Volume 65, Issue 1, 2021

Journal of Scientific Research

Institute of Science,

Banaras Hindu University, Varanasi, India.

How Strong is the Epidemiological Evidence for a Role of Vitamin D Levels on COVID-19 Infections and Mortality?

Samer Singh*[1], Sangram Singh[2], Rajinder Kaur[1], Amita Diwaker[3], and Dhiraj Kishore[4]

[1]Centre of Experimental Medicine & Surgery, Institute of Medical Sciences, Banaras Hindu University, Varanasi - 221005, India.
samer.singh10@bhu.ac.in*, rajinderkaurabt@gmail.com
[2]Department of Biochemistry, Dr. RML Awadh University, Ayodhya – 224001. sangram_rml@yahoo.co.in
[3]Department of Obstetrics & Gynecology, Institute of Medical Sciences, Banaras Hindu University, Varanasi – 221005, India.
amitadiwakarkg@gmail.com
[4]Department of General Medicine, Institute of Medical Sciences, Banaras Hindu University, Varanasi – 221005, India.
dhirajkishore@gmail.com

Abstract: **Protective role for vitamin D serum levels on overall COVID-19 impact on populations had been suggested previously based upon single-point cross-sectional analysis of 8 April 2020 data (an early pre-peak-of-infections stage) from 20 European countries assuming comparable confounding variables for these populations. Comparative time-series cross-sectional analysis of the COVID-19 data from the relevant phase of pandemic [12 March (early pre-peak) to 26 July (late 1st post-peak-of-infections) 2020] was performed to reassess the strength of the assertion. The study subjects included 1,829,634 COVID-19 cases (11.11% global) and 179,135 associated deaths (27.45% global) on 26 July 2020. Previously suggested cross-sectional study design and methodology could not consistently and significantly (*p*-value≥0.05) support the notion of the potential protective role for vitamin D levels on COVID-19 incidence and mortality. However, the exponential correlative or the alternative simple regression analysis on *ln*/*log*$_{10}$ COVID-19 data for the time period indicated consistently negative association with vitamin D levels and also improved the overall predictive potential (*adjusted*-R^2 by 1.33 - 2.47 fold, *p*-value= 0.0457-0.0035, for cases/million; *adjusted*-R^2 by 2.21 - 3.74 fold, *p*-value= 0.0049-0.0228, for deaths/million). Considering vitamin D's role in immunity and the strong association observed with reduced mortality, randomized controlled trials are suggested to ascertain its potential in reducing the COVID-19 impact on populations.**

Index Terms: **COVID-19, SARS-CoV-2, Correlation, Vitamin D, Europe, Time-series, Cross-sectional, Epidemiological, Ecological**

I. INTRODUCTION

The urgency of the COVID-19 pandemic caused by SARS-CoV-2 had suddenly posed a great majority of researchers to focus on understanding COVID-19 and explore ways to mitigate its effects on the worldwide human population. By 26 July 2020, more than 16.4 million cases of COVID-19 had been reported with more than 652 thousand lives lost (Worldometers, 2020). European countries (including Turkey) had been disproportionately affected by COVID-19, accounting for 18.34% cases and 31.73% of deaths. Many potential protective variables for the populations have been suggested based upon statistical analyses, *e.g.*, Trained immunity, BCG vaccination coverage/policy, Vitamin D levels, Zinc levels, etc. (Berg, Yu, Salvador, *et al.* 2020; Escobara, Molina-Cruzb & Barillas-Mury, 2020; Ilie, Stefanescu & Smith, 2020; Singh, 2020a; Singh, 2020b and References therein). Vitamin D serum levels in European countries had been suggested to be potentially playing a protective role based upon a one-point cross-sectional analysis of the data by Ilie *et al*. (Ilie, Stefanescu & Smith, 2020). However, cautionary notes by us (Singh, Kaur & Singh, 2020) and more recently by Maruotti *et al*. (Maruotti, Belloc & Nicita, 2020) have been raised on the methodology and analysis. The letter by Maruotti *et al*. beautifully articulates the existing general fascination of interpreting any exploratory correlation as to have a cause-and-effect relationship by people at large. From time-to-time various bodies keep reminding the general audience on the same (References in Maruotti, Belloc & Nicita, 2020). In the light of suggestions made by us (Singh, Kaur & Singh, 2020) and others (Maruotti, Belloc & Nicita, 2020), we have analyzed the available data for the role of populations' mean vitamin D levels on COVID-19 incidence and mortality, and present a detailed time-series analysis of the COVID-19 data of the same twenty European countries whose mean serum vitamin D levels had been reported previously (Ilie, Stefanescu & Smith, 2020).

We hypothesized that if indeed vitamin D levels may have any potential protective role in COVID-19 as suggested

[*] Corresponding Author

DOI: 10.37398/JSR.2021.650120

previously, the impact of COVID-19 on the select European countries with different vitamin D levels would consistently negatively covary rather than appearing as a random fluctuation just observable on 8 April 2020 or as a result of seasonality. The methodology previously employed (Ilie, Stefanescu & Smith, 2020) failed to consistently support the notion of vitamin D levels as being a potential protective variable over the relevant study duration. However, the methodology suggested by us (Singh, Kaur & Singh, 2020) consistently supports the potential protective role of vitamin D levels on COVID-19 incidences and mortality during the study period (26 March 2020 to 26 July 2020).

II. Material And Methods

Twenty European countries for which average vitamin D levels of the populations are reported in the literature (Ilie, Stefanescu & Smith, 2020), have comparable exposure to UV-rays but have displayed variable COVID-19 impact, and have been analyzed in the recent past for the covariation of the vitamin D levels in the populations with the overall COVID-19 impact (Ilie, Stefanescu & Smith, 2020; Singh, Kaur & Singh, 2020) were selected for the current analysis. The time-series data of total COVID-19 cases and deaths reported per million populations for the time period starting from 12 March to 26 July 2020 (Table 1) were taken from a previous publication (Ilie, Stefanescu & Smith, 2020) and the coronavirus pandemic data portal https://www.worldometers.info/coronavirus/ (Worldometers, 2020). The country-specific mean serum vitamin D levels (nmol/L) data were taken from the previous publication (Ilie, Stefanescu & Smith, 2020). The study subjects included a total of 1,829,634 cases and 179,135 deaths accounting for 11.11% cases and 27.45% deaths from COVID-19 worldwide. Data transformation (*log₁₀* and *ln*) and basic statistics calculations (Pearson correlation coefficient, Regression analysis, curve-fitting - calculation of best-fit trend line) were performed using Microsoft Excel as done previously (Singh, 2020a; Singh, Kaur & Singh, 2020).

III. Results

The simple linear regression analysis previously used to propose potential protective role of the mean vitamin D levels of the populations on COVID-19 impact based upon a single day data (Ilie, Stefanescu & Smith, 2020) was found inadequate to support such an assertion when retested and applied over a period of time at the previously employed significance *p*-value cutoff (<0.05) (Table 2 and Fig. 1 A; refer to Table 1 for COVID-19 incidence and mortality data along with estimates of vitamin D for the population, basic statistical estimates). The negative correlation between vitamin D levels and COVID-19 infections observed in per million population was found to be

Table 1: Estimates of COVID-19 incidence (CpM) and mortality per million (DpM) during the study period (12 March to 26 July 2020) and Vitamin D levels (nmol/L)

Countries	CpM 12 Mar	DpM 12 Mar	CpM 26 Mar	DpM 26 Mar	*CpM 8 Apr	*DpM 8 Apr	CpM 12 Apr	DpM 12 Apr	CpM 26 Apr	DpM 26 Apr	CpM 12 May	DpM 12 May	CpM 26 May	DpM 26 May	CpM 12 Jun	DpM 12 Jun	CpM 26 Jun	DpM 26 Jun	CpM 12 Jul	DpM 12 Jul	CpM 26 Jul	DpM 26 Jul	Population	Vit. D levels nmol/L
Spain	665.43	1.84	3118.24	93.35	3137	314	4667.48	368.05	5243.71	495.96	5493.08	574.36	5622.58	595.47	5738.28	604.71	5819.68	606.06	6019.13	607.49	6584.73	608.11	46757733	42.5
Iceland	342.51	0.00	2347.79	5.85	4736	18	4979.54	23.42	5245.93	29.27	5272.28	29.27	5281.06	29.27	5289.84	29.27	5327.90	29.27	5365.96	29.27	5406.94	29.27	341598	57.0
Ireland	14.15	0.20	367.73	3.84	1230	48	1950.46	67.52	3892.64	218.94	4697.25	300.01	4999.08	325.68	5103.19	343.88	5136.34	348.93	5181.02	352.98	5232.17	356.21	4946514	56.4
Belgium	34.40	0.34	537.60	41.30	2019	193	2556.27	370.16	3977.84	627.10	4637.01	753.16	4953.97	797.05	5157.80	821.97	5268.77	829.30	5398.11	833.52	5667.21	836.63	11597764	49.3
UK	7.89	0.13	155.54	12.92	895	105	1123.82	180.71	2038.02	352.76	3019.69	472.95	3536.57	532.41	3906.24	577.36	4125.05	592.38	4274.81	601.16	4406.99	605.08	67943420	47.4
Italy	250.17	16.87	1333.08	136.33	2306	292	2586.60	330.19	3269.98	441.89	3659.54	512.71	3814.12	546.69	3909.39	567.70	3970.11	576.73	4021.39	580.80	4071.91	583.34	60447245	50.0
Switzerland	100.18	0.81	1363.11	22.16	2686	103	2933.15	127.64	3353.93	185.81	3506.16	215.47	3550.13	221.01	3584.98	223.66	3633.80	226.43	3795.03	227.13	3971.49	228.17	8664751	46.0
Sweden	76.27	0.10	296.36	13.26	834	68	1079.30	134.03	1865.89	253.13	2775.15	369.85	3523.17	438.21	4982.21	500.53	6450.65	536.53	7209.05	558.89	7504.02	568.68	10109375	73.5
Portugal	7.65	0.00	347.73	5.89	1289	37	1627.29	49.45	2324.22	88.60	2738.77	114.11	3042.34	131.67	3549.91	147.67	4009.69	152.57	4563.66	162.88	4921.99	168.47	10191814	39.0
France	44.04	0.93	446.50	25.97	1671	167	1461.06	220.22	1907.82	349.69	2147.53	412.99	2229.12	436.54	2393.48	449.47	2495.31	455.66	2635.48	459.23	2795.92	462.50	65296963	60.0
Netherlands	35.82	0.29	433.52	25.32	1199	131	1492.73	159.68	2207.86	261.07	2507.67	321.45	2659.00	341.64	2827.19	353.13	2917.27	356.05	2976.60	357.91	3088.85	358.20	17141028	59.5
Germany	32.75	0.07	524.15	3.19	1309	25	1525.21	36.05	1882.08	71.29	2065.81	92.31	2162.64	101.38	2233.77	105.73	2319.04	107.67	2385.26	108.96	2466.27	109.79	83827321	50.1
Denmark	116.30	0.00	323.87	7.07	933	38	1065.31	47.11	1479.60	72.82	1827.45	90.93	1971.87	97.14	2072.12	102.49	2187.04	104.22	2233.80	105.08	2318.69	105.77	5795502	65.0
Turkey	0.01	0.00	42.96	0.89	453	10	674.17	14.18	1303.58	33.20	1674.60	46.09	1879.22	52.05	2074.01	56.56	2302.37	59.95	2521.14	63.48	2676.28	66.44	84482851	51.8
Norway	147.38	0.18	621.22	2.58	1123	19	1202.10	23.58	1386.69	37.03	1502.76	42.00	1544.40	43.29	1588.06	44.58	1627.11	45.87	1654.56	46.43	1679.62	46.98	5428015	65.0
Estonia	20.35	0.00	405.52	0.75	893	18	986.67	18.84	1238.43	36.93	1316.07	45.98	1382.40	47.49	1484.91	47.49	1496.97	47.49	1518.08	47.49	1533.15	47.49	1326680	51.0
Finland	19.67	0.00	172.86	0.90	449	7	536.62	10.10	825.68	34.28	1083.16	49.62	1195.93	56.30	1276.23	58.64	1297.52	59.18	1315.92	59.18	1333.97	59.36	5542120	67.7
Czechia	10.92	0.00	189.32	0.84	488	9	561.32	12.88	693.23	20.54	769.50	26.42	846.98	29.59	929.87	30.71	1032.56	32.39	1231.77	32.86	1432.48	34.63	10712210	62.5
Hungary	1.66	0.00	27.03	1.04	93	6	146.02	10.25	258.90	28.17	343.09	44.01	390.52	51.68	419.73	57.48	427.39	59.86	438.47	61.62	459.28	61.72	9656316	60.6
Slovakia	3.85	0.00	41.39	0.00	125	0.4	135.90	0.37	252.56	3.30	268.31	4.94	277.10	5.13	282.41	5.13	300.91	5.13	348.16	5.13	399.08	5.13	5460073	81.5
Average	**96.57**	**1.09**	**654.78**	**20.17**	**1393.4**	**80.42**	**1664.55**	**110.22**	**2232.43**	**182.09**	**2565.24**	**225.93**	**2743.11**	**243.98**	**2940.18**	**256.41**	**3107.27**	**261.58**	**3254.37**	**265.07**	**3397.55**	**267.10**	**25783464.65**	**56.79**
STDEV	**161.25**	**3.74**	**805.92**	**34.92**	**1129.98**	**94.615**	**1322.18**	**123.97**	**1486.62**	**186.11**	**1575.98**	**223.81**	**1633.18**	**239.23**	**1720.83**	**249.95**	**1850.68**	**254.23**	**1950.42**	**256.69**	**2037.58**	**257.91**	**29608125.08**	**10.61**

Note: Cases per million: CpM; Deaths Per million (DpM); Vitamin D levels, *CpM and *DpM estimates from Ilie, Stefanescu & Smith, 2020, others from Worldometers, 2020

Table 2. Time-series correlation and regression analysis of vitamin D mean serum level (nmol/L) and COVID-19 incidence (top panel) and mortality per million (bottom panel) data of European countries.

		COVID-19 CASES per MILLION (CpM) vs POPULATIONS' Vit. D LEVELS (nmol/L										
	Data-Date	12 Mar	26 Mar	8 Apr	12 Apr	26 Apr	12 May	26 May	12 Jun	26 Jun	12 Jul	26 Jul
Simple linear Regression	**Corr r(20)/R value**	-0.2760	-0.4080	-0.4435	-0.5072	-0.5448	-0.5412	-0.5265	-0.4727	-0.4054	-0.3908	-0.4034
	R Square	0.0762	0.1665	0.1967	0.2572	0.2968	0.2929	0.2772	0.2235	0.1644	0.1527	0.1627
	Adjusted R Square	0.0248	0.1202	0.1521	0.2160	0.2577	0.2536	0.2370	0.1803	0.1180	0.1057	0.1162
	Standard Error	159.2335	755.9334	1040.5152	1170.7413	1280.8313	1361.5566	1426.5450	1557.9696	1738.0987	1844.5105	1915.5266
	F-value	1.4837	3.5957	4.4079	**6.2335**	**7.5960**	**7.4557**	**6.9031**	**5.1799**	3.5411	3.2446	3.4985
	p-value	**0.2389**	**0.0741**	**0.0501**	**0.0225**	**0.0130**	**0.0137**	**0.0171**	**0.0353**	**0.0761**	**0.0884**	**0.0778**
		LOG TRANSFORMED (LOG$_{10}$) CASES per MILLION (CpM) vs POPULATIONS' Vit. D LEVELS (nmol/L)										
		12 Mar	*26 Mar*	*8 Apr*	*12 Apr*	*26 Apr*	*12 May*	*26 May*	*12 Jun*	*26 Jun*	*12 Jul*	*26 Jul*
Exponential Model or Log$_{10}$ Transformed Varaible- Linear Regression	**Corr r(20)/R value**	-0.0526	-0.4515	-0.5685	-0.6205	-0.6202	-0.6101	-0.6015	-0.5786	-0.5544	-0.5476	-0.5477
	R-square	0.0028	**0.2039**	**0.3232**	**0.3850**	**0.3846**	**0.3722**	**0.3618**	**0.3348**	**0.3074**	**0.2999**	**0.3000**
	Adjusted R Square	-0.0526	0.1597	0.2856	0.3508	0.3505	0.3373	0.3263	0.2978	0.2689	0.2610	0.2611
	Standard Error	1.0500	0.4996	0.3591	0.3320	0.2983	0.2951	0.2952	0.3035	0.3116	0.3081	0.3030
	F-value	0.0499	**4.6101**	**8.5972**	**11.2685**	**11.2514**	**10.6713**	**10.2045**	**9.0578**	**7.9891**	**7.7107**	**7.7129**
	p-value	**0.8258**	**0.0457**	**0.0089**	**0.0035**	**0.0035**	**0.0043**	**0.0050**	**0.0075**	**0.0112**	**0.0124**	**0.0124**

		DEATHS per MILLION (DpM) vs POPULATIONS' Vit. D LEVELS (nmol/L)										
	Data-Date	12 Mar	*26 Mar*	*8 Apr*	*12 Apr*	*26 Apr*	*12 May*	*26 May*	*12 Jun*	*26 Jun*	*12 Jul*	*26 Jul*
Simple linear Regression Analysis	**Corr r(20)/R value**	-0.1948	-0.3683	-0.4378	-0.4193	-0.3990	-0.3745	-0.3596	-0.3448	-0.3340	-0.3303	-0.3291
	R Square	0.0379	0.1357	0.1917	0.1758	0.1592	0.1403	0.1293	0.1189	0.1116	0.1091	0.1083
	Adjusted R Square	-0.0155	0.0876	0.1468	0.1300	0.1125	0.0925	0.0809	0.0699	0.0622	0.0596	0.0588
	Standard Error	3.7722	33.3574	87.3958	115.6291	175.3307	213.2015	229.3434	241.0568	246.1939	248.9276	250.2122
	F-value	0.7099	2.8252	4.2684	3.8389	3.4087	2.9372	2.6728	2.4285	2.2609	2.2037	2.1864
	p-value	**0.4105**	**0.1101**	**0.0535**	**0.0658**	**0.0814**	**0.1037**	**0.1194**	**0.1366**	**0.1500**	**0.1550**	**0.1565**
		LOG TRANSFORMED (LOG$_{10}$) DEATHS per MILLION (DpM) vs POPULATIONS' Vit. D LEVELS (nmol/L)										
	Data-Date	12 Mar	*26 Mar*	*8 Apr*	*12 Apr*	*26 Apr*	*12 May*	*26 May*	*12 Jun*	*26 Jun*	*12 Jul*	*26 Jul*
Exponential Model or Log$_{10}$ Transformed Varaible- Linear Regression	**Corr r(20)/R value**	-0.3149	-0.6024	-0.6023	-0.5899	-0.5372	-0.5156	-0.5114	-0.5085	-0.5061	-0.5084	-0.5102
	R-square	0.0991	0.3629	0.3627	0.3480	0.2886	0.2659	0.2615	0.2585	0.2562	0.2585	0.2603
	Adjusted R Square	0.0491	0.3275	0.3273	0.3118	0.2491	0.2251	0.2205	0.2173	0.2148	0.2173	0.2192
	Standard Error	1.3951	0.9007	0.5767	0.6042	0.5164	0.5120	0.5150	0.5197	0.5208	0.5204	0.5190
	F-value	1.9807	**10.2533**	**10.2452**	**9.6064**	**7.3023**	**6.5184**	**6.3749**	**6.2764**	**6.1988**	**6.2747**	**6.3330**
	p-value	**0.1763**	**0.0049**	**0.0050**	**0.0062**	**0.0146**	**0.0200**	**0.0212**	**0.0221**	**0.0228**	**0.0221**	**0.0216**

Note: The estimates of correlation *r*(20)/R-value, R-square (R^2), *adjusted-R^2*, the *p*-value for the exponential model, and the linear regression of *ln* transformed data will be same as that displayed for *log_{10}* transformed data but they would not suppress the apparent variability in data allowing easy variable discovery (compare Fig. 1D with 1C and 1B, see discussion). The highlighted 8 April 2020 COVID-19 data was previously analyzed in a cross-sectional study (Ilie, Stefanescu & Smith, 2020) to propose vitamins D's protective role.

statistically significant (*p*-value <0.05) only between 12 April to 12 June 2020, and remained non-significant thereafter upto the end of the current analysis period, *i.e.*, 26 July 2020. The correlation between deaths per million population with vitamin D levels never reached the suggested statistical significance level (*p*-value <0.05) during the whole study period, *i.e.*, 8 April to 26 July 2020, (*p-value* for correlation varied from 0.0535 to 0.1550). The presence of heteroscedasticity in the early infection data (data not shown) further makes the regression analysis prone to over/under-estimation.

The exponential correlative modeling of the data (Fig. 1 B) as suggested by us for the time period improved the correlative predictive potential of populations vitamin D levels, *i.e.*, R^2 by 1.27-1.96 fold and *adjusted-R^2* by 1.33-2.47 fold for the COVID-19 cases per million, while the change in R^2 by 1.81-2.67 fold and *adjusted-R^2* by 2.21-3.74 fold for deaths per million population upto the end of the current analysis period, *i.e.*, 26 March (early stage) to 26 July 2020 (late post-peak of infections), was observed. Its implementation over the simple linear regression was suggested by us for the exploratory statistical analysis of the time-series data preferably from post-peak of infections to arrive at more dependable estimates (Singh, Kaur & Singh, 2020). The linear regression modeling of the *log*-transformed data (natural (*ln*) as well as *log_{10}*) for assessing its potential predictive value (assuming cause and effect relationship is established in the future) displayed the same stable significant negative covariation (*p*-value<0.05) of the COVID-19 incidences and associated mortality/deaths in the per million population starting from 26 March (early stage) to 26 July 2020 (late post-peak of infections) stage of the pandemic (Fig 1C and D and Table 2), with concomitant reduction in heteroscedasticity in the data set as expected (data not shown).

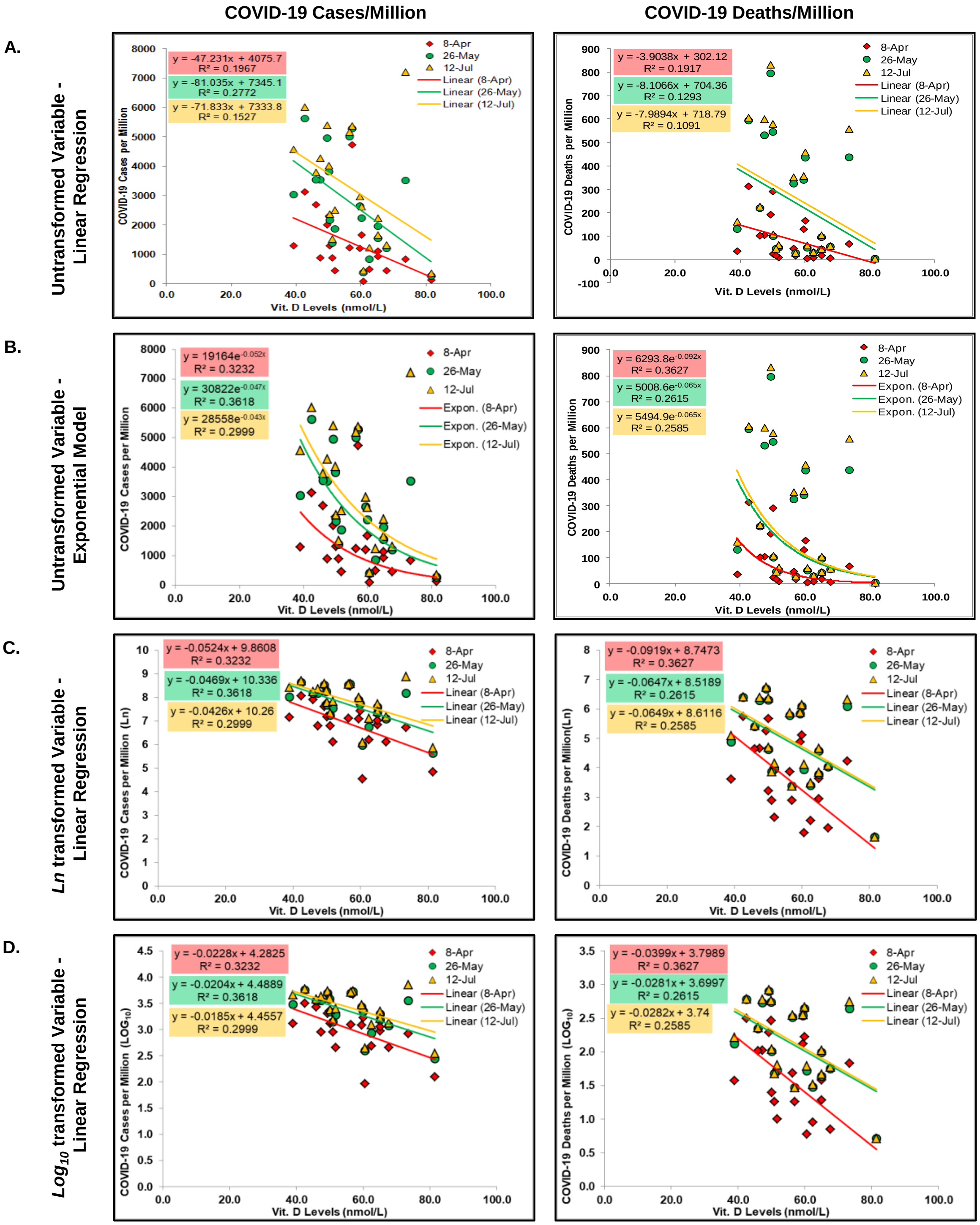

Figure 1: Time-series Regression analysis of COVID-19 impact and mean serum vitamin D levels in European populations. The vitamin D levels negatively correlated with the COVID-19 cases (left) and deaths per million populations (right) of the European countries during different time points of the study period. Representative analysis for 8 April, 26 May, and 12 July 2020 data is shown [See Table 2 for analysis of the period extending from 12 March (early pre-peak-of-infections) to 26 July (late post-peak-of-infections) 2020]. The correlation and regression estimates obtained on simple linear regression (A), improved on the application of exponential model (B) or linear regression model on transformed COVID-19 estimates of the population - *ln* (C) and log_{10} (D). The transformation of data progressively suppressed the apparent variability/scatter in the data (compare 1C and 1D with 1B). The simple linear regression model without log COVID-19 estimate did not statistically support (significance/ *p*-value<0.05) the negative covariation of the COVID-19 incidences or mortality with vitamin D levels on 8th April 2020 as well as at other time points analyzed (see Table 2).

Refer to Fig. 1 for the visual comparison of the data covariation (trend)/ regression analysis performed, *i.e.*, simple linear regression analysis, a more appropriate exponential curve fitting for the exploratory analysis, and the linear regression modeling/analysis of the data post *ln* and log_{10} transformation of COVID-19 incidence and mortality per million populations. Note the transformation of data leads to visual suppression of underlying anomalies and any apparent heteroscedasticity that could have also indicated the presence of unidentified variable, with a progressively increased chance of missing out the important variable/anomalies that could have been discovered (compare Fig 1B with Fig. 1C and D).

IV. Discussion & Conclusion

The race to discover potential protective variables for COVID-19 incidence and mortality is ongoing with a range of things being proposed to be correlated with COVID-19 impact (Berg, Yu, Salvador, *et al.* 2020; Escobara, Molina-Cruzb & Barillas-Mury, 2020; Ilie, Stefanescu & Smith, 2020; Maruotti, Belloc & Nicita, 2020; O'Neill & Netea, 2020; Singh, Kaur & Singh, 2020; Singh, Maurya & Singh, 2020; Singh, 2020a; Singh, 2020b; and references therein) though not all necessarily having a potential protective role and without the necessary rigor for such assertions. Based upon a correlation and regression analysis of a single day COVID-19 incidence and mortality data with serum vitamin D levels in populations (Ilie, Stefanescu & Smith, 2020), the vitamin D had been proposed as a potential protective variable that may be explored further through dedicated studies. As indicated by us and others the one-day data analysis had weaknesses (Maruotti, Belloc & Nicita, 2020; Singh, Kaur & Singh, 2020) and it would have been better had the analysis been not constrained by commonly used *p*-value cutoffs and improper model application on the biological data set. Preferably the estimation of the correlation should have been performed at multiple time-points using the phase-matched post-peak of infections data to reduce the effect of potential data reporting delays in infection and adverse outcome and arrive at potentially dependable estimates.

We now have COVID-19 data for the European countries for a larger time frame that by 26 July 2020 included a total of 1,829,634 cases and 179135 deaths accounting for the worldwide 11.11% cases and 27.45% deaths (Worldometers, 2020). The reanalysis of the data for an extended period indicates the potential problem of the analysis and model presented by Ilie *et al.* (Ilie, Stefanescu & Smith, 2020) which was previously predicted/indicated by us and others (Maruotti, Belloc & Nicita, 2020; Singh, Kaur & Singh, 2020). The simple linear regression modeling of the covariation of vitamin D levels with COVID-19 cases per million as suggested by the authors did not indicate a statistically significant correlation (p-value ≥ 0.05) upto 8 April, which became statistically significant (at p-value cut off < 0.05) by 12 April and stayed that way till 12 June then again become insignificant and remained so till the end of the analysis period, *i.e.*, 26 July 2020. The correlation between deaths per million and vitamin D levels *never* ever reached the commonly used statistical significance level suggested by Ilie *et al.* during the whole study period starting from 12 March to 26 July including that on the 8th April 2020 when it was the closest to the significant *p*-value cutoff of 0.05 [$r(20)$ - 0.4378, *p*-value 0.0535]. In the study period, the *p*-value had actually progressively increased from 0.0535 on 8th April to 0.1565 on 26 July 2020. Thus, the reanalysis of indicated 8th April data as well as the time-series data of the period extending upto 26 July 2020 using previously suggested methodology to model or evaluate the vitamin D levels and COVID-19 impact on the populations, could not endorse a correlational potentially protective role for the populations mean serum vitamin D levels.

However, exponential modeling of the covariation (or the linear regression analysis of vitamin D levels and ln / log_{10} transformed COVID-data) improved the correlation as well as correlative predictive potential over the simple linear regression analysis upto the end of the current analysis period, *i.e.*, 26 March to 26 July 2020. For the COVID-19 cases per million population the R^2 and *adjusted*-R^2 increased by 1.27-1.96 fold and 1.33-2.47 fold, respectively. The R^2 and *adjusted*-R^2 for deaths per million population, increased by 1.81-2.67 fold and 2.21-3.74 fold, respectively. As apparent from table 2 the standard error term, F-value and *p*-values combined show a drastic transition between 12 March (early stage of COVID-19 pandemic) to 8-12 April 2020 then stabilizes and remains so. More importantly, the analyzed parameters (correlation, R^2, adjusted-R^2, *p*-value) remain more stable post-peak of infections (12 May 2020) as predicted and suggested previously by us, suggesting the potential existence of a cause-and-effect relationship between vitamin D levels and COVID-19 impact that may be tested in dedicated future studies.

However, some of the limitations about the available data and hence the analysis presented must be borne in mind while designing experiments/trials to explore the protective potential of serum vitamin D in COVID-19. The presence of data on the overall deficiency or sufficiency variation in vitamin D levels of the affected populations or the individuals who got infected or had an adverse outcome in the current pandemic would have significantly improved the confidence in analysis. Robustness of the analysis could have been further improved if the data of mean serum vitamin D levels of other European countries that have had moderate impact of COVID-19 would have been available and included in the analysis. The countries currently included in the analysis, seemingly, are from affected extremes (severe and low) accounting for 11.11% of the cases but 27.45% deaths worldwide opposed to 18.34% cases and 31.73% deaths accounted by European countries combined together on 26 July 2020 (Worldometers, 2020). It makes the analysis prone to biases due to relative disparate data sets (Singh, 2020b). Nevertheless, the potential protective impact of higher (or lower) vitamin D levels on COVID-19 incidence and adverse outcome possibility could be ascertained through controlled trials evaluating the adverse impact due to prevailing vitamin D levels or alternatively that may be resulting from over or under supplementation of vitamin D with or without a medical prescription in the populations.

The analysis as such presented, also highlights the pitfalls of our fascination with p-value cutoffs, potential data dredging knowingly or unknowingly, and the need for more rigor exercised by authors, reviewers, and editors of the journal alike in the current testing times (References in Maruotti, Belloc & Nicita, 2020). The ongoing urge or fascination to go on to fit the data to unprecedented level in literature and the apparent data dredging seen during the current COVID-19 pandemic has to be balanced. It has been observed, besides the necessary exploratory statistical analysis of the existing data to find any variable that could be potentially protective or be making individuals vulnerable that could be evaluated clinically or through controlled trials, more efforts appear to be invested in making the correlations look significant by inclusion of the unequal data points, the inclusion of variables logically not seemed to have cause and effect relationship, the transformation of the variables to derive higher association/covariation/correlation with the highly significant 'lower unrealistic *p*-value' without any metrics for size effect, wherever possible. We do agree that statistical modeling showing highly correlated variables should be evaluated with caution and the journals/editors may take a lead in informing/educating the audience and reviewers alike as had been advocated in the recent past (References in Maruotti, Belloc & Nicita, 2020). In the same vein as indicated by Maruotti *et al*., we have presented the perils of the one-point data analysis in the exploratory epidemiological study and indicate how the data analysis methodology employed by Ilie *et al*. may have been inadequate to derive the conclusions. We would submit that the senior editors of esteemed journals may take a more cautionary approach in educating the junior editors and reviewers about the statistical analysis, perils of one-point cross-sectional analyses, the fascination of unrealistic *p*-values that may make something appear relevant when it is not as the *p*-values drastically change with the number of data points in a small data set, the inclusion of selective disparate data sets, *etc*. Equally important would be to avoid visual depiction/presentation of only transformed data as much as possible in the exploratory analysis so as to allow the discovery of other potential variables that may otherwise remain hidden in the plain sight on generally employed data transformations such as ln/log_{10}, square root, etc. The least transformed (minimum necessary) data presentation could allow discovery of potential protective variables not necessarily by the authors of the study who may have already missed it anyway or had reasons to believe it was not important, but by onlookers who may have access to knowledge of other variables. For instance, in the data set presented the population of Sweden (mean vitamin D level 73.5 nmol/L) that was apparently more protected from COVID-19 infection on 8 April 2020 as compared to many countries, *e.g*., Spain, Iceland, Italy, Switzerland, Denmark, Norway, had surpassed all nations by 12 July 2020, potentially due to a variable only applicable to Sweden. Nonetheless, the stark outlier could be more easily identified in the untransformed data presentation by someone who may know that Sweden as a nation among all the countries analyzed had one of the lowest stringent measures in place to prevent the spread of COVID-19, so increasing the chance of further refinement of the analysis presented. Similarly, other readers/scientists who may be short on time could also contribute to the identification of other variables which may be helpful in the long run to find out a solution to the problem in hand. So, in our view, for the exploratory statistical analysis, the data may be presented with the least transformations employed to increase the chances of the discovery of new variables. The heteroscedasticity may be allowed to stay in plain sight, rather modeling parameters could be changed to show the potential relationship and in no way the exploratory regression analysis be construed as a one depicting cause and effect relationship. We like to add, for all the exploratory analyses performed the authors may also proactively reassert at quite a few places to indicate the analysis being exploratory in nature or some consensus among the scientific fraternity could be reached to clearly label such studies.

In conclusion, our exploratory study using European countries COVID-19 incidence and mortality data starting from 26 March to 26 July 2020 indicates consistently negative (potentially protective) covariation with serum vitamin D levels, and suggest vitamin D as one of the potential candidates for evaluation by dedicated trials/studies using matched (*i.e*., age, comorbidities, sex, genetic background, disease severity, etc.) control and test groups to once and for all establish the role of vitamin D levels on COVID-19 incidences or the adverse outcome, if any, whether negative or positive. The controlled trials initiated in different countries for evaluating the link between vitamin D levels and COVID-19 outcomes, *e.g*., Argentina (NCT04411446), France (NCT04344041), Iran (IRCT20200324046850N1), Spain (NCT04334005), could provide the necessary evidence for the same (Mohan, Cherian & Sharma, 2020).

Author Contributions Statement: SS[1] conceived the idea, collected and analyzed the data, and written the manuscript. SS[2], RK, AD and DK helped with data collection and provided inputs.

Conflict of interest: There is no conflict of interest to disclose.

Ethical statement: The study is compliant with the ethical standards. Considering the design of the study no human or animal rights were infringed upon.

Informed consent: Considering the design of the study no informed consent was necessary.

Note: A version of manuscript is deposited on MedRxiv preprint server https://doi.org/10.1101/2020.11.20.20235705

Acknowledgment

No specific source of funding was utilized for the current study. However, SS[1] acknowledges the funding support to his laboratory from Institute of Eminence (IoE) seed grant, Banaras Hindu University.

References

Berg MK, Yu Q, Salvador CE, Melani I, Kitayama S. (2020) Mandated Bacillus Calmette-Guérin (BCG) vaccination predicts flattened curves for the spread of COVID-19. Sci Adv. 6, eabc1463

Escobara L.E., Molina-Cruzb A., Barillas-Mury C. (2020) BCG vaccine protection from severe coronavirus disease 2019 (COVID-19). PNAS, 117, 17720-17726; www.pnas.org/cgi/doi/10.1073/pnas.2008410117

Ilie, P.C., Stefanescu, S. & Smith, L. (2020) The role of vitamin D in the prevention of coronavirus disease 2019 infection and mortality. Aging Clin Exp Res 32, 1195–1198. https://doi.org/10.1007/s40520-020-01570-8

Maruotti, A., Belloc, F. & Nicita, A. Comments on: The role of vitamin D in the prevention of coronavirus disease 2019 infection and mortality. Aging Clin Exp Res 32, 1621–1623 (2020). https://doi.org/10.1007/s40520-020-01618-9

Mohan, M., Cherian, J.J., Sharma, A. (2020) Exploring links between vitamin D deficiency and COVID-19. PLoS Pathog, 16(9): e1008874. https://doi.org/10.1371/journal.ppat.1008874

O'Neill L.A.J., Netea M.G. (2020) BCG-induced trained immunity: can it offer protection against COVID-19? Nat Rev Immunol, 20, 335–337. https://doi.org/10.1038/s41577-020-0337-y

Singh S. (2020a) Covariation of Zinc Deficiency with COVID19 Infections and Mortality in European Countries: Is Zinc Deficiency a Risk Factor for COVID-19? Journal of Scientific Research, 64(2): 153-157. http://dx.doi.org/10.37398/JSR.2020.640222

Singh S. (2020b) BCG Vaccines May Not Reduce Covid-19 Mortality Rates. medRxiv2020.04.11.20062232. https://doi.org/10.1101/2020.04.11.20062232

Singh S., Kaur R., Singh R.K. (2020) Revisiting the role of vitamin D levels in the prevention of COVID-19 infection and mortality in European countries post infections peak. Aging Clin Exp Res, 32, 1609–1612. https://doi.org/10.1007/s40520-020-01619-8

Singh, S., Maurya, R.P., Singh, R.K. (2020) 'Trained immunity' from *Mycobacterium* spp. exposure or BCG vaccination and COVID-19 outcomes. PLoS Pathog, 16(10), e1008969. https://journals.plos.org/plospathogens/article?id=10.1371/journal.ppat.1008969

Worldometers, (2020) COVID-19 Coronavirus Pandemic. https://www.worldometers.info/coronavirus/

***