

1 Title: Influenza-Negative Influenza-Like Illness (fnILI) Z-Score as a Proxy for Incidence and
2 Mortality of COVID-19

3 Running Title: fnILI Z-Score Proxy for COVID-19

4 *Brief Report – Journal of Infectious Disease*

5

6 Fatima N. Mirza, MPH¹, Aryn A. Malik, MBBS, MPH, PhD^{1,2}, Saad B. Omer, MBBS, MPH,

7 PhD^{1,2,3,4}

8 (1) Department of Internal Medicine, Yale School of Medicine, Yale University, New Haven, CT 06510, USA

9 (2) Yale Institute for Global Health, Yale University, New Haven, CT 06510, USA

10 (3) Department of Epidemiology of Microbial Diseases, Yale School of Public Health, Yale University, New Haven, CT 06510, USA

11 (4) Yale School of Nursing, Yale University, Orange, CT 06477, USA

12

13 **Corresponding Author:**

14 1 Church St

15 New Haven, CT 06510

16 saad.omer@yale.edu

17 Phone: (203) 432-3656

18

19 Word Counts: *Abstract:* 96/100 *Text:* 1309/2000

20 References: 9 Figures: 1/1 Tables: 1/1

21

22

23

24

25

26

27

28 **Abstract**

29

30 Though ideal for determining the burden of disease, SARS-CoV2 test shortages preclude its
31 implementation as a robust surveillance system in the US. We correlated the use of the derivative
32 influenza-negative influenza-like illness (fnILI) z-score from the CDC as a proxy for incident
33 cases and disease-specific deaths. For every unit increase of fnILI z-score, the number of cases
34 increased by 70.2 (95% CI[5.1,135.3]) and number of deaths increased by 2.1 (95% CI[1.0,3.2]).
35 FnILI data may serve as an accurate outcome measurement to track the spread of the and allow
36 for informed and timely decision-making on public health interventions.

37

38 **Key Words:** COVID-19, ILI, CDC, influenza

39

40

41

42

43

44

45

46

47

48

49

50

51

52 **Background**

53

54 Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) has spread exponentially since
55 December 2019, transforming from a localized outbreak in Wuhan, China to a global pandemic.
56 As of mid-April, roughly 2 million cases have been reported(1), with one-fifth of cases requiring
57 inpatient hospitalization and a case-fatality at an estimated 2%(2).

58

59 As the number of individuals infected rapidly climbs, the testing capacity has been outpaced by
60 the need for such tests. In the United States, this has posed a challenge to physicians and public
61 health professionals at large, particularly as it relates to accurately tracking the spread of disease.
62 Assessing the intensity of the epidemic nationally in a given region is the backbone of allocating
63 resources at the federal and state level, and inform the implementation or relaxing of public
64 health restrictions (e.g. initiating or easing a lockdown).

65

66 Given the rapid increase in cases in the previous weeks without parallel expansion in testing
67 capacity and unclear specificity/sensitivity, this problem will only continue to be exacerbated
68 until a nationwide program is made available and further validation studies have been completed
69 (3). In the interim, there is an urgent need to identify proxies for disease incidence that are
70 routinely collected through available infrastructure in the United States in order to guide the
71 evolving public health response in this country(4).

72

73 The Centers for Disease Control and Prevention (CDC) centrally collates data using the U.S.
74 Outpatient Influenza-like Illness Surveillance Network (ILINet) and the National Respiratory
75 and Enteric Virus Surveillance System (NREVSS) (5). We believe that combining both sources

76 of this publicly-available, routinely collected data may serve as a reliable proxy for SARS-CoV-
77 2 incidence and mortality. In this study, we used influenza-negative ILI (fnILI) z-scores and
78 compared them against the reported COVID-19 cases and deaths by week to document trends
79 over time.

80

81 **Methods**

82

83 We downloaded flu negative influenza-like illness (fnILI) data derived from the Center for
84 Disease Control and Prevention's ILINet and NREVSS data for states (5,6). ILINet consists of
85 records from outpatient healthcare providers in all 50 states and reported 60 million patient visits
86 in the 2018-2019 season. Weekly, approximately 2,600 outpatient healthcare providers around
87 the country report the total number of patients as well as those with influenza-like illness,
88 defined as a temperature of 100°F or greater alongside a cough and/or sore throat, as well as
89 regional baseline are reported. These data are weighted by state population, and percentage of
90 flu-positive influenza-like illness is compared to regional estimates and a historical nationwide
91 baseline of 2.4%. NREVSS provides virologic surveillance data weekly from approximately 100
92 public health and over 300 clinical laboratories throughout the United States, including the total
93 number of respiratory specimens tested for influenza, the number positive for influenza viruses,
94 and the percent positive by influenza virus type. Some states may have limited data or have
95 delays in reporting that may not make this information immediately available.

96

97 Reich et al (6) reviewed twenty-three seasons of influenza data, beginning in 1997, and ten
98 seasons of statewide data, beginning in 2010 and calculated fnILI from the CDC (5). fnILI was
99 determined using weighted influenza-like illness (wILI) from ILINet – which represents the

100 percentage of doctor’s office visits that presented with a primary complaint of fever and one
101 additional influenza-like symptoms – and percentages of positive influenza specimens from
102 NREVSS data, compared to a baseline calculated from prior all seasons of data extracted as
103 described above. fnILI was calculated as:

$$fnILI = (1 - \textit{proportion of tests positive for influenza}) \times \frac{wILI}{\textit{baseline level for ILI}}$$

104
105 These data included a z-score that represents the degree to which a given fnILI observation was
106 significantly lower or higher than expected based on past trends at similar times during prior
107 years. Z-score was calculated as:

$$Z = \frac{fnILI - \overline{fnILI}}{sdfnILI}$$

108
109 with as the average weekly observations for the past nine years with one week on either
110 side and as associated standard deviation.

111
112 We merged this dataset with the CDC-reported SARS-CoV-2 cases and disease-specific deaths.
113 We graphically represented the fnILI z-score, cumulative cases, and cumulative deaths for the
114 contiguous U.S for the month of March 2020. We used a mixed effects linear model accounting
115 for clustering at the state level using random effects to determine the relationship between
116 weekly case counts or deaths and fnILI z-score, and a lag term to account for delay between
117 onset of symptoms and confirmation of diagnosis. We selected the model with the lowest Akaike
118 Information Criteria (AIC) and Bayesian Information Criteria (BIC) (7) while maintaining
119 statistical significance. Median fnILI z-score across all states were determined by week and

120 plotted against total nationwide cases and/or deaths per week. All data was analyzed using Stata
121 v15.0 (StataCorp, College Station, Texas).

122

123 **Results**

124

125 fnILI data was available for all states except Florida, New Hampshire, New Jersey, and Rhode
126 Island. Over the course of the study period, September 29th 2019 to March 29th 2020, there were
127 123,509 reported COVID-19 cases and 2,303 deaths, representing a case-fatality rate of 1.8%.
128 This represented 36,749,916 unique patient visits across the US with an average of 3,040
129 providers nationwide per week. The rate of fnILI across the study period was 85.8% and the
130 average flu positive baseline across the country was 1.4%.

131

132 There is an apparent tracking between fnILI z-score and cases or deaths by state over the course
133 of the study period. This phenomenon is particularly pronounced when comparing these indices
134 during the month of March (Figure 1A).

135

136 When assessing the correlation over time between fnILI z-score and either new cases or deaths,
137 we observed a z-score peak prior to an increase in cases or deaths. Therefore, we used a lag
138 variable of two weeks for incidence and one week for mortality to better fit the model (Table 1).

139 On the mixed effects linear model accounting for clustering at the state level using random
140 effects, we found that for every unit increase of fnILI z-score two weeks prior, the number of
141 cases increased by 70.2 (95% CI [5.1, 135.3]). Similarly, we found that for every unit increase of
142 fnILI z-score one week prior, the number of deaths increased by 2.1 (95% CI [1.0, 3.2]), also
143 when correcting for regional effects. When plotting the median nation-wide z-score two week or

144 one week prior, respectively, versus new cases (Figure 1B) or deaths (Figure 1C), the two
145 measures tracked. Nationwide, one unit increase of fnILI z-score correlated with 49243.4 (95%
146 CI[34250.8, 64235.9]) and 608.3 deaths (95% CI[477.3, 739.4]).

147

148 **Discussion**

149

150 There is increasing evidence that current testing may significantly underestimate the burden of
151 disease (8), necessitating alternative methods to accurately assess these trends over time. Our
152 results suggest that the fnILI z-score data can be used as a proxy for the trajectory of disease
153 incidence and mortality in the United States. In the context of limited resources in a rapidly
154 changing field, it becomes increasingly necessary to innovatively utilize available infrastructure
155 to tackle the apparent gaps in knowledge quickly. To our knowledge, this is the first academic
156 study to use fnILI z-scores from ILINet and NVRESS data in order to model and potentially
157 predict the burden of COVID-19 over time.

158

159 This report demonstrates the important potential of such a proxy, and validates its correlation
160 with incidence and mortality. Importantly, we present the optimal model for such a prediction by
161 building in a lag term. This two week lag term is likely necessary for incidence due to the known
162 incubation period of this disease and because of a delay in testing (2). For mortality, our one
163 week lag term likely represents the rapid escalation to fatality (9). These lag terms also may
164 allow our model to function as an early warning system for rise in cases, similar to ILINet.

165

166 As there is already a robust infrastructure in place to collect these data, validating the use of such
167 data is extremely valuable, especially in the setting of limited availability and capacity of testing

168 kits. fnILI provides a good proxy in the absence of testing to evaluate the results of public health
169 interventions and make timely decisions to change course.

170

171

172

173

174

175

176

177

178

179

180

181

182

183

184

185

186

187

188

189

190

191

192 **Footnote**

193

194 Disclosures: The authors have no conflicts of interests to disclose.

195 Funding/Support: There is no funding for this study.

196 Statement of Prior Presentation: This information has not been previously presented.

197 Correspondence and Requests:

198 Dr. Saad B. Omer

199 1 Church St

200 New Haven, CT 06510

201 saad.omer@yale.edu

202 Phone: (203) 432-3656

203

204

205

206

207

208

209

210

211

212

213

214

215

216 **References**

- 217 1. Dong E, Du H, Gardner L. An interactive web-based dashboard to track COVID-19 in real
218 time. *The Lancet Infectious Diseases*. Lancet Publishing Group; 2020.
- 219 2. Guan W, Ni Z, Hu Y, Liang W, Ou C, He J, et al. Clinical Characteristics of Coronavirus
220 Disease 2019 in China. *N Engl J Med*. 2020 Feb 28;
- 221 3. Di Gennaro F, Pizzol D, Marotta C, Antunes M, Racalbutto V, Veronese N, et al.
222 Coronavirus Diseases (COVID-19) Current Status and Future Perspectives: A Narrative
223 Review. *Int J Environ Res Public Health* [Internet]. 2020 Apr 14 [cited 2020 Apr
224 19];17(8):2690. Available from: <https://www.mdpi.com/1660-4601/17/8/2690>
- 225 4. Ienca M, Vayena E. On the responsible use of digital data to tackle the COVID-19
226 pandemic. *Nat Med*. 2020 Mar 27;26(4).
- 227 5. U.S. Influenza Surveillance System: Purpose and Methods | CDC [Internet]. [cited 2020
228 Apr 19]. Available from: <https://www.cdc.gov/flu/weekly/overview.htm>
- 229 6. Reich NG, Ray EL, Gibson GC, Cramer E RC. Looking for evidence of a high burden of
230 COVID-19 in the United States from influenza-like illness data. Github. 2020.
- 231 7. Konishi S, Kitagawa G (Genshiro). *Information criteria and statistical modeling*. Springer;
232 2007.
- 233 8. Bendavid E, Mulaney B, Sood N, Shah S, Ling E, Bromley-Dulfano R, et al. COVID-19
234 Antibody Seroprevalence in Santa Clara County, California. medRxiv. 2020 Apr
235 17;2020.04.14.20062463.
- 236 9. Jung S, Akhmetzhanov AR, Hayashi K, Linton NM, Yang Y, Yuan B, et al. Real-Time
237 Estimation of the Risk of Death from Novel Coronavirus (COVID-19) Infection:
238 Inference Using Exported Cases. *J Clin Med*. 2020 Feb 14;9(2):523.

239 **Tables**

240 **Table 1. Possible Models for fnILI Z-Score and Incidence or Mortality.**

241

Incidence	Lag	Coef.	95% CI		p	AIC	BIC
			Lower	Upper			
	None	128.9	79.2	178.6	<0.01	20516.2	20536.5
	1 Week	129.9	75.5	184.3	<0.01	19834.7	19854.8
	2 Week	70.2	5.1	135.3	0.03	19095.9	19115.9
	3 Week	1.97	-68.4	72.3	0.9	18364.2	18384
Mortality	None	2.1	1.1	3.1	<0.01	11328.3	11348.6
	1 Week	2.1	1.0	3.2	<0.01	10970.8	10991.0
	2 Week	1.2	-0.08	2.5	0.06	10585.1	10605.1

242

243

244

245

246

247

248

249

250

251

252 **Figure Legend**

253

254 **Figure 1. Overall state-specific and median nationwide fnILI Z-Score versus COVID-19**

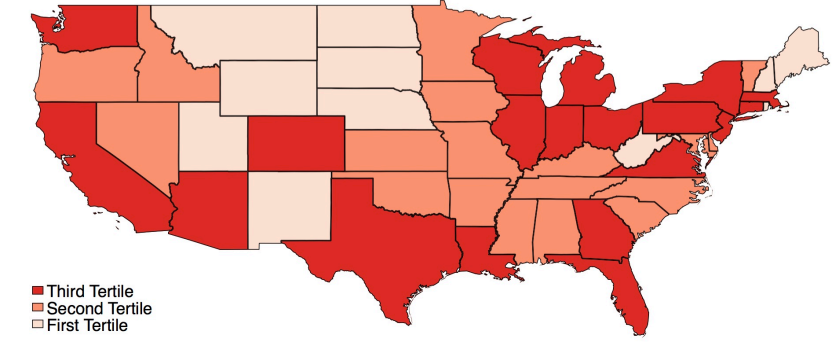
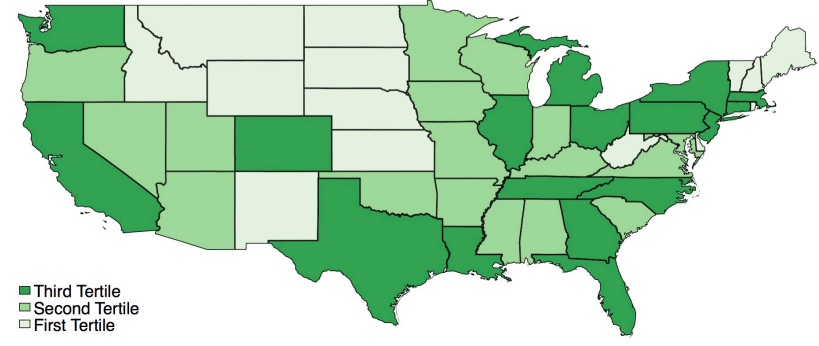
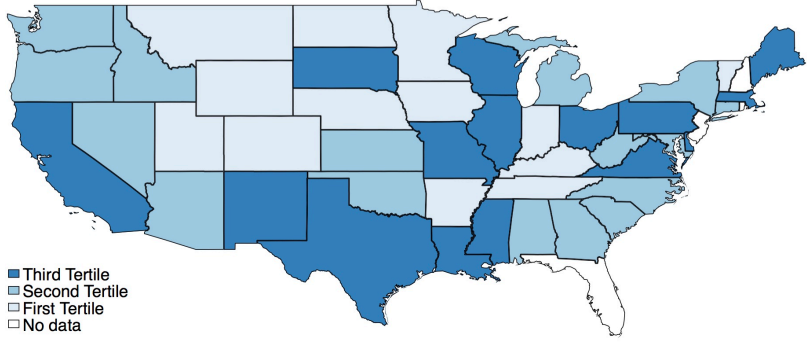
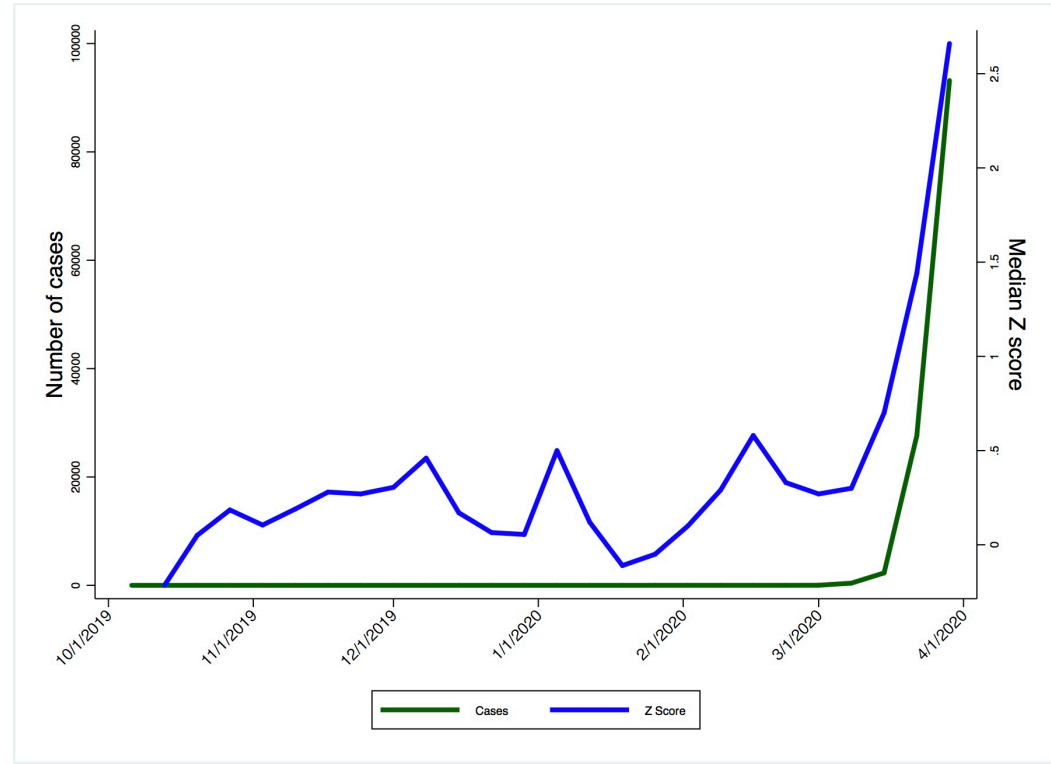
255 **cases or deaths.** (A) fnILI z-scores and cumulative cases/deaths for individual states in the

256 month of March. (B) Nationwide weekly median fnILI z-scores versus new cases nationwide.

257 (C) Nationwide weekly median fnILI z-scores versus deaths nationwide.

258

259

A**Z Score fnILI****Cumulative Cases****Cumulative Deaths****B****C**