990` `1991` `1992` `1993` `1994`  
 <chr> <chr> <chr> <chr> <chr> <chr> <chr> <chr> <chr> <chr>   
1 VNM Viet Nam L L L L L L L L   
2 VIR Virgin I… H H H H H H H H   
3 PSE West Ban… .. .. .. .. .. .. .. LM   
4 YEM Yemen, R… LM\* LM\* LM LM L L L L   
5 ZMB Zambia L L L L L L L L   
6 ZWE Zimbabwe LM LM LM LM L L L L   
# ℹ 29 more variables: `1995` <chr>, `1996` <chr>, `1997` <chr>, `1998` <chr>,  
# `1999` <chr>, `2000` <chr>, `2001` <chr>, `2002` <chr>, `2003` <chr>,  
# `2004` <chr>, `2005` <chr>, `2006` <chr>, `2007` <chr>, `2008` <chr>,  
# `2009` <chr>, `2010` <chr>, `2011` <chr>, `2012` <chr>, `2013` <chr>,  
# `2014` <chr>, `2015` <chr>, `2016` <chr>, `2017` <chr>, `2018` <chr>,  
# `2019` <chr>, `2020` <chr>, `2021` <chr>, `2022` <chr>, `2023` <chr>

## 3.4 Statistical analysis

*Explain anything related to your statistical analyses.*

# 4. Results

## 4.1 Exploratory/Descriptive analysis

*Use a combination of text/tables/figures to explore and describe your data. Show the most important descriptive results here. Additional ones should go in the supplement. Even more can be in the R and Quarto files that are part of your project.*

[Table 1](#tbl-summarytable) shows a summary of the data.

Note the loading of the data providing a **relative** path using the ../../ notation. (Two dots means a folder up). You never want to specify an **absolute** path like C:\ahandel\myproject\results\ because if you share this with someone, it won’t work for them since they don’t have that path. You can also use the here R package to create paths. See examples of that below. I generally recommend the here package.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Table 1: Data summary table.   | skim\_type | skim\_variable | n\_missing | complete\_rate | factor.ordered | factor.n\_unique | factor.top\_counts | numeric.mean | numeric.sd | numeric.p0 | numeric.p25 | numeric.p50 | numeric.p75 | numeric.p100 | numeric.hist | | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | | factor | Gender | 0 | 1 | FALSE | 3 | M: 4, F: 3, O: 2 | NA | NA | NA | NA | NA | NA | NA | NA | | numeric | Height | 0 | 1 | NA | NA | NA | 165.66667 | 15.97655 | 133 | 156 | 166 | 178 | 183 | ▂▁▃▃▇ | | numeric | Weight | 0 | 1 | NA | NA | NA | 70.11111 | 21.24526 | 45 | 55 | 70 | 80 | 110 | ▇▂▃▂▂ | |

## 4.2 Basic statistical analysis

*To get some further insight into your data, if reasonable you could compute simple statistics (e.g. simple models with 1 predictor) to look for associations between your outcome(s) and each individual predictor variable. Though note that unless you pre-specified the outcome and main exposure, any “p<0.05 means statistical significance” interpretation is not valid.*

[Figure 2](#fig-result) shows a scatterplot figure produced by one of the R scripts.

|  |
| --- |
| Figure 2: Height and weight stratified by gender. |

## 4.3 Full analysis

*Use one or several suitable statistical/machine learning methods to analyze your data and to produce meaningful figures, tables, etc. This might again be code that is best placed in one or several separate R scripts that need to be well documented. You want the code to produce figures and data ready for display as tables, and save those. Then you load them here.*

Example [Table 2](#tbl-resulttable2) shows a summary of a linear model fit.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Table 2: Linear model fit table.   | term | estimate | std.error | statistic | p.value | | --- | --- | --- | --- | --- | | (Intercept) | 149.2726967 | 23.3823360 | 6.3839942 | 0.0013962 | | Weight | 0.2623972 | 0.3512436 | 0.7470519 | 0.4886517 | | GenderM | -2.1244913 | 15.5488953 | -0.1366329 | 0.8966520 | | GenderO | -4.7644739 | 19.0114155 | -0.2506112 | 0.8120871 | |

# 5. Discussion

## 5.1 Summary and Interpretation

*Summarize what you did, what you found and what it means.*

## 5.2 Strengths and Limitations

*Discuss what you perceive as strengths and limitations of your analysis.*

## 5.3 Conclusions

*What are the main take-home messages?*

*Include citations in your Rmd file using bibtex, the list of references will automatically be placed at the end*

This paper (1) discusses types of analyses.

These papers (2,3) are good examples of papers published using a fully reproducible setup similar to the one shown in this template.

Note that this cited reference will show up at the end of the document, the reference formatting is determined by the CSL file specified in the YAML header. Many more style files for almost any journal [are available](https://www.zotero.org/styles). You also specify the location of your bibtex reference file in the YAML. You can call your reference file anything you like.

# 6. References

To cite other work (important everywhere, but likely happens first in introduction), make sure your references are in the bibtex file specified in the YAML header above and have the right bibtex key. Then you can include like this:

Examples of reproducible research projects can for instance be found in (2,3).

1. Leek JT, Peng RD. [Statistics. What is the question?](https://doi.org/10.1126/science.aaa6146) *Science (New York, N.Y.)*. 2015;347(6228):1314–1315.

2. McKay B, Ebell M, Billings WZ, et al. [Associations Between Relative Viral Load at Diagnosis and Influenza A Symptoms and Recovery.](https://doi.org/10.1093/ofid/ofaa494) *Open forum infectious diseases*. 2020;7(11):ofaa494.

3. McKay B, Ebell M, Dale AP, et al. [Virulence-mediated infectiousness and activity trade-offs and their impact on transmission potential of influenza patients.](https://doi.org/10.1098/rspb.2020.0496) *Proceedings. Biological sciences*. 2020;287(1927):20200496.