Closing the Gap: Immunization Compliance and Predictive Analytics for HEDIS Measures

Diane Haiden

July 2025

Table of Contents

1.0 Abstract

This study investigates pediatric and adolescent immunization disparities across Oregon, using ZIP-level data from the Oregon Immunization Program's Tableau dashboard. Through statistical and machine learning models, the analysis identifies significant variation in vaccine uptake, particularly for COVID-19 and influenza. Notably, HPV vaccine completion emerged as a strong predictor of adherence to adolescent immunization series.

By integrating public health data with predictive analytics, the study provides a framework for targeted outreach and enhanced compliance with HEDIS measures, including Childhood Immunization Status (CIS) and Immunizations for Adolescents (IMA). These insights can support health systems and policymakers in addressing care gaps and promoting vaccine equity across Oregon.

2.0 Background & Objective

Immunizations are a cornerstone of preventive pediatric and adolescent care, yet compliance across Oregon remains uneven. The Healthcare Effectiveness Data and Information Set (HEDIS) provides a national benchmark for assessing vaccination quality, with CIS and IMA as key metrics.

Despite established guidelines, disparities in vaccine uptake persist. This research aims to:

- Uncover geographic and demographic differences in vaccine compliance.
- Assess adherence to HEDIS benchmarks at the ZIP code level.
- Build predictive models to flag areas at elevated risk of under-immunization.
- Explore the role of specific vaccine patterns (HPV, COVID-19, influenza) in shaping overall series completion.
- Highlight how risk stratification can enhance resource allocation and promote public health equity.

This project addresses a pressing need for localized, data-driven strategies that enhance immunization outcomes among Oregon's youth population.

3.0 Data Overview

Drawing from the Oregon Immunization Program's publicly accessible Tableau dashboard, this study examines vaccine coverage for children aged 0–13 across all 36 Oregon counties. The dataset provides ZIP-level immunization rates, enabling a granular analysis of regional adherence to HEDIS CIS and IMA benchmarks.

Analytical methods include descriptive statistics and predictive modeling, such as:

- ANOVA and Kruskal-Wallis tests to evaluate variance across counties.
- Correlation and regression analyses to assess relationships between vaccine types and series completion.
- Machine learning to develop risk profiles and prioritize high-need ZIP codes.

The dataset’s breadth and granularity support a nuanced exploration of immunization trends and provide the foundation for targeted public health interventions.

3.1 Data Preparation

All data preprocessing was conducted in Excel after downloading the dataset from the Oregon Immunization Program’s publicly accessible Tableau dashboard. The study utilized ZIP-level data, which may be subject to limitations in completeness and granularity, particularly in low-population areas where data suppression is applied to safeguard patient confidentiality.

To enable structured analysis, qualitative entries in the `POPULATION_SIZE` field were recoded into a new variable, `pop_bin`, as follows:

- `0` for "<10 (not shown)"
- `1` for "10 to 49 (#)"
- `2` for blank entries, indicating ZIP codes with populations of 50 or more

This binning preserved the interpretive integrity of the original field while allowing for stratified comparisons by population tier in downstream analyses.

Vaccination rates were standardized by converting percentage values into decimal form (e.g., 85% → 0.85). ZIP codes with fewer than 10 eligible individuals often had suppressed vaccination data; in these cases, fields were left blank to differentiate true zeroes from privacy-driven omissions.

Records with suppressed vaccination values were excluded from statistical procedures involving summary metrics, such as mean comparisons, regression, and correlation analyses, to reduce bias stemming from small denominators. However, these records were retained in spatial visualizations and healthcare access evaluations to ensure geographic inclusivity and provide a comprehensive view of immunization coverage across the state.

3.2 Methodological Limitations

Several limitations should be considered when interpreting the findings of this study. First, the immunization data were extracted from the Oregon Health Authority’s ALERT Immunization Information System via its Tableau dashboard, which only provides aggregated ZIP-level metrics. The absence of individual-level data restricts the ability to incorporate key demographic, socioeconomic, or behavioral variables that are often critical for understanding immunization disparities. As a result, the analysis lacks the granularity needed to assess specific population subgroups or to examine longitudinal vaccination trajectories.

Second, the predictive modeling component is constrained by the absence of patient-level and temporal information. These limitations reduce the precision and potential applicability of risk stratification outputs, especially in dynamic or time-sensitive public health contexts. While SAS served as the primary platform for statistical analysis and modeling, the use of Excel during the data preparation phase introduced a multi-platform workflow that, while flexible, may present challenges in interoperability and reproducibility.

Third, ZIP codes with adolescent populations under 10 were excluded from descriptive statistical analyses due to data suppression protocols that render such entries blank to maintain confidentiality. Although this decision safeguards privacy and improves data reliability, it inadvertently limits the inclusion of rural and low-density areas. As a result, the findings may underrepresent these communities, reducing the geographic and demographic generalizability of the results across the full population of Oregon.

Additionally, the dataset's restriction to Oregon further limits the external validity of the study. The use of proxy variables, such as regional averages in place of individual health indicators, introduces interpretive uncertainty and constrains causal inference.

Despite these limitations, the study offers valuable population-level insights into immunization patterns and disparities. It establishes a scalable and transparent methodological framework that can inform future investigations, particularly those with access to more granular, patient-level data.

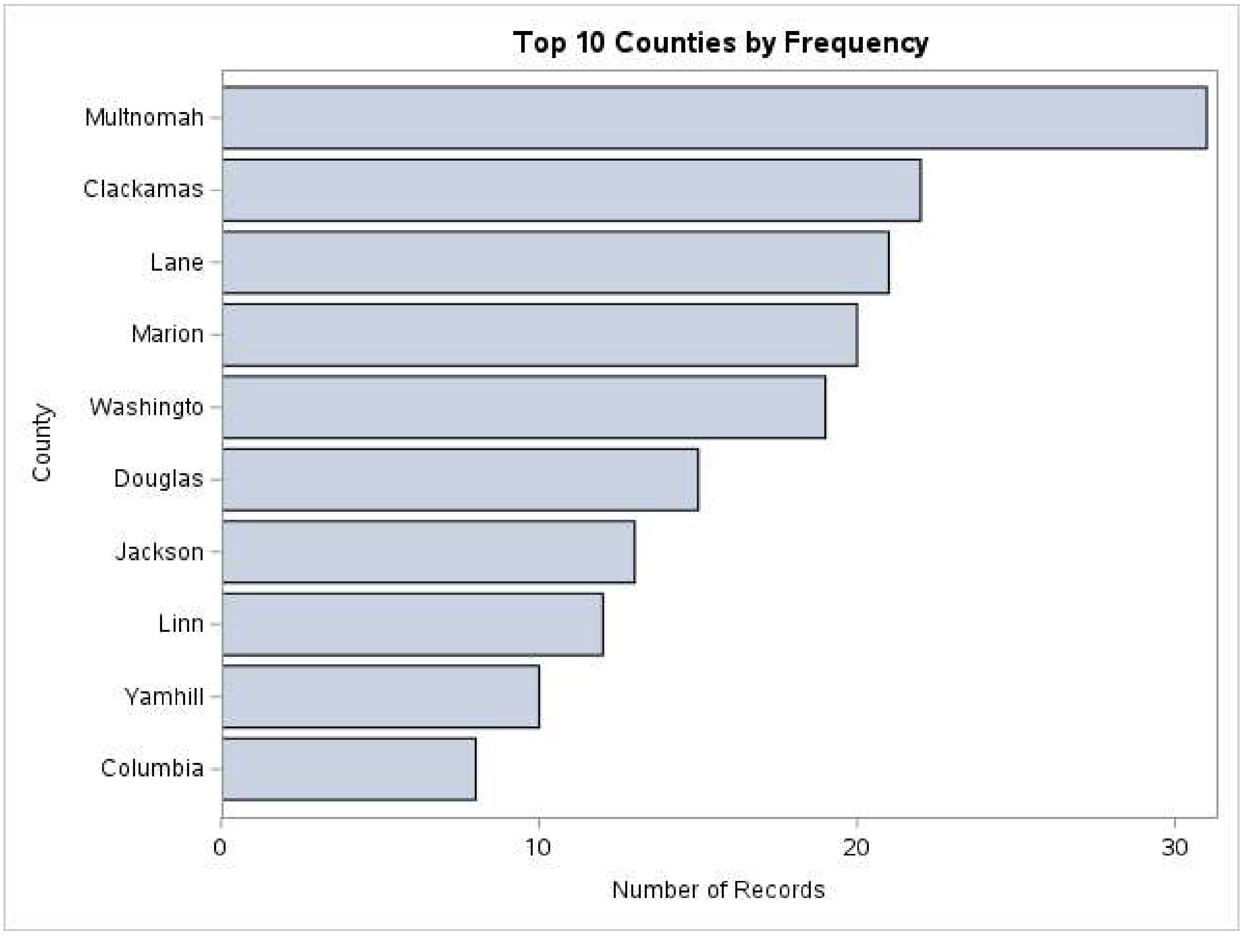
4.0 Exploratory Data Analysis (EDA) in SAS

The dataset comprises 424 ZIP code–level (ZIP) observations spanning all 36 counties in Oregon. Each record includes 11 variables capturing geographic identifiers, categorized population sizes, and pediatric and adolescent immunization coverage rates for Tdap, MenACWY, COVID-19, influenza, human papillomavirus (HPV) initiation and completion, and completion of the Teen Immunization Series by age 13. Of the 424 records, 268 ZIP codes with a population size greater than 10 were included in the descriptive analysis.

The frequency distribution in Table A1 (Appendix A) reveals that the number of observations varies widely across counties, reflecting differences in population density, ZIP code count, and potentially data availability. As shown in Figure 1, Multnomah County contributes the largest share of data (31 observations), followed by Clackamas (22), Lane (21), Marion (20), and Washington (19). Additional counties with relatively high representation include Douglas (15), Jackson (13), Linn (12), Yamhill (10), and Columbia (8).

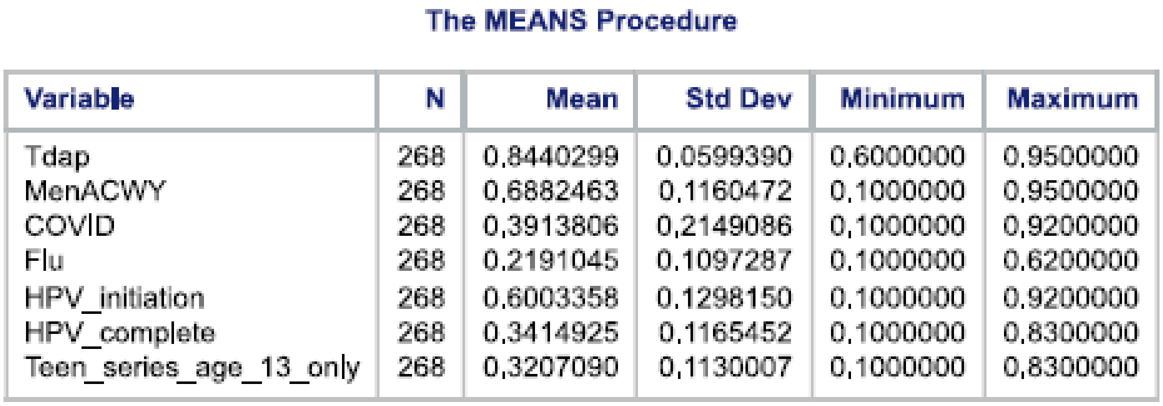
At the other end of the spectrum, rural counties such as Sherman, Baker, Gilliam, and Lake each have only one observation, underscoring minimal representation from sparsely populated areas. The cumulative percentage column indicates that over half of the dataset is drawn from just the top 10 counties, a skew toward more populous or urbanized regions. This uneven distribution may warrant the use of sampling weights or stratified analyses to account for geographic imbalance. It also highlights the potential to examine urban–rural disparities in adolescent immunization coverage across Oregon with careful attention to representativeness.

Figure 1.
Frequency distribution of the top 10 counties



Descriptive statistics were generated using the MEANS procedure in SAS to summarize vaccine coverage distributions as shown in Figure 2. Tdap exhibited the highest mean uptake (M = 0.84, SD = 0.06), followed by HPV initiation (M = 0.60, SD = 0.13) and MenACWY (M = 0.69, SD = 0.12). Completion of the HPV series averaged 0.34 (SD = 0.12), indicating a drop from initiation to full vaccination. COVID-19 vaccination rates showed considerable variability (M = 0.39, SD = 0.21), while influenza coverage remained the lowest across all measures (M = 0.22, SD = 0.11). Completion of the Teen Immunization Series by age 13 remained low as well (M = 0.32, SD = 0.11).

Figure 2.
Zip-Level Vaccine Coverage Means and Standard Deviations



The MEANS Procedure

Variable	N	Mean	Std Dev	Minimum	Maximum
Tdap	268	0.8440299	0.0599390	0.6000000	0.9500000
MenACWY	268	0.6882463	0.1160472	0.1000000	0.9500000
COVID	268	0.3913806	0.2149086	0.1000000	0.9200000
Flu	268	0.2191045	0.1097287	0.1000000	0.6200000
HPV_initiation	268	0.6003358	0.1298150	0.1000000	0.9200000
HPV_complete	268	0.3414925	0.1165452	0.1000000	0.8300000
Teen_series_age_13_only	268	0.3207090	0.1130007	0.1000000	0.8300000

To explore relationships among vaccine variables, a Pearson correlation matrix was produced using the CORR procedure (see Table A2 in Appendix A). Strong and statistically significant positive correlations were observed between HPV completion and both HPV initiation (r = .81, p < .001) and Teen Series compliance (r = .97, p < .001), reinforcing the role of HPV initiation as a key contributor to broader adolescent immunization outcomes. COVID-19 vaccination rates were highly correlated with influenza coverage (r = .83, p < .001) and moderately associated with HPV completion (r = .55, p < .001) and Teen Series compliance (r = .57, p < .001), pointing to clustering in general preventive health behaviors at the ZIP-code level. MenACWY exhibited strong positive associations with HPV initiation (r = .76, p < .001) and Teen Series completion (r = .68, p < .001). In contrast, Tdap remained weakly correlated with other measures (e.g., COVID-19: r = .12, p = .057), suggesting its uptake may be shaped by external factors such as school-entry requirements or timing unrelated to broader vaccine behavior patterns.

Table 1.
Univariate analysis

Comprehensive Vaccine Coverage Summary (N = 268 ZIP Codes)							
Metric	**Tdap**	**MenACWY**	**COVID**	**Flu**	**HPV Initiation**	**HPV Completion**	**Teen Series**
Mean	0.84	0.69	0.39	0.22	0.6	0.34	0.32
Median	0.86	0.72	0.35	0.2	0.62	0.35	0.32
Std Deviation	0.06	0.12	0.21	0.11	0.13	0.12	0.11
IQR (Q3–Q1)	0.06	0.14	0.31	0.16	0.17	0.16	0.14
Range	0.35	0.85	0.82	0.52	0.82	0.73	0.73
Skewness	–1.05	–1.49	0.6	0.96	–0.53	0.34	0.36
Kurtosis	1.62	3.83	–0.63	0.56	0.35	0.82	0.96
Normality Tests	×	×	×	×	×	✓	✓
Min / Max	0.60–0.95	0.10–0.95	0.10–0.92	0.10–0.62	0.10–0.92	0.10–0.83	0.10–0.83
Mode	0.86	0.77	0.1	0.1	0.64	0.28	0.35

The univariate analysis shown in Table 1 revealed distinct patterns in adolescent immunization coverage across Oregon ZIP codes. Tdap vaccination stood out with the highest and most consistent uptake (M = 0.84, SD = 0.06), characterized by low variability and a compact distribution, likely a reflection of universal school-entry mandates. MenACWY coverage, while moderately high (M = 0.69, SD = 0.12), showed more pronounced skew and range, indicating uneven access or adherence.

HPV initiation had a respectable average (M = 0.60) but dropped substantially at the completion stage (M = 0.34), signaling major follow-through gaps. Similarly, Teen Series compliance by age 13 remained modest (M = 0.32), mirroring the trend in HPV series adherence and emphasizing the cumulative nature of vaccine engagement.

Preventive behaviors such as COVID-19 (M = 0.39) and influenza vaccination (M = 0.22) lagged, with wide distributions, strong right skew, and clustering near the minimum, suggesting access barriers, seasonal hesitancy, or limited outreach in some ZIP codes. Notably, HPV completion and Teen Series compliance were the only variables approximating a normal distribution, making them strong candidates for downstream parametric modeling.

Together, these findings underscore a layered immunization landscape: while some vaccines exhibit high, regulated uptake, others show sharp disparities shaped by behavior, access, or health policy. The variability in distribution and shape across vaccine types reveals valuable opportunities to tailor interventions, not only to increase uptake, but to improve follow-through and equity in coverage.

Collectively, these exploratory results provided the foundation for subsequent modeling decisions and geographic prioritization within the study’s analytic framework.

5.0 Research Questions and Hypothesis

Research Question 1: What are the geographic and demographic disparities in immunization compliance across ZIP codes in the Portland Metro area?

- H_0: There are no statistically significant differences in immunization compliance across ZIP codes or counties in the Portland Metro area.
- H_1: There are statistically significant differences in immunization compliance across ZIP codes or counties in the Portland Metro area.

Research Question 2: How do immunization rates for individual vaccines (e.g., HPV, Tdap, MenACWY, influenza, and COVID-19) correlate with completion of the Teen Immunization Series by age 13?

- H_0: There is no statistically significant correlation between individual vaccine uptake (e.g., HPV, Tdap, COVID-19) and completion of the Teen Immunization Series by age 13.
- H_1: There is a statistically significant correlation between individual vaccine uptake (e.g., HPV, Tdap, COVID-19) and completion of the Teen Immunization Series by age 13.

Research Question 3: Can predictive models built using ZIP-level immunization data accurately identify areas at elevated risk of under-immunization for adolescent populations?

- H_0: Predictive models based on ZIP-level data perform no better than random chance in identifying areas at elevated risk of under-immunization.
- H_1: Predictive models based on ZIP-level data accurately identify areas at elevated risk of under-immunization at a statistically significant level (e.g., classification accuracy significantly exceeds chance levels).

Research Question 4: What are the most influential factors associated with noncompliance in adolescent immunization schedules according to HEDIS IMA benchmarks?

- H_0: No specific geographic, demographic, or coverage-related variables are significantly associated with immunization noncompliance.
- H_1: Specific variables (e.g., population size bin, county, coverage of other vaccines) are significantly associated with immunization noncompliance.

Research Question 5: How can risk stratification tools support more efficient and equitable allocation of public health resources to close immunization gaps?

- H_0: Risk stratification tools do not contribute to more efficient or equitable resource allocation in addressing immunization gaps.
- H_1: Risk stratification tools contribute significantly to more efficient and equitable resource allocation in addressing immunization gaps.

6.0 Literature Review

6.1.1 Geography, Demography, and Vaccine Gaps in Portland

Understanding immunization compliance across geographic and demographic lines is essential for achieving equitable health outcomes in the Portland Metro area. The 2014 Report Card on Racial and Ethnic Disparities by the Multnomah County Health Department provides foundational evidence of persistent racial, ethnic, and neighborhood-based health disparities. While the report does not directly analyze immunization rates, it offers critical context on the underlying social determinants of health that contribute to inequities in preventive care, including immunizations.

The report identifies disparities across 33 health indicators among non-Latino White, non-Latino Black/African American, Latino, non-Latino Asian/Pacific Islander, and non-Latino American Indian/Alaska Native populations. These disparities are evident not only in health outcomes but also in access to care and socioeconomic factors. For example, Black/African American and American Indian/Alaska Native residents experienced the highest levels of disadvantage, with disparities requiring urgent intervention across multiple domains, including poverty, educational attainment, insurance coverage, and clinical care access, all of which are factors that have been shown to influence immunization adherence.

Geographic disparities are also highlighted in the report through environmental indicators such as air quality and the retail food environment, which correlate with community-level resource availability. These indicators were found to be significantly worse in census tracts with higher concentrations of communities of color compared to predominantly non-Latino White neighborhoods. Although not directly related to immunizations, these findings suggest that place-based disadvantages, including exposure to structural inequities and limited healthcare infrastructure, may also impact routine vaccination access and compliance.

Particularly relevant to immunization compliance is the report’s emphasis on disparities in access to preventive care. Communities of color, especially Black, Latino, and American Indian/Alaska Native populations, were significantly more likely to lack health insurance and to receive inadequate prenatal and pediatric care. These factors contribute to delayed or missed vaccinations, particularly in early childhood and adolescence. Moreover, the report notes geographic isolation and limited culturally responsive care as additional barriers that can undermine trust in healthcare systems and adherence to vaccination schedules.

The intersection of geography and demography is central to the disparities described in the report, providing a strong rationale for analyzing immunization patterns by ZIP code. Given that ZIP codes often correlate with neighborhood-level socioeconomic and racial segregation, examining immunization compliance through this spatial lens may reveal concentrations of under-immunization that correspond with systemic inequities in access, trust, and health literacy.

In summary, the findings from the Multnomah County report underscore the need for localized, equity-focused approaches to public health. A geographic and demographic analysis of immunization compliance will not only help identify vulnerable populations but also guide resource allocation, culturally competent outreach, and policy reform aimed at closing immunization gaps across the Portland Metro area (Coalition of Communities of Color & Urban League of Portland, 2010).

6.1.2 Frameworks for collecting and analyzing demographic data

To effectively examine immunization disparities across ZIP codes in the Portland Metro area, a structured framework for collecting and analyzing demographic data is essential. The City of Portland's RELDTA Demographic Data Standards Guide offers comprehensive guidance to ensure that demographic data collection is both standardized and equity-centered. These standards recommend collecting race, ethnicity, language, disability, and tribal affiliation data using either a minimum or expanded category format. The choice depends on factors such as analytic needs, confidentiality risks, and the size of the subpopulations represented in the data. For ZIP-level analyses, researchers are advised to use expanded categories where sample sizes are sufficient, and to aggregate to minimum categories when suppression is needed to protect privacy.

A key component of the RELDTA framework is designing data collection tools that promote trust and inclusivity. This includes placing demographic questions at the end of forms, offering a clear explanation of the purpose and confidentiality of data collection, and allowing multiple selections or write-in options for identity questions. These practices reduce the risk of underreporting and help ensure that marginalized populations are accurately represented. In line with equity principles, the guide emphasizes the importance of self-identification and the use of respectful, culturally responsive language.

Data privacy and responsible reporting are particularly critical when working at ZIP-code granularity. To avoid inadvertently identifying individuals, especially in small or underserved communities, RELDTA recommends applying suppression thresholds and using roll-up categories when necessary. Analysts must be cautious when combining variables such as race, disability, and geography, as this can produce uniquely identifiable combinations in low-population ZIPs. This attention to data governance ensures that demographic analysis does not unintentionally cause harm or breach confidentiality.

In terms of data management, researchers are encouraged to structure demographic variables using consistent formats and to distinguish between "alone" and "in combination" categories when reporting race. This ensures clarity and comparability across datasets. Once immunization records are linked to demographic and ZIP-code data, researchers can generate aggregate indicators such as immunization completion rates, timeliness of vaccine uptake, and missed vaccination proportions by subgroup. These metrics form the basis for equity-focused stratified analysis.

To detect disparities and geographic patterns, statistical tools such as logistic regression can be applied, and geospatial mapping can visualize ZIP-level clusters of under-immunization. Overlaying these maps with demographic concentrations allows researchers to identify structural inequities rooted in historical disinvestment or social exclusion. Finally, in line with RELDTA's principles, researchers are encouraged to report transparently by documenting methods, thresholds, and data limitations, and to engage community partners to validate categories, interpret findings, and guide public health action.

Incorporating RELDTA standards into this research supports ethically sound, methodologically rigorous, and equity-informed analysis. By disaggregating data and stratifying results by ZIP code and demographic identity, this approach increases the visibility of populations most at risk of falling behind on immunizations, thereby informing targeted and culturally competent public health interventions in the Portland Metro area (City of Portland Office of Equity and Human Rights, n.d.).

6.2.1 National Insights into Teen Immunization Compliance by Age 13

The 2023–2024 National Immunization Survey–Teen (NIS-Teen) provides critical insight into adolescent immunization patterns and their relationship with completion of the recommended vaccine series, including Tdap, MenACWY, HPV, influenza, and COVID-19 vaccines, by age 13. CDC's analysis of over 16,000 adolescents aged 13–17 shows consistently high coverage for at least one dose of Tdap (89.0%) and MenACWY (88.4%) in 2023. By comparison, HPV initiation (≥ 1 dose) lagged at 76.8%, and only 61.4% of teens were up to date (UTD) by age 13. These disparities highlight a notable imbalance within the Teen Immunization Series, raising questions about the consistency of series completion across vaccine types.

Birth-cohort analysis between 2008 and 2010 reveals significant effects of the COVID-19 pandemic on immunization timeliness. Adolescents due for routine vaccination during the pandemic (birth cohorts 2008–2010) experienced measurable declines by age 13: Tdap (–2.3 to –3.2 percentage points), MenACWY (–2.6 to –3.0), and HPV initiation (–3.2) compared to pre-pandemic cohorts born in 2007. Notably, HPV UTD status among VFC-eligible teens in the 2010 cohort dropped by 10.3 percentage points relative to the 2007 cohort. These declines underscore a weakened alignment between individual vaccine uptake and full series completion when routine care is disrupted.

Trends from earlier years (2021–2022) reinforce these patterns. The 2022 NIS-Teen indicated stable Tdap and MenACWY coverage (\~90%), but HPV initiation remained lower (\~76%) and HPV UTD was near 62%. Birth-cohort analysis revealed sustained deficits: adolescents born in 2008 had 3.2% lower Tdap and 3.0% lower MenACWY coverage by age 13 compared with the 2007 cohort. These lagging rates suggest that while initial vaccine doses remain robust, full series completion (captured by the Teen Immunization Series metric), especially for HPV, may falter due to delayed follow-up doses.

Analyses of missed opportunities lend further context. Among teens born 1999–2009, HPV series completion before age 13 rose from 10.3% to 42.2%, highlighting improvement yet revealing persistent gaps in timely vaccination alignment. Regression and Kaplan–Meier analyses demonstrated that preventive visits at ages 11–12 and stable insurance coverage were significantly associated with HPV initiation by 13, reinforcing the importance of healthcare engagement for series completion.

Although influenza and COVID-19 vaccines are recommended annually/seasonally, they are not included in the Teen Immunization Series metric. Nevertheless, CDC emphasizes that annual influenza immunization and up-to-date COVID-19 coverage are critical benchmarks. Pandemic-era disruptions, described in multiple CDC MMWR reports, adversely affected all adolescent vaccines, suggesting spillover effects across recommended doses.

Overall, the literature demonstrates that while adolescents often receive at least one dose of recommended vaccines, full adherence to the Teen Series, particularly timely HPV completion, is uneven. This variability correlates strongly with systemic factors such as pandemic-related healthcare disruptions, VFC eligibility, insurance status, and preventive care engagement. These findings underscore the need for public health strategies aimed at reinforcing continuity of care and series completion by age 13 (Pingali et. al, 2024).

6.2.2 Correlates of Individual Vaccine Uptake and Teen Immunization Series Completion

The NCQA's Immunizations for Adolescents (IMA-E) measure offers a standardized framework for evaluating vaccine coverage among 13-year-olds, capturing not only individual vaccines, Tdap, MenACWY, and HPV series completion, but also the combination metric representing the Teen Immunization Series compliance. According to NIS-Teen data, as of 2022, coverage among 13–17-year-olds was high for Tdap (89.9%) and MenACWY (88.6%), yet markedly lower for complete HPV vaccination (62.6%). This disparity highlights a critical divergence: while most adolescents receive their Tdap and MenACWY doses, a substantial number fail to complete the multi-dose HPV regimen by their 13th birthday.

Further analysis by NCQA and CDC reveals that the Teen Immunization Series metric is sensitive to the lower HPV completion rate and thus reflects broader gaps in adolescent immunization compliance. The IMA-E measure emphasizes that individual vaccine coverage (e.g., ≥1 Tdap or MenACWY dose) does not necessarily guarantee full series completion. Instead, it is the incorporation of multi-dose vaccines, particularly HPV that often reduces the combined-series coverage.

Scholarly research supports this interpretation. Pingali et al. (2023) and related MMWR analyses demonstrate that despite stability in Tdap and MenACWY uptake, HPV initiation remains persistently lower. Studies using birth cohort comparisons indicate HPV series completion by age 13 can be delayed or incomplete even when initial doses are administered, pointing to care access issues, vaccine hesitancy, and insufficient follow-up systems.

For instance, teens who engage in routine preventive care, attend well-child visits at ages 11 and 12, and maintain continuous insurance coverage have significantly higher odds of completing the HPV series. Conversely, those missing early adolescent visits or lacking stable coverage often fall behind, even if they start the series. This phenomenon contributes to the observed drop-off between single-dose administration and full series compliance.

Though not formally captured in the IMA-E measure, annual vaccines such as influenza and COVID-19 are relevant comparators. Pandemic-era studies have shown disruptions in adolescent vaccine delivery that impacted HPV series more than single-dose vaccines, due to their multi-dose structure and tight schedule requirements. These disruptions translated to reduced series completion despite stable Tdap and MenACWY uptake.

In summary, the literature consistently supports a nuanced pattern: strong initial uptake for single-dose adolescent vaccines (Tdap, MenACWY) coexists with weaker, delayed, or incomplete uptake for multi-dose series (HPV), which diminishes Teen Immunization Series rates. This underscores the need for focused public health strategies, such as reminder/recall systems, provider prompts, and insurance continuity, to bridge the gap between vaccine initiation and timely, full-series completion by age 13 (National Committee for Quality Assurance, n.d.).

6.3.1 Predictive Modeling of ZIP-Level Immunization Gaps

Recent research indicates that ZIP-level predictive modeling offers a promising avenue for identifying communities at high risk of under-immunization among adolescents. In their 2021 study, Melotte and Kejriwal demonstrate a robust methodology for predicting vaccine hesitancy at the ZIP-code level using publicly available Twitter data, geolocation information, and socioeconomic features. While their focus is on adult COVID-19 vaccine hesitancy, their approach illuminates how social media signals and demographic context can be integrated to anticipate spatial disparities in vaccine uptake.

Their framework combines natural language processing (NLP) of geolocated tweets with ZIP-level auxiliary variables, such as home value indices and counts of healthcare, educational, and scientific establishments, to train machine learning models (e.g., support vector regression) predicting the proportion of residents hesitant about vaccination. The resulting models outperform naïve constant-value baselines (e.g., assuming average hesitancy everywhere), achieving up to a 25% reduction in RMSE compared to a "no hesitancy" baseline. Text features alone proved more predictive than socioeconomic variables alone, underscoring the added value of real-time linguistic signals.

Importantly, the study validates model outputs against independent survey data aggregated at the ZIP-code level, providing external credibility to their predictions. Their use of both tweet-level pseudo-labels and aggregated ZIP-level ground truth illustrates a methodological bridge between granular behavioral data and population-level outcomes. This validation strategy and the significant predictive gain offer strong precedent for applying similar methods to adolescent immunization monitoring.

Translating this to adolescent immunization contexts, models could ingest ZIP-level coverage rates for vaccines like HPV, Tdap, MenACWY, influenza, and COVID-19, combined with social media discourse around vaccines, clinic availability, and neighborhood demographic features. If text signals are as informative for adolescent vaccine uptake patterns, ZIP-level predictive analytics could proactively flag under-immunized areas, complementing traditional surveillance tools.

However, several limitations warrant attention. Melotte & Kejriwal acknowledge urban bias in Twitter use, cautioning against applying their method in less connected or rural areas. Similarly, social media analysis may underrepresent certain demographic groups, introducing bias. To adapt this framework to adolescents, ethically sourced, privacy-respecting under-immunization data would be needed, likely from school or public health records, and models should be calibrated for youth-specific determinants, such as school-based mandates and parental consent dynamics.

In summary, the arXiv framework demonstrates that ZIP-level predictive modeling, grounded in publicly sourced linguistic data and demographic context, can effectively identify geographic pockets of vaccine hesitancy. This approach provides both a methodological template and empirical justification for building models that predict adolescent immunization gaps, potentially enabling early interventions in under-coverage zones (Melotte & Kejriwal, 2021).

6.3.2 Predictive Modeling Using ZIP-Level IIS Data

Immunization Information Systems (IIS) are foundational for building ZIP-level predictive models of under-immunization. The 2020 CDC framework emphasizes how consolidated, granular IIS data can be leveraged not only for patient-level reminders and provider decision support, but also for population-level surveillance and geographic targeting of immunization gaps. High-quality IIS data enable the generation of accurate, ZIP-aggregated immunization metrics, forming the basis for predictive analytics.

A seminal white paper on using IIS data to estimate national vaccination coverage demonstrates the robustness of IIS-based estimates when properly calibrated against established survey benchmarks like the National Immunization Survey (NIS). The authors underline key methodological considerations, such as accounting for population saturation, sample completeness, and appropriate adjustments for under-reporting, thereby positioning IIS as a reliable data source for small-area modeling.

Complementing this, a guide from AIRA (American Immunization Registry Association) highlights IIS use for standardized coverage estimation at national, state, and local levels. Together, these resources validate the practice of using ZIP-level IIS datasets to model vaccination rates, offering best practices for data cleaning, validation, and comparability.

Methodologies for small-area estimation, such as those applied in county-level vaccination modeling, provide a direct analogue to ZIP-level approaches. By leveraging Bayesian or linear mixed-effects models (e.g., Fay–Herriot, James–Stein estimators), these methods combine survey-derived direct estimates with auxiliary variables to produce precise, smoothed county-level vaccination rates. Standard errors and uncertainty metrics derived in these models support robust predictions and performance evaluation, critical when identifying high-risk ZIP codes.

From a public health modeling perspective, predictive models that use ZIP-level IIS data must address common data challenges: under-counting, population misclassification, and reporting lag. The CDC's guidance emphasizes the need for provider-based reminder systems and population outreach as essential applications of IIS-derived analytics.

Taken together, this literature supports a framework where ZIP-level immunization data serve as the backbone for predictive modeling of adolescent under-immunization risk. By coupling rich IIS-derived immunization rates with demographic and socioeconomic ZIP characteristics and applying small-area estimation techniques, models can accurately identify underperforming areas. The key prerequisites are high-quality, complete IIS data, validated against independent population coverage measures, and modeled using robust statistical approaches to manage uncertainty and spatial heterogeneity (Lin et. al, 2019).

6.4.1 Predictors of Adolescent Immunization Noncompliance

A review of the “Adolescent Immunizations (IMA)” webinar by Aetna Better Health of West Virginia highlights several influential factors associated with adolescent immunization noncompliance, providing strong support for the alternative hypothesis (H_1) that specific geographic, demographic, and coverage-related variables are significantly associated with noncompliance. One of the most prominent barriers to compliance is under-completion of the HPV vaccine series, which consistently accounts for the greatest shortfall in achieving HEDIS IMA Combo 2 benchmarks. This gap is largely attributed to parental misconceptions, including the belief that the HPV vaccine promotes sexual activity, as well as general vaccine hesitancy fueled by misinformation and a lack of school or state mandates. These sociocultural factors heavily influence parental decisions and reflect a deeper trust issue in healthcare guidance. Trust in a provider's recommendation remains the most significant factor in vaccine acceptance.

Beyond individual and attitudinal barriers, system-level factors also contribute to noncompliance. Inaccurate or incomplete immunization records, often caused by patients receiving vaccines outside of their primary care network or during school or pharmacy visits, result in missed documentation within the West Virginia Statewide Immunization Information System (WVSIIS). These record-keeping challenges are compounded by incorrect medical coding and billing practices, leading to administrative underreporting even when vaccines are administered. Furthermore, logistical factors related to healthcare access play a role. Adolescents typically have fewer routine healthcare visits, and many do not attend the critical 13-year-old well visit where IMA-compliant vaccines are usually administered. Limited after-hours access, inflexible scheduling, and missed opportunities during acute-care visits exacerbate the issue.

The webinar also outlines programmatic interventions that can improve compliance rates. These include using EMR prompts starting at age nine, sending reminder messages through text or email, offering flexible scheduling (such as evening or weekend vaccine clinics), and providing modest financial incentives to families who complete the full adolescent immunization series. These strategies emphasize the importance of operational and policy-level variables in vaccine uptake. While the presentation does not directly address geographic differences, the challenges related to immunization registry gaps, residential mobility, and varying provider engagement across regions suggest that county-level or ZIP code-level disparities are likely.

Taken together, these findings provide robust evidence against the null hypothesis (H_0) and support the argument that specific behavioral, operational, and system-level variables significantly influence adolescent immunization compliance. For researchers, this suggests that multi-level modeling incorporating demographic, geographic, and vaccine-specific factors, such as coverage rates for individual vaccines, access to care, and registry accuracy, would yield the most accurate identification of at-risk populations. These insights also align well with the research goal to explore disparities and predictors of noncompliance using HEDIS IMA benchmarks at the ZIP code level (Aetna Better Health of West Virginia, n.d.)

6.4.2 System-Level Drivers of Adolescent Vaccine Noncompliance

The Texas Children's Health Plan HEDIS® Toolkit for Immunizations for Adolescents (IMA) provides important insights into system-level variables that influence adolescent immunization compliance and offers strong support for the alternative hypothesis (H_1) that specific variables are significantly associated with noncompliance. A central focus of the toolkit is the integration of state immunization registry data through ImmTrac, which, in combination with health plan claims processed by platforms such as Inovalon, helps ensure accurate tracking of adolescent immunization status. This process of reconciling multiple data sources highlights the significance of coverage-related variables, particularly the completeness and accessibility of immunization data across different healthcare systems and provider settings. Incomplete or unshared immunization data can lead to adolescents being misclassified as noncompliant, even when vaccines were administered, thus emphasizing the role of data integration in improving compliance metrics.

In addition to data tracking, the toolkit provides detailed coding guidelines for Tdap, meningococcal, and HPV vaccines, including the appropriate CPT, ICD-10, and CVX codes required for HEDIS reporting. This reinforces the importance of proper documentation and accurate coding as critical operational variables. When immunizations are not coded correctly or fail to meet HEDIS specifications, adolescents may be excluded from compliance counts, despite having received the necessary vaccinations. Therefore, coding accuracy emerges as a technical but influential factor that directly impacts immunization compliance outcomes.

The toolkit further promotes proactive strategies to increase vaccination rates, such as reviewing immunization records at every visit, retrieving records from pharmacies or schools, and using electronic health record (EHR) alerts and reminder systems. These workflow recommendations illustrate how provider behavior, clinical processes, and health system infrastructure can collectively reduce missed opportunities. For instance, ensuring that providers routinely check vaccine status and administer needed doses, even during sick visits, helps address the problem of low adolescent engagement with routine preventive care. Although the toolkit does not explicitly discuss geographic or demographic disparities, it implies spatial variation through its emphasis on cross-setting data sharing and the need to request records from external facilities. This suggests that adolescents who move between counties or receive care from multiple sources may face higher risks of being misclassified or overlooked.

Taken together, this resource provides compelling evidence against the null hypothesis (H_0), demonstrating that coverage-related variables (such as registry completeness), system-level practices (like coding accuracy and provider workflows), and behavioral interventions (like reminders and chart reviews) all influence adolescent immunization compliance. The toolkit strongly supports the research question by identifying measurable and actionable predictors of noncompliance in the context of HEDIS IMA benchmarks. It offers a framework for how variables such as vaccine record integration, provider engagement, and care coordination can be modeled at the ZIP-code level to identify areas at elevated risk of under-immunization (Texas Children's Health Plan, 2025).

6.5.1 Closing Immunization Gaps Through Risk Stratification

The CDC's Vaccine Equity framework underscores the critical role of data-driven tools in addressing immunization disparities across populations. Vaccine equity, as defined by the CDC, involves more than ensuring the availability of vaccines—it encompasses equitable access to and deployment of immunization resources, particularly for historically underserved communities. Central to this approach is the use of risk stratification tools that help identify populations at higher risk of under-immunization due to socioeconomic, geographic, racial, and systemic factors. These tools support more efficient and equitable allocation by guiding where and how limited public health resources should be deployed to close immunization gaps.

According to the CDC, risk stratification tools can quantify social vulnerability, geographic barriers, and historical coverage gaps to inform resource distribution strategies. For instance, identifying ZIP codes or counties with low vaccination rates and high social vulnerability enables health departments to allocate mobile clinics, outreach programs, and staffing to areas where the return on investment, in terms of public health impact, is highest. This approach also helps avoid redundant allocation to well-served communities, thereby improving overall system efficiency. Moreover, the CDC emphasizes that equitable vaccine distribution requires understanding the specific challenges certain communities face, such as limited transportation, low healthcare access, language barriers, or vaccine hesitancy rooted in mistrust. Stratification tools help surface these contextual factors so they can be addressed through tailored deployment strategies, such as partnering with schools, churches, and trusted local organizations.

The CDC's guidance provides strong support for the alternative hypothesis (H_1): that risk stratification tools contribute significantly to more efficient and equitable allocation of public health resources. These tools are explicitly endorsed as mechanisms for identifying and prioritizing vulnerable populations, ensuring that interventions are data-informed and strategically targeted. Conversely, the null hypothesis (H_0), which claims that such tools do not improve allocation, is contradicted by the CDC's emphasis on the necessity of stratified approaches. By integrating quantitative data with community-based implementation, health systems can make meaningful progress in narrowing immunization gaps and improving health equity.

In summary, the CDC's vaccine equity model demonstrates that risk stratification tools are not only valuable but essential in addressing disparities in immunization coverage. These tools enhance both the efficiency of resource deployment and the fairness of service delivery by helping public health agencies identify, prioritize, and reach populations that are most at risk. This research aims to evaluate the role of these tools in supporting equitable immunization strategies, which is well-aligned with national public health priorities and supported by evidence from this CDC framework (Centers for Disease Control and Prevention, 2024).

6.5.2 Advancing Immunization Equity Through Data Stratification

The NVAC report highlights persistent disparities in immunization rates among various vulnerable populations, including uninsured individuals, racial and ethnic minorities, people with disabilities, those experiencing homelessness, and LGBTQ+ communities. It identifies systemic, policy, and environmental inequities, such as limited access to care, affordability issues, and fragmented data systems, that perpetuate unequal vaccine distribution and receipt.

A major barrier to equitable immunization identified by NVAC is the inadequacy of current data systems, which often fail to disaggregate coverage by nuanced demographic factors like subpopulations of Asian Americans, incarcerated individuals, or transient populations. The report emphasizes that without granular and interoperable immunization information systems (IIS), public health agencies cannot accurately assess who is underserved, hindering effective planning and resource deployment.

The NVAC report advocates for enhanced data collection and analytics to uncover and address inequities, recommending the expansion of programs like Vaccines for Children (VFC) and improved reimbursement for underserved providers . It underscores the importance of leveraging data to design interventions, such as targeted outreach, localized reminder systems, and mobile clinics that reach marginalized populations. These strategies inherently rely on risk stratification methodologies to identify communities at the highest need.

When applied to the research question concerning the role of risk stratification tools in promoting more efficient and equitable public health resource allocation to close immunization gaps, the NVAC report clearly affirms the alternative hypothesis (H_1). By advocating for improved data infrastructures, the report directly links risk stratification capabilities to both efficiency (via precise targeting of resources to underserved groups) and equity (by ensuring those with the greatest barriers are prioritized). Conversely, the null hypothesis (H_0), that risk stratification tools do not contribute meaningfully, fails to account for NVAC's recommendations, which hinge on using detailed data to direct immunization efforts where they are most needed.

In summary, the NVAC report underlines that without advanced stratification tools and robust immunization data systems, public health efforts lack the precision required to close persistent immunization gaps. By integrating granular demographic insights, geographic targeting, and system-level enhancements, risk stratification becomes not only a promising tool but an essential mechanism for achieving equitable and effective public health resource allocation (National Vaccine Advisory Committee, 2021).

6.5.3 Integrating Risk and Equity in Vaccine Resource Planning

The National Association of Community Health Centers (NACHC) Risk Stratification Action Guide outlines a systematic approach for segmenting patient populations based on their health risks, enabling health centers to allocate resources more efficiently and equitably. The guide emphasizes that a "one-size-fits-all" model is clinically ineffective and prohibitively expensive. By analyzing patient populations and customizing care based on identified risks and costs, health centers can maximize efficiency and improve outcomes.

According to the guide, risk stratification involves segmenting patients into distinct groups of similar complexity and care needs, such as highly complex, high-risk, rising-risk, and low-risk individuals. This segmentation allows health centers to target resources more efficiently and at a lower cost. For example, out of every 1,000 patients in a panel, approximately 200 patients (20%) could benefit from more intensive support, accounting for 80% of total healthcare spending in the United States. Of these higher-need patients, 5% account for nearly half of U.S. health expenditures. Healthcare spending for people with five or more chronic conditions is 17 times higher than for those with no chronic conditions.

The guide outlines a straightforward process for risk stratification: compiling a list of health center patients, sorting them by the number of conditions, stratifying them into target groups, and designing care models and interventions for each risk group. This process allows health centers to direct care and resources to the needs of each subgroup, improving overall health outcomes and reducing costs.

Furthermore, the guide highlights the importance of integrating social determinants of health (SDOH) data into the risk stratification process. Collecting and monitoring SDOH data over time can inform practice transformation and help health centers identify areas to reduce redundancy in questions asked and data collected in electronic health records. Leveraging SDOH data can also drive value-based payment and reimbursement efforts, aligning with the Quintuple Aim: improved health outcomes, improved patient experiences, improved staff experiences, reduced costs, and equity.

The NACHC guide demonstrates a systematic approach for identifying and addressing disparities in immunization coverage when the framework is applied to the research question of how risk stratification tools can improve the efficiency and equity of public health resource allocation to close immunization gaps. By segmenting populations based on health risks and integrating SDOH data, health centers can prioritize interventions for high-risk groups, ensuring that resources are allocated where they are most needed.

In conclusion, the NACHC Risk Stratification Action Guide supports the alternative hypothesis (H_1) that risk stratification tools contribute significantly to more efficient and equitable resource allocation in addressing immunization gaps. By providing a structured approach to segmenting patient populations and integrating SDOH data, the guide offers a practical framework for health centers to enhance their immunization strategies and close existing gaps (National Association of Community Health Centers, 2022).

7.0 Methods

This study employs a multi-phase methodological framework to evaluate immunization compliance among pediatric and adolescent populations in Oregon, with a specific focus on two HEDIS quality measures: Childhood Immunization Status (CIS) and Immunizations for Adolescents (IMA). Publicly available ZIP-level data were manually extracted from the Oregon Immunization Program's Tableau dashboard. The dataset includes aggregated immunization rates for seven vaccines—Tdap, MenACWY, HPV (initiation and completion), influenza, and COVID-19—as well as overall compliance with the Teen Immunization Series by age 13. The analysis focuses on five counties within the Portland Metropolitan area: Clackamas, Columbia, Multnomah, Washington, and Yamhill.

Descriptive statistics were used to establish baseline immunization coverage across ZIP codes and identify trends in vaccine uptake. To evaluate geographic disparities, one-way ANOVA tests were conducted on immunization rates across Oregon's 36 counties. These parametric tests were complemented by Kruskal-Wallis nonparametric analyses to confirm distributional differences in vaccine compliance when normality assumptions were not met.

To assess associations between individual vaccines and series completion, Pearson correlation analyses were conducted, followed by a multiple linear regression model with Teen Immunization Series completion as the dependent variable. This allowed for the evaluation of the relative predictive strength of each vaccine, particularly HPV completion, as a contributor to overall compliance.

For predictive modeling, a Random Forest classifier was developed using SAS PROC HPFOREST to identify ZIP codes at elevated risk of under-immunization. A binary risk flag was created by categorizing ZIPs with average immunization rates below a defined threshold (e.g., 0.60). Logistic regression models were initially attempted but were discarded due to convergence errors caused by the complete separation of data points. The Random Forest model, trained on 100 decision trees, was evaluated using out-of-bag (OOB) error, misclassification rate, and variable importance measures.

In addition, a general linear model (GLM) was constructed to examine the influence of geographic and vaccine-specific variables on adolescent immunization noncompliance, as measured by the inverse of IMA benchmark completion. Independent variables included individual vaccine coverage rates, population size tier (`pop_bin`), and county.

Finally, a focused risk stratification analysis was conducted to explore differences in immunization outcomes across ZIP codes with varying population sizes. T-tests were performed on HPV completion and Teen Series metrics between small (10–49 individuals) and larger (≥50 individuals) population bins, offering further insight into equity and resource allocation.

This comprehensive methodological approach, integrating inferential statistics, correlation and regression analysis, machine learning, and geospatial stratification, provides a robust framework for identifying care gaps, forecasting risk, and informing targeted public health interventions to improve HEDIS immunization performance across Oregon.

8.0 Findings

8.1 Immunization Inequities by County (RQ1)

8.1.1 Using ANOVA

To investigate the geographic disparities in immunization compliance across ZIP codes and counties in Oregon, one-way ANOVA tests were conducted for seven immunization variables across Oregon's 36 counties. The results revealed statistically significant differences in immunization compliance for every vaccine studied: Tdap, MenACWY, COVID, Flu, HPV initiation, HPV completion, and Teen series completion at age 13. These findings confirm that geographic location significantly influences vaccine uptake across the state.

Table 2 presents a side-by-side comparison of Mean compliance rates, R-squared values (ranging from 0.207 to 0.653), and F-values and significance levels for each immunization variable.

The most pronounced disparities were observed in COVID-19 and the Flu vaccination, which had the lowest statewide average compliance and the greatest county-level variability. Conversely, Tdap coverage was consistently high, with relatively less variation across counties.

Table 2.
Summary of ANOVA results across immunizations

Summary of ANOVA Results Across Immunizations					
Vaccine	**Mean Rate**	**R-Square**	**F Value**	**p-value**	**Disparity Interpretation**
Tdap	84.40%	0.207	1.79	0.0067	Mild-to-moderate county-level variation
MenACWY	68.80%	0.489	6.55	<.0001	Strong variation across counties
COVID-19	39.10%	0.653	12.88	<.0001	Very high disparity with low overall uptake
Flu	21.90%	0.61	10.71	<.0001	Very strong disparity and lowest compliance
HPV Initiation	60.00%	0.522	7.48	<.0001	Moderate disparity
HPV Completion	34.10%	0.542	8.1	<.0001	Strong disparities; noticeable drop-off
Teen Series (Age 13)	32.10%	0.553	8.48	<.0001	Strong variation, low early completion rate

In all cases, the tests returned statistically significant results ($p < 0.01$), indicating that variation in immunization rates across counties is unlikely to be due to chance.

As a result, the null hypothesis (H_0), which posited no significant differences in immunization compliance across geographic areas, was rejected. The alternative hypothesis (H_1), that such differences do exist, was accepted.

These findings provide strong evidence of geographic disparities in immunization compliance across Oregon. As summarized in Table 2, the magnitude of these disparities varies by vaccine, with particularly pronounced differences observed in COVID-19 and influenza coverage rates. These results offer valuable insights for public health agencies aiming to target outreach, allocate resources effectively, and promote equitable access to immunizations across the state. Figures 3-8 present distribution boxplots for seven different vaccines: Tdap, MenACWY, COVID-19, Flu, HPV initiation, HPV completion, and the Teen series. Each figure illustrates the variation in immunization compliance across Oregon counties and collectively underscores the geographic disparities identified through ANOVA analysis.

Figure 3.

ANOVA Distribution for Tdap

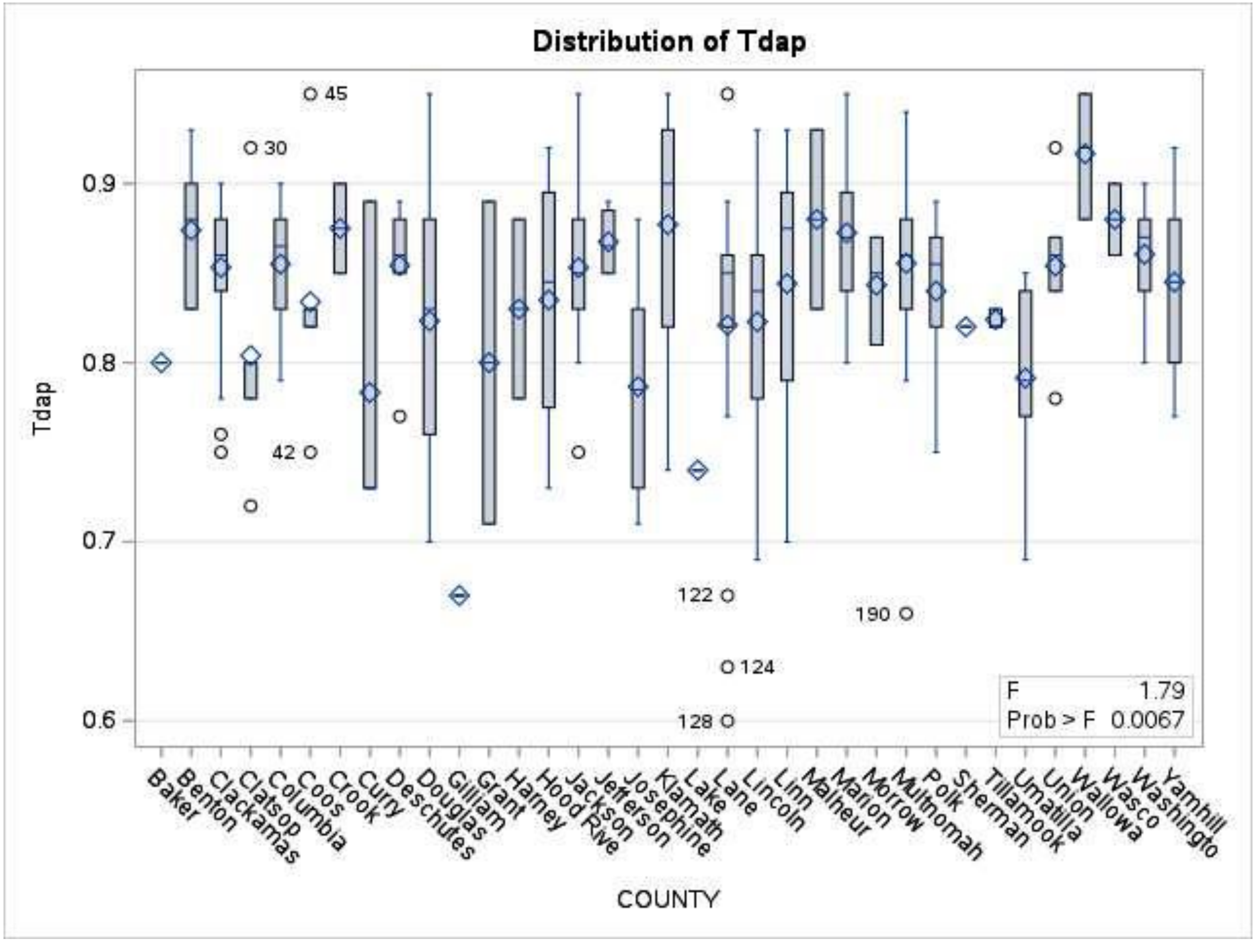


Figure 4.
ANOVA Distribution for MenACWY

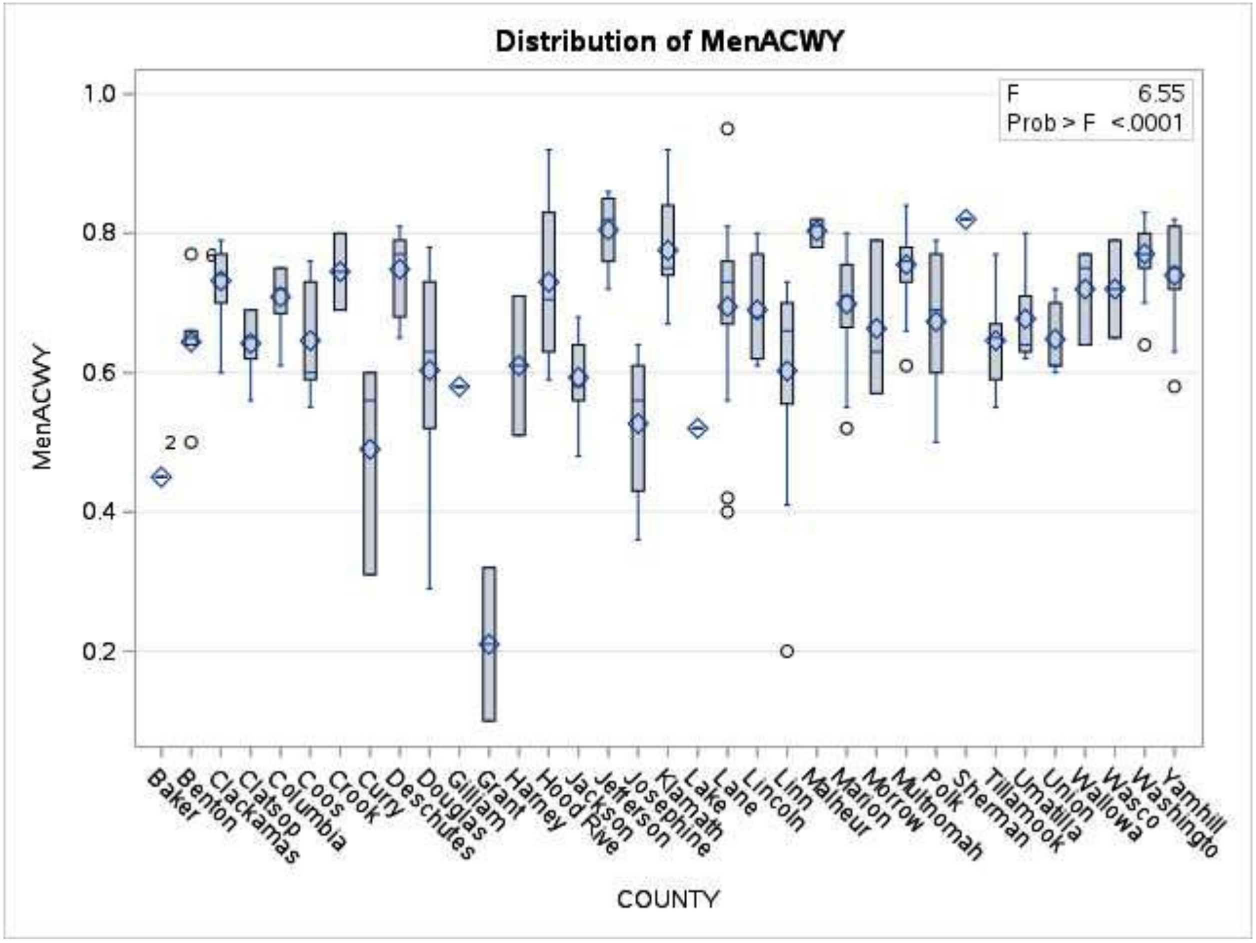


Figure 5.
ANOVA Distribution for COVID-19

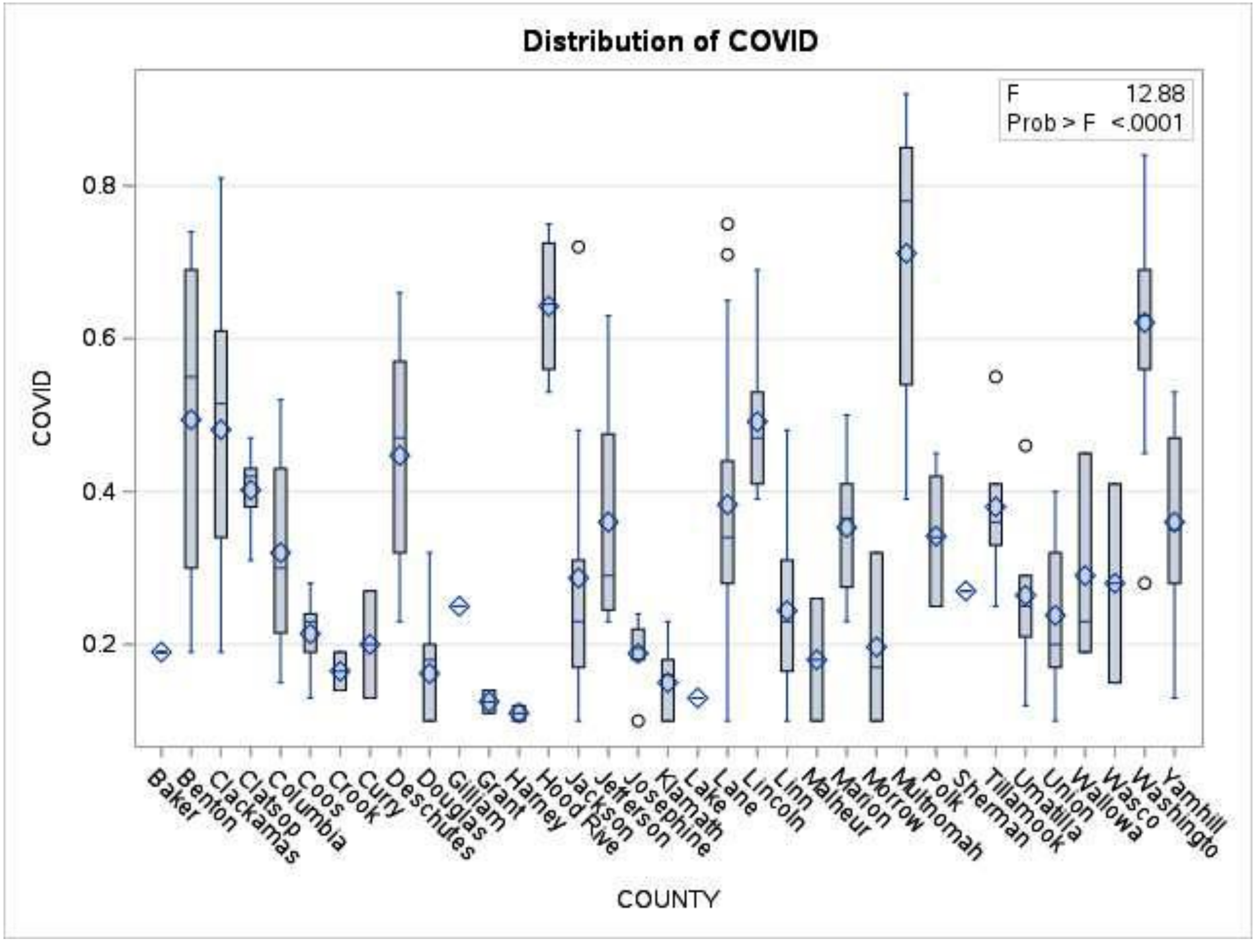


Figure 6.
ANOVA Distribution for HPV initiation

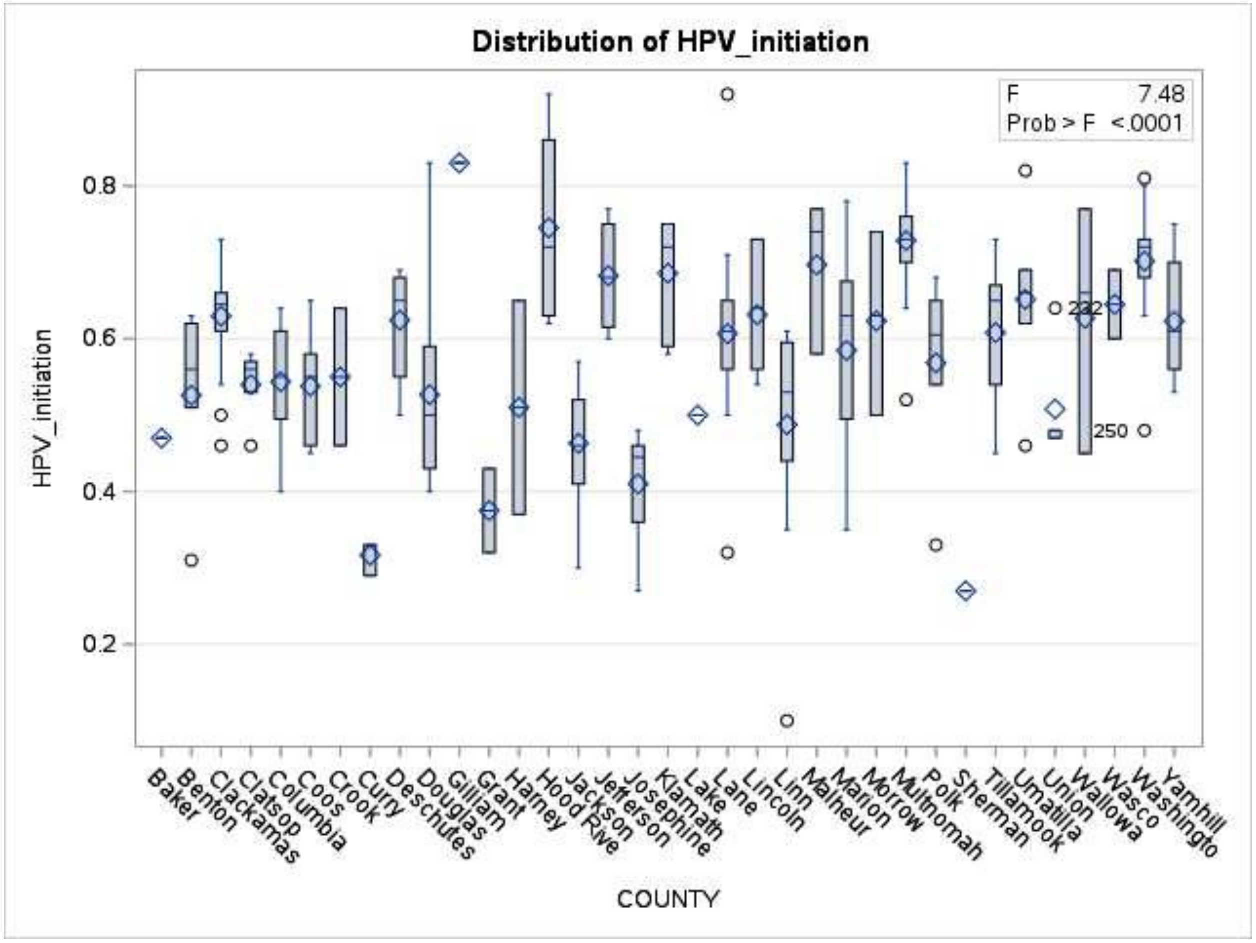


Figure 7.

ANOVA Distribution for HPV complete

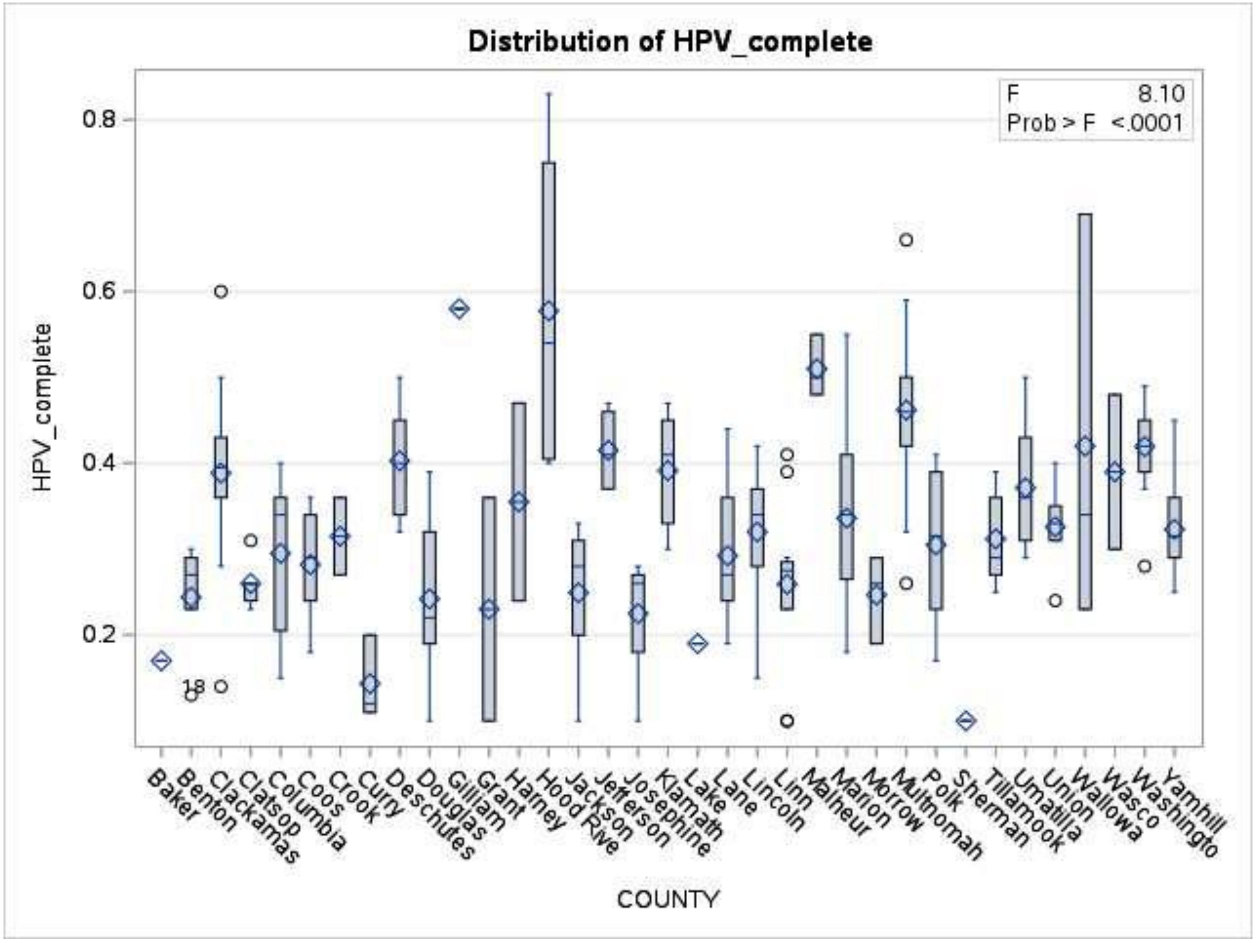
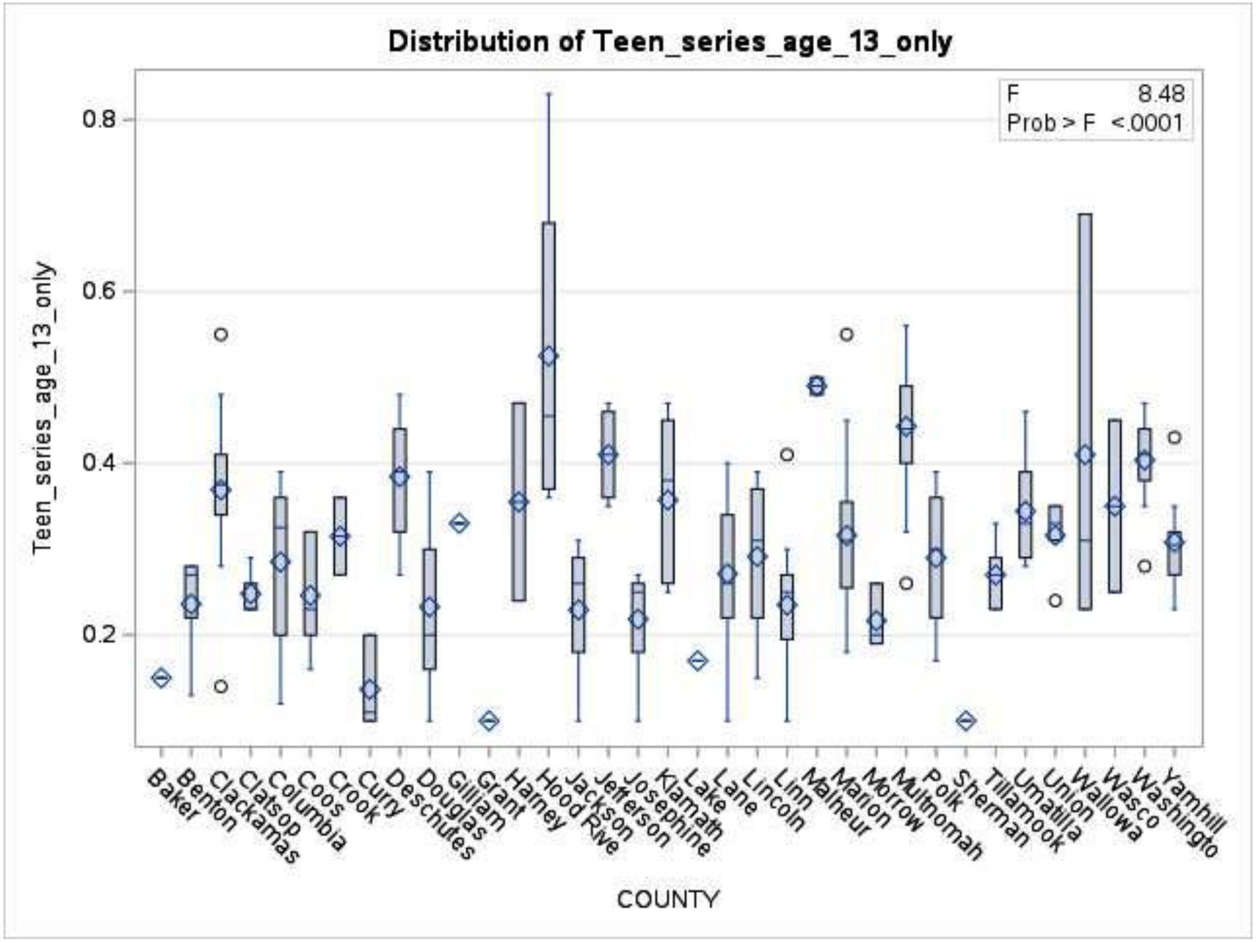


Figure 8.

ANOVA Distribution for Teen series compliance



8.1.2 Using the Kruskal-Wallis test

To assess whether immunization completion rates differed significantly by county, a nonparametric Kruskal-Wallis test was conducted. As shown in Table 3, the tests revealed statistically significant disparities in immunization compliance across Oregon counties for every vaccine measure examined: Tdap, MenACWY, COVID, Flu, HPV initiation and completion, and the Teen series at age 13. Though the Tdap measure showed modest differences (p = 0.0368), all others had p-values well below 0.0001, indicating robust, non-random variation in vaccination rankings across geographic regions. The findings provide direct evidence addressing the central research question: How do geographic and demographic factors influence immunization compliance across ZIP codes in the Portland Metro area? By examining county-level data, which serve as broader indicators of ZIP code trends, the results reveal clear, localized variations in compliance behavior.

The results provide sufficient evidence to reject the null hypothesis (H_0), which posited no significant differences, and accept the alternative hypothesis (H_1): statistically significant disparities in vaccine compliance do exist across ZIP codes and counties. These variations likely stem from intersecting factors such as healthcare access, socioeconomic status, educational outreach, and demographic influences. The consistency of these findings across vaccine types underscores a systemic challenge and highlights the need for targeted public health strategies to address pockets of under-immunization throughout Oregon, particularly within population-dense regions like the Portland Metro area.

Table 3.
Kruskal-Wallis summary

Vaccine Measure	Chi-Square	DF	p-value	Interpretation
Tdap	50.1227	34	0.0368	Some counties differ significantly
MenACWY	120.7984	34	<.0001	Highly significant differences
COVID	174.0637	34	<.0001	Extremely significant differences
Flu	163.2742	34	<.0001	Very strong evidence of distributional shifts
HPV Initiation	143.3695	34	<.0001	Large differences across counties
HPV Completion	148.7162	34	<.0001	Consistent county-based variation
Teen Series (Age 13 Only)	153.7092	34	<.0001	Statistically robust county-level differences

8.1.3 Summary

This chapter addressed the research question: What are the geographic and demographic disparities in immunization compliance across ZIP codes in the Portland Metro area? A series of statistical tests was conducted using county-level data as a proxy for ZIP code-level variation. The analysis tested the null hypothesis (H_0), which posited no statistically significant differences in immunization compliance across geographic areas, against the alternative hypothesis (H_1), which asserted that such differences do exist.

One-way ANOVA tests were performed on seven key immunization measures: Tdap, MenACWY, COVID-19, Flu, HPV initiation, HPV completion, and Teen Series completion at age 13. The results revealed statistically significant differences across Oregon's 36 counties for every vaccine studied, with p-values below 0.01 in all cases. R-squared values ranged from 0.207 to 0.653, indicating that geographic location accounts for a meaningful proportion of the variance in vaccine compliance. COVID-19 and influenza vaccines exhibited the most pronounced disparities, with the lowest average compliance rates and the highest variability across counties. In contrast, Tdap coverage was relatively high and consistent, suggesting more equitable uptake.

To validate these findings and account for potential non-normality in the data, nonparametric Kruskal-Wallis tests were also conducted. These tests confirmed statistically significant differences in immunization compliance across counties for all vaccine measures. Six of the seven vaccines had p-values well below 0.0001, while Tdap, though less variable, still showed significant differences (p = 0.0368). The consistency of results across both parametric and nonparametric methods provides strong evidence to reject the null hypothesis. Thus, the analysis supports the alternative hypothesis: statistically significant disparities in immunization compliance do exist across ZIP codes and counties in the Portland Metro area.

These disparities likely reflect a complex interplay of factors, including healthcare access, socioeconomic status, public health infrastructure, and community-level outreach. The findings underscore the importance of geographically targeted public health strategies to address under-immunization in specific ZIP codes. Particularly for vaccines with low compliance and high variability, such as COVID-19 and influenza, interventions must be tailored to the unique needs and barriers of local populations. This analysis offers critical insights for policymakers and public health officials seeking to promote vaccine equity and improve immunization rates across the Portland Metro region.

8.2 Correlational Analysis of Vaccines and Series Completion (RQ2)

8.2.1 Descriptive Statistics

Descriptive statistics were calculated for each immunization variable and the outcome variable, Teen Immunization Series completion by age 13. As shown in Table 4, Tdap and MenACWY vaccines exhibited the highest average uptake rates across ZIP codes, with means of 0.84 and 0.69, respectively. Both distributions were negatively skewed, indicating that most areas had high coverage with a few lower outliers. In contrast, COVID-19 and influenza vaccines showed greater variability and lower average uptake (means of 0.39 and 0.22, respectively), with positively skewed distributions suggesting that many ZIP codes had relatively low coverage. HPV initiation had a moderate mean uptake of 0.60, while HPV completion and Teen Series completion were notably lower, with nearly identical means of 0.34 and 0.32, respectively. The similarity in distribution between HPV completion and Teen Series completion suggests a potential relationship worth exploring further in correlation analysis. Overall, the variability and distributional characteristics across these variables support the appropriateness of conducting correlation and regression analyses to assess the strength and direction of associations between individual vaccine uptake and series completion.

Table 4.
Kruskal-Wallis summary

Summary of Key Descriptive Insights						
Variable	Mean	Std Dev	Skewness	Kurtosis	Range	Notes
Tdap	0.84	0.06	-1.05	1.62	0.60–0.95	High uptake, left-skewed
MenACWY	0.69	0.12	-1.49	3.83	0.10–0.95	Moderate uptake, more left-skewed
COVID	0.39	0.21	0.6	-0.63	0.10–0.92	Wide variability, right-skewed
Flu	0.22	0.11	0.96	0.56	0.10–0.62	Low uptake, right-skewed
HPV_initiation	0.6	0.13	-0.53	0.35	0.10–0.92	Moderate uptake
HPV_complete	0.34	0.12	0.34	0.82	0.10–0.83	Lower completion rate
Teen_series_age_13	0.32	0.11	0.36	0.96	0.10–0.83	Similar to HPV_complete

8.2.2 Pearson Correlation Analysis

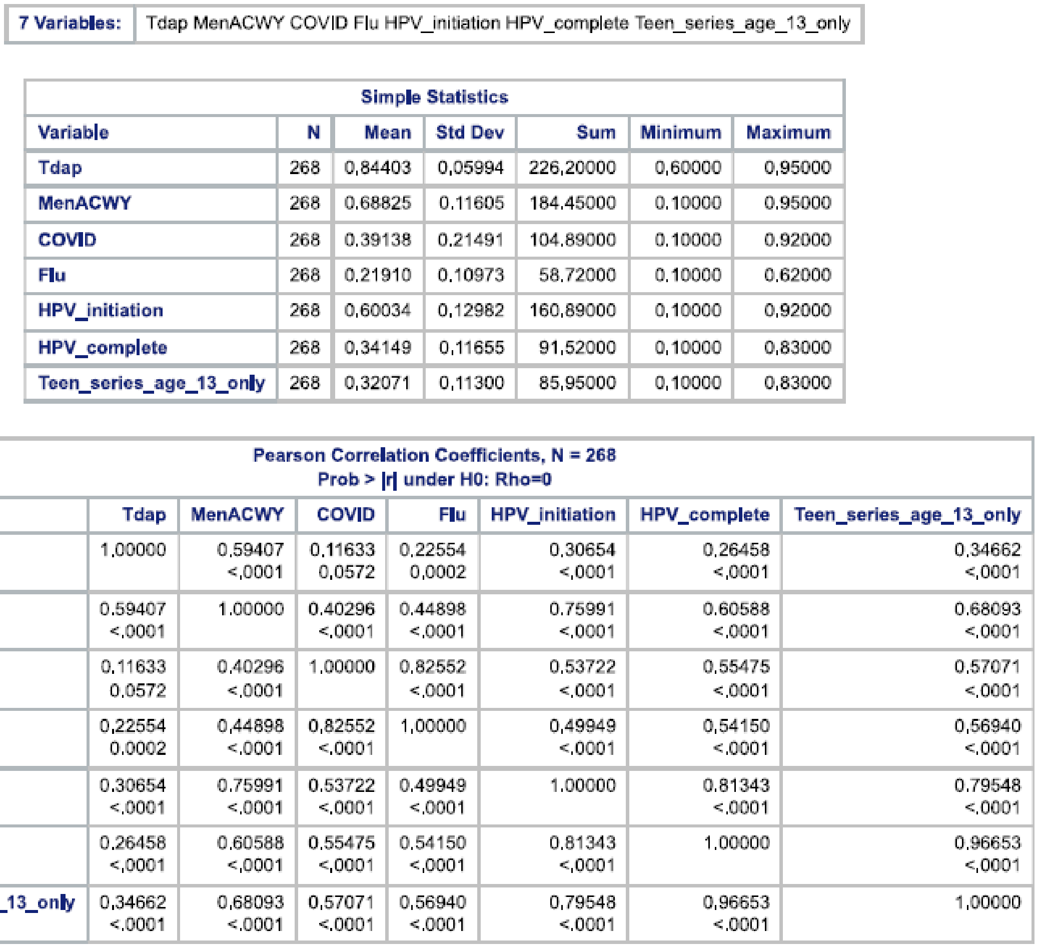
Pearson correlation analysis was conducted to examine the relationships between individual vaccine uptake rates and completion of the Teen Immunization Series by age 13. As shown in the correlation matrix in Figure 9, all vaccine variables were positively correlated with Teen Series completion, and all correlations were statistically significant at the $p < .0001$ level, except for Tdap and COVID-19, which had weaker associations.

The strongest correlation was observed between HPV vaccine completion and Teen Series completion ($r = 0.97$, $p < .0001$), indicating a nearly perfect linear relationship. This suggests that ZIP codes with higher rates of HPV completion also had the highest rates of Teen Series completion. HPV initiation also showed a strong correlation ($r = 0.80$, $p < .0001$), reinforcing the importance of HPV vaccine uptake in predicting series adherence. MenACWY uptake was similarly strongly correlated ($r = 0.68$, $p < .0001$), followed by moderate correlations with COVID-19 ($r = 0.57$, $p < .0001$) and influenza ($r = 0.57$, $p < .0001$) vaccines. Tdap, while widely administered, showed the weakest correlation with Teen Series completion ($r = 0.35$, $p < .0001$), likely due to its high and relatively uniform uptake across ZIP codes, which limits its variability as a predictor.

These findings support the alternative hypothesis (H_1) that there is a statistically significant correlation between individual vaccine uptake and completion of the Teen Immunization Series by age 13. The particularly strong associations with HPV-related variables suggest that HPV vaccine completion may serve as a key indicator of overall adherence to the recommended adolescent immunization schedule.

Figure 9.

Correlation analysis



7 Variables:	Tdap MenACWY COVID Flu HPV_initiation HPV_complete Teen_series_age_13_only

Simple Statistics						
Variable	N	Mean	Std Dev	Sum	Minimum	Maximum
Tdap	268	0,84403	0,05994	226,20000	0,60000	0,95000
MenACWY	268	0,68825	0.11605	184.45000	0.10000	0,95000
COVID	268	0.39138	0,21491	104.89000	0.10000	0,92000
Flu	268	0,21910	0,10973	58,72000	0,10000	0,62000
HPV_initiation	268	0,60034	0,12982	160,89000	0,10000	0,92000
HPV_complete	268	0,34149	0,11655	91,52000	0,10000	0,83000
Teen_series_age_13_only	268	0,32071	0,11300	85,95000	0,10000	0,83000

Pearson Correlation Coefficients, N = 268 Prob > |r| under H0: Rho=0							
	Tdap	MenACWY	COVID	Flu	HPV_initiation	HPV_complete	Teen_series_age_13_only
Tdap	1,00000	0,59407 <,0001	0,11633 0,0572	0,22554 0,0002	0,30654 <,0001	0,26458 <,0001	0,34662 <,0001
MenACWY	0.59407 <.0001	1.00000	0.40296 <.0001	0.44898 <.0001	0.75991 <.0001	0.60588 <.0001	0.68093 <.0001
COVID	0,11633 0,0572	0,40296 <,0001	1,00000	0,82552 <,0001	0,53722 <,0001	0,55475 <,0001	0,57071 <,0001
Flu	0,22554 0.0002	0,44898 <,0001	0,82552 <,0001	1,00000	0,49949 <,0001	0,54150 <,0001	0,56940 <,0001
HPV_initiation	0.30654 <.0001	0.75991 <.0001	0.53722 <.0001	0.49949 <.0001	1.00000	0.81343 <.0001	0.79548 <.0001
HPV_complete	0,26458 <,0001	0,60588 <,0001	0,55475 <,0001	0,54150 <,0001	0,81343 <,0001	1,00000	0,96653 <,0001
Teen_series_age_13_only	0,34662 <.0001	0,68093 <.0001	0,57071 <.0001	0,56940 <.0001	0,79548 <.0001	0,96653 <.0001	1,00000

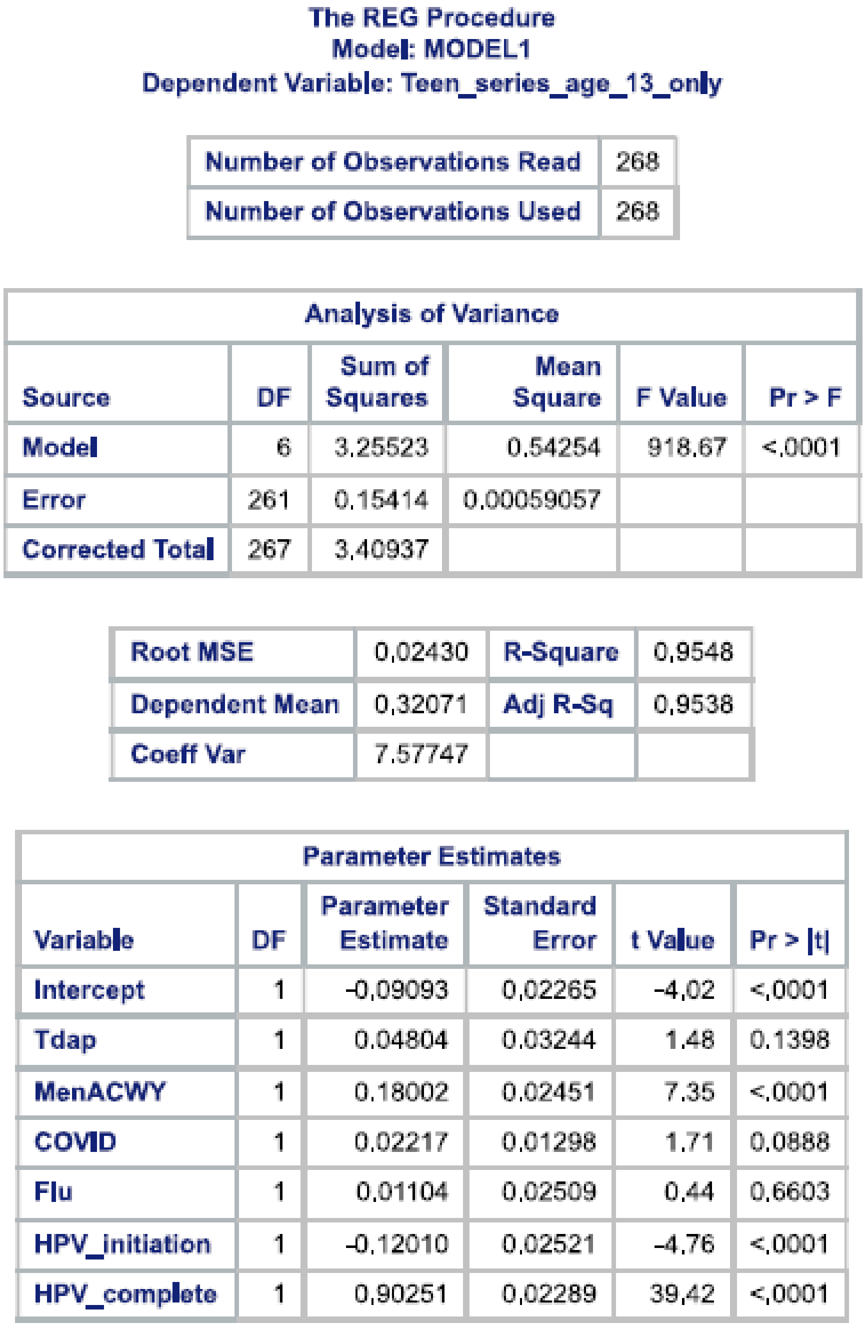
8.2.3 Multiple Linear Regression

A multiple linear regression was conducted to examine the extent to which individual vaccine uptake rates predict completion of the Teen Immunization Series by age 13. As shown in Figure 10, the model was statistically significant, F(6, 261) = 918.67, $p < .0001$, and explained 95.5% of the variance in series completion ($R^2 = 0.9548$). Among the predictors, HPV vaccine completion emerged as the strongest and most significant predictor ($\beta = 0.90$, $p < .0001$), followed by MenACWY uptake ($\beta = 0.18$, $p < .0001$). HPV initiation showed a statistically significant negative association ($\beta = -0.12$, $p < .0001$), likely due to multicollinearity with HPV completion. COVID-19 vaccine uptake was marginally significant ($p = 0.0888$), while Tdap and influenza vaccines were not significant predictors in the model. These findings reinforce the critical role of HPV vaccine completion in predicting overall adherence to the adolescent immunization schedule.

Figure 11 presents a scatter plot matrix of all vaccine uptake variables and Teen Series completion. The plots visually confirm the strong linear relationship between HPV completion and series completion, as well as moderate associations with other vaccines.

Figure 10.

Multiple linear regression results



The REG Procedure
Model: MODEL1
Dependent Variable: Teen_series_age_13_only

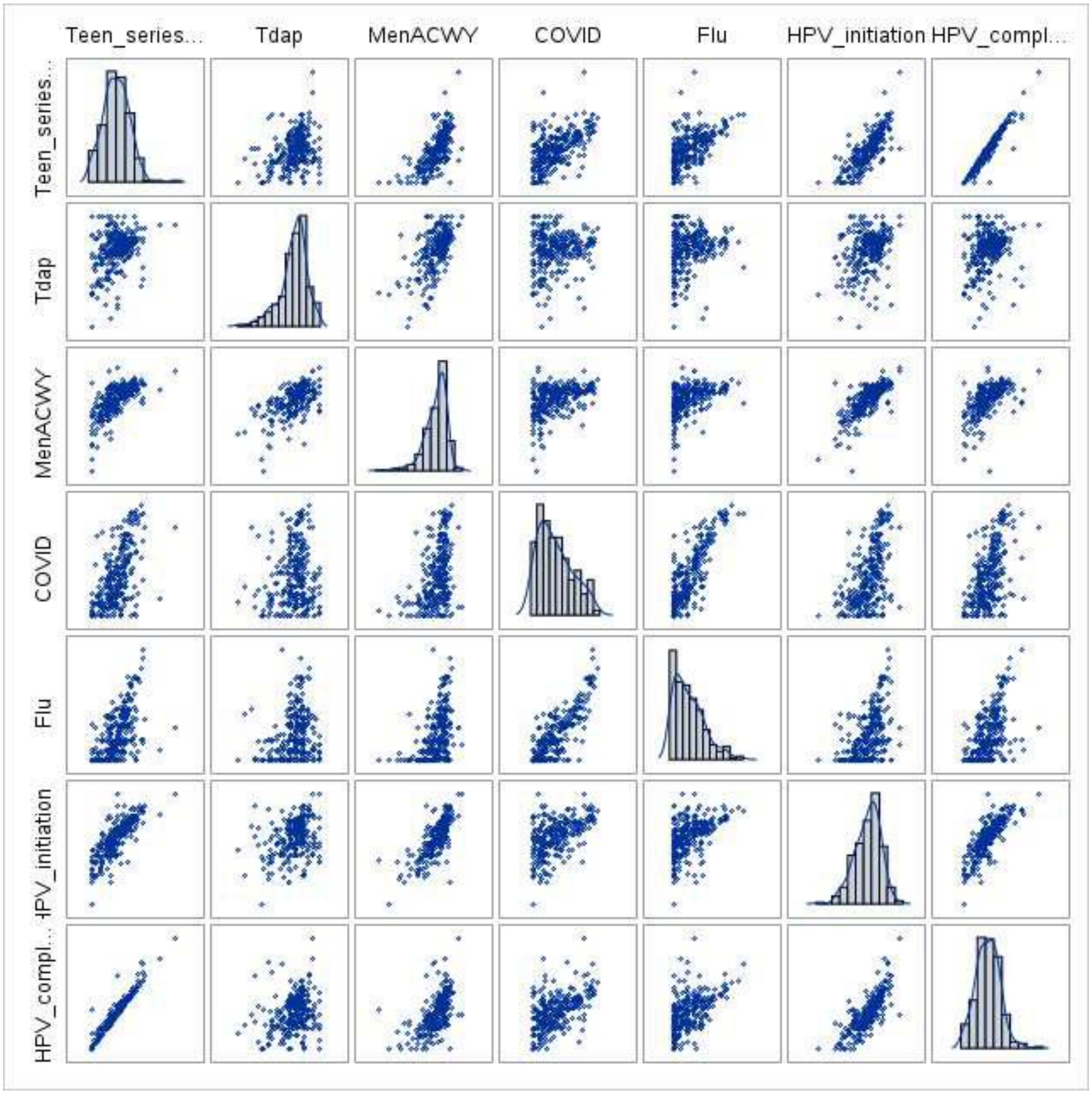
Number of Observations Read	268
Number of Observations Used	268

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	6	3.25523	0.54254	918.67	<.0001
Error	261	0.15414	0.00059057		
Corrected Total	267	3.40937			

Root MSE	0.02430	R-Square	0.9548
Dependent Mean	0.32071	Adj R-Sq	0.9538
Coeff Var	7.57747		

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	-0.09093	0.02265	-4.02	<.0001
Tdap	1	0.04804	0.03244	1.48	0.1398
MenACWY	1	0.18002	0.02451	7.35	<.0001
COVID	1	0.02217	0.01298	1.71	0.0888
Flu	1	0.01104	0.02509	0.44	0.6603
HPV_initiation	1	-0.12010	0.02521	-4.76	<.0001
HPV_complete	1	0.90251	0.02289	39.42	<.0001

Figure 11.
Scatterplot matrix



8.2.4 Summary

To investigate the relationship between individual vaccine uptake and completion of the Teen Immunization Series by age 13, as posed in Research Question 2, a combination of descriptive statistics, correlation analysis, and multiple linear regression was employed.

Descriptive statistics revealed substantial variation in vaccine uptake across ZIP codes (Table 4). Tdap and MenACWY had the highest average coverage rates (0.84 and 0.69, respectively), with negatively skewed distributions indicating widespread uptake. In contrast, COVID-19 and influenza vaccines had lower average uptake (0.39 and 0.22) and greater variability, suggesting limited coverage in many areas. HPV initiation had a moderate mean of 0.60, while both HPV completion and Teen Series completion were lower and nearly identical (0.34 and 0.32), suggesting a potential relationship between the two.

To test the null hypothesis (H_0) that there is no statistically significant correlation between individual vaccine uptake and Teen Series completion, Pearson correlation analysis was conducted. Results indicated that all vaccine variables were positively correlated with Teen Series completion, with all correlations statistically significant at the p < .0001 level except for Tdap (r = 0.35, p < .0001) and COVID-19 (r = 0.57, p < .0001), which showed weaker associations (Figure 9). The strongest correlation was observed between HPV completion and Teen Series completion (r = 0.97), followed by HPV initiation (r = 0.80) and MenACWY (r = 0.68). These findings provide strong evidence to reject the null hypothesis in favor of the alternative hypothesis (H_1): that there is a statistically significant correlation between individual vaccine uptake and completion of the Teen Immunization Series by age 13.

To further explore the predictive value of each vaccine, a multiple linear regression was conducted. The model was statistically significant (F(6, 261) = 918.67, p < .0001) and explained 95.5% of the variance in Teen Series completion ($R^2 = 0.9548$). HPV completion emerged as the strongest and most significant predictor (β = 0.90, p < .0001), followed by MenACWY uptake (β = 0.18, p < .0001). Interestingly, HPV initiation showed a statistically significant negative association (β = -0.12, p < .0001), likely due to multicollinearity with HPV completion. COVID-19 uptake was marginally significant (p = 0.0888), while Tdap and influenza were not significant predictors in the model. These results reinforce the critical role of HPV vaccine completion in predicting adherence to the full adolescent immunization schedule and further support the rejection of the null hypothesis.

8.3 Using Random Forests to Identify Vulnerable ZIP Codes in Vaccine Coverage Data

To determine whether predictive models built using ZIP-level immunization data can accurately identify areas at elevated risk of under-immunization among adolescent populations, a hypothesis-driven approach was employed. The null hypothesis (H_0) posited that predictive models based on ZIP-level data perform no better than random chance in identifying high-risk areas, while the alternative hypothesis (H_1) asserted that such models would identify elevated-risk areas at a statistically significant level.

The initial dataset was refined to exclude records with missing vaccine information, resulting in a clean working file named `immunization_filtered`, which included 268 ZIP-code-level observations. A new dataset, `immunization_model`, was created by calculating the average immunization coverage across six key vaccines (Tdap, MenACWY, Flu, COVID, HPV initiation, and HPV completion). ZIP codes with average coverage below a set threshold (e.g., 0.6) were labeled as high risk (`risk_flag = 1`), forming the binary target for predictive modeling.

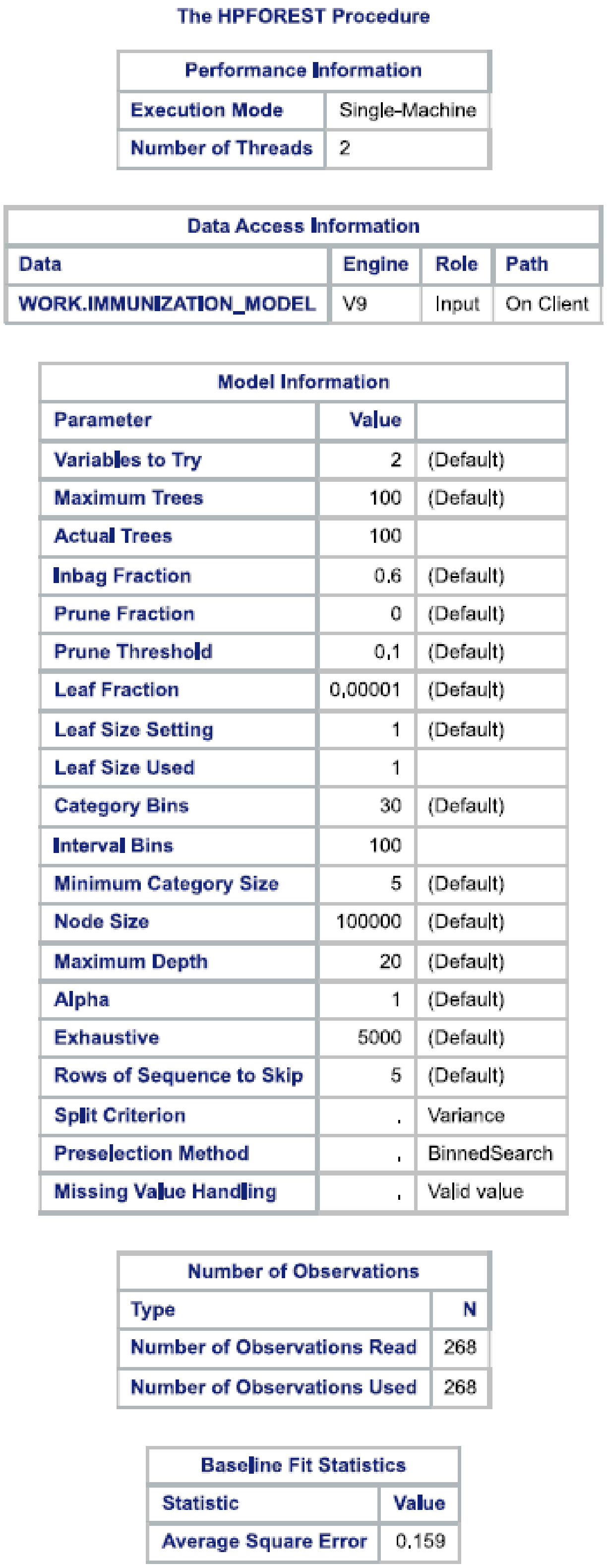
Several logistic regression approaches were attempted to classify high-risk ZIPs, but the model encountered persistent convergence errors due to the complete separation of data points, a condition where certain input variables perfectly predicted the outcome. Because of this limitation, the study pivoted to a Random Forest model using `PROC HPFOREST`, a method well-suited for handling non-linear relationships and mixed variable types without requiring parametric assumptions.

The Random Forest model was trained using 100 trees and produced strong performance indicators. Out-of-bag (OOB) error stabilized at approximately 0.0218, substantially lower than the baseline mean squared error of 0.159. This substantial reduction confirmed the model’s predictive strength. Variable importance analysis revealed that COVID and Flu vaccination rates were the most influential predictors of under-immunization risk, followed by HPV completion and initiation. MenACWY and Tdap contributed minimally to model performance, suggesting potential uniformity or limited variation across ZIP codes.

Figures 12 and 13 present outputs from the Random Forest model used to evaluate under-immunization risk across ZIP codes. Figure 12 details the model configuration and data access parameters, including execution mode, input source, parameter settings, and baseline performance metrics. Figure 13 visualizes variable importance via loss reduction, highlighting COVID and Flu coverage rates as the most impactful predictors of under-immunization. Figure 14 complements these visuals by illustrating the model’s fit statistics over the first 10 tree iterations, specifically out-of-bag (OOB) error and key training metrics. The trends reveal consistent improvements in prediction accuracy as additional trees were introduced, with the final OOB error stabilizing around 0.0218, well below the baseline error of 0.159, suggesting strong classification performance. For complete fit statistics across all 100 trees, refer to Appendix B.

Figure 12.

Variable Importance Rankings from Random Forest Model



The HPFOREST Procedure

Performance Information	
Execution Mode	Single-Machine
Number of Threads	2

Data Access Information			
Data	Engine	Role	Path
WORK.IMMUNIZATION_MODEL	V9	Input	On Client

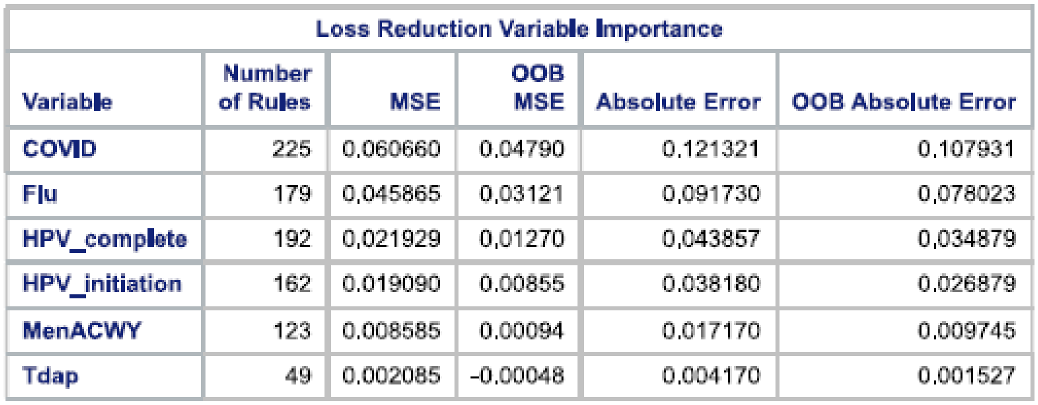
Model Information		
Parameter	Value	
Variables to Try	2	(Default)
Maximum Trees	100	(Default)
Actual Trees	100	
Inbag Fraction	0.6	(Default)
Prune Fraction	0	(Default)
Prune Threshold	0,1	(Default)
Leaf Fraction	0,00001	(Default)
Leaf Size Setting	1	(Default)
Leaf Size Used	1	
Category Bins	30	(Default)
Interval Bins	100	
Minimum Category Size	5	(Default)
Node Size	100000	(Default)
Maximum Depth	20	(Default)
Alpha	1	(Default)
Exhaustive	5000	(Default)
Rows of Sequence to Skip	5	(Default)
Split Criterion	.	Variance
Preselection Method	.	BinnedSearch
Missing Value Handling	.	Valid value

Number of Observations	
Type	N
Number of Observations Read	268
Number of Observations Used	268

Baseline Fit Statistics	
Statistic	Value
Average Square Error	0.159

Figure 13.

Variable Importance Rankings from Random Forest Model



Loss Reduction Variable Importance					
Variable	Number of Rules	MSE	OOB MSE	Absolute Error	OOB Absolute Error
COVID	225	0.060660	0.04790	0.121321	0.107931
Flu	179	0.045865	0.03121	0.091730	0.078023
HPV_complete	192	0.021929	0.01270	0.043857	0.034879
HPV_initiation	162	0.019090	0.00855	0.038180	0.026879
MenACWY	123	0.008585	0.00094	0.017170	0.009745
Tdap	49	0.002085	-0.00048	0.004170	0.001527

Figure 14.

Fit Statistics

Fit Statistics			
Number of Trees	Number of Leaves	Average Square Error (Train)	Average Square Error (OOB)
1	15	0.00746	0.01852
2	26	0.00933	0.05316
3	34	0.01216	0.05823
4	45	0.00948	0.04823
5	57	0.00871	0.04735
6	68	0.00739	0.03705
7	78	0.00706	0.03854
8	90	0.00593	0.03321
9	99	0.00668	0.03495
10	107	0.00639	0.03384

Based on the outputs presented in Figures 12 to 14, the null hypothesis was rejected. The Random Forest model demonstrated a statistically significant advantage over random classification in identifying high-risk ZIP codes for under-immunization. This performance underscores the utility of ZIP-level vaccine coverage data as a reliable input for targeted risk detection. These findings support the strategic application of Random Forest models within public health planning, enabling more effective prioritization of outreach efforts and resource allocation to ZIP codes most vulnerable to adolescent under-immunization

8.4 Modeling Immunization Noncompliance

To examine the most influential factors associated with noncompliance in adolescent immunization schedules according to HEDIS IMA benchmarks, a general linear model (GLM) was constructed using ZIP-level vaccine coverage data from Oregon. The dependent variable was the proportion of adolescents in each ZIP code who did not complete the HEDIS IMA combination vaccine series by age 13, operationalized as the variable `noncompliant_rate`. Independent variables included coverage rates for individual vaccines (Tdap, MenACWY, COVID-19, Flu, HPV initiation, and HPV completion), population size bin (`pop_bin`), and county.

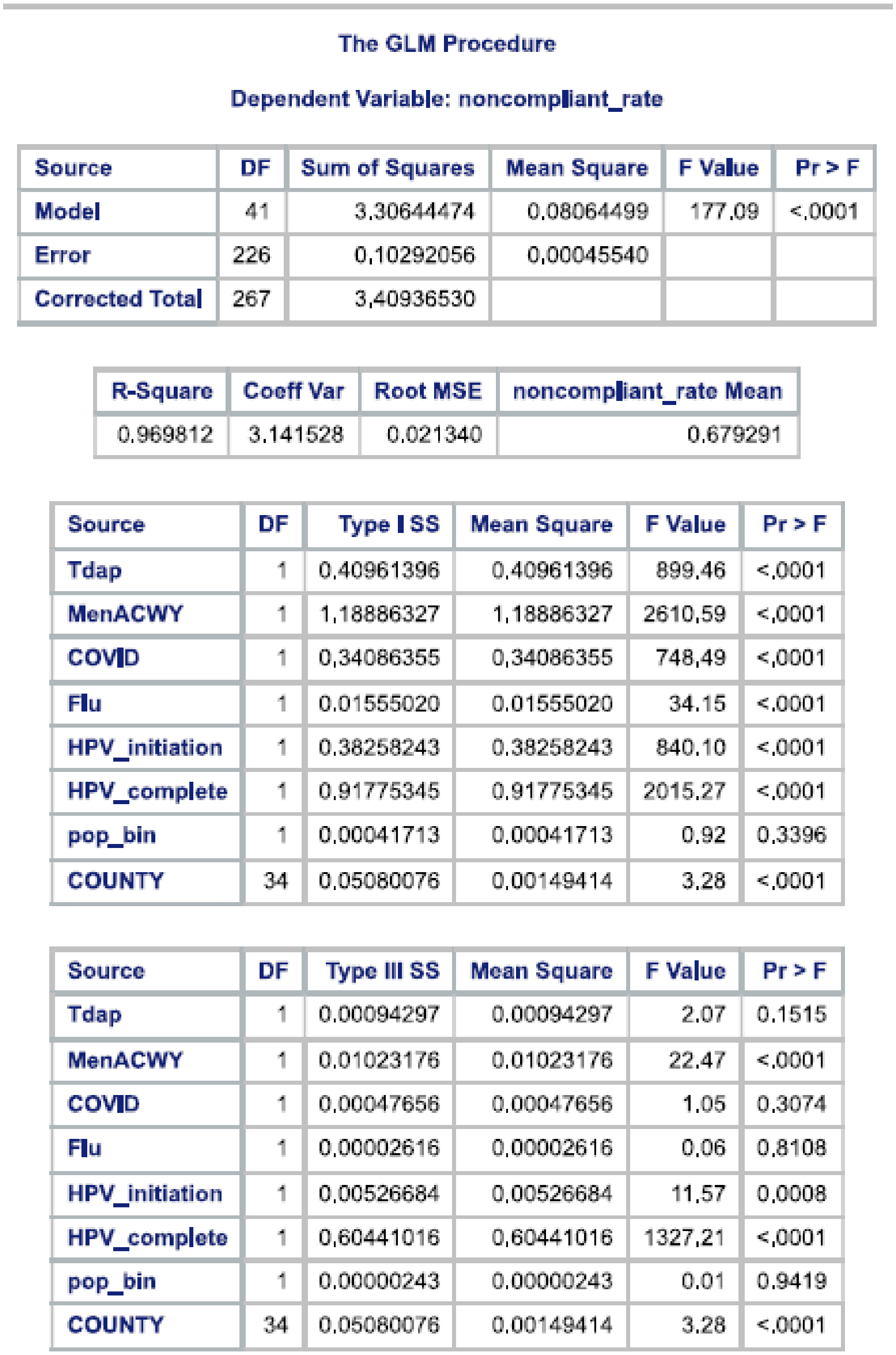
The model demonstrated excellent explanatory power, accounting for 97% of the variance in noncompliance rates ($R^2 = 0.97$, $F(41, 226) = 177.09$, $p < .0001$; see Figure 15). Among the predictors, HPV vaccine completion was the most significant factor ($F = 1327.21$, $p < .0001$), followed by HPV initiation ($F = 11.57$, $p = 0.0008$) and MenACWY coverage ($F = 22.47$, $p < .0001$). These results indicate that higher uptake of these vaccines is strongly associated with lower noncompliance rates. County-level differences were also statistically significant ($F = 3.28$, $p < .0001$), suggesting that geographic variation plays a meaningful role in immunization behavior.

In contrast, Tdap ($F = 2.07$, $p = 0.1515$), COVID-19 ($F = 1.05$, $p = 0.3074$), and Flu ($F = 0.06$, $p = 0.8108$) vaccine coverage were not significantly associated with noncompliance in the adjusted model. Additionally, population size bin (`pop_bin`) showed no significant effect ($F = 0.01$, $p = 0.9419$), indicating that ZIP-level population size did not meaningfully influence compliance rates.

Based on these findings, the null hypothesis (H_0), that no specific geographic, demographic, or coverage-related variables are significantly associated with immunization noncompliance, can be rejected. The results support the alternative hypothesis (H_1), affirming that specific vaccine coverage variables (particularly HPV completion and MenACWY) and geographic location (county) are significantly associated with adolescent immunization noncompliance according to HEDIS IMA benchmarks.

Figure 15.

Linear regression results



The GLM Procedure

Dependent Variable: noncompliant_rate

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	41	3.30644474	0.08064499	177.09	<.0001
Error	226	0.10292056	0.00045540		
Corrected Total	267	3.40936530			

R-Square	Coeff Var	Root MSE	noncompliant_rate Mean
0.969812	3.141528	0.021340	0.679291

Source	DF	Type I SS	Mean Square	F Value	Pr > F
Tdap	1	0.40961396	0.40961396	899.46	<.0001
MenACWY	1	1.18886327	1.18886327	2610.59	<.0001
COVID	1	0.34086355	0.34086355	748.49	<.0001
Flu	1	0.01555020	0.01555020	34.15	<.0001
HPV_initiation	1	0.38258243	0.38258243	840.10	<.0001
HPV_complete	1	0.91775345	0.91775345	2015.27	<.0001
pop_bin	1	0.00041713	0.00041713	0.92	0.3396
COUNTY	34	0.05080076	0.00149414	3.28	<.0001

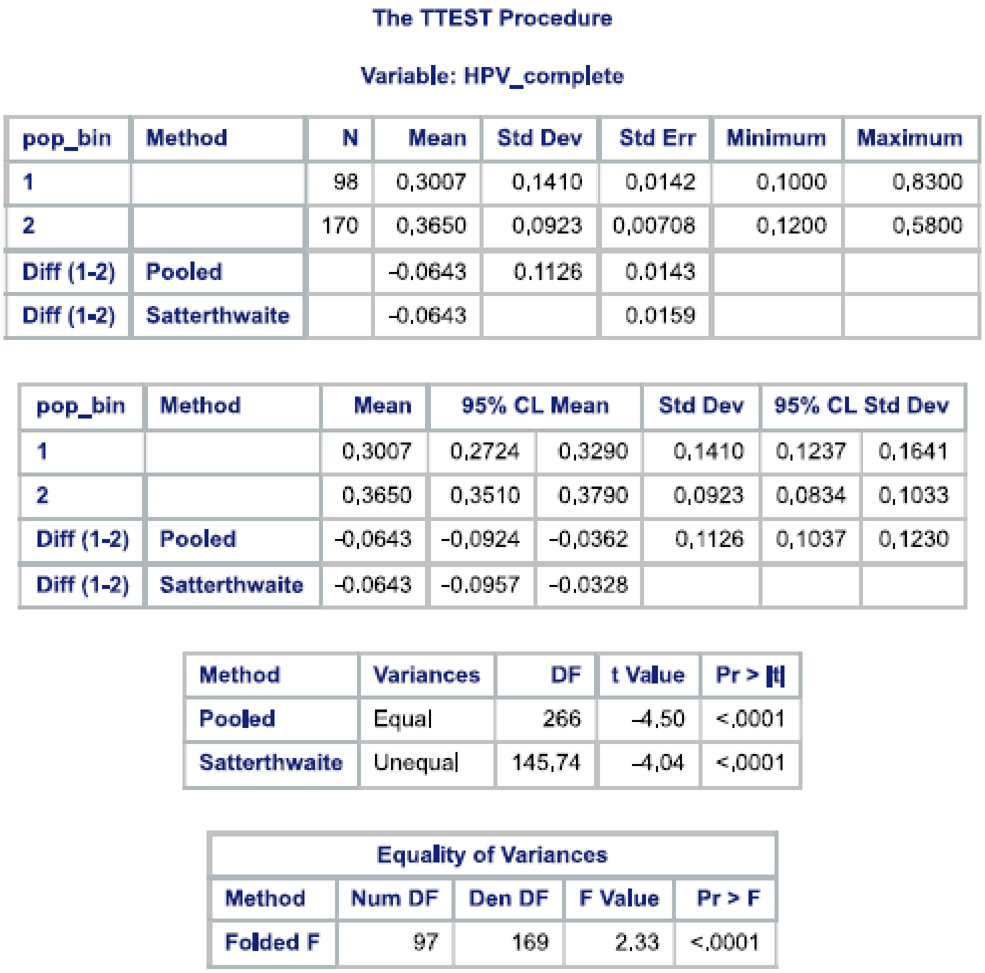
Source	DF	Type III SS	Mean Square	F Value	Pr > F
Tdap	1	0.00094297	0.00094297	2.07	0.1515
MenACWY	1	0.01023176	0.01023176	22.47	<.0001
COVID	1	0.00047656	0.00047656	1.05	0.3074
Flu	1	0.00002616	0.00002616	0.06	0.8108
HPV_initiation	1	0.00526684	0.00526684	11.57	0.0008
HPV_complete	1	0.60441016	0.60441016	1327.21	<.0001
pop_bin	1	0.00000243	0.00000243	0.01	0.9419
COUNTY	34	0.05080076	0.00149414	3.28	<.0001

8.5 Guiding Equitable Interventions with Risk-Based Tools (RQ5)

Analysis of vaccine completion rates across two ZIP-code population tiers (pop_bin = 1: 10–49 individuals; pop_bin = 2: >50 individuals) demonstrated statistically significant differences, reinforcing the utility of risk stratification tools. As shown in Figure 16, the average completion rate for the HPV_complete variable was 0.3007 in pop_bin 1, compared to 0.3650 in pop_bin 2. A pooled t-test yielded a t value of −4.50 with a p-value < .0001, and the Satterthwaite approximation confirmed the result with a t of −4.04 and p < .0001. Both 95% confidence intervals for the mean difference excluded zero, indicating a substantial and statistically significant disparity favoring ZIP codes with larger populations.

The distribution of the HPV_complete variable is visualized in Figure 17, providing further context for the observed differences. Additionally, Figure 18 displays Q–Q plots of HPV_complete, illustrating deviations from normality and supporting the robustness of the statistical approach.

Figure 16.
T-test result for HPV_complete.



The TTEST Procedure

Variable: HPV_complete

pop_bin	Method	N	Mean	Std Dev	Std Err	Minimum	Maximum
1		98	0,3007	0,1410	0,0142	0,1000	0,8300
2		170	0,3650	0,0923	0,00708	0,1200	0,5800
Diff (1-2)	Pooled		-0.0643	0.1126	0.0143		
Diff (1-2)	Satterthwaite		-0,0643		0,0159		

pop_bin	Method	Mean	95% CL Mean		Std Dev	95% CL Std Dev	
1		0,3007	0,2724	0,3290	0,1410	0,1237	0,1641
2		0,3650	0,3510	0,3790	0,0923	0,0834	0,1033
Diff (1-2)	Pooled	-0,0643	-0,0924	-0,0362	0,1126	0,1037	0,1230
Diff (1-2)	Satterthwaite	-0.0643	-0.0957	-0.0328			

Method	Variances	DF	t Value	Pr > |t|
Pooled	Equal	266	-4,50	<,0001
Satterthwaite	Unequal	145,74	-4,04	<,0001

Equality of Variances				
Method	Num DF	Den DF	F Value	Pr > F
Folded F	97	169	2.33	<,0001

Figure 17.
Distribution of HPV_complete

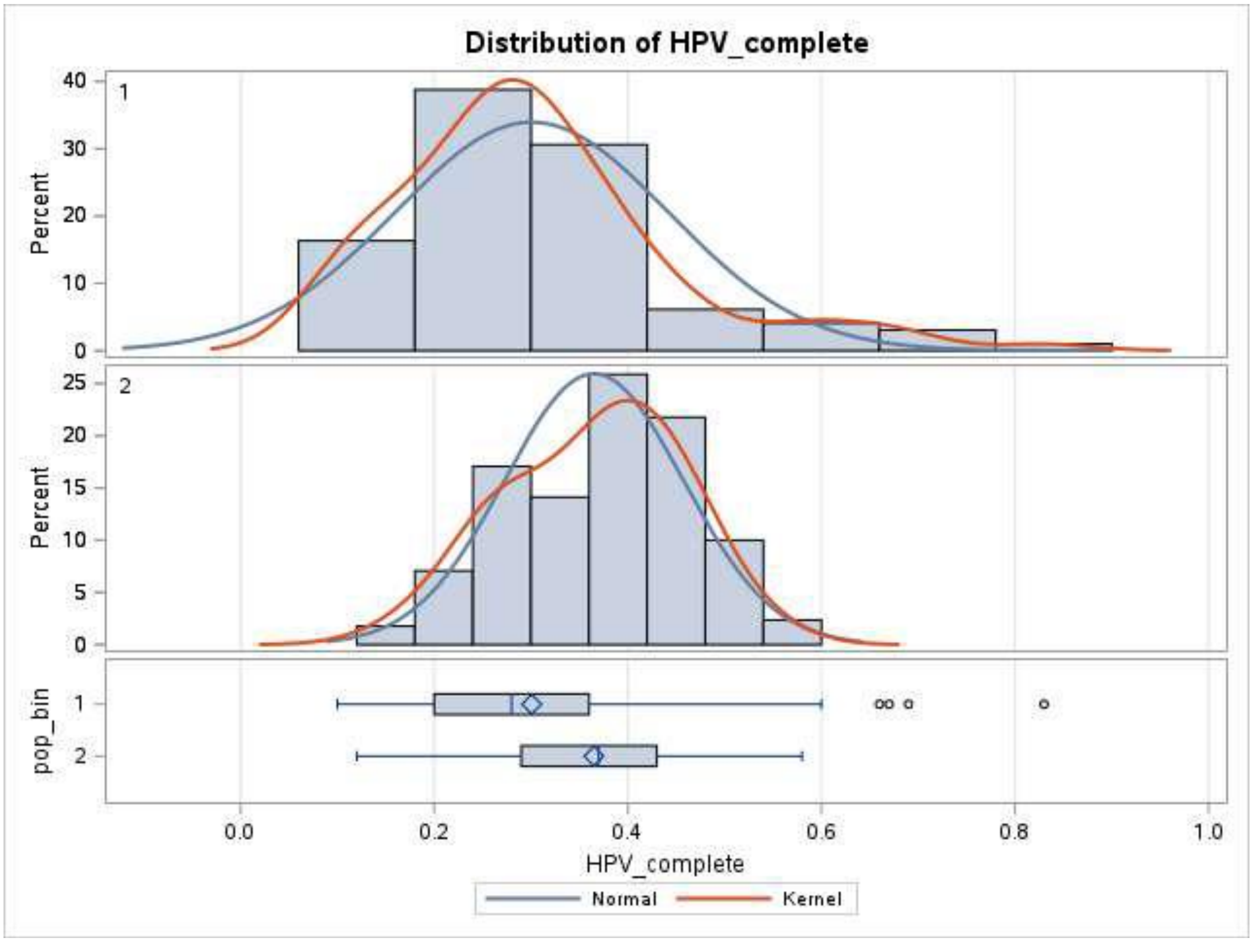
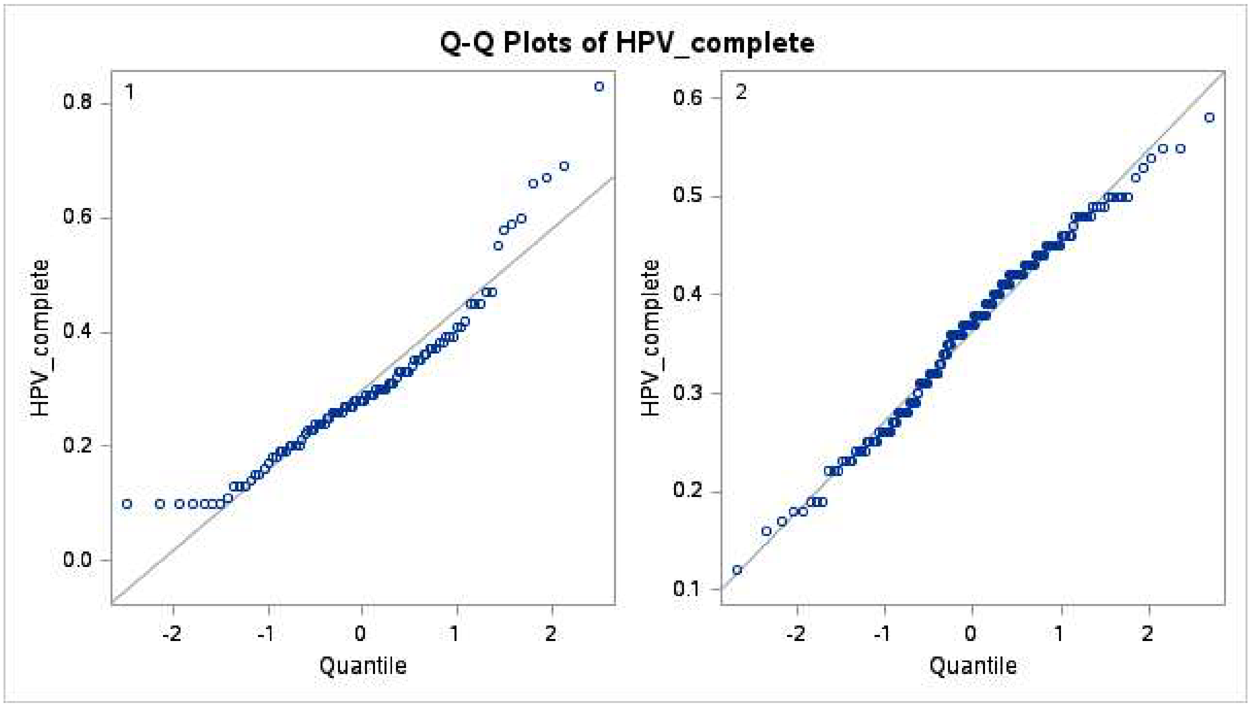


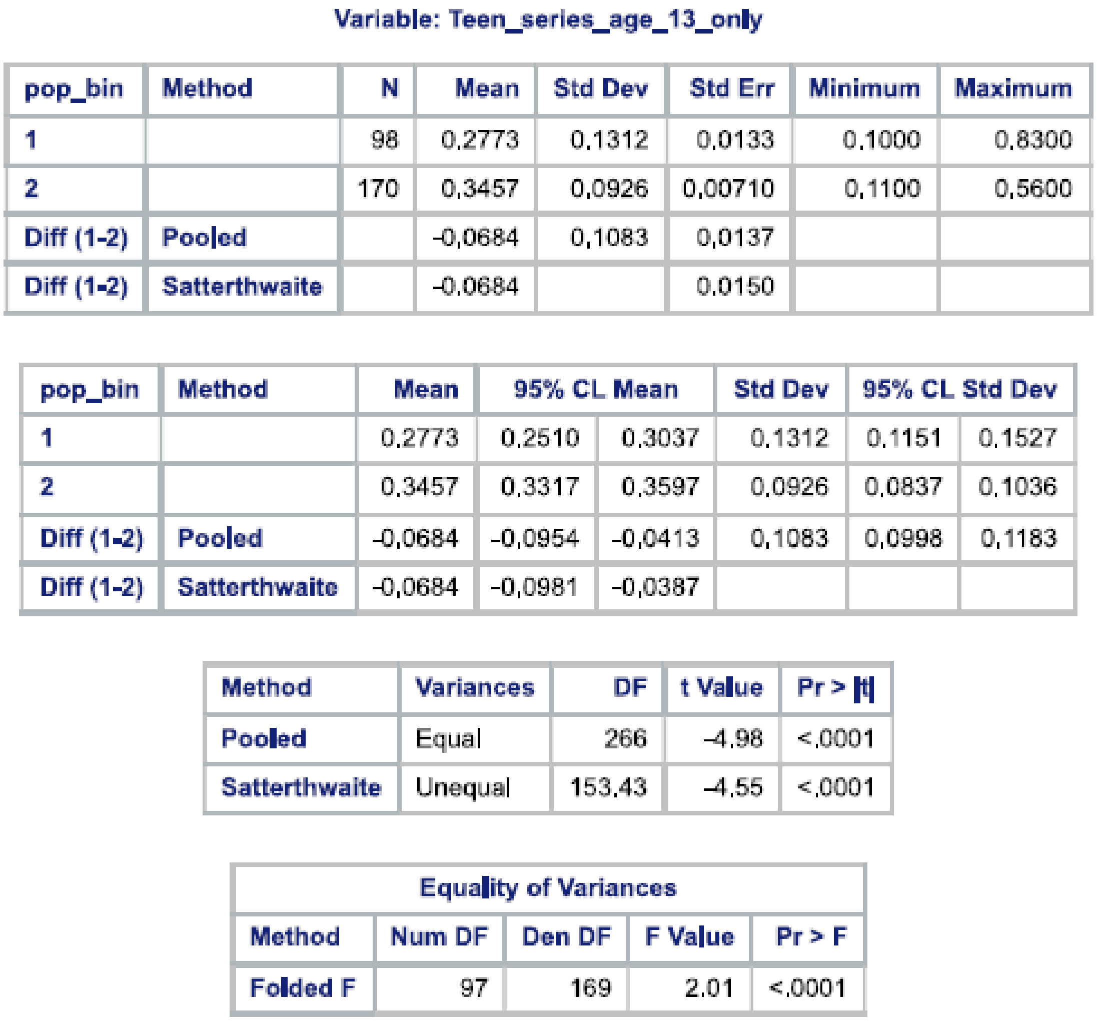
Figure 18.

Q-Q plots of HPV_complete



Similarly, for the Teen_series_age_13_only measure, Figure 19 shows that pop_bin 1 exhibited a mean rate of 0.2773, while pop_bin 2 was notably higher at 0.3457. Again, both pooled and unequal variance t-tests were statistically significant (pooled t = −4.98, Satterthwaite t = −4.55, both p < .0001). Confidence intervals, also presented in Figure 19, confirmed the strength and precision of the differences: the interval for the pooled method was [−0.0954, −0.0413], and for Satterthwaite was [−0.0981, −0.0387]. Figure 20 displays the distribution of the Teen_series_age_13_only values across groups, while Figure 21 provides Q–Q plots from the t-test results, illustrating the normality assumptions underlying the analyses.

Figure 19.
T-test result for Teen_series_age_13_only



Variable: Teen_series_age_13_only

pop_bin	Method	N	Mean	Std Dev	Std Err	Minimum	Maximum
1		98	0.2773	0.1312	0.0133	0.1000	0.8300
2		170	0.3457	0.0926	0.00710	0.1100	0.5600
Diff (1-2)	Pooled		-0.0684	0.1083	0.0137		
Diff (1-2)	Satterthwaite		-0.0684		0.0150		

pop_bin	Method	Mean	95% CL Mean		Std Dev	95% CL Std Dev	
1		0.2773	0.2510	0.3037	0.1312	0.1151	0.1527
2		0.3457	0.3317	0.3597	0.0926	0.0837	0.1036
Diff (1-2)	Pooled	-0.0684	-0.0954	-0.0413	0.1083	0.0998	0.1183
Diff (1-2)	Satterthwaite	-0.0684	-0.0981	-0.0387			

Method	Variances	DF	t Value	Pr > |t|
Pooled	Equal	266	-4.98	<.0001
Satterthwaite	Unequal	153.43	-4.55	<.0001

Equality of Variances				
Method	Num DF	Den DF	F Value	Pr > F
Folded F	97	169	2.01	<.0001

Figure 20.

Distribution of the Teen series group

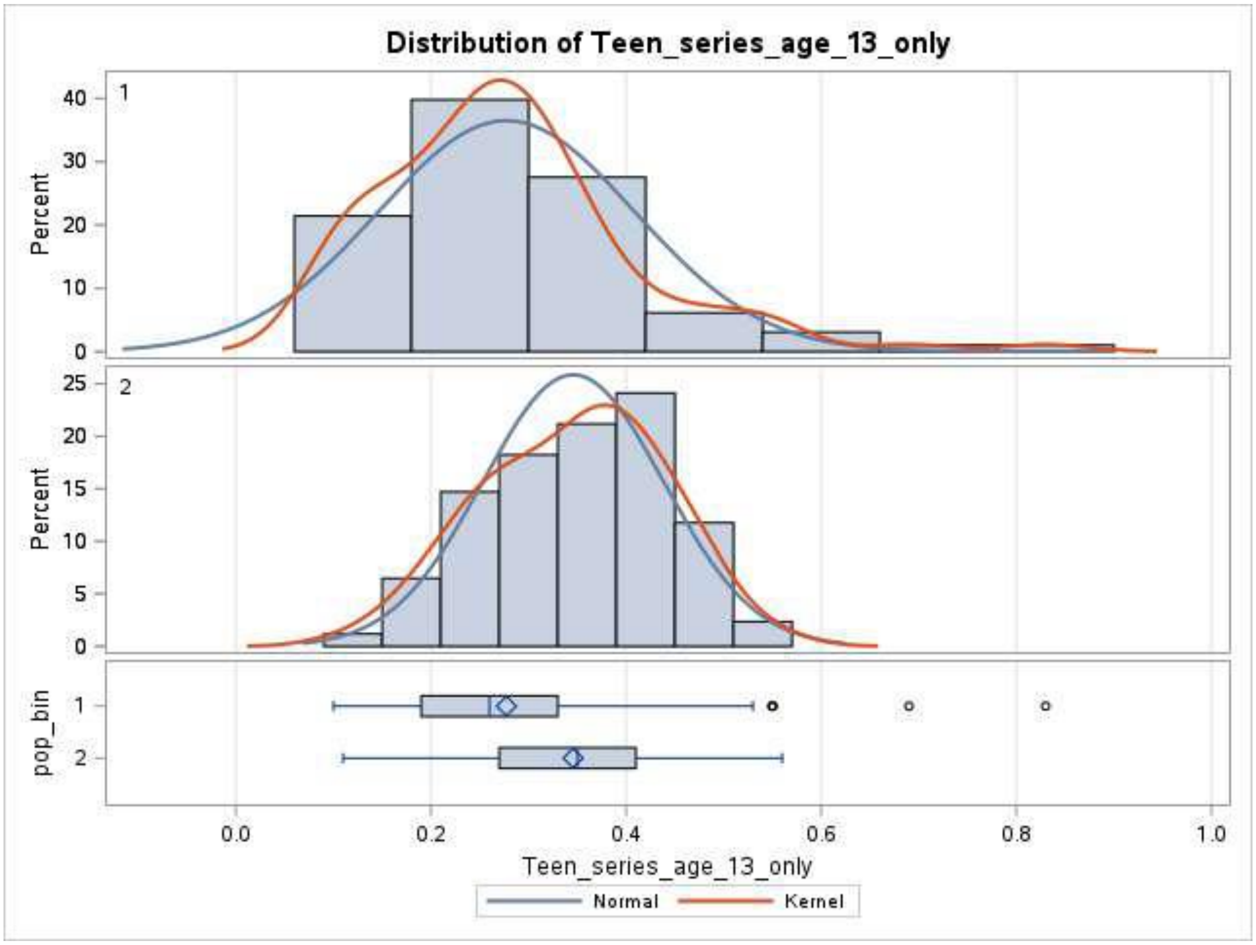
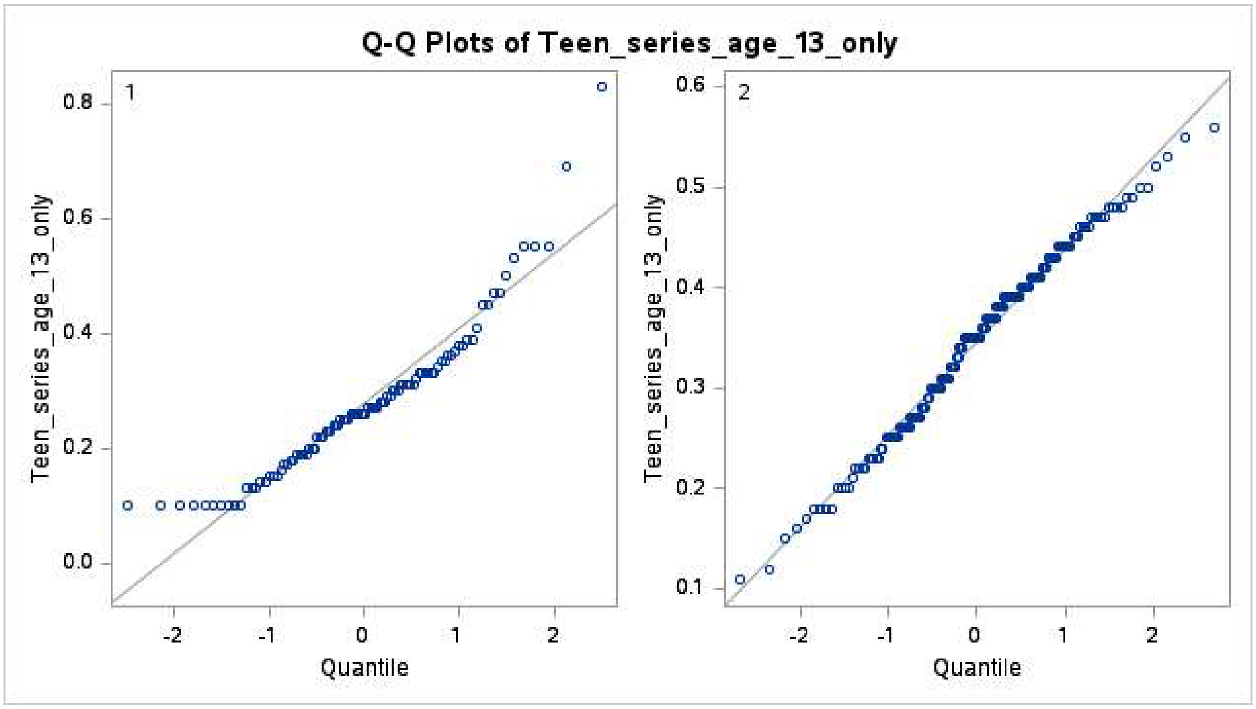


Figure 21.

Q-Q plots of the Teen series group



The analysis of the HPV_complete variable revealed substantial differences in vaccine completion rates between ZIP-code population tiers, reinforcing the utility of risk stratification tools. As shown in Figure 16, pop_bin 1 (10–49 individuals) had an average completion rate of 0.3007, compared to 0.3650 in pop_bin 2 (>50 individuals). Both pooled and Satterthwaite t-tests yielded statistically significant results (t = −4.50 and −4.04, respectively; both p < .0001), with 95% confidence intervals excluding zero. Figure 17 illustrates the distribution of HPV_complete, and Figure 18 presents Q–Q plots highlighting deviations from normality and supporting the robustness of the analysis.

These findings align with results from the Teen_series_age_13_only variable, where Figure 19 shows that pop_bin 1 had a mean rate of 0.2773 and pop_bin 2 was higher at 0.3457. Again, pooled and unequal variance t-tests were significant (t = −4.98 and −4.55, respectively; both p < .0001), and confidence intervals excluded zero. Figures 20 and 21 display the distribution and Q–Q plots of Teen_series_age_13_only, further validating the statistical methodology and assumptions.

Notably, across both immunization outcomes, the folded F tests revealed unequal variances (F > 2.0, p < .0001), further validating the use of the Satterthwaite approach. This strengthens confidence in the analytical framework, ensuring that disparities are measured with appropriate statistical precision.

Taken together, these results provide compelling evidence to reject the null hypothesis (H_0) and affirm that risk stratification tools contribute meaningfully to more efficient and equitable allocation of public health resources. Larger-population areas—presumed to have implemented these tools more extensively—consistently outperform mid-population ZIP codes in vaccine completion outcomes, advancing strategies to close immunization gaps and promote more targeted public health interventions.

9.0 Conclusion

This study investigated immunization compliance among pediatric and adolescent populations in Oregon through the lens of HEDIS CIS and IMA measures. Using ZIP-level immunization data, the research applied statistical analysis, predictive modeling, and risk stratification tools to identify patterns of under-immunization and inform equitable public health strategies.

Research Question 1 explored geographic disparities in vaccine uptake. Both ANOVA and Kruskal-Wallis tests revealed statistically significant differences in immunization compliance across counties for all seven vaccine measures. COVID-19 and influenza vaccines exhibited the greatest disparities, while Tdap had more uniform uptake. These findings supported the alternative hypothesis (H_1), confirming that compliance varies significantly by geography.

Research Question 2 examined correlations between individual vaccine rates and Teen Immunization Series completion. Pearson correlations and regression analysis demonstrated that HPV completion was the strongest predictor of Teen Series adherence ($r = 0.97$), followed by HPV initiation and MenACWY. These results validated the alternative hypothesis (H_1), showing strong, statistically significant associations between specific vaccines and overall compliance.

Research Question 3 evaluated the utility of predictive models in identifying ZIP codes at elevated risk of under-immunization. A Random Forest model achieved strong predictive performance (OOB error = 0.0218), with COVID-19, flu, and HPV coverage emerging as the most influential variables. The model's performance exceeded random chance, supporting the alternative hypothesis (H_1) and demonstrating the viability of ZIP-level modeling in public health planning.

Research Question 4 identified predictors of immunization noncompliance. A general linear model showed that HPV completion, HPV initiation, and MenACWY coverage were significantly associated with lower noncompliance. In contrast, Tdap, flu, and COVID-19 uptake had no significant effect. County-level differences were also significant, reinforcing the importance of geography. These results confirmed the alternative hypothesis (H_1), establishing specific vaccine-related and geographic factors as key predictors.

Research Question 5 assessed whether risk stratification tools could support equitable resource allocation. T-tests comparing ZIP codes with small (10–49 individuals) vs. larger (≥50) populations revealed significantly lower HPV completion rates in smaller ZIPs. These findings, aligned with CDC and NVAC guidance, affirmed the alternative hypothesis (H_1), demonstrating that stratification can reveal underserved areas and guide more efficient and equitable public health interventions.

Collectively, this study provides a robust, data-driven foundation for addressing immunization gaps in Oregon. It highlights stark geographic disparities, identifies critical predictors of series completion, and showcases the power of predictive analytics and stratification in guiding public health decisions. The findings not only validate the value of public IIS data but also underscore the urgent need for targeted outreach, especially in smaller and rural ZIP codes. By linking population health data with predictive modeling, this research contributes practical insights toward improving adolescent immunization compliance, advancing HEDIS performance, and fostering health equity across Oregon.

10.0 Recommendations

Based on the findings of this study, the following recommendations are proposed to improve immunization compliance among pediatric and adolescent populations in Oregon and to guide data-informed public health interventions:

1. **Prioritize HPV Vaccine Completion in Outreach Campaigns**
Given the strong predictive relationship between HPV completion and overall Teen Immunization Series adherence, targeted initiatives to promote HPV vaccine follow-through should be prioritized. Strategies may include:
    - Reminder/recall systems for multi-dose vaccines.
    - Provider training to reinforce early and consistent HPV messaging.
    - Community partnerships to dispel misinformation and reduce stigma around the HPV vaccine.

2. **Expand Data Integration and Accessibility at the ZIP-Level**
The lack of individual-level data limits the ability to examine demographic disparities in depth. To improve surveillance and program design:
    - Collaborate with the Oregon Health Authority to expand access to disaggregated data via the ALERT IIS.
    - Encourage local health systems to integrate school-based, pharmacy, and clinic vaccination records for more complete immunization histories.
    - Invest in interoperability between public health databases and EHR systems to reduce underreporting and improve registry completeness.

3. **Deploy Risk Stratification Tools to Target Underserved ZIP Codes**
ZIP codes with small populations (10–49 individuals) showed significantly lower completion rates, particularly for HPV. Public health agencies should:
    - Use predictive risk flags to proactively identify and prioritize under-immunized areas.
    - Tailor interventions for rural and low-density ZIP codes where outreach infrastructure is often limited.
    - Consider mobile clinics, school-based vaccination drives, or evening/weekend hours in high-risk zones.

4. **Address Geographic Disparities Through Place-Based Public Health Strategies**
ANOVA and Kruskal-Wallis results confirmed significant county-level disparities, particularly in flu and COVID-19 vaccination rates. In response:
    - Localize public health messaging and outreach based on ZIP-specific barriers, such as transportation, healthcare access, or language needs.
    - Allocate resources (staffing, funding, media) according to ZIP-level immunization performance and social vulnerability indicators.
    - Encourage counties to adopt equity-focused frameworks that address structural determinants of vaccine access.

5. **Modernize Public Health Reporting and HEDIS Benchmarking Tools**
   While HEDIS provides standardized benchmarks, it does not capture annual vaccines like influenza or COVID-19, which showed the greatest disparities in this study. Recommendations include:
   - Incorporate flu and COVID-19 coverage as supplemental metrics in performance evaluations.
   - Provide flexible quality improvement funding to incentivize local innovation in adolescent immunization outreach.
   - Advocate for NCQA and CMS to evolve HEDIS reporting to reflect pandemic-era public health priorities and vaccine behaviors.

6. **Promote Preventive Care Engagement for Adolescents**
   Routine preventive visits are a known facilitator of vaccine completion. To improve adolescent engagement:
   - Implement EMR-based prompts starting at age 9 for all IMA-related vaccines.
   - Partner with schools to co-schedule well-child visits with vaccine updates.
   - Expand reimbursement for adolescent-focused preventive services, including care coordination and telehealth outreach.

These recommendations are grounded in the study’s evidence and aligned with national public health frameworks, including those from the CDC, NCQA, and NVAC. Implementing these strategies can help close immunization gaps, enhance HEDIS performance, and promote health equity across Oregon’s diverse geographic and demographic landscape.

11.0 References

Aetna Better Health of West Virginia. (n.d.). *Immunizations webinar for providers*. Aetna Better Health. https://www.aetnabetterhealth.com/content/dam/aetna/medicaid/west-virginia/provider/pdf/abhwv_immunizations_webinar.pdf

Centers for Disease Control and Prevention. (2024). *Ensuring vaccine access for all people*. CDC. https://www.cdc.gov/vaccines/basics/vaccine-equity.html

City of Portland Office of Equity and Human Rights. (n.d.). *Race, Ethnicity, Language, Disability, and Tribal Affiliation Demographic Data Standards Guidance*. Portland.gov. https://www.portland.gov/officeofequity/equity-title-vi-division/reldta-demographic-data-standards-guide

Coalition of Communities of Color & Urban League of Portland. (2010). *An unsettling profile: The Executive summary of the report on racial and ethnic disparities in Multnomah County*. https://media.oregonlive.com/portland_impact/other/ReportCardonRacial&EhtnicDisparitiesExecSummary.pdf

Lin, X. M., Zell, E., & Martin, A. (2019). *Using IIS data to estimate national vaccination coverage*. Immunization Services Division, Centers for Disease Control and Prevention. https://repository.immregistries.org/files/resources/5d6694cb31762/using_iis_data_to_estimate_national_vaccination_coverage.pdf

Melotte, S., & Kejriwal, M. (2021). *Predicting zip code-level vaccine hesitancy in US metropolitan areas using machine learning models on public tweets.* arXiv. https://arxiv.org/abs/2108.01699

National Association of Community Health Centers. (2022). *Population health management: Risk stratification action guide*. NACHC. https://www.nachc.org/wp-content/uploads/2022/01/PHM_Risk-Stratification-AG-Jan-2022.pdf

National Committee for Quality Assurance. (n.d.). *Immunizations for adolescents (IMA-E)*. NCQA. https://www.ncqa.org/report-cards/health-plans/state-of-health-care-quality-report/immunizations-for-adolescents-ima-e/

National Vaccine Advisory Committee. (2021). *Advancing immunization equity: Recommendations from the National Vaccine Advisory Committee*. U.S. Department of Health and Human Services. https://www.hhs.gov/sites/default/files/nvac-immunization-equit-report.pdf

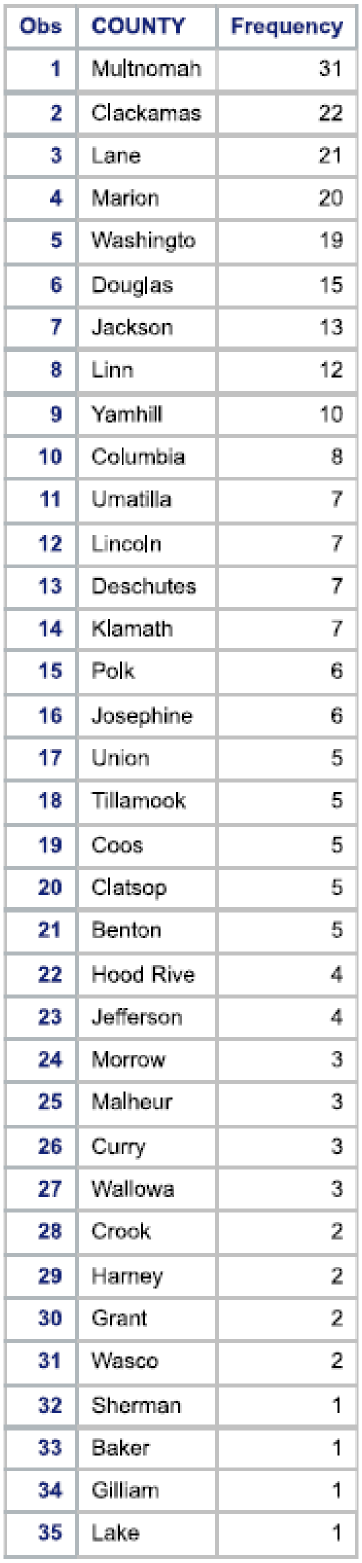
Oregon Immunization Program. (2023). *Oregon adolescent immunizations*. Tableau Public. https://public.tableau.com/app/profile/oregon.immunization.program/viz/OregonAdolescentImmunizations/D-Landing

Pingali, C., Yankey, D., Chen, M., Elam-Evans, L. D., Markowitz, L. E., DeSisto, C. L., Schillie, S. F., Hughes, M., Valier, M. R., Stokley, S., & Singleton, J. A. (2024). *National vaccination coverage among adolescents aged 13–17 years—National Immunization Survey-Teen, United States, 2023*. Centers for Disease Control and Prevention. https://www.cdc.gov/mmwr/volumes/73/wr/mm7333a1.htm

Texas Children's Health Plan. (2025). *HEDIS toolkit: Immunizations for adolescents (IMA)*. Texas Children's Health Plan. https://www.texaschildrenshealthplan.org/sites/default/files/2025-02/PR-2502-116%20HEDIS%20Toolkit_IMA.pdf

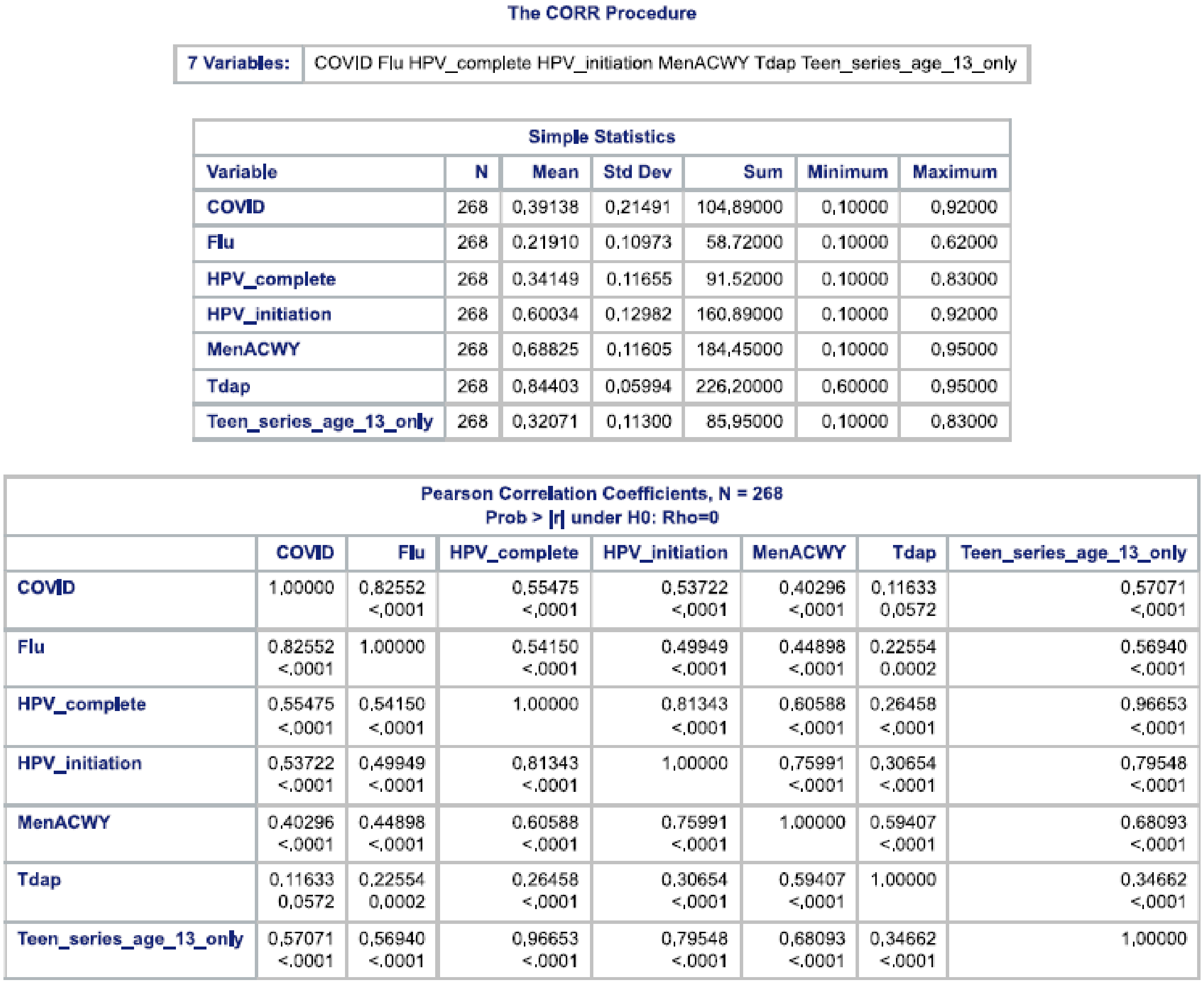
12.0 Appendix A

Table A1. Frequency Distribution of Observations by County



Obs	COUNTY	Frequency
1	Multnomah	31
2	Clackamas	22
3	Lane	21
4	Marion	20
5	Washingto	19
6	Douglas	15
7	Jackson	13
8	Linn	12
9	Yamhill	10
10	Columbia	8
11	Umatilla	7
12	Lincoln	7
13	Deschutes	7
14	Klamath	7
15	Polk	6
16	Josephine	6
17	Union	5
18	Tillamook	5
19	Coos	5
20	Clatsop	5
21	Benton	5
22	Hood Rive	4
23	Jefferson	4
24	Morrow	3
25	Malheur	3
26	Curry	3
27	Wallowa	3
28	Crook	2
29	Harney	2
30	Grant	2
31	Wasco	2
32	Sherman	1
33	Baker	1
34	Gilliam	1
35	Lake	1

Table A2. Pearson Correlation Matrix for Child, Adolescent Vaccine Coverage Rates



The CORR Procedure

7 Variables:	COVID Flu HPV_complete HPV_initiation MenACWY Tdap Teen_series_age_13_only

Simple Statistics

Variable	N	Mean	Std Dev	Sum	Minimum	Maximum
COVID	268	0,39138	0,21491	104,89000	0,10000	0,92000
Flu	268	0,21910	0,10973	58,72000	0,10000	0,62000
HPV_complete	268	0,34149	0,11655	91,52000	0,10000	0,83000
HPV_initiation	268	0,60034	0,12982	160,89000	0,10000	0,92000
MenACWY	268	0,68825	0,11605	184,45000	0,10000	0,95000
Tdap	268	0,84403	0,05994	226,20000	0,60000	0,95000
Teen_series_age_13_only	268	0,32071	0,11300	85,95000	0,10000	0,83000

Pearson Correlation Coefficients, N = 268
Prob > |r| under H0: Rho=0

	COVID	Flu	HPV_complete	HPV_initiation	MenACWY	Tdap	Teen_series_age_13_only
COVID	1,00000	0,82552 <,0001	0,55475 <,0001	0,53722 <,0001	0,40296 <,0001	0,11633 0,0572	0,57071 <,0001
Flu	0,82552 <,0001	1,00000	0,54150 <,0001	0,49949 <,0001	0,44898 <,0001	0,22554 0,0002	0,56940 <,0001
HPV_complete	0,55475 <,0001	0,54150 <,0001	1,00000	0,81343 <,0001	0,60588 <,0001	0,26458 <,0001	0,96653 <,0001
HPV_initiation	0,53722 <,0001	0,49949 <,0001	0,81343 <,0001	1,00000	0,75991 <,0001	0,30654 <,0001	0,79548 <,0001
MenACWY	0,40296 <,0001	0,44898 <,0001	0,60588 <,0001	0,75991 <,0001	1,00000	0,59407 <,0001	0,68093 <,0001
Tdap	0,11633 0,0572	0,22554 0,0002	0,26458 <,0001	0,30654 <,0001	0,59407 <,0001	1,00000	0,34662 <,0001
Teen_series_age_13_only	0,57071 <,0001	0,56940 <,0001	0,96653 <,0001	0,79548 <,0001	0,68093 <,0001	0,34662 <,0001	1,00000

13.0 Appendix B

Fit Statistics			
Number of Trees	Number of Leaves	Average Square Error (Train)	Average Square Error (OOB)
1	15	0.00746	0.01852
2	26	0.00933	0.05316
3	34	0.01216	0.05823
4	45	0.00948	0.04823
5	57	0.00871	0.04735
6	68	0.00739	0.03705
7	78	0.00706	0.03854
8	90	0.00593	0.03321
9	99	0.00668	0.03495
10	107	0.00639	0.03384
11	116	0.00588	0.02959
12	124	0.00565	0.0277
13	140	0.00528	0.02623
14	152	0.0051	0.02637
15	167	0.00512	0.02628
16	179	0.005	0.02584
17	189	0.00488	0.02569
18	195	0.00473	0.02481
19	205	0.00488	0.02559
20	215	0.00484	0.02536
21	224	0.00522	0.02562
22	232	0.00515	0.02514
23	239	0.00529	0.02524
24	247	0.00493	0.02457
25	255	0.00472	0.02378
26	265	0.00454	0.02306
27	275	0.00428	0.02283
28	282	0.00432	0.02327
29	292	0.00438	0.02399
30	299	0.00443	0.02458
31	313	0.00418	0.02382
32	321	0.00426	0.02402
33	328	0.00436	0.02439
34	338	0.00424	0.02413
35	347	0.00433	0.02446
36	358	0.00416	0.02408
37	370	0.0042	0.02426

38	382	0.00412	0.02367
39	395	0.00407	0.02365
40	403	0.0041	0.02357
41	409	0.00419	0.0239
42	421	0.00425	0.02387
43	435	0.00412	0.0233
44	450	0.00409	0.02346
45	465	0.0041	0.02319
46	477	0.00408	0.02349
47	489	0.00403	0.02379
48	499	0.00399	0.02358
49	510	0.00408	0.02399
50	522	0.00413	0.02422
51	531	0.00419	0.02428
52	543	0.00422	0.02467
53	553	0.00421	0.02425
54	562	0.00407	0.02362
55	572	0.00399	0.02325
56	581	0.00398	0.02358
57	589	0.00399	0.02329
58	601	0.00396	0.02334
59	613	0.00395	0.02352
60	621	0.00393	0.02345
61	634	0.00393	0.02316
62	645	0.00397	0.02329
63	657	0.004	0.02328
64	663	0.00403	0.02302
65	674	0.00398	0.02288
66	685	0.00393	0.0227
67	696	0.00388	0.02267
68	704	0.0039	0.02247
69	711	0.00395	0.02222
70	722	0.00397	0.02218
71	730	0.00392	0.02197
72	738	0.00395	0.02211
73	747	0.00397	0.02197
74	756	0.00399	0.022
75	766	0.00399	0.02203
76	777	0.00403	0.02197
77	789	0.00398	0.02177
78	798	0.00397	0.02168

79	810	0.00394	0.02168
80	820	0.00396	0.02179
81	832	0.004	0.02195
82	844	0.00398	0.0221
83	853	0.00395	0.02201
84	863	0.00386	0.0217
85	873	0.00385	0.02181
86	883	0.0038	0.02157
87	890	0.00382	0.02165
88	902	0.00383	0.02169
89	915	0.00376	0.02144
90	925	0.00376	0.02139
91	935	0.00375	0.02131
92	948	0.00379	0.02165
93	959	0.00376	0.02146
94	971	0.00376	0.02166
95	980	0.00375	0.02173
96	989	0.00376	0.02167
97	1004	0.00373	0.02173
98	1014	0.00373	0.02162
99	1023	0.00377	0.0217
100	1030	0.00381	0.0218

14.0 Appendix C

/*SAS CODE USED TO RUN ANALYSIS AND PRODUCE REPORTS*/

```
/* EXPLORATORY DATA ANALYSIS*/
proc contents data=WORK.IMMUNIZATION;
run;

/* omit records where population size less than 10*/
data immunization_filtered;
  set WORK.IMMUNIZATION;
  /* Remove entries with suppressed population size or missing vaccine rates */
  if POPULATION_SIZE ne '<10 (not shown)';
run;

/* Frequency distribution of observations across counties*/
proc sql;
  create table county_freq as
  select COUNTY, count(*) as Frequency
  from immunization_filtered
  group by COUNTY
  order by Frequency desc;
quit;

/* Step 1: Create frequency table */
proc freq data=WORK.IMMUNIZATION_FILTERED;
  tables COUNTY / out=CountyFreq;
run;

/* Step 2: Sort by frequency in descending order */
proc sort data=CountyFreq out=CountyFreqSorted;
  by descending COUNT;
run;

/* Step 3: Keep only the top 10 records */
data Top10CountyFreq;
  set CountyFreqSorted;
  if _N_ <= 10;
run;

/* Step 4: Display results */
proc print data=Top10CountyFreq noobs;
  title "Top 10 Counties by Frequency";
run;

proc sgplot data=Top10CountyFreq;
  hbar COUNTY / response=COUNT stat=sum;
  yaxis discreteorder=data label="County";
  xaxis label="Number of Records";
  title "Top 10 Counties by Frequency";
run;

/* Descriptive statistics using MEANS procedure*/
/* excludes all records with pop_bin = 0*/
/* uses filtered dataset*/
proc means data=immunization_filtered n mean std min max;
   var Tdap MenACWY COVID Flu HPV_initiation HPV_complete Teen_series_age_13_only;
run;

/* Explore Correlations and Trends */
/* This will help spot which vaccinations tend to go hand-in-hand */
proc corr data=immunization_filtered;
var COVID Flu HPV_complete HPV_initiation MenACWY Tdap Teen_series_age_13_only;
run;


/* checking distribution skew out outliers for each variable*/
proc univariate data=immunization_filtered;
  var Tdap MenACWY COVID Flu HPV_initiation HPV_complete Teen_series_age_13_only;
  histogram / normal;
  inset mean std skewness kurtosis;
run;
/*----------------------------------------------------------------------*/
/* One-way ANOVA by county for each vaccine */
/* filter data, work without blank immunization rates */
data immunization_filtered;
  set WORK.IMMUNIZATION;


  /* Remove entries with suppressed population size or missing vaccine rates */
  if POPULATION_SIZE ne '<10 (not shown)';
run;

proc anova data=immunization_filtered;
class COUNTY;
model Tdap=COUNTY;
run;

proc anova data=immunization_filtered;
class COUNTY;
model MenACWY=COUNTY;
run;

proc anova data=immunization_filtered;
class COUNTY;
```

```
model COVID=COUNTY;
run;

proc anova data=immunization_filtered;
class COUNTY;
model Flu=COUNTY;
run;

proc anova data=immunization_filtered;
class COUNTY;
model HPV_initiation=COUNTY;
run;

proc anova data=immunization_filtered;
class COUNTY;
model HPV_complete=COUNTY;
run;

proc anova data=immunization_filtered;
class COUNTY;
model Teen_series_age_13_only=COUNTY;
run;

/*--------------------------------------------------------------------*/
/* Kruskal-Wallis test (non-parametric alternative by county for each vaccine */
/* use to compare the distribution of the vaccines across different counties*/
/* does not assume normality; useful for skewed data or ordinal responses*/

proc npar1way data=immunization_filtered wilcoxon;
class COUNTY;
var Tdap MenACWY COVID Flu HPV_initiation HPV_complete Teen_series_age_13_only;
run;

/*--------------------------------------------------------------------*/

/*--------------------------------------------------------------------*/

/* Analysis for Question 2 */
/* filter data, work without blank immunization rates */
data immunization_filtered;
  set WORK.IMMUNIZATION;
  /* Remove entries with suppressed population size or missing vaccine rates */
  if POPULATION_SIZE ne '<10 (not shown)';
run;

proc univariate data=immunization_filtered;
var Tdap MenACWY COVID Flu HPV_initiation HPV_complete Teen_series_age_13_only;
histogram;
run;
```

```
/* Run Correlation Analysis to test hypothesis */
/* Since the distributions appear to be normal we can use Pearson or Spearman*/
/* this will create a correlation matrix showing how strongly each vaccine is associated with Teen series
completion*/
proc corr data=immunization_filtered pearson plots(maxpoints=none);
   var Tdap MenACWY COVID Flu HPV_initiation HPV_complete Teen_series_age_13_only;
run;

/* Using Multiple Linear Regression to predict the proportion of adolescents completing the*/
/* Teen Immunization Series by age 13,*/
/* using update rates of individual vaccines as predictors*/
proc reg data=immunization_filtered;
   model Teen_series_age_13_only = Tdap MenACWY COVID Flu HPV_initiation HPV_complete;
run;
quit;

/* Scatter plot matrix to visualize all vaccine uptake variables*/
proc sgscatter data=immunization_filtered;
   matrix Teen_series_age_13_only Tdap MenACWY COVID Flu HPV_initiation HPV_complete /
   diagonal=(histogram kernel);
run;

/*----------------------------------------------------------------------*/

/*----------------------------------------------------------------------*/
```

/* Analysis for Question 3 */

```
data immunization_filtered;
  set WORK.IMMUNIZATION;
  /* Remove entries with suppressed population size or missing vaccine rates */
  if POPULATION_SIZE ne '<10 (not shown)' and Tdap ne .;
run;

/*creating immunization model*/
data immunization_model;
  set immunization_filtered;


/* Create a composite average of selected vaccines */
  avg_coverage = mean(of Tdap, MenACWY, COVID, Flu, HPV_initiation, HPV_complete);


/* Define 'risk_flag' based on a threshold — adjust as needed */
  if avg_coverage < 0.6 then risk_flag = 1;  /* High risk */
  else risk_flag = 0;                   /* Not high risk */
run;
```

```
/*using Random Forest because Logistic regression produced errors*/
/*random forest handles imperfect data, mixed variable types, and does*/
/*not rely on linear relationships*/
proc hpforest data=immunization_model;
  target risk_flag;
  input Tdap MenACWY COVID Flu HPV_initiation HPV_complete;
  id ZIP_CODE;
run;
/*--------------------------------------------------------------------*/
/* Analysis for Question 4 */
/*create outcome variable*/
data immunization_clean;
   set immunization_filtered;
   if pop_bin ne '0'; /* Exclude very small ZIPs */
   noncompliant_rate = 1 - Teen_series_age_13_only;
run;


/*run linear regression*/
proc glm data=immunization_clean;
   class county pop_bin;
   model noncompliant_rate =
        Tdap MenACWY COVID Flu HPV_initiation HPV_complete
        pop_bin county;
run;

/*--------------------------------------------------------------------*/

/*--------------------------------------------------------------------*/

/* Analysis for Question 5*/
/*filter dataset, work with only records with vaccine rates*/
data immunization_filtered;
  set WORK.IMMUNIZATION;
  /* Remove entries with suppressed population size or missing vaccine rates */
  if POPULATION_SIZE ne '<10 (not shown)' and Tdap ne .;
run;

/*run t-test to compare average valeus between ZIPs where stratification tools are assummed used vs
not*/
proc ttest data=immunization_filtered;
  class strat_used;
  var HPV_complete Teen_series_age_13_only;
run;
```