COVID-19: The Increasing Popularity of Homeschooling

Submitted By: Antara Saha   **Date**: 16-Oct-2022

POLICY ISSUE

Children are not the face of this pandemic, but undoubtedly, they are amongst its biggest victims. The COVID-19 pandemic has impacted children socially, economically, and mentally across all ages and countries, and sometimes, the mitigation measures taken to reduce the COVID-19 impact have inadvertently done more harm than good for the children.

One such aspect is the learning crisis for today's generation. The Covid-19 pandemic drove most schools and colleges to close their doors. At the peak of the pandemic, 188 countries imposed countrywide school closures, affecting more than 1.6 billion children [1], and three years into the COVID-19, with vaccines now readily available, 23 countries – home to nearly 405 million schoolchildren – are yet to fully open schools [1].

In the US, the analysis shows a significant impact on K-12 students, resulting in an average achievement gap in reading and math of four and five months, respectively [2]. As a result, some parents have started opting for alternatives to school, like homeschooling. In the study conducted by the US Census Bureau, the % of homeschooled children almost doubled to ~11% at the start of the 2020 – 2021 school year compared to the prior year [3].

But an increased number of homeschoolers is terrible news for parents & children. Some of the demerits of homeschooling include the absence of a certified educational professional, limited social interaction, a slower pace of learning, and an increased financial burden for the parents.

PROBLEM STATEMENT

While there are several reasons behind increased homeschooling, we will focus on the rate of immunization. States mandate that parents immunize their children before allowing them to enroll in public or private schools to safeguard public health [4]. But for homeschooled children, there are no such rules. Some parents may be opting for homeschooling if they are worried about vaccine safety or are, in general, vaccine skeptics. Hence, in this paper, I will examine the relationship between Homeschooled kids and their Covid-19 vaccination status. I will also investigate how the preference for being Homeschooled varies with Race and Household Education Attainment.

Understanding the relationship will assist the government in formulating the right policies to raise awareness and address parental concerns related to vaccines to make them feel safe enough to enroll their children in school.

METHODOLOGY

This section will outline the procedures used to examine the relationship stated above.

Data Utility:

To perform this analysis, I am using the Household Pulse Survey developed by the U.S. Census Bureau to capture the social and economic effects of the Covid-19 pandemic on American households. This survey is pertinent to this research because it includes information on households' K–12 enrollment decisions, whether they opted for public/private or homeschooling, and their vaccination decisions. Further, the demographic criteria used in this survey are thorough and take a broad view of society. It includes all the major races (White, Black, Asian) and the level of educational attainment of households that responded to the survey.

I'll be using data from Phases 3.4 and 3.5 of the Household Pulse Survey since both the age groups of 5-11 and 12-17 had vaccines available by then. The data is available at the URLs below:
- [Phase 3.5 PUF Releases](#)
- [Phase 3.4 PUF Releases](#)

Note: During the Pulse Survey Phase 3.4 and 3.5, both the age groups had vaccines available.

Analysis:

The analysis will comprise four sections. It will be performed on the dependent/independent variables outlined in Table 1:

1. **DESCRIPTIVE STATISTICS:** Proportion of children Homeschooled by their vaccination status, race, and household education attainment.
2. **DISPROPORTIONALITY INDICES:** Disproportionality indices for the homeschooled children by their vaccination status, race, and household education attainment.
3. **STATISTICAL SIGNIFICANCE**: As per the analysis topology, I will perform Multi-variate logistic regression to test whether the association of the Nominal/dependent variable (Homeschooling) is statistically significant with the nominal/independent variables (Vaccination status, race, and household education attainment)
   The hypothesis to be tested:
   H_0: *Homeschooling is independent of Vaccination Status, Race and Education Attainment*
   H_1: *Homeschooling is dependent on Vaccination Status, Race, and Education Attainment*
   NOTE: For the purposes of this analysis, I'll use the reference categories of
   - "White" for race
   - "Well-Educated" for educational attainment
   - "Vaccinated" for Vaccination Statuss
4. **ODD RATIO**: Lastly, the odds of Homeschooling for above stated factors.

Sno.	Variable Type	Variable Name	Possible Values
1	Dependent/Nominal	HMSCH	1 : Homeschooled 0: Not Homeschooled
2	Independent/Nominal	KIDDOSESRV_5_11Y KIDDOSESRV_12_17Y	1: Vaccinated 2: Not Vaccinated
3	Independent/Nominal	RRACE	1: White 2: Black 3: Asian 4:Others
4	Independent/Nominal	EEDUC	1: Less than high school 2: Some high school 3: High school graduate or equivalent (for example GED) 4: Some college, but degree not received or is in progress 5: Associate's degree (for example AA, AS) 6: Bachelor's degree (for example BA, BS, AB) 7: Graduate degree

Table 1: Dependent & Independent Variables used for the analysis

Next, we will prepare the data in the necessary format before conducting the analysis. The procedure for loading and cleaning the data into R is as follows:

1. Phases 3.4 and 3.5 have a combined six files, with three files from each Phase.
2. Replace the missing values, identified as -88 and -99, with NA.
3. Select the required fields from the data.
4. Filter the data to households with kids only.
5. Add a new column with the label HMSCH to indicate households with homeschooled kids. 1 is equivalent to Yes, and 0 is No.
6. Files from different phases have different column names. Hence, rename the fields to be the same across the files and append the data.
7. Remove records where the Vaccination Status for kids is anything other than Yes or No.
8. Create two data frames for two different age groups of kids 5 to 11 and 12 to 17.
9. Convert Vaccination Status, Race, and Education attainment into factors.
10. For the Education attainment, to keep it simple, I have grouped the values in the Education attainment field to form two categories – Less Educated and Well Educated.
    Less Educated:
    1. Less than high school
    2. Some high school
    3. High school graduate or equivalent (Ex: GED)
    4. Some college, but degree not received or is in progress
    Well Educated:
    1. Associate degree (Ex: AA, AS)
    2. Bachelor's degree (Ex: BA, BS, AB)
    3. Graduate degree (Ex: Master, Professional, Doctorate)

Data Limitation

While the Pulse Survey data is very rich, there are a few data limitations. Firstly, the survey has homeschooling information at the household level. So, if a family has more than one child and they are not homeschooling all children, there is no way to know that from the survey. Further, while the survey captures vaccine rates by age groups 5-11 and 12-17, the homeschooling information does not have that distinction. Hence, we will work under the assumption that all children in a household are either homeschooled or otherwise.

There can be multiple reasons for homeschooling. For example, schools are closed, better education prospects at home, and safety reasons (even after vaccination). However, the parents who opted for homeschooling gave no justification for their decision. Understanding the cause is crucial because it might not just be a matter of immunization.

Lastly, Hispanic or Latino cannot be accounted for in the Race column. The survey captures that information in another field called Hispanic Origin. The Hispanic Origin field overlaps significantly with the "Race" column. Hence Hispanics or Latinos cannot be isolated in the data.

RESULTS

We will analyze the data by the age groups of 5–11-year-old and 12-17 years old. We will start of the analysis with 5-11 years old.

Descriptive Statistics

In Figure 1, vaccination status, race, and household education attainment were separately analyzed for homeschooling. The households with un-vaccinated kids are more likely to prefer homeschooling by almost 60% compared with vaccinated kids. It could be due to the vaccination mandates in many US states to enter the school system. It is also interesting to see that ~7% of the children are in homeschooling even after vaccinations, indicating that there are reasons beyond just the vaccinations to keep children at home. Further, Other and White children are the highest racial groups that prefer homeschooling. Asians are the lowest at 6%. Lastly, there doesn't seem to be much difference in the household education attainment. It could be because well-educated households may have better means for homeschooling, while less educated may prefer homeschooling due to the lack of resources.

Percent of Kids Homeschool by: Kids Vaccination Status,Race,Education Attainment
Age Group: 5_11Yrs Old
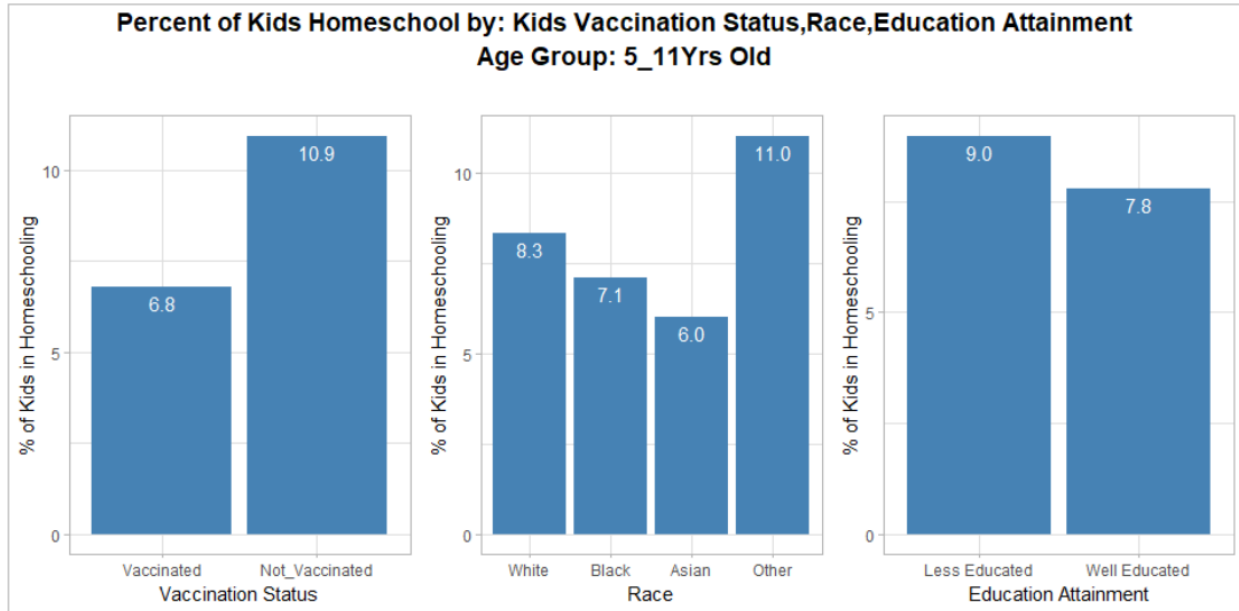
Figure 1

Disproportionality indices

In Table 2, we can see a similar picture. The households with un-vaccinated children experience the highest disproportionality. They make up ~34% of the population, but ~45% of un-vaccinated prefer homeschooling.

Similarly, the other races experience the highest disproportionality among various racial groups. They make up ~6.3% of the overall population, but ~8.5% of Others prefer homeschooling. While the White racial group is ranked 2, the disproportionality index is ~1, in line with the overall White population in the US. The Black and Asian racial groups are well below the disproportionality index of 1, indicating that these groups are less inclined towards homeschooling compared to others.

Lastly, looking at household education attainment, less educated households experience higher disproportionality compared to well-educated. They make up for ~32.7% of the population, but ~36% of less-educated households prefer homeschooling.

Disproportionality Indices for Kids Homeschooling For 5_11 Yrs Old				
category	share_condition	share_population	disproportionality	Rank
Not_Vaccinated	0.4501856	0.3364678	**1.3379752**	1
Vaccinated	0.5498144	0.6635322	0.8286176	2
category	share_condition	share_population	disproportionality	Rank
Other	0.0850699	0.0631402	**1.347318**	1
White	0.7687696	0.7560025	1.0168877	2
Black	0.0876392	0.1013137	0.865028	3
Asian	0.0585213	0.0795436	0.7357131	4
category	share_condition	share_population	disproportionality	Rank
Less Educated	0.3596917	0.3273678	**1.0987389**	1
Well Educated	0.6403083	0.6726322	0.9519441	2

Table 2

Logistic Regression and Odds Ratio:

Now, we will create a binomial logistic regression model to check our hypothesis that homeschooling is dependent on the independent variables of vaccination status, race, and household education attainment.

In Figure 2a, we can see that the logistics regression on homeschooling was generally statistically significant across the dependent variables. Most variables have p-values smaller than 1%, and hence statistically significant at a 99% confidence level.

In Figure 2b, we can see the odds ratios are similar to the disproportionality indices we calculated above. The odds of households with un-vaccinated children opting for homeschooling are ~1.66 compared to vaccinated children. Similarly, the odds of Other racial groups opting for homeschooling are ~1.34 compared to the White racial group. The Black and Asian groups have lower odds of homeschooling at ~0.79 compared to the White racial group. While the odds of households with less-education attainment are marginally higher compared to the well-educated group, household education attainment is not a statistically significant predictor.

In Figure 2c, we can see that the $X2(4) = 13,353$ and the p-value of the Chi-Square statistic is <2.2e-16, we can conclude that the multivariate model can fit the data better compared to the model with just an intercept term.

With this, we can also reject the null hypothesis that Homeschooling is independent of Vaccination Status, Race, and Education Attainment for age group 5-11 years old.

```
## 
## Call:
## glm(formula = HMSCH ~ KIDDOSESRV_5_11Y + relevel(RRACE, 1) +
##     relevel(EEDUC, 2), family = "binomial", data = fnl_dt_5_11)
## 
## Deviance Residuals:
##     Min      1Q   Median      3Q      Max
## -0.5559  -0.4370  -0.3775  -0.3775   2.4072
## 
## Coefficients:
##                              Estimate Std. Error z value Pr(>|z|)
## (Intercept)                  -2.60563    0.02725 -95.604  < 2e-16 ***
## KIDDOSESRV_5_11YNot_Vaccinated 0.51165    0.03696  13.842  < 2e-16 ***
## relevel(RRACE, 1)Black       -0.22835    0.06302  -3.624  0.00029 ***
## relevel(RRACE, 1)Asian       -0.23500    0.07548  -3.113  0.00185 **
## relevel(RRACE, 1)Other        0.28796    0.06516   4.420 9.89e-06 ***
## relevel(EEDUC, 2)Less Educated 0.01689   0.03846   0.439  0.66056
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## (Dispersion parameter for binomial family taken to be 1)
## 
##     Null deviance: 24256  on 42856  degrees of freedom
## Residual deviance: 23997  on 42851  degrees of freedom
## AIC: 24009
## 
## Number of Fisher Scoring iterations: 5
```

```
##                  (Intercept) KIDDOSESRV_5_11YNot_Vaccinated
##                   0.07385628                     1.66804893
##         relevel(RRACE, 1)Black        relevel(RRACE, 1)Asian
##                   0.79584268                     0.79056755
##         relevel(RRACE, 1)Other relevel(EEDUC, 2)Less Educated
##                   1.33369855                     1.01703480
```

```
## Likelihood ratio test
## 
## Model 1: HMSCH ~ 1
## Model 2: HMSCH ~ KIDDOSESRV_5_11Y + relevel(RRACE, 1) + relevel(EEDUC,
##     2)
##   #Df   LogLik Df Chisq Pr(>Chisq)
## 1   2  -5321.9
## 2   6 -11998.5  4 13353  < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figure 2A, 2B, 2C Respectively

Next, we will carry out a similar analysis for the age group 12-17 years old.

Descriptive Statistics

In figure 3, we have the descriptive statistics for homeschooling by vaccination status, race, and household education attainment. Households with un-vaccinated children are ~85% more likely to prefer homeschooling compared to households with vaccinated children. It is an even more stark difference compared to the age group of 5-11 years old.

Looking at the stats by race, we can see that they are very similar to age group 5-11 years. The Others and White are the higher racial groups, followed by Black and Asians. The Asians in this age group are almost at ~4.5%.

Lastly, looking at household education attainment, the difference between the less and well-educated groups is much wider in the age group of 12-17 compared to 5-11.
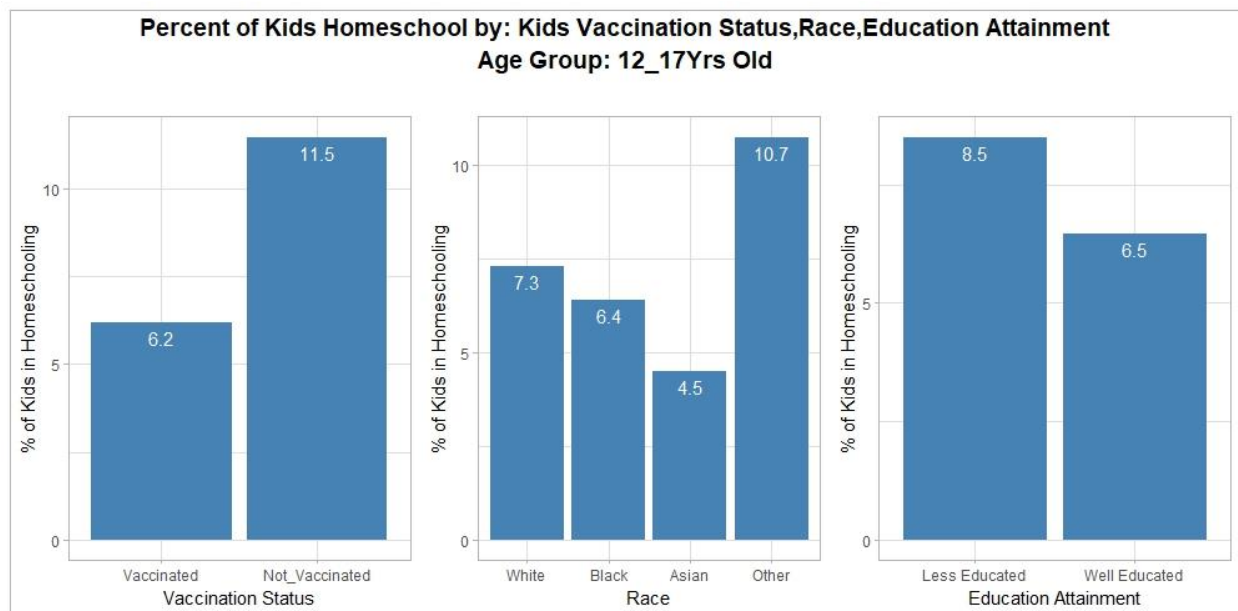
Figure 3

Disproportionality indices

In Table 3, we see that the disproportionality indices are like the age group of 5-11 years old. But it is worth noting that the gap between the high and the low end is wider.

Households with un-vaccinated children experience the highest disproportionality in homeschooling. They make up 19% of the population, but ~30% of households with unvaccinated children homeschool. Since the vaccines for the age group 12-17 came earlier than the 5-11 years old, we can also see a trend that homeschooling rates will be higher in un-vaccinated households.

Looking at race, the Other racial group experience the highest disproportionality. The Other racial group makes up for 6.2% of the population, but ~9.2% of the Other racial group prefers homeschooling. The White racial group is at par with their population as their

disproportionality index is 1. Asians, on the other hand, are at the extreme low end. Asians make up for ~6.9% of the population, but ~4.3% of Asians prefer homeschooling.

For household education attainment, less-educated households make up 37% of the population, but ~43.6% of less-educated households prefer homeschooling.

Disproportionality Indices for Kids Homeschooling For 12_17 Yrs Old				
category	share_condition	share_population	disproportionality	Rank
Not_Vaccinated	0.3060109	0.1923061	**1.59127**	1
Vaccinated	0.6939891	0.8076939	0.8592229	2
category	share_condition	share_population	disproportionality	Rank
Other	0.0926086	0.0622613	**1.487419**	1
White	0.7679034	0.7593215	1.011302	2
Black	0.0960598	0.1085368	0.8850439	3
Asian	0.0434282	0.0698804	0.6214651	4
category	share_condition	share_population	disproportionality	Rank
Less Educated	0.4368709	0.3707648	**1.1782964**	1
Well Educated	0.5631291	0.6292352	0.8949422	2

Table 3

Logistic Regression and Odds Ratio:

Next, we will create a binomial logistics regression model for the age group 12-17 years old. Our objective remains to check whether homeschooling is dependent on the independent variables of vaccination status, race, and household education attainment.

In Figure 4a, we can also see that the logistics regression on homeschooling for the age group 12-17 is statistically significant across all the dependent variables. All the variables have p-values smaller than 1%. Hence our results are statistically significant at a 99% confidence level.

In Figure 4b, we can also see that the odds ratios are similar to the disproportionality indices we calculated above. The odds of households with un-vaccinated children opting for homeschooling are ~1.85 compared to vaccinated children. Similarly, the odds of Other racial groups opting for homeschooling are ~1.46 compared to the White racial group. The Black and Asian groups have lower odds of homeschooling at ~0.84 and ~0.68, respectively, compared to the White racial group.

Further, the odds of households with less-education attainment are ~1.19 compared to the well-educated group. Unlike the age group 5-11 years old, the results are statistically significant at a 99% confidence level.

In Figure 4c, we can see that the X2(4) = 18,123 and the p-value of the Chi-Square statistic is <2.2e-16, we can conclude that the multivariate model can fit the data better compared to the model with just an intercept term.

With this, we can also reject the null hypothesis that Homeschooling is independent of Vaccination Status, Race, and Education Attainment for the age group 12-17.

```
##
## Call:
## glm(formula = HMSCH ~ KIDDOSESRV_12_17Y + relevel(RRACE, 1) +
##     relevel(EEDUC, 2), family = "binomial", data = fnl_dt_12_17)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -0.6092  -0.3815  -0.3501  -0.3501   2.5245
##
## Coefficients:
##                                Estimate Std. Error  z value Pr(>|z|)
## (Intercept)                    -2.76143    0.02664 -103.643  < 2e-16 ***
## KIDDOSESRV_12_17YNot_Vaccinated 0.61302    0.03971   15.439  < 2e-16 ***
## relevel(RRACE, 1)Black         -0.17520    0.06038   -2.902  0.00371 **
## relevel(RRACE, 1)Asian         -0.38301    0.08622   -4.442 8.90e-06 ***
## relevel(RRACE, 1)Other          0.38074    0.06279    6.064 1.33e-09 ***
## relevel(EEDUC, 2)Less Educated  0.17751    0.03664    4.845 1.27e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 24976  on 48167  degrees of freedom
## Residual deviance: 24598  on 48162  degrees of freedom
## AIC: 24610
##
## Number of Fisher Scoring iterations: 5
```

```
##               (Intercept) KIDDOSESRV_12_17YNot_Vaccinated
##                0.06320158                      1.84599307
##       relevel(RRACE, 1)Black          relevel(RRACE, 1)Asian
##                0.83929131                      0.68180428
##       relevel(RRACE, 1)Other  relevel(EEDUC, 2)Less Educated
##                1.46337327                      1.19424334
```

```
## Likelihood ratio test
##
## Model 1: HMSCH ~ 1
## Model 2: HMSCH ~ KIDDOSESRV_12_17Y + relevel(RRACE, 1) + relevel(EEDUC,
##     2)
##   #Df   LogLik Df Chisq Pr(>Chisq)
## 1   2  -3237.6
## 2   6 -12298.9  4 18123  < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figure 4A, 4B, 4C Respectively

Conclusion, Limitations & Next Steps

In this analysis, we used multivariate logistic regression to reject the null hypothesis that the rate of homeschooling is independent of vaccination rate, race, and household education attainment. We established that for both the age groups of 5-11 years old and 12-17 years old, the odds of homeschooling increase significantly for un-vaccinated children. These odds increase further if those children belong to the Other racial group (other than White, Black, and Asian) or belong to a household with less-educated parents. Hence, for any back-to-school government policy, the government should first address parents' concerns regarding COVID-19 vaccines and specifically target Other racial and less-educated groups for a bigger impact.

There are a few data limitations that we have identified along the way, such as homeschooling information not available by age groups or lack of information on Latino racial groups. Further, this paper only looked at homeschooling trends at the national level. Since the vaccine mandates and availability varies by state, future work could include more demographics, especially to represent the state-level nuances and variations.

Also, as stated in the data limitation, the survey lacked qualitative insights into parents' decisions to homeschool. It is a significant aspect to understand, especially when we know that ~6% - 7% of vaccinated children also stay out of school for homeschooling.

References

[1] https://data.unicef.org/covid-19-and-children/
[2] https://www.mckinsey.com/industries/education/our-insights/covid-19-and-education-the-lingering-effects-of-unfinished-learning
[3] https://www.census.gov/library/stories/2021/03/homeschooling-on-the-rise-during-covid-19-pandemic.html
[4] Khalili, Donya, and Arthur Caplan. "Off the Grid: Vaccinations among Homeschooled Children." Journal of Law, Medicine & Ethics 35, no. 3 (2007): 471–77. doi:10.1111/j.1748-720X.2007.00169.x.