Supplement - Socioeconomic bias in influenza surveillance

Samuel V. Scarpino[1,2,3,4], James G. Scott[5], Rosalind M. Eggo[6], Bruce Clements[7], Nedialko B. Dimitrov[5], and Lauren Ancel Meyers[5,8,*]

[1]Network Science Institute, Northeastern University, Boston, MA, 02115, USA
[2]Marine & Environmental Sciences, Northeastern University, Boston, MA, 02115, USA
[3]Physics, Northeastern University, Boston, MA, 02115, USA
[4]ISI Foundation, 10126 Turin, Italy
[5]The University of Texas at Austin, Austin, TX, USA
[6]London School of Hygiene and Tropical Medicine, London, UK
[7]Texas Department of State Health Services, Austin, TX, USA
[8]Santa Fe Institute, Santa Fe, New Mexico, USA
[*]address general correspondence to laurenmeyers@austin.utexas.edu

Supplemental Tables

Surveillance Data Sources	Poverty Quartile 1st quartile (lowest poverty)	2nd quartile	3rd quartile	4th quartile (highest poverty)	Aggregate
ILINet	1.45	1.81	2.63	4.04	2.20
BioSense 2.0	1.69	2.05	2.66	4.23	2.52
GFT	1.33	1.64	2.61	3.63	2.00
ILINet+ BioSense 2.0	1.55	1.81	2.49	3.88	2.18
ILINet+ GFT	1.33	1.59	2.65	3.29	2.00
BioSense 2.0 + GFT	1.39	1.91	2.59	4.03	2.35
ILINet+ BioSense 2.0 + GFT	1.39	1.91	2.52	4.03	2.35

Table 1. Out-of-sample (leave-one-out) root mean-squared error (ORMSE) for each Poisson generalized additive model. Values are normalized by the population size of each zip code quartile and then multiplied by 10^6 to obtain ORMSE per one million residents. The rightmost column gives aggregate ORMSE across all Texas zip codes.

Surveillance Data Sources	Poverty Quartile				Aggregate
	1st quartile (lowest poverty)	2nd quartile	3rd quartile	4th quartile (highest poverty)	
ILI	-243.20	-268.55	-323.38	-449.58	-191.60
Biosense	-252.47	-278.49	-326.32	-451.25	-211.57
GFT	-227.45	-241.87	-311.60	-375.69	-160.17
ILI + Biosense	-243.19	-261.42	-301.16	-416.81	-173.76
ILI + GFT	-223.72	-240.27	-322.68	-354.08	-154.03
Biosense + GFT	-231.61	-257.83	-317.07	-415.65	-182.16
ILI + Biosense + GFT	-231.61	-257.83	-297.48	-415.65	-182.16

Table 2. The table shows out-of-sample (leave-one-out) Poisson log-likelihood using a Poisson generalized additive model. The values were normalized by the population size of each quantile and then multiplied by 10^6, so are Poisson log-likelihood per one million residents. The quartiles have approximately the same number of residents, so the RMSE values are very nearly the Poisson log-likelihood for each time series.

Surveillance Data Sources	Poverty Quartile				Aggregate
	1st quartile (lowest poverty)	2nd quartile	3rd quartile	4th quartile (highest poverty)	
ILI	1.74	2.16	3.01	4.35	2.56
Biosense	2.32	2.87	3.34	3.84	2.91
GFT	1.81	2.02	2.89	3.78	2.51
ILI + Biosense	1.93	2.87	3.18	3.92	3.09
ILI + GFT	1.83	2.26	3.30	4.27	2.76
Biosense + GFT	1.49	2.01	3.78	4.01	2.86
ILI + Biosense + GFT	1.49	2.01	3.51	4.01	2.86

Table 3. The table shows out-of-sample (leave-one-out) root mean-squared error (RMSE) using a Gaussian generalized additive model.The values were normalized by the population size of each quantile and then multiplied by 10^6, so are Poisson log-likelihood per one million residents. The quartiles have approximately the same number of residents, so the RMSE values are very nearly the Poisson log-likelihood for each time series.

Surveillance Data Sources	Poverty Quartile				
	1st quartile (lowest poverty)	2nd quartile	3rd quartile	4th quartile (highest poverty)	Aggregate
ILI	-250.01	-273.16	-322.71	-447.12	-192.58
Biosense	-253.13	-278.45	-346.70	-480.60	-199.57
GFT	-231.79	-240.43	-311.57	-382.54	-159.35
ILI + Biosense	-255.92	-271.01	-324.76	-438.27	-177.71
ILI + GFT	-226.35	-250.27	-326.13	-405.46	-176.10
Biosense + GFT	-228.51	-251.79	-324.07	-393.63	-165.57
ILI + Biosense + GFT	-228.51	-251.79	-303.11	-393.63	-165.57

Table 4. The table shows out-of-sample (leave-one-out) Poisson log-likelihood using a Gaussian generalized additive model. The values were normalized by the population size of each quantile and then multiplied by 10^6, so are Poisson log-likelihood per one million residents. The quartiles have approximately the same number of residents, so the RMSE values are very nearly the Poisson log-likelihood for each time series.

Surveillance Data Sources	Poverty Quartile				
	1st quartile (lowest poverty)	2nd quartile	3rd quartile	4th quartile (highest poverty)	Aggregate
ILI	-102.28	-117.46	-139.78	-190.16	-89.88
Biosense	-97.83	-112.08	-130.92	-155.79	-79.14
GFT	-86.75	-96.78	-127.28	-150.90	-72.93
ILI + Biosense	-93.39	-106.36	-130.75	-164.27	-77.57
ILI + GFT	-88.13	-92.55	-128.24	-153.46	-73.13
Biosense + GFT	-93.66	-102.27	-122.34	-151.50	-73.37
ILI + Biosense + GFT	-93.66	-100.77	-122.59	-151.50	-72.55

Table 5. The table shows out-of-sample (60/40 training/testing) Poisson log-likelihood using a Poisson generalized additive model. The values were normalized by the population size of each quantile and then multiplied by 10^6, so are Poisson log-likelihood per one million residents. The quartiles have approximately the same number of residents, so the RMSE values are very nearly the Poisson log-likelihood for each time series.

	Mean	Upper 95th percentile	Lower 95th percentile
1st quartile	0.49	0.60	0.38
2nd quartile	0.50	0.61	0.39
3rd quartile	0.47	0.60	0.36
4the quartile	0.48	0.62	0.36

Table 6. The mean and 95% confidence intervals for the proportion of the residual error that was positive (over prediction) determined by bootstrapping the residuals 1,000 times for the best-fit model in each quartile. Across all four quartiles, with the 1st having the lowest proportion below poverty and 4th the highest, the best-fit models were unbiased. A value of 0.5 would indicate that exactly half of the error was associated with under-prediction and half of the error was associated with over-prediction.

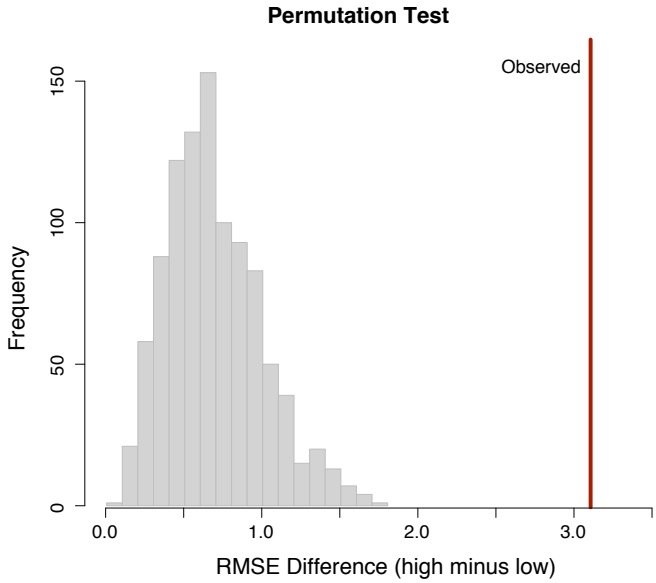


Figure 1. Supplemental Figure 1 - Result of the permutation test for the ILI + Biosense + GFT model across 10,000 Monte Carlo samples. The vertical red line is at 3.3, the observed value based on the poverty grouping. The results indicate that it is unlikely for the observed value to arise by chance. The Monte Carlo *p*-value is 0.0001, with only of our randomized permutations yielding an ORMSE gap at least as large as 3.3.

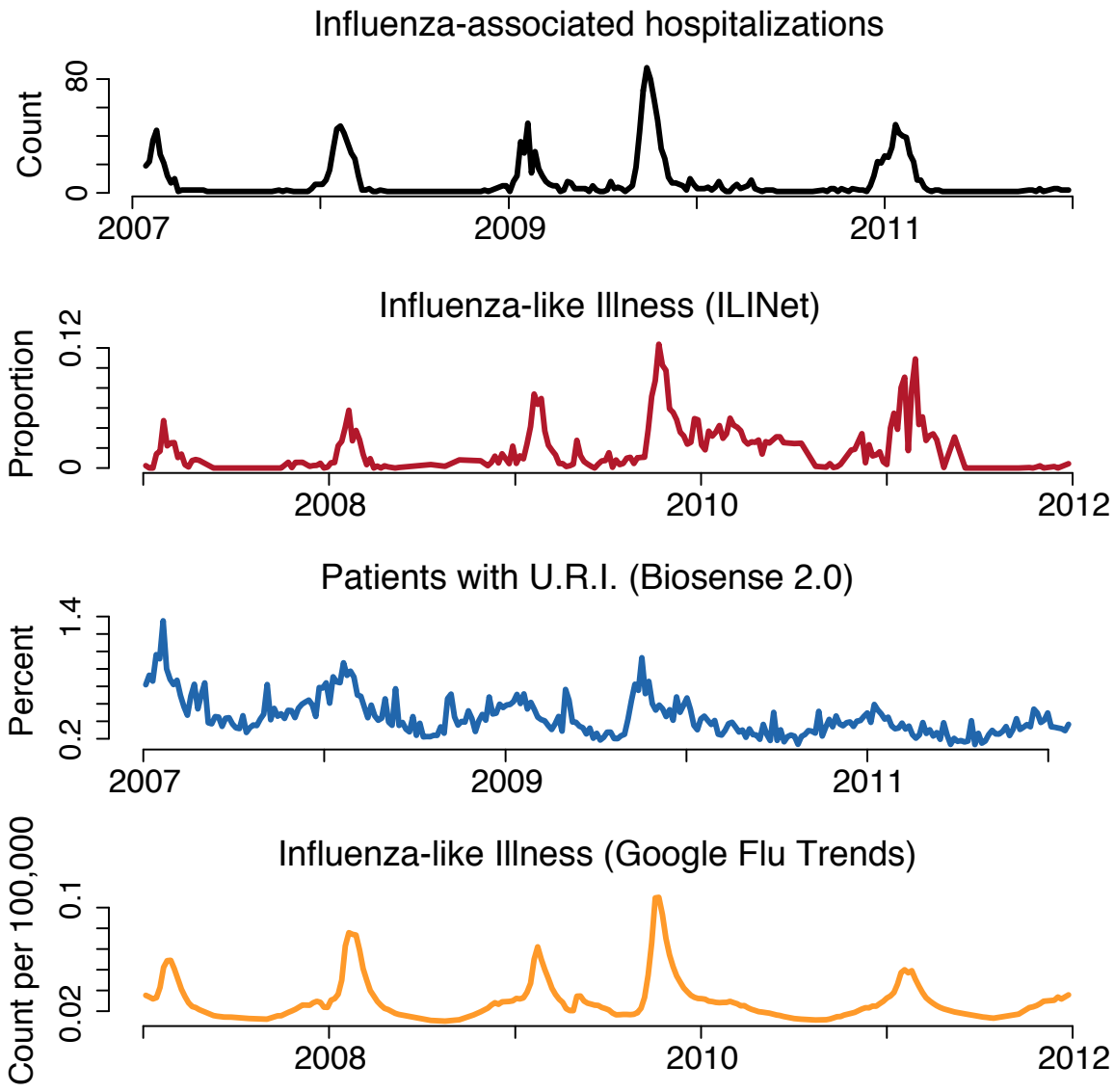


Figure 2. Supplemental Figure 2 - Key Datasets used in the analysis. The top panel shows influenza-associated inpatient hospitalizations in black, as defined by ICD9 codes 486 and 487, the next panel shows ILINet in red where the proportion of doctor visits are for influenza-like illness, the next panel shows BioSense 2.0 in blue, which is the proportion of ED visits per week that are for URI. The final panel shows the Google Flu Trends estimate, in orange, of the number of influenza-like-illness cases per 100,000 people.

Principal competent analysis to detect synchrony

For each poverty quartile, we also performed a principal-component analysis of zip code level hospitalization counts. That is, we calculated the principal components of the matrix $Y^{(i)}$ of hospitalization counts whose rows are weeks, and whose columns are the zip codes within poverty quartile i. The highest poverty quartile has the highest percent variation explained by these leading components, indicating greater synchrony in influenza trends within poorer populations. Thus, we do not believe that the reduced performance in lower socioeconomic groups stems from greater variation in temporal flu trends.

Analysis of Hospitalization & Surveillance Detection Rates

Lower income groups often have a higher hospitalization rates than higher income groups (as illustrated in our Figure 2, Thompson et al. 2004, Sloan et al. 2015, and Tam et al. 2014). However, a higher case hospitalization rate, on its own, cannot explain the observed disparity in predictability of the time series across different poverty quartiles. Therefore, other factors must be at work, such as reduced rates of ILI primary care in lower socioeconomic groups (Shi et al. 1999; Wilkinson and Pickett 2006), lower correlation between ILI-related internet searches and actual ILI in lower socioeconomic groups (Richiardi et al. 2014), socioeconomic differences in vaccination levels (Fiscella et al. 2000; CDC 2001; Galea et al. 2005), or socioeconomic differences in underlying health conditions (Adler and Newman 2002).

To demonstrate that a higher case hospitalization rate alone cannot explain the observed disparity in predictions between high poverty and low poverty quartiles, we conduct a simple simulation experiment. The simulation demonstrates that, when all else is equal, a higher hospitalization rate should instead lead to greater prediction accuracy. The larger number of hospitalizations increases the statistical power of the models. We now describe this analysis.

We used the following numerical simulation to assess whether a higher hospitalization rate in disadvantaged populations might explain the poorer surveillance:

1. We generated a time series of Influenza-Like-Illness (ILI) cases reported per day, S. This time series was based on a standard Susceptible-Infectious-Recovered, i.e. SIR, model with a population of 100,000 individuals, where the force of infection, β, was 0.5, and the recovery rate, γ, was 0.4. The choice of these parameter values does not impact our qualitative results. (Figure 3)

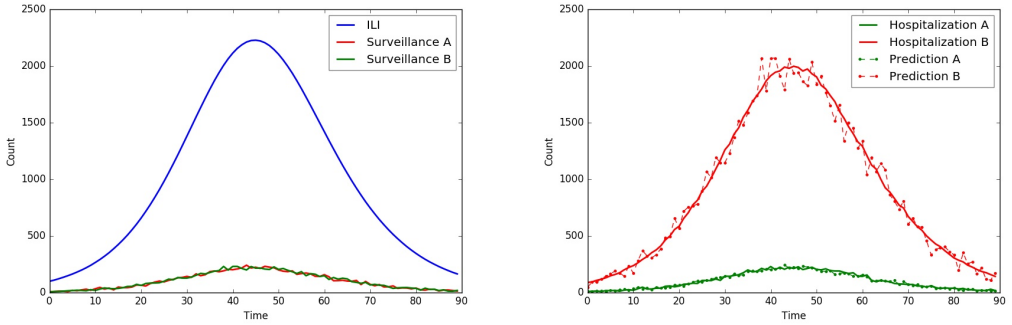
2. We create two populations, A and B, each with ILI trends that perfectly match S.

3. We simulate two noisy surveillance time series, one for population A and one for population B. For each day in each time series, we generate a binomial random variable, with parameters n (number of trials) equal to the total number of ILI cases reported on that day and binomial probability p equal to 0.1. This assumes that the surveillance systems in both populations pick up cases stochastically at the same rate (10%). We refer to this parameter as the surveillance detection rate.

4. We simulate two hospitalization time series, one for population A and one for population B. As in step 3, for each day in each time series, we generate a binomial random variable with n (number of trials) equal to the total number of ILI cases reported on that day. In this case, however, the binomial probability models the hospitalization rate rather than the surveillance detection rate, and we assume that it differs between the two populations. Specifically, we use p=0.1 and p=0.9 in populations A and B, respectively.

5. We then attempt to predict the hospitalizations in both groups using the same regression model as in the main analysis.

We performed 10,000 stochastic simulations and found that predictions were better, on average, for the population with the higher case hospitalization rate than the population with the lower case hospitalization rate (Supplement Figure 3(b)). Specifically, when comparing the predicted hospitalization time series to the actual hospitalization time series, the average R^2 was 0.9830 and 0.9914, for populations A and B, respectively – a statistically significant difference with $p < 0.0001$.



(a) Generated ILI and signals

(b) Generated hospitalizations and predictions

Figure 3. Simulating Disparate Case Hospitalization Rates. The curves illustrate a typical simulation. Figure 3(a) depicts the Influenza-Like-Illness (ILI) time series (blue) for populations A and B, and surveillance time series for A (red) and B (green) derived by stochastically sampling the ILI time series, assuming that 10% of cases are detected by the system (for example, via internet use or physician visits). Figure 3(b) depicts the hospitalization time series and predicted hospitalizations for populations A and B, which had hospitalization rates of 0.1 and 0.9, respectively. The hospitalization curves were generated by stochastically sampling the ILI curve in Figure 3(a) and the predictions were created using the same regression model as in the main analysis. The average R^2 over 10,000 simulations for these predictions are 0.9986 and 0.9993, for A and B, respectively.

We modified the simulation described above to assess, more generally, the combined impact of the hospitalization rate and the surveillance detection rate on the prediction error. Specifically, we ran a total of 10,000 single population simulations, varying the case hospitalization rate and the surveillance detection rate from 0.1 to 0.9, assuming $\beta = 0.076$ and $\gamma = 0.07$. We find a marked and predictable decline in precision, as the hospitalization rate and surveillance detection rate decline (Supplement Figure 4), suggesting low sampling by a surveillance system (for example, stemming from reduced use of primary care and internet resources in disadvantaged populations), with or without an increase in hospitalization rate, could impede situational awareness and short-term forecasts.

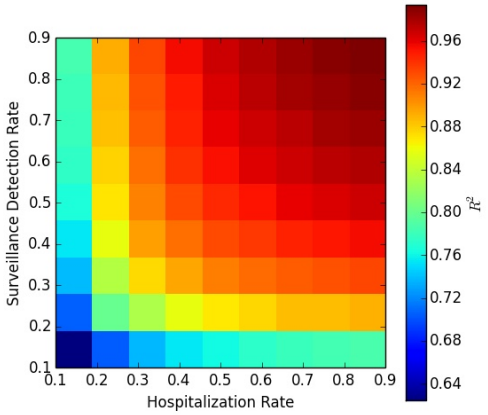


Figure 4. As the hospitalization rate or the surveillance detection rate drops, the predictions become less precise. For each combination of surveillance detection rate and hospitalization rate, we run 100 simulations to estimate the expected R^2. These simulations are conducted assuming $\beta = 0.076$ and $\gamma = 0.07$ and the results are qualitatively the same for other values of these parameters.

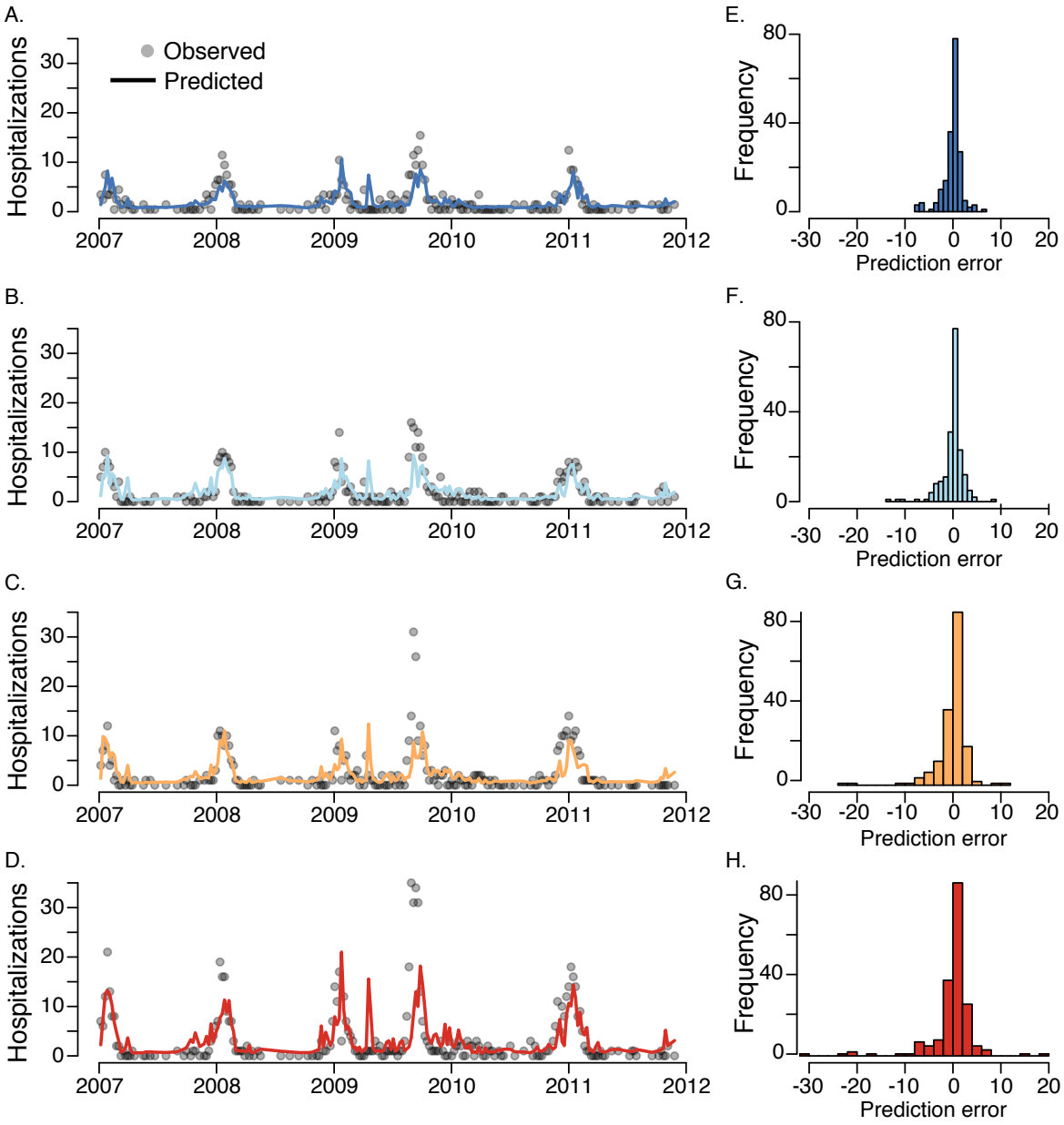


Figure 5. Comparison between one-week ahead model predictions and the total number of weekly observed influenza hospitalizations for each of the four poverty quartiles (A) Upper quartile, (B) Upper-middle quartile, (C) Lower-middle quartile, (D) Lowest quartile and the distribution of out-of-sample prediction errors (observed - predicted) for the (E) Upper quartile, (F) Upper-middle quartile, (G) Lower-middle quartile, and (H) Lowest quartile. Across all four quartiles, the model was unbiased according to a resampling test on the residuals, see Supplement.