

1 **Cases and deaths of COVID-19 in the United States:**
2 **Taylor's law in a plague year**

3 **Joel E. Cohen**

4 Laboratory of Populations, The Rockefeller University and Columbia University, 1230 York
5 Avenue, New York, NY 10065, U.S.A., Email cohen@rockefeller.edu;
6 Earth Institute and Department of Statistics, Columbia University;
7 Department of Statistics, University of Chicago

8 **ABSTRACT**

9 We discover, describe, and analyze large-scale, persistent patterns in the variation over space and
10 time of the cumulative numbers of COVID-19 cases and deaths in U.S. states and counties from
11 April 2020 to March 2021. Taylor's law of fluctuation scaling describes these patterns.

12 KEYWORDS: COVID-19, spatial variance, Taylor's law

13 **EXTENDED SUMMARY**

14 **Theoretical focus**

15 The numbers of COVID-19 cases and deaths vary from one country to another. Within the
16 United States, the numbers of COVID-19 cases and deaths vary from one state (the primary
17 political subdivision) to another, and within each state, from one county (the secondary political
18 subdivision) to another. Are there large-scale, persistent patterns in the variation, over space and
19 time, of the cumulative numbers of COVID-19 cases and deaths in the U.S.? How can these
20 patterns be used to improve preparation for, and responses to, this pandemic and others?

21 **Data**

22 On March 27, 2021, we downloaded the *New York Times* historical data file with cumulative
23 daily counts of coronavirus cases and deaths, by county and state, from the first reported case on
24 January 21, 2020, to March 26, 2021 [U.S. County-Level Data]([us-counties.csv](https://raw.githubusercontent.com/nytimes/covid-19-data/master/us-counties.csv))[Raw
25 CSV](<https://raw.githubusercontent.com/nytimes/covid-19-data/master/us-counties.csv>).
26 According to the documentation by the *New York Times*, "We are compiling this time series data
27 from state and local governments and health departments in an attempt to provide a complete
28 record of the ongoing outbreak." Moreover, "the numbers in the historical files are the final
29 counts at the end of each day," so final counts are not available until several days after the
30 nominal date of record; e.g., on March 27, 2021, final death counts were available for early
31 March but not including March 26, 2021. No attempt is made here to adjust the reported numbers
32 of cases and deaths for undercounts or overcounts. Henceforth "state" will refer to a state,
33 Washington, DC, or a U.S. territory or possession, and "county" will refer to the largest political
34 subunit of the "state" (such as a parish).

35 The data file has one line of data for each county in each U.S. state for each day with at least one
36 cumulative case or at least one cumulative death (or both), with a total of 1,160,632 lines of data
37 by county and day. (As a plausibility check, if the data covered 400 days \times 50 states \times 60
38 counties per state, there would be 1.2 million lines, not far from the actual number of lines.) This
39 file was too large to load into Microsoft Excel. All calculations used Matlab (version 2021a).
40 The columns of [us-counties.csv](#) are date (yyyy-mm-dd), county (a character string), state (a

41 character string), fips (a positive integer, identifying the location), cases (a nonnegative integer)
42 and deaths (a nonnegative integer).

43 **Research methods**

44 We analyzed only the subset of data pertaining to the first day of each month from 2020-02-01
45 (February 1, 2020) through 2021-03-01 (March 1, 2021), a total of 14 first days, and only those
46 counties with positive cumulative counts (of cases or of deaths) on that date. Some days had
47 counties with no cases but positive deaths and some days had counties with cases but no deaths.
48 The 37226 county-days that had positive counts of cases differed from the 28331 county-days
49 that had positive counts of deaths. (As a plausibility check, ignoring February and March 2020
50 when infection was in the early stages of spreading, if the data covered 12 days \times 50 states \times 60
51 counties per state, there would be 36,000 lines, not far from the actual number of lines.)

52 For each selected day and each state, we calculated the mean and the variance of the cumulative
53 counts (of cases or of deaths, respectively) on that day. We excluded any state with variance
54 equal to zero. A zero variance could arise in a state, for example, if all its counties with a positive
55 count (of cases or of deaths) happened to have equal counts. Such an improbable event happened
56 (rarely) early in the pandemic. These calculations produced a list of dates, states, the number of
57 counties in that state with a positive count, and the mean and the variance (when the variance
58 was positive) of the count over counties in that state with positive counts on that date.

59 The numbers of cases or deaths were low on February 1, 2020, and March 1, 2020. Therefore we
60 restricted further analysis to the remaining 12 first days, April 1, 2020 to March 1, 2021. For
61 each date, we plotted, on log-log coordinates, each state's variance of counts (vertical axis) as a
62 function of the state's mean of counts (horizontal axis), one \times marker per state, separately for
63 cases (Figure 1) and for deaths (Figure 2).

64 We compared these data points with the predictions of Taylor's law (TL). Widely applicable in
65 ecology (Taylor 1961, Eisler et al. 2008, Taylor 2019), demography (Cohen et al. 2013, 2018),
66 and epidemiology (Keeling and Grenfell 1999), TL predicts that a state's variance of counts
67 (variance computed over its counties with positive counts) should be proportional to some power
68 of the mean counts of that state (mean computed over its counties with positive counts). Thus on
69 a given fixed first day of month, for every state labeled j , TL predicts that

$$70 \quad (1) \quad \textit{variance of state } j \approx a \times (\textit{mean of state } j)^b, a > 0, j = 1, 2, \dots$$

71 We estimated the parameters a and b by fitting a straight line to the log-log transform of (1) by
72 ordinary least squares:

$$73 \quad (2) \quad \log_{10} \textit{variance of state } j = \log_{10} a + b \log_{10} \textit{mean of state } j.$$

74 We also calculated 95% confidence intervals for each estimate of a and b . If a state's counts on a
75 given day were randomly (Poisson) distributed over counties, then a straight line of slope 1
76 through the origin (or equivalently TL with $a = 1, b = 1$) should describe approximately the
77 observed (mean, variance) pairs on log-log coordinates.

78 **Initial findings**

79 In each of the 12 panels of Figure 1 (for cumulative cases) and Figure 2 (for cumulative deaths),
80 on log-log coordinates, each blue \times marker shows the (mean, variance) of counts for one state on

81 the given date; each yellow line shows the fitted TL (1), and each orange line with filled orange
82 circles shows TL with $a = 1, b = 1$ as predicted by a Poisson distribution of counts. The ranges
83 of the axes of all 12 panels are the same within each figure to make it easy see the shift, as time
84 passes, of the observed (mean, variance) from the lower left corner to the upper right corner.
85 Visual inspection of Figures 1 and 2 shows that:

86 (a) Poisson random variation is consistently a poor description of the relation of variance to
87 mean of cumulative cases and cumulative deaths by state over these 12 months in the U.S.

88 (b) TL is a plausible first approximation to (adequate description of) the relation of variance to
89 mean of cumulative cases and cumulative deaths by state over these 12 months in the U.S. All 24
90 values of the coefficient of determination R^2 exceed 0.90 for both cases and deaths.

91 Tabulations of estimates of a and b and their confidence intervals (which are not included here
92 because of the page limit for the extended summary) further show that:

93 (c) Of the 24 point estimates of the slope b of TL (12 estimates for cases, 12 for deaths), 23 were
94 close to 2 and had 95% confidence intervals that included 2. Only April 1, 2020, had a TL slope
95 $b = 2.9223$ with 95% confidence interval (2.632, 3.2126) that excluded 2.

96 (d) For cases, the point estimate of the TL intercept a initially rises sharply from April 1, 2020 to
97 peak at May 1, 2020, then declines gradually and deeply. For deaths, the estimated a initially
98 rises slowly, peaks at October 1, 2020, and declines slowly and more moderately.

99 Further research and the full manuscript will address why TL is a successful model of the
100 variability of cumulative cases and deaths by county in states of the U.S. and what can be learned
101 scientifically and practically from the above patterns.

102 REFERENCES

103 Cohen, Joel E., Xu, Meng and Brunborg, Helge 2013 Taylor's law applies to spatial variation in
104 a human population. *Genus* 69(1):25-60. DOI 10.4402/genus-491.
105 <http://scistat.cilea.it/index.php/genus/article/view/491>.

106 Cohen, Joel E., Bohk-Ewald, Christina, Rau, Roland 2018 Gompertz, Makeham, and Siler
107 models explain Taylor's law in human mortality data. *Demographic Research* 38(29):773-842,
108 March. DOI: 10.4054/DemRes.2018.38.29

109 Eisler, Zoltán, Bartos, Imre and Kertész, János 2008 Fluctuation scaling in complex systems:
110 Taylor's law and beyond. *Advances in Physics* 57(1):89-142. DOI:
111 10.1080/00018730801893043 Cite as: arXiv:0708.2053v2 [physics.soc-ph]

112 Keeling, M. J. and Grenfell, B. T. 1999 Stochastic dynamics and a power law for measles
113 variability. *Phil. Trans. R. Soc. Lond. B* 354:769-776. DOI 10.1098/rstb.1999.0429

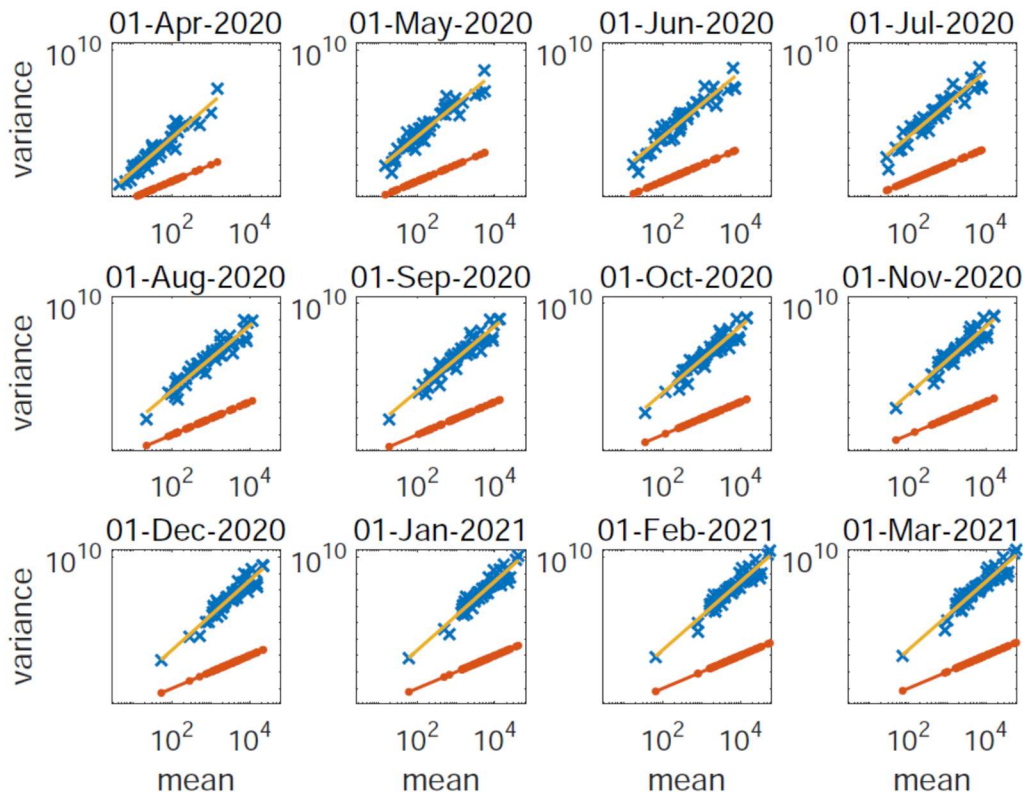
114 MATLAB 2021 (R2010a). Natick, Massachusetts: The MathWorks Inc.

115 Taylor, L. R. 1961 Aggregation, variance and the mean. *Nature* 189(4766):732-735, March 4.
116 doi:10.1038/189732a0

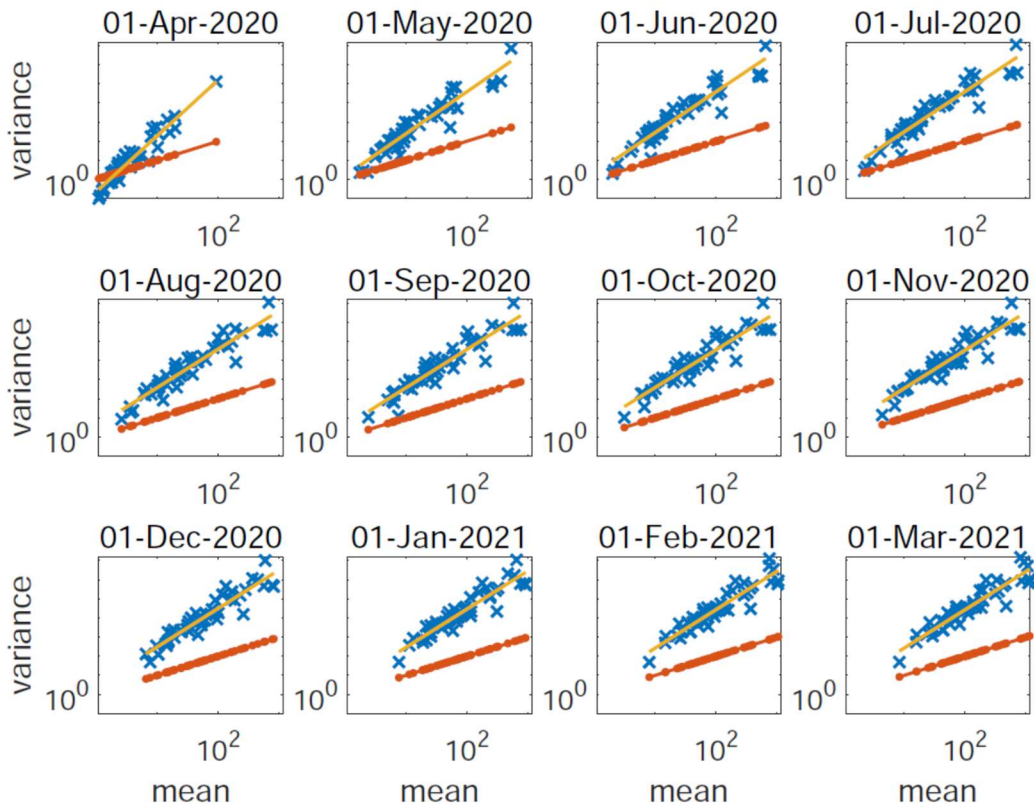
117 Taylor, R. A. J. 2019 *Taylor's Power Law: Order and Pattern in Nature*. Elsevier Academic
118 Press, Cambridge, MA.

119

120 **Figure 1. Covid-19 cumulative cases/day/county, grouped by U.S. state. See text for details.**



121 **Figure 2. Covid-19 cumulative deaths/day/county, grouped by U.S. state. See text for details.**



123