**Supplemental Methods Section of the paper “System Factors to Explain 2009 Pandemic H1N1 State Vaccination Rates for Children and High-Risk Adults in US Emergency Response to Pandemic”**

This document complements the Methods section (Analysis sub-section) of the paper “System Factors to Explain 2009 Pandemic H1N1 State Vaccination Rates for Children and High-Risk Adults in US Emergency Response to Pandemic”, and the analysis presented by Davila-Payan et al. [[1](#_ENREF_1)] related with a similar analysis performed on adults.

**Data: High risk population**

For one of our measured outcomes, the state-specific 2009 H1N1 estimation for the percentage of high-risk adults vaccinated, calculated by the CDC, participants on BRFSS and NHFS surveys administered during November 2009-February 2010 were asked if they (or their children) had received an H1N1 vaccine during October 2009-January 2010. High-risk conditions in the surveys include anemia (including sickle cell), asthma, diabetes, coronary heart disease, kidney problems, lung problems other than asthma, myocardial infarction, and a weakened immune system caused by illness or medicines [[2](#_ENREF_2), [3](#_ENREF_3)].

**Data: Shipping data, Simultaneous Tracking Project, and definition of Medically Underserved**

*Shipping data*: From the detailed shipping information we calculated the average number of shipments per location (the total number of shipments divided by the total number of ship-to-sites per state). Performing targeted queries, we also categorized shipments by type of provider, showing that vaccine went to pediatricians or children’s clinics, primary care facilities, county health departments, internists, specialists (not including specialists for children or women), long-term care, veterans, urgent care, hospitals, clinics, pharmacies, and other venues that could be more closely related to the vaccination of children and high-risk adults. We also combined some of these categories in subgroupings to see which had a greater impact on these populations. For example: the targeted access group included doses sent to long term care facilities, internists, specialists, nursing homes, and children; the general access group include doses sent to primary care, MDs, counties, hospitals, urgent care, clinics, or pharmacies. Information was adequate to categorize more than 75% of the overall shipments.

*Simultaneous Tracking Project*: When analyzing data from the Simultaneous Tracking Project led by Clark [[4](#_ENREF_4)], we did not include in the analysis those variables that had a considerable number of missing data points, given that missingness can deeply affect the adequacy of out models.

*Definition of Medically Underserved*: is the as the proportion of population living in primary care health professional shortage areas [[5](#_ENREF_5)].

**Analysis**

Given that we considered the proportion of population vaccinated in each state, and not the number of people vaccinated, the primary technique used for modeling was multivariate linear regression (ordinary least squares) with some logarithmic variable transformations specified when used. In particular, we used a logarithmic transformation of the vaccination rate for children, to better approximate normality.

**Logarithmic transformation and 0-1 scaling of data**

Coverage of 2009 H1N1 vaccination was measured as the percentage (0%-100%) of population who reported having been vaccinated. The outcome for high-risk adults was used as the dependent variable for the linear regression model built for that sub-group. The outcome for children was first transformed by calculating the natural logarithm of the coverage percentages. For example, a 30% coverage for the high-risk population, would have been assigned the value 30 for modeling, while a 30% coverage in the children model was transformed to ln(30) = 3.401 for modeling.

To favor computational stability, all variables, both dependent and independent, were linearly scaled to values between 0 and 1, before calculating the coefficients of the regression model. The minimum value of each of the variables was set to 0, while the maximum value was set to 1. All other values were scaled accordingly. For example, the coverage for high risk adults ranged from 10.4% to 47.2%. We will refer to the difference between the maximum and the minimum values in a variable as the range of the variable, in this case: 47.2 – 10.4 = 36.8 is the range of the high-risk adult’s coverage. Then, when scaling, 10.4 becomes 0, and 47.2 becomes 1, and all other points in this variable take values according to their difference with the minimum value in proportion to the range. In this case, a coverage of 30% would become after scaling: (30 – 10.4)/(47.2 – 10.4) = 0.53. Following the same example, the values for the coverage in children ranged from 21.3% to 84.7% before the logarithmic transformation. Then, the transformed values range from ln(21.3) to ln(84.7). This is, from 3.059 to 4.439. Scaling these values between 0 and 1, the minimum 3.059 becomes 0, while the maximum 4.439 becomes 1. In this case, a coverage of 30% for children would be transformed to ln(30) = 3.401, and this value would become (3.401 – 3.059)/(4.439 – 3.059) = 0.248 in a 0-1 scale.

**Interpretation of results**

Since all of the variables were scaled between 0 and 1 before calculating the linear regression coefficients, each of the coefficients represents a proportion between the expected change (in terms of the range) of the dependent variable and a change (in terms of the range) of its independent variable, given that all other variables remain unchanged. Since the values of the high-risk model were only scaled (and not transformed), this interpretation extends directly to the un-scaled (real) values. As an example, the coefficient for the previous seasonal influenza coverage has a value of 0.36, and the seasonal influenza variable has a range of 43.9 (55.4 maximum minus 11.5 minimum). According to our model, if a state would increase its seasonal influenza coverage by 10% of the range for this variable (a percent change of 4.39% in the proportion of people vaccinated for seasonal influenza) it would expect to increase by 0.36\*0.1\*(range of dependent variable), which equals 0.36\*0.1\*36.8 = 1.32% the high-risk adult 2009 H1N1vaccination coverage, if all other variables remain unchanged. Notice that we can address the effect of a change without necessarily considering the initial value for the variables. This is possible due to the linearity of the model.

For the children’s model, the logarithmic transformation requires a more careful interpretation. The dependent variable of the regression model is the logarithmic transformation of the real dependent variable, the 2009 H1N1 vaccination coverage for children. Given that the transformation is logarithmic and not linear, we need to consider the value of the dependent and the independent variables, as well as the change, to estimate the new value of the true dependent variable. For example, the coefficient for the % of children population is -0.18817. The range of this variable is 22.1, and the range of the logarithmic transformation of coverage is 1.38. Suppose that a certain state had an H1N1 coverage of 30% for children, with a 10% of its population younger than 18 years. In this example, an assumed increment of 2.21% in the proportion of children population (10% of its range) would cause a decrease of 1.88% of the range of H1N1 coverage for children in that state, if all other variables remain unchanged. The transformation of the 30% coverage is ln(30) = 3.401, and the size of the change in this transformed variable is -0.0188\*1.38 = -0.026. Therefore, the final value of the transformed variable is 3.401 – 0.026 = 3.375. The real value of the coverage will be exp(3.375) = 29.2%, representing a real change in H1N1 coverage of only 0.8%.

**Approximation to normality in dependent variables.**

One of the main assumptions to achieve optimality of the coefficients’ estimation in linear regression is that the distribution of the dependent variable, conditional on the independent variables, is normal. In other words, the error term of the regression model is assumed normal. This assumption does not imply normality of the independent variables, which is not required for linear regression. Additionally, since normality of the conditional distribution of the dependent variable is rarely found, approximation to normality is usually accepted, understanding that such approximation will provide good coefficient estimators, but likely not optimal [[6](#_ENREF_6)]. Figure A presents the Q-Q plot for the residuals of the children’s model, and Figure B the Q-Q plot for the residuals of the high-risk adults’ model. The closer the plotted points are from the straight line, the better its distribution approximates normality.



**References:**

[1] Davila-Payan C, Swann J, Wortley PM. System factors to explain H1N1 state vaccination rates for adults in US emergency response to pandemic. Vaccine. 2013 [cited 2013 July 7]; Available from: <http://www.sciencedirect.com/science/article/pii/S0264410X13006841>.

[2] Centers for Disease Control and Prevention. Interim results: state-specific influenza A (H1N1) 2009 monovalent vaccination coverage - United States, October 2009-January 2010. MMWR Morbidity and Mortality Weekly Report2010 April 2;59(12):363-8.

[3] Singleton J, Santibanez T, Lu P, Ding H, Euler G, Armstrong G, Bell B, Town M, Balluz L. Interim results: influenza A (H1N1) 2009 monovalent vaccination coverage-United States, October-December 2009. MMWR2010;59:44-8.

[4] Clark S. H1N1 Vaccine Implementation Simultaneous Tracking Project - Summary of Findings weeks 09-28-09 to 10-2-09, 10-05-09 to 10-09-09, 10-19-09 to 10- 23-09, 10-26-09 to 10-30-09, 11-02-09 to 11-06-09, 11-09-09 to 11-13-09, 11-23-09 to 11-28-09, 11-30-09 to 12-04-09, 12-14-09 to 12-18-09, 01-04-10 to 01-08-10, and 01-11-10 to 01-15-10. University of Michigan2010.

[5] The Henry J. Kaiser Family Foundation - Statehealthfacts.org. 50 State Comparisons. [cited 2010 July 12]; Available from: <http://www.statehealthfacts.org/compare.jsp>.

[6] Kleinbaum DG. Applied regression analysis and multivariable methods: CengageBrain. com; 2007.