

## Supplementary Appendix

This appendix has been provided by the authors to give readers additional information about their work.

Supplement to: Daniel W, Nivet M, Warner J, Podolsky DK. Early evidence of the effect of SARS-CoV-2 vaccine at one medical center. *N Engl J Med*. DOI: [10.1056/NEJMc2102153](https://doi.org/10.1056/NEJMc2102153)

## Table of Contents

Forecasting of COVID-19 Cases among UTSW Employees Sans Vaccine (Figure 1B)	1
Count of Total Employees in Self Isolation/Quarantine	2
Vaccine Eligible UTSW Employees Stratified by COVID-19 Vaccination Status (N=23,234)	3

## Forecasting of COVID-19 Cases among UTSW Employees Sans Vaccine (Figure 1B)

### ARIMA Model

The Automatic Regressive Integrated Moving Average (ARIMA) model<sup>1</sup> was used to forecast the number of positive COVID-19 cases among UTSW Employees without the vaccination. On 12/15/2020, UTSW started vaccinating employees. The forecasting is based on the number of positive cases from 10/01/2020 to 12/14/2020.

The ARIMA (p, d, q) consists of three components – AR (p) (Autoregression), I (d) (Integrated), and MA (q) (Moving Average):

AR (p) models the current value  $Y_t$  linearly depends on its lagged values  $Y_{t-1}, Y_{t-2}, \dots, Y_{t-p}$  and the current residuals  $\varepsilon_t$ , as shown in Eq. (1).

$$Y_t = \alpha + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \dots + \beta_p Y_{t-p} + \varepsilon_t \quad (1)$$

MA (q) models the current value  $Y_t$  linearly depends on the lag residuals  $\varepsilon_{t-1}, \varepsilon_{t-2}, \dots, \varepsilon_{t-q}$ , which were generated from the autoregressive models of the respective lags, as shown in Eq. (2).

$$Y_t = \alpha + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q} + \varepsilon_t \quad (2)$$

I (d) transformed the non-stationary time series into stationary by taking the difference process d times. Thus, the ARIMA model is the combination of AR and MA after the time series was differenced and stationary in Eq. (3):

$$Y_t = \alpha + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \dots + \beta_p Y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q} \quad (3)$$

Partial auto correlation (PACF) and auto correlation (ACF) were used to determine the parameters p and q. The Akaike Information Criteria (AIC) was further calculate to obtain a combination of parameters that have a better model performance, i.e., the lower the AIC, the better the performance. In this study, ARIMA (0, 1, 4) showed the best performance.

### Description of Figure 1B

The upward forecasting line in Figure 1B showed that COVID-19 infections would trend upward without vaccination. Yet, after vaccination began on December 15, 2020, the actual number of COVID-19 infections started to deviate from the projected trend, going downward.

### Reference

1. Ceylan Z (2020) Estimation of COVID-19 prevalence in Italy, Spain, and France. Science of The Total Environment, 729, 138817.



Weekly count of UTSW employees (N = 23,234) in self-isolation (SI) or self-quarantine (SQ), 11/30/2020 – 2/28/2021. SI refers to a 10-day isolation period starting from the employee’s first positive COVID-19 test. The SI time frame will be extended to 20 days if the employee is already immunocompromised from prior diagnosis. SQ refers to a 10-day home quarantine if the employee is exposed to a COVID-19 positive individual. People in SQ will switch to SI if they have a positive COVID-19 test.

**Supplemental Table 1: Vaccine Eligible UTSW Employees Stratified by COVID-19 Vaccination Status (N=23,234)**

Demographic Factor		COVID-19 Vaccination Status					
		Fully Vaccinated		Not Vaccinated		Partially Vaccinated	
		n (%)		n (%)			
<b>Age Group (years)</b>	<=24	437	(5.4%)	609	(6.8%)	736	(12.0%)
	>24- 44	4330	(53.3%)	5314	(59.2%)	3244	(52.8%)
	>44-64	2990	(36.8%)	2550	(28.4%)	1903	(31.0%)
	>64	334	(4.1%)	258	(2.9%)	238	(3.9%)
	Unknown	30	(0.4%)	238	(2.7%)	23	(0.4%)
<b>Gender</b>	Female	5181	(63.8%)	6282	(70.0%)	3891	(63.3%)
	Male	2924	(36.0%)	2558	(28.5%)	2232	(36.3%)
	Unknown	16	(0.2%)	129	(1.4%)	21	(0.3%)
<b>Race</b>	American Indian or Alaska Native	42	(0.5%)	41	(0.5%)	22	(0.4%)
	Asian	1223	(15.1%)	750	(8.4%)	1066	(17.4%)
	Black or African American	531	(6.5%)	1661	(18.5%)	616	(10.0%)
	Native Hawaiian/ Pacific Islander	14	(0.2%)	16	(0.2%)	10	(0.2%)
	White	3362	(41.4%)	2516	(28.1%)	2269	(36.9%)
	Multi-race	23	(0.3%)	23	(0.3%)	25	(0.4%)
	Some other race	277	(3.4%)	202	(2.3%)	220	(3.6%)
	Declined	52	(0.6%)	38	(0.4%)	33	(0.5%)
Unknown Race	2597	(32.0%)	3722	(41.5%)	1883	(30.6%)	
<b>Ethnicity</b>	Hispanic or Latino	761	(9.4%)	855	(9.5%)	692	(11.3%)
	Non-Hispanic/Latino	4845	(59.7%)	4556	(50.8%)	3608	(58.7%)
	Unknown	2515	(31.0%)	3558	(39.7%)	1844	(30.0%)
<b>Area Deprivation Index (ADI)</b>	1-3	4645	(57.2%)	2128	(23.7%)	2897	(47.2%)
	4-7	1794	(22.1%)	1401	(15.6%)	1914	(31.2%)
	8-10	286	(3.5%)	401	(4.5%)	257	(4.2%)
	Other	123	(1.5%)	84	(0.9%)	147	(2.4%)
	Unknown	1273	(15.7%)	4955	(55.2%)	929	(15.1%)
<b>Education</b>	HS or less	409	(5.0%)	1137	(12.7%)	404	(6.6%)
	Associate	1266	(15.6%)	2126	(23.7%)	946	(15.4%)
	Bachelor	2254	(27.8%)	1668	(18.6%)	1698	(27.6%)
	Doctorate & Postdoc	2249	(27.7%)	1750	(19.5%)	1071	(17.4%)
	Master	1337	(16.5%)	1151	(12.8%)	1003	(16.3%)
	Unknown/ Not Indicated	606	(7.5%)	1137	(12.7%)	1022	(16.6%)
<b>Language</b>	English	6401	(78.8%)	6022	(67.1%)	4828	(78.6%)
	Non-English	105	(1.3%)	102	(1.1%)	109	(1.8%)
	Unknown/ Not Available	1615	(19.9%)	2845	(31.7%)	1207	(19.6%)
<b>Employment Duration (years)</b>	<=3	3557	(43.8%)	4985	(55.6%)	2651	(43.1%)
	>3-6	1653	(20.4%)	1582	(17.6%)	1224	(19.9%)
	>6-9	892	(11.0%)	779	(8.7%)	619	(10.1%)
	>9	1901	(23.4%)	1327	(14.8%)	1217	(19.8%)
	Unknown	118	(1.5%)	296	(3.3%)	433	(7.0%)
<b>Telework Status</b>	Blended Schedule	1267	(15.6%)	883	(9.8%)	1357	(22.1%)
	Remote	772	(9.5%)	1755	(19.6%)	1902	(31.0%)
	On Campus	5585	(68.8%)	4739	(52.8%)	2053	(33.4%)
	Unknown/ Unmapped	497	(6.1%)	1592	(17.8%)	832	(13.5%)
<b>Total</b>		<b>8121</b>		<b>8969</b>		<b>6144</b>	