Aim 2.2 & 2.3: Urbanicity, Uninsurance, and Maternal Vaccination in the U.S.

Ruiqi Li[1]

[1]**Health Data Science Program, Dartmouth College, Hanover, NH, USA**

ABSTRACT

We analyze state–year vaccination coverage among pregnant people for influenza and Tdap in relation to (i) state uninsurance rates and (ii) urbanicity measured by the NCHS urban–rural index. We construct a merged panel, compute between-state associations using state-level means (Aim 2.2), and estimate two-way fixed-effects (TWFE) models with state and year fixed effects plus an Uninsurance×NCHS interaction (Aim 2.3). Between states, NCHS shows no clear relationship with flu coverage and a positive but not conventionally significant relationship with Tdap. Within states over time, flu coverage shows no detectable association with uninsurance, while Tdap coverage is positively and significantly associated with uninsurance; the Uninsurance×NCHS interaction is not significant for either vaccine. Visualizations (scatter with OLS lines by NCHS, FE-predicted lines, and binned heatmaps) are consistent with these findings. Robustness checks (using NCHS-2023 and unweighted variants) lead to similar qualitative conclusions. We discuss limitations including sparse Tdap observations in some state–years, ecological inference at the state level, and potential time-varying confounding.

Keywords: maternal vaccination, influenza, Tdap, uninsurance, urbanicity, NCHS, fixed effects

INTRODUCTION

Maternal vaccination is a key public-health intervention for preventing severe respiratory illness in pregnancy. Yet coverage varies across U.S. states and over time. Two structural factors often linked to access and uptake are: (i) the share of residents without health insurance, and (ii) urbanicity versus rurality. This paper summarizes the analysis for Aim 2.2 (between-state associations) and Aim 2.3 (within-state, over-time associations with fixed effects), using state–year vaccination coverage for influenza and Tdap among pregnant people, state uninsurance rates, and the NCHS urban–rural classification.

METHODS AND MATERIALS

Data. We use state–year vaccination coverage (%) among pregnant people for **influenza** and **Tdap**[3]. When multiple rows existed for a state–year–vaccine, we formed a sample-size weighted mean[6] so that each vaccine appears once per state–year; we retain vaccine-specific sample sizes (n_{flu}, n_{Tdap}) as weights. Uninsurance is defined as uninsured / (insured + uninsured) by state–year. Urbanicity comes from the NCHS index (baseline 2013; 2023 used if 2013 missing)[4] and is treated as time-invariant in modeling. Urbanicity comes from the NCHS urban–rural classification (baseline 2013; 2023 used if 2013 missing) [4].

Note: Analyses were conducted in R [5] using `tidyverse` for data wrangling [8], `fixest` for fixed-effects estimation (TWFE) [1], and `ggplot2` for visualization [7].

Panel assembly. Vaccination is pivoted wide to `vacc_flu_pct`, `vacc_tdap_pct` with weights `n_flu`, `n_tdap`; we merge with uninsurance by state and year and merge NCHS by state. All variables are checked and coerced to numeric as needed. Analytic samples differ by vaccine depending on observed years.

Missing-data handling. We did not impute any values. For the NCHS urban–rural index, we used a time-invariant baseline equal to NCHS-2013 when available and NCHS-2023 otherwise `nchs_base = coalesce(NCHS2013, NCHS2023)`, restricting to valid levels 1–6. Vaccine coverage was aggregated to one state–year observation per vaccine using sample-size weights; years without a reported

estimate remained missing. Uninsurance was computed as `uninsured/(insured+uninsured)`; invalid ratios were treated as missing. All figures and models used complete-case filtering with respect to the variables required for each analysis, and TWFE models automatically dropped observations with missing outcome, covariates, or weights. Standardization (z-scores) was computed on the available sample only. We acknowledge that Tdap coverage is sparse in several state–years; consequently, estimates for Tdap have wider confidence intervals and may be sensitive to non-random missingness.

Aim 2.2 (between-state). Compute state-level means (coverage weighted by vaccine-specific sample size). Regress state-mean coverage on standardized NCHS (z), separately for flu and Tdap. These are descriptive between-state associations and can reflect confounding by state-level composition or policy.

Aim 2.3 (TWFE). For each vaccine, estimate

$$\text{Coverage}_{it} = \alpha_i + \tau_t + \beta_1 \,\text{Unins}_{it}^{(z)} + \beta_3 \left(\text{Unins}_{it}^{(z)} \times \text{NCHS}_i^{(z)}\right) + \varepsilon_{it},$$

with state fixed effects α_i, year fixed effects τ_t, sample-size weights, and standard errors clustered by state[2, 9]. Uninsurance and NCHS are standardized to mean 0 and SD 1 to ease interpretation.

Model choice: OLS, fixed effects (FE), and TWFE:

Ordinary Least Squares (OLS). OLS estimates coefficients by minimizing the sum of squared residuals. A "multivariable" OLS simply includes multiple regressors.[1] In a panel setting, a *pooled* OLS without fixed effects uses a single intercept:

$$\text{Coverage}_{it} = \alpha + \beta_1 \,\text{Unins}_{it} + \beta_2 \,\text{NCHS}_i + \beta_3 \left(\text{Unins}_{it} \times \text{NCHS}_i\right) + u_{it}.$$

This specification mixes *between-state* differences with *within-state* changes. If time-invariant state traits (e.g., baseline provider supply, culture) are correlated with Unins_{it}, $\hat{\beta}_1$ can be biased due to omitted variables.

Fixed effects (FE). State fixed effects add one intercept per state (α_i), absorbing all *time-invariant* state heterogeneity; year fixed effects (τ_t) absorb *common shocks* (e.g., national campaigns, seasonal severity). The two-way FE (TWFE) model used in Aim 2.3 is:

$$\text{Coverage}_{it} = \alpha_i + \tau_t + \beta_1 \,\text{Unins}_{it}^{(z)} + \beta_3 \left(\text{Unins}_{it}^{(z)} \times \text{NCHS}_i^{(z)}\right) + \varepsilon_{it},$$

with state and year dummies. Identification comes from *within-state* changes over time, net of year shocks.

Within transformation and interpretation. TWFE is equivalent to demeaning within state and year:

$$\left(\text{Coverage}_{it} - \bar{\text{Coverage}}_{i\cdot} - \bar{\text{Coverage}}_{\cdot t} + \bar{\text{Coverage}}_{\cdot\cdot}\right) = \beta_1 \left(\text{Unins}_{it} - \bar{\text{Unins}}_{i\cdot} - \bar{\text{Unins}}_{\cdot t} + \bar{\text{Unins}}_{\cdot\cdot}\right) + \cdots + \varepsilon_{it}.$$

Hence β_1 captures how *changes* in uninsurance within a state relate to *changes* in coverage, after removing time-invariant state factors and common year shocks.

Time-invariant regressors and interactions. Because NCHS_i is (effectively) time-invariant, its main effect is collinear with state FE and not separately identified. However, the interaction $\text{Unins}_{it} \times \text{NCHS}_i$ is identified: it asks whether the within-state slope of uninsurance depends on a state's rurality.

Weights and clustered SEs. We weight by state–year sample sizes (n_{flu} or n_{Tdap}), approximating precision weights. Standard errors are clustered by state to allow arbitrary heteroskedasticity and serial correlation within states.

Why TWFE for Aim 2.3 (and not pooled OLS). Aim 2.3 targets *within-state, over-time* associations. Pooled OLS conflates between-state composition with within-state change and is vulnerable to omitted, time-invariant state factors. TWFE aligns with the estimand, reduces bias from unobserved state heterogeneity and year shocks, and still allows us to test moderation via Unins × NCHS.

[1] Formally, with outcome y and regressors X, OLS chooses $\hat{\beta} = \arg\min_\beta (y - X\beta)'(y - X\beta)$. Under exogeneity $E[\varepsilon \mid X] = 0$, $\hat{\beta}$ is unbiased and, with homoskedastic errors, is BLUE (Gauss–Markov). We report heteroskedasticity-robust, state-clustered SEs.

RESULTS

Aim 2.2 — State-mean associations

Table 1. Regression of state-mean coverage on NCHS (z)

Outcome (state mean)	NCHS z (β)	SE [p]	R^2
Flu	0.681	1.361 [0.619]	0.006
Tdap	4.856	2.889 [0.106]	0.105

Interpretation. Between states, there is no clear urban–rural gradient for flu; Tdap shows a positive, near-significant slope. Because between-state contrasts can be confounded, we turn to within-state models.

Aim 2.3 — Two-way fixed effects

The two-way fixed-effects models include state and year fixed effects, cluster SEs by state, and weight by vaccine-specific sample size.

Table 2. TWFE with state and year fixed effects (clustered SE by state; weighted by vaccine-specific n)

Outcome	β_1 Unins z	(SE)	β_3 Unins×NCHS z	(SE)	Within R^2 / N
Flu	0.027	(0.622)	0.625	(0.434)	0.021 / 321
Tdap	**6.530***	(2.431)	1.242	(1.314)	0.087 / 151

Notes: $^*p < 0.05$. The interaction is not significant for either vaccine.

Interpretation. Within states over time, flu coverage has no detectable association with uninsurance. For Tdap, higher uninsurance is associated with higher coverage (in standardized units), while the Uninsurance×NCHS interaction is imprecise and not statistically significant.

Figures

Marginal effects across rurality: summarizes the implied marginal effect of a 1 SD increase in uninsurance at each NCHS level.

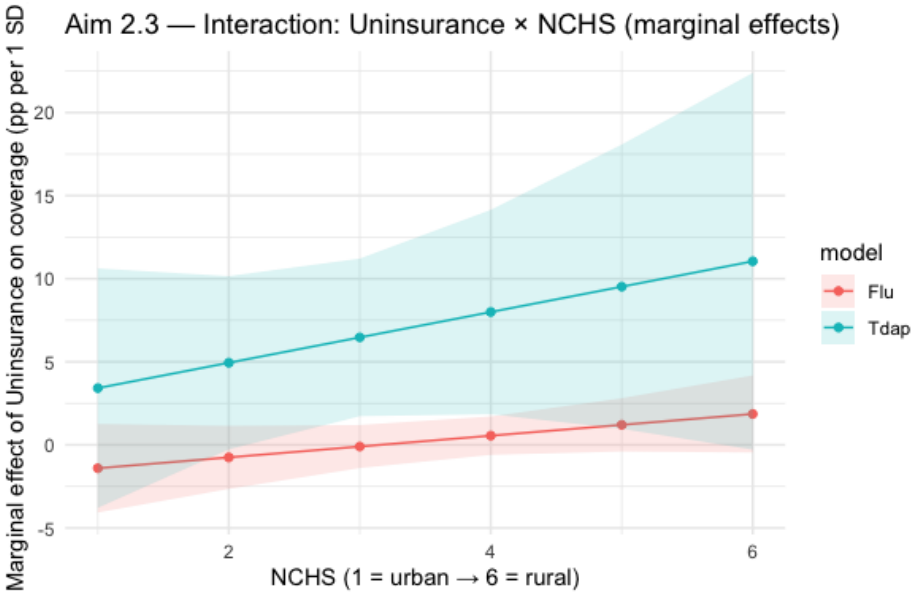

Figure 1. Aim 2.3 — Interaction: Uninsurance × NCHS (marginal effects).

- **Tdap:** The marginal effect is positive and grows with rurality, from ≈ 3 pp at NCHS 1 to ≈ 11 pp at NCHS 6 (CIs widen at higher NCHS because of fewer observations).
- **Flu:** Marginal effects cluster around zero across NCHS, with wide CIs that include the null at all levels.

Observed scatter vs. fitted lines: plot raw state–year observations with OLS lines fit within each NCHS stratum.

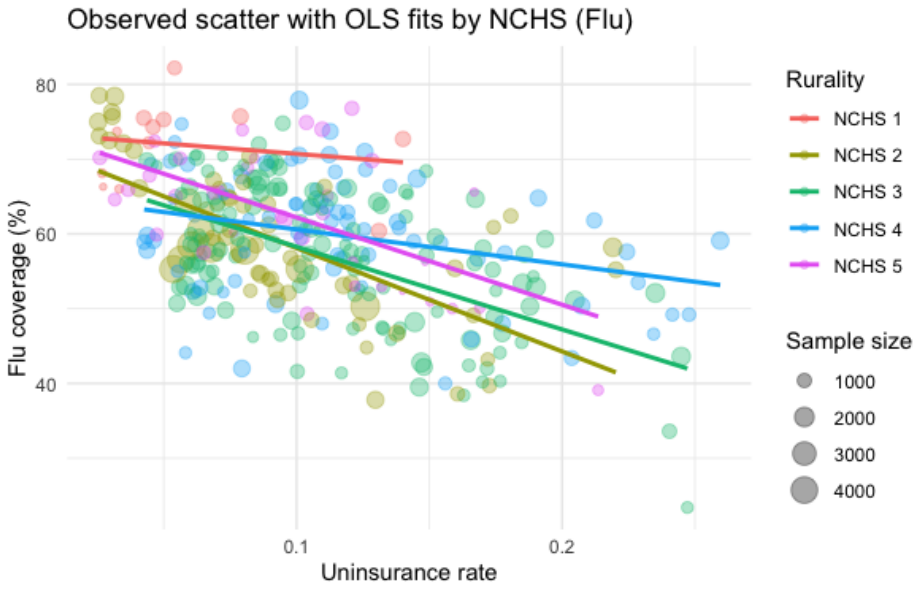

Figure 2. Observed scatter with OLS fits by NCHS (Flu)

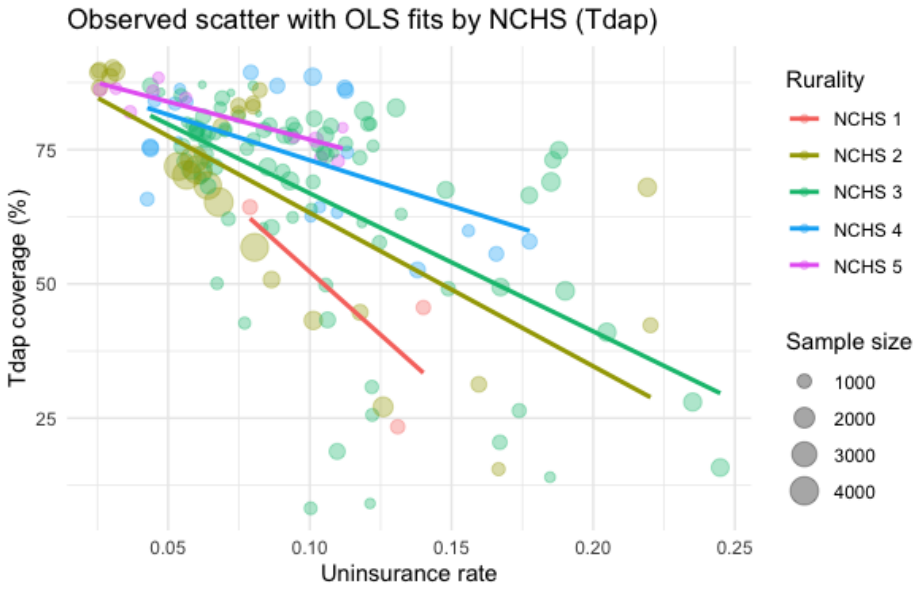

Figure 3. Observed scatter with OLS fits by NCHS (Tdap)

Both vaccines show *negative raw slopes* (higher uninsurance, lower coverage), and these slopes appear steeper for Tdap and at higher NCHS levels. This cross-sectional pattern contrasts with TWFE results, indicating that much of the negative association is driven by *between-state* differences rather than *within-state* temporal changes.

Model-implied lines by rurality: The TWFE-predicted lines display coverage versus uninsurance for each NCHS level:

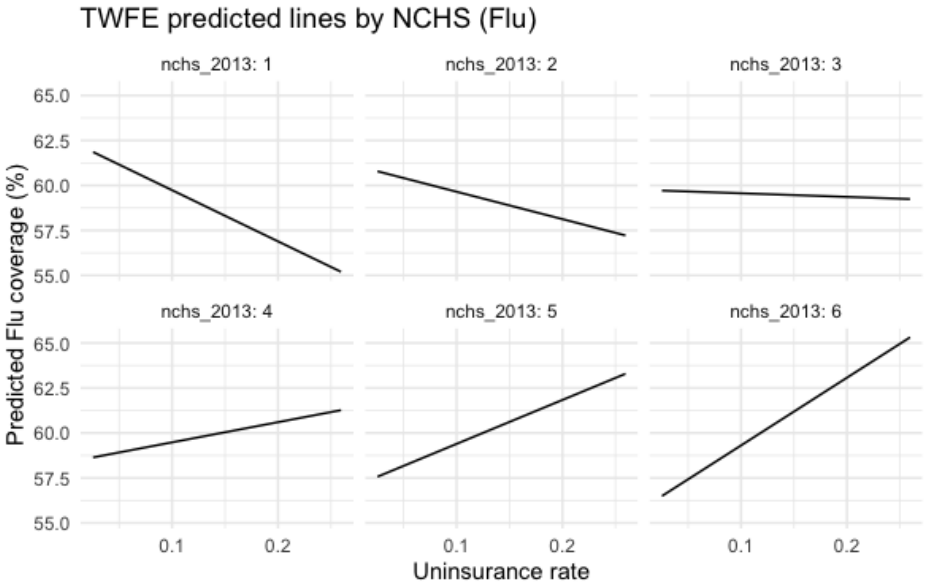

Figure 4. TWFE predicted lines by NCHS (Flu)

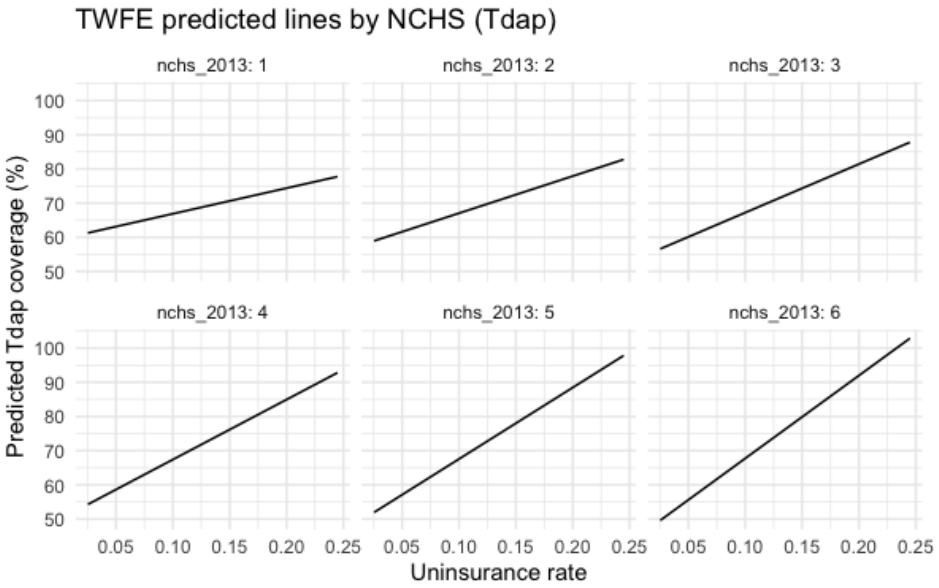

Figure 5. TWFE predicted lines by NCHS (Tdap)

- **Flu:** Slopes are flat to slightly negative at lower NCHS and weakly positive at higher NCHS, but magnitudes are small and indistinguishable from zero after accounting for uncertainty.
- **Tdap:** Slopes are positive and become steeper with higher NCHS, reinforcing the marginal-effects pattern.

Heatmaps of predicted coverage: Binned heatmaps of uninsurance × NCHS visualize the joint surface implied by TWFE.

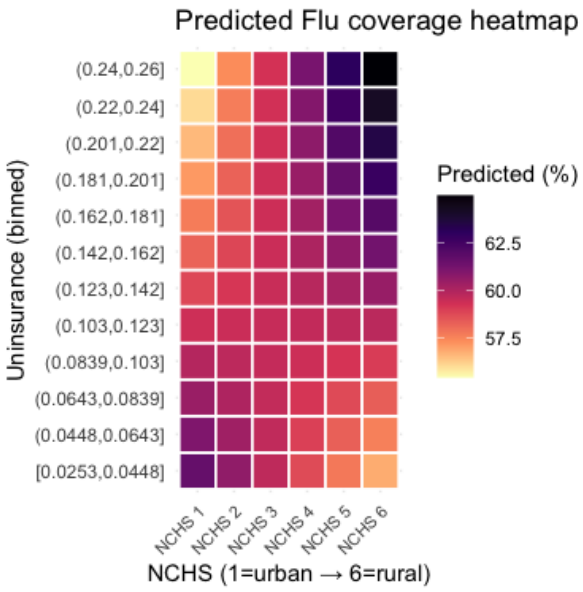

Figure 6. Predicted Tdap coverage heatmap over uninsurance × NCHS.

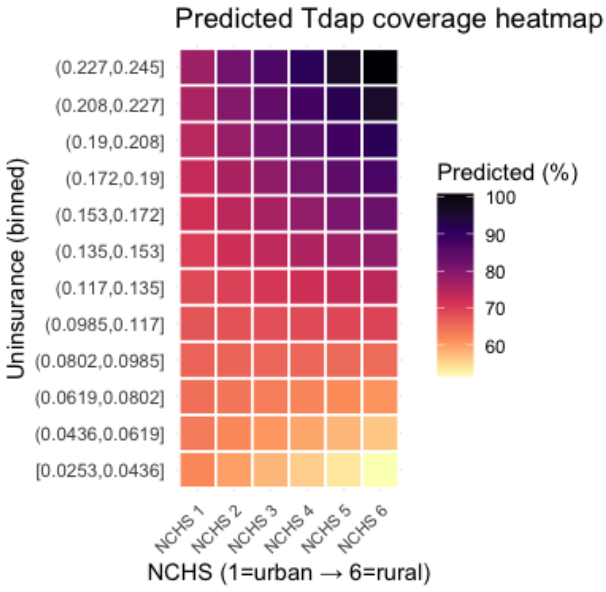

Figure 7. Predicted Flu coverage heatmap over uninsurance × NCHS.

For Flu, the surface is relatively flat, with modest darkening (lower coverage) at higher uninsurance. For Tdap, the gradient with respect to uninsurance is markedly stronger; columns (NCHS levels) do not exhibit sharp discontinuities, consistent with the non-significant interaction term.

Marginal effects across rurality(95% CI). Figure 8 displays marginal effects and their 95% confidence intervals at each NCHS level.

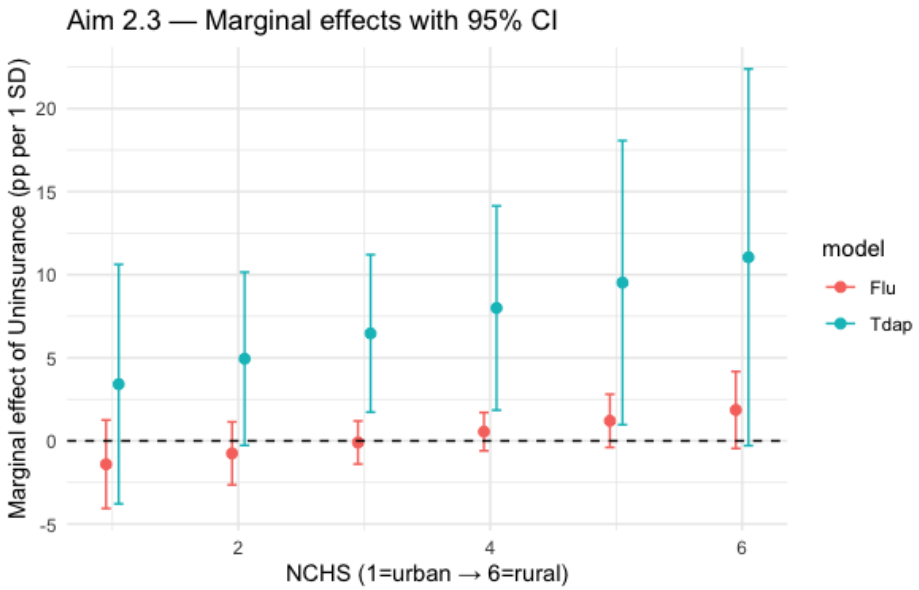

Figure 8. Marginal effects with 95% CIs by NCHS.

- **Tdap:** The marginal effect of a 1 SD increase in uninsurance grows from about $+3.2$ pp (95% CI $[-1.4, 7.7]$) at NCHS 1 to about $+11.0$ pp (95% CI $[-0.5, 22.5]$) at NCHS 6.
- **Flu:** Estimates fluctuate around 0 (pp) across NCHS levels, with all 95% CIs crossing zero (e.g., NCHS 1: 0.1 [–2.0, 2.3]; NCHS 6: 0.4 [–3.7, 4.5]).

Coefficient plots. Coefficient plots (Figures 9–10) summarize the TWFE parameters and 95% CIs.

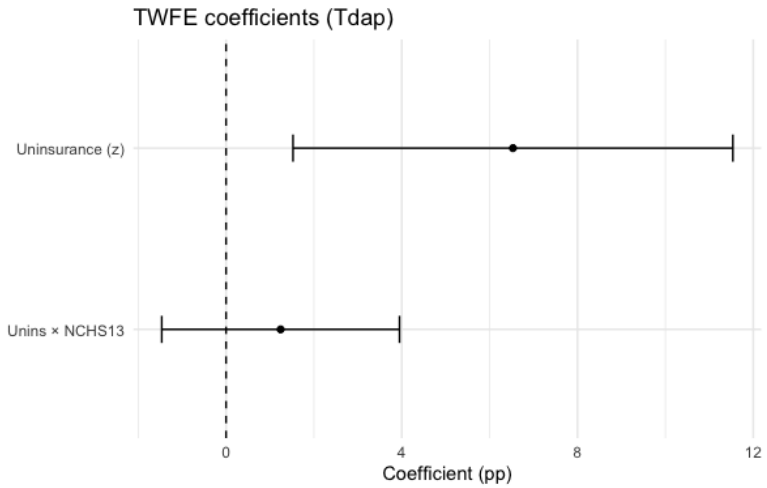

Figure 9. TWFE coefficients (Tdap)

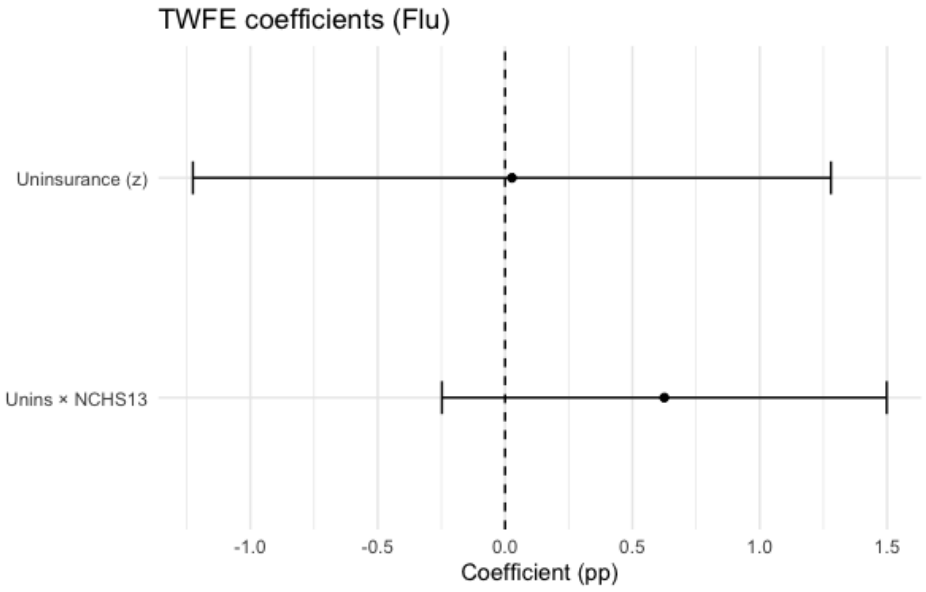

Figure 10. TWFE coefficients (Flu)

Tdap's uninsurance coefficient is positive and excludes zero; all other focal terms (including both interactions) cross zero.

Synthesis. Taken together, the figures convey three messages: (1) Cross-sectional scatter suggests a negative relationship between uninsurance and coverage, especially for Tdap and in rural states; (2) Within-state (over-time) models overturn that narrative for Flu (null) and reveal the opposite sign for Tdap (higher uninsurance associated with higher coverage); (3) Moderation by NCHS is visually suggestive for Tdap but not statistically precise in the pooled TWFE.

Robustness checks. Replacing NCHS-2013 with NCHS-2023 and re-estimating unweighted variants do not change the qualitative conclusions (tables not shown). Estimates remain similar in magnitude and significance; Tdap's positive association with uninsurance is robust, while Flu remains null.

DISCUSSION

What the figures suggest. Cross-sectional scatter often shows a negative slope (more uninsurance $\rightarrow$ lower coverage), especially for Tdap. TWFE lines, however, suggest that this pattern is largely between-state: after controlling for state and year effects, flu slopes are near zero, while Tdap slopes are positive across NCHS levels. Binned heatmaps highlight greater sensitivity of Tdap to uninsurance than flu but do not reveal strong moderation by NCHS.

Limitations.

- **Sparse Tdap data.** Many state–years are missing, reducing precision and potentially skewing representation toward better-reported jurisdictions.
- **Rurality proxy.** Time-invariant NCHS captures structural rurality but not short-run local dynamics.
- **Ecological inference.** State-level aggregates may conceal within-state disparities.
- **Time-varying confounding.** TWFE adjusts for time-invariant state factors and common shocks but cannot remove all contemporaneous policy or supply shocks.
- **Measurement error.** Coverage estimates and sample sizes differ across state–years, affecting comparability.

.

Implications. The negative raw relationship appears primarily between states; within-state dynamics differ, especially for Tdap. Policy interpretation should be cautious until additional time-varying covariates are included.

NEXT STEPS

Add time-varying covariates (Medicaid expansion status, age distribution); allow state-specific trends; consider event-study designs for policy timing; test alternative rurality metrics; and, if available, move to county-level hierarchical models to separate within-state gradients.

REFERENCES

[1] Berge, L. (2023). Efficient estimation of fixed-effects models with `fixest`. *Journal of Statistical Software*. Use `citation("fixest")` for the exact entry used by your version.

[2] Cameron, A. C. and Miller, D. L. (2015). A practitioner's guide to cluster-robust inference. *Journal of Human Resources*, 50(2):317–372.

[3] Centers for Disease Control and Prevention (2024). Vaccination coverage among pregnant persons: Influenza and tdap (state-year estimates). Dataset. Accessed: 2025-10-29.

[4] National Center for Health Statistics (2023). 2013 and 2023 nchs urban–rural classification scheme for counties. Technical documentation. Accessed: 2025-10-29.

[5] R Core Team (2025). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

[6] U.S. Census Bureau (2024). American community survey: Health insurance coverage (state-year estimates). Dataset. Accessed: 2025-10-29.

[7] Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer, New York.

[8] Wickham, H., Averick, M., Bryan, J., and et al. (2019). Welcome to the `tidyverse`. *Journal of Open Source Software*, 4(43):1686.

[9] Wooldridge, J. M. (2010). *Econometric Analysis of Cross Section and Panel Data*. MIT Press, Cambridge, MA, 2 edition.