Evidence Supports a Causal Model for Vitamin D in COVID-19 Outcomes

Davies G, Garami AR, Byers J
Draft 1.0 - preprint submission draft

Abstract

We analyse global data for COVID-19 deaths and recoveries and show that outbreak severity displays a striking latitude relationship with a northern hemisphere bias. Transmission rates can be explained by seasonal weather conditions, but this does not account for observed variations in fatality rates. Many factors point to Vitamin D as a candidate explanation but historical controversy surrounding Vitamin D studies and the lack of a coherent framework for causal inference has hampered acceptance of this explanation despite a wealth of evidence in its favour.

We analyse global COVID-19 data using Causal Inference, constructing two contrasting directed acyclic graph (DAG) models, one causal and one acausal, and set out clearly multiple predictions made by each model. We show that observed data strongly match predictions made by the causal model but largely contradict those of the acausal model. We explore historic evidence further supporting the causal model.

We review biochemical mechanisms that may explain the various ways in which vitamin D acts. We detail the mechanisms by which the SARS-Cov-2 virus causes the disease and known pathways that involve Vitamin D and show how these both protect against viral infection, as well as ameliorating disease symptoms in COVID-19 and other respiratory diseases.

We examine the factors that govern confidence in causal inference models and conclude that a high level of confidence in a causal beneficial role for Vitamin D is justified.

Introduction

Severe COVID-19 outbreaks show a striking latitude relationship with severe outbreaks occurring exclusively in locations above the 30°N latitude line. Global reports of deaths and recoveries reveal that transmission rates and fatality rates from January to March 2020 were significantly determined by latitude.

Past coronaviruses and influenza viruses have displayed very strong seasonality [1] and since latitude governs the seasons this goes some way towards an explanation. Winter provides a more favourable environment for viral transmission: enveloped viruses are fragile and more easily destroyed by heat and UV light than other viruses [2], but this only accounts

for lower transmission rates. It does not explain why fatality rates also strongly depended on latitude.

A helpful role of Vitamin D in the response to the COVID-19 pandemic has been discussed [3]. Vitamin D is a logical candidate hypothesis since it is a steroid hormone with cell receptors in most tissue cell types [8] and plays an important role in immune system health. Deficiency is known to be seasonal since it is produced in the skin in strong sunlight. Risk factors for low vitamin D are old age, winter, darker skin pigmentation , less sunlight exposure, dietary habits and absence of vitamin D fortification [4]. The country worst affected in Europe by COVID-19, Italy, has very high levels of vitamin D deficiency [5].

Methods

Data Sources, Processing and Code

COVID-19 data on cases, recoveries and deaths was provided by John Hopkins CSSE's public github repository [6]. Some CSV files were converted to Excel format for previsualisation checks before being imported into Matlab.

Data for population by latitude was provided by the Center for International Earth Science Information Network, Columbia University [7].

Data and code are available online at github: https://github.com/gruffdavies/GD-COVID-19.

Latitude Analysis

We analysed COVID-19 fatalities by latitude up to 28th March 2020 aggregating total deaths for each location into data bins of size five degrees each over the range 40°S to 70°N. We aggregated world population latitude data into corresponding bins and calculated Deaths per Million as a function of latitude.

We compared severity by location with the known timeline of spread of COVID-19 around the globe to show that timing of infection was not involved in determining outbreak severity.

We then performed a detailed analysis of outbreak severity by location for all 239 global reporting locations over the same period. Outbreak severity was judged using the ratio of reported recoveries to deaths and the total number of deaths. We calculated the Epidemic Severity Index (ESI) [8] with a reference severity defined by a hospital survival ratio, S, of 6.5, corresponding to a disease severity approximately twice that of season flu in the US.

We plot ESI scores using geobubble plots using both linearised and log scales. In the linearised ESI plots bubble area is directly comparable to the total number of deaths. The logarithmic ESI form is is used to show locations with severity index values too small to be seen on the linearised plots.

We look in detail at locations that reported infections within six weeks of the initial outbreak in Hubei and identify outliers that break the latitude dependence pattern.

Causal Models

In order to interpret the results of the latitude analyses as well as other available global data we use a formal causal inference (CI) framework and model causal relationships using directed acyclic graphs (DAGs). The vast scale of the COVID-19 pandemic provides a unique opportunity for high-level causal models involving a wide set of independent inputs ("exogenous variables", or root causes) and this makes it possible to do a simple but powerful *qualitative* analysis of model predictions from which we draw firm causal conclusions by comparing those predictions with observed data.

We categorised variables by their expected causation effect to simplify the model relationships. We list specific members of each category, in order to highlight possible confounders that would otherwise potentially be masked by this categorisation.

Traditional CI methods construct a a single model in which the strength of causal relationships is to be determined quantitatively, but since the question we wish answer is binary - whether Vitamin D plays a causal role in COVID-19 disease outcomes or not - we express the causal/not-causal dichotomy with two contrasting models (figure 1 and figure 2). Technically, these are two views of the same model with binary values for the strength of the causal connection. The two models generate different testable predictions about the outcomes we should expect to see.

We use three categories for exogenous variables (root causes) influencing COVID-19 outcomes:

1. Sources of Vitamin D ("D Increasers")
2. Causes of Vitamin D deficiency ("D Reducers")
3. Causes of Morbidities with no Vitamin D relationship ("D-neutral Morbidities Causes")

Morbidities are categorised in a similar positive/negative/neutral sense:

A. "D-reducing" (illnesses that lower D serum levels)
B. "D-impacted" (illness caused by or exacerbated by low Vitamin D levels)
C. "D-neutral" (no causal relationship with serum D levels)

A (non-exhaustive) set of examples for each variable category have been included, with lines drawn in red so they are visually distinct from the causal paths of interest. Most of these variables are self-evidently independent.

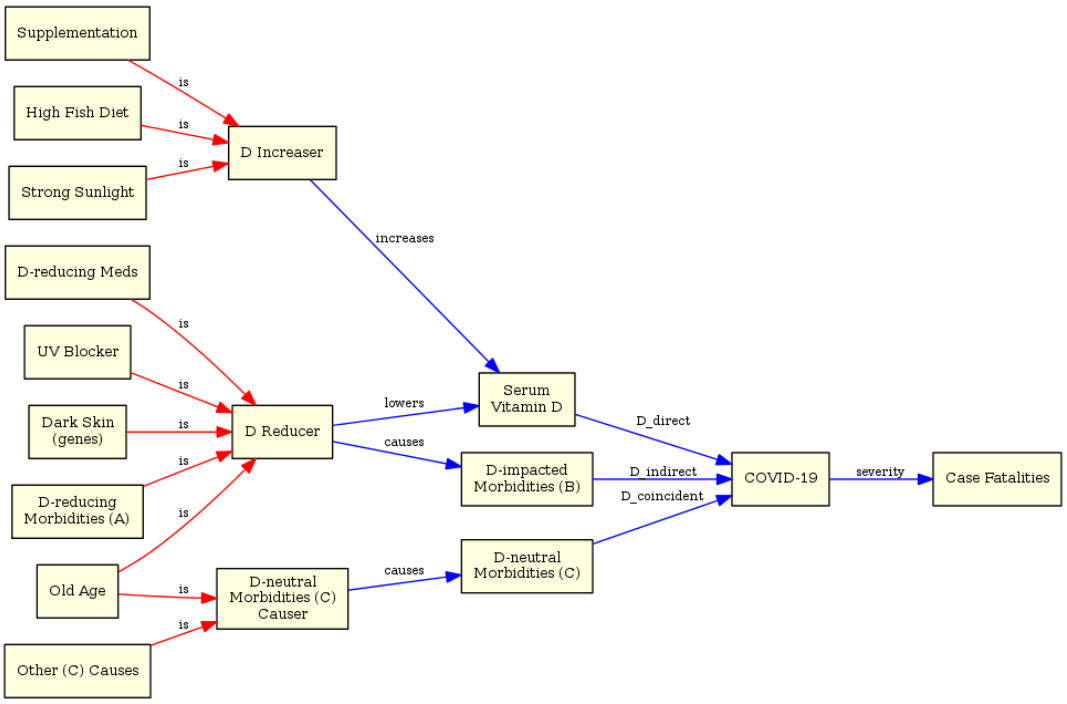

Figure 1 Model A (casual) - serum Vitamin D levels play direct and indirect ("mediated") causal roles in disease severity and outcomes

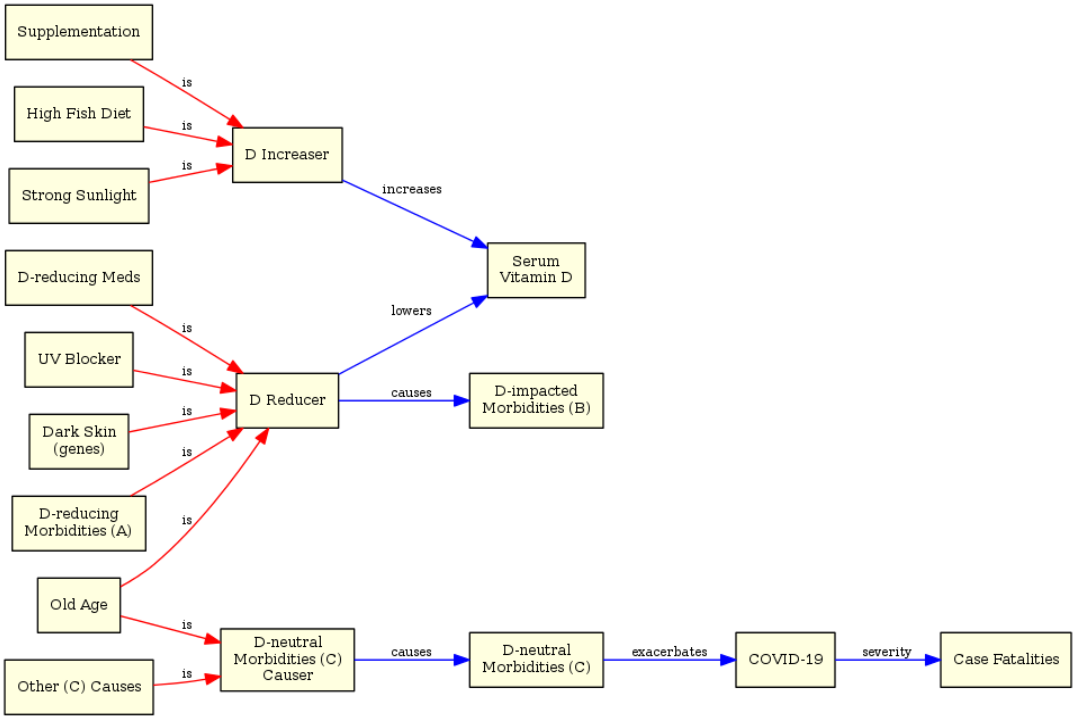

Figure 2 - Model B (acausal or "bystander") - direct and mediated causal links between serum Vitamin D levels and COVID-19 have been removed

We analysed the model for potential biases and confounders and then evaluated the predictions made by both models, tabulating these to highlight differences.

Finally, we cross-reference available observed data and compare these with model predictions.

Results

Severity Analysis

Severity by Latitude

COVID-19 fatalities display striking a correlation with latitude. Figure 3 shows deaths by latitude, human population by latitude, and finally deaths per million people by latitude. Severe outbreaks with large fatalities have occurred exclusively above the +30ºN latitude line, in the winter hemisphere, where deaths per million ranged from 3% up to 37% with an mean average of 11% for latitudes in the range 30ºN to 55ºN. By contrast, outbreaks in the tropics and southern summer hemisphere were very mild with an average of just 0.2% deaths per million.

This pattern is not explained by the timeline of spread of infection (see figure 4). The first death occurred on 9th Jan 2020 in Wuhan, China. Systematic reporting began on 22 January once the severity of COVID-19 became apparent. Within two weeks, 35 neighbouring locations in Asia had reported at least one death or recovery, including Japan, Thailand, Hong Kong, and the Philippines, and one location in Australia (New South Wales). By the end of February, air travel had spread infections to 65 locations across Europe, North America, Oceania and South Asia. Three weeks later, the majority of populous locations in the world had reported at least one recovery or death.

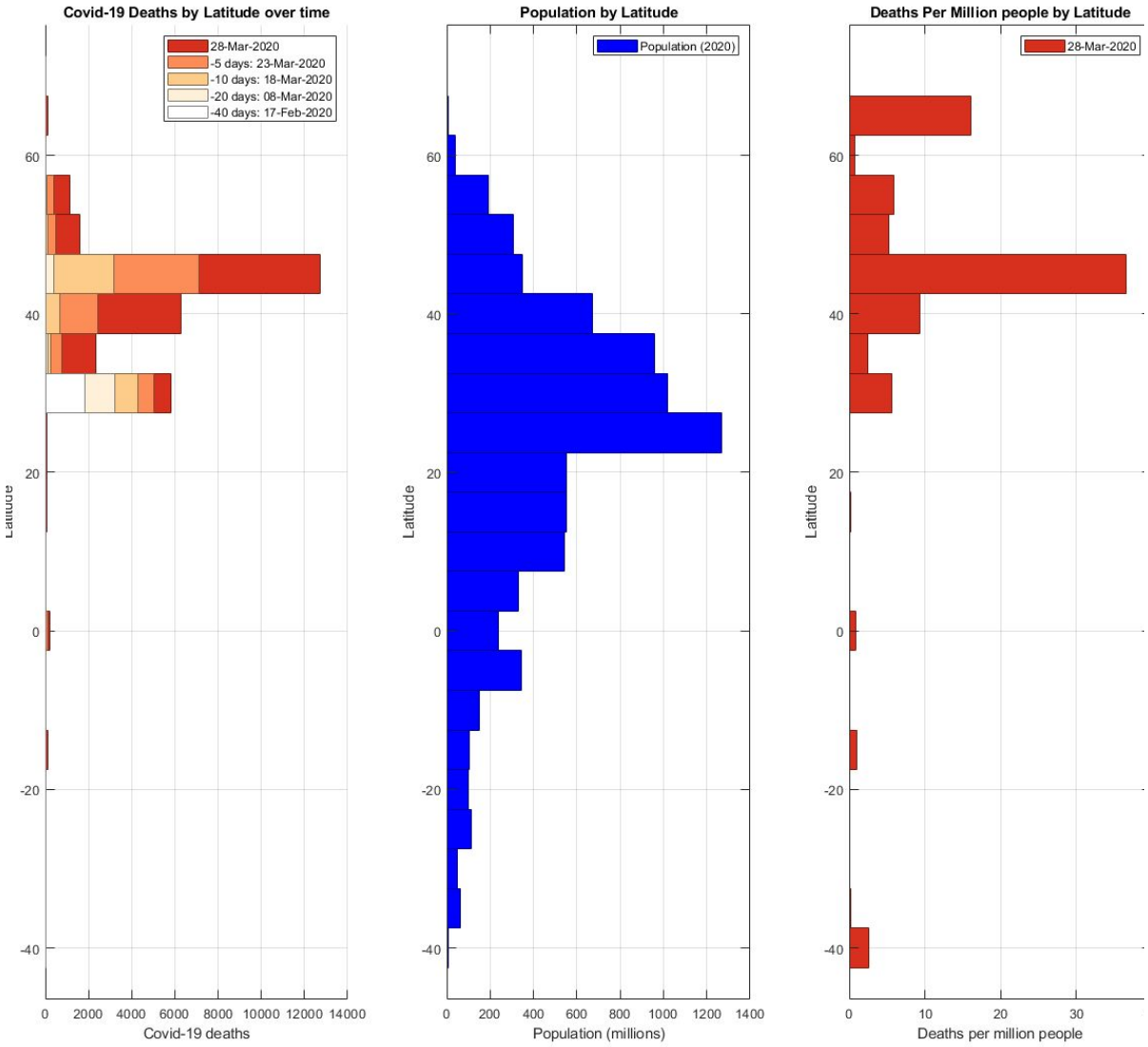

Figure 3 - (left) COVID-19 fatalities by latitude and over time; (middle) 2020 population by latitude; (right) COVID-19 fatalities per million people by latitude. Note: the Deaths per Million figure at -40ºS is a statistical artefact due to dividing two small numbers and may be ignored.

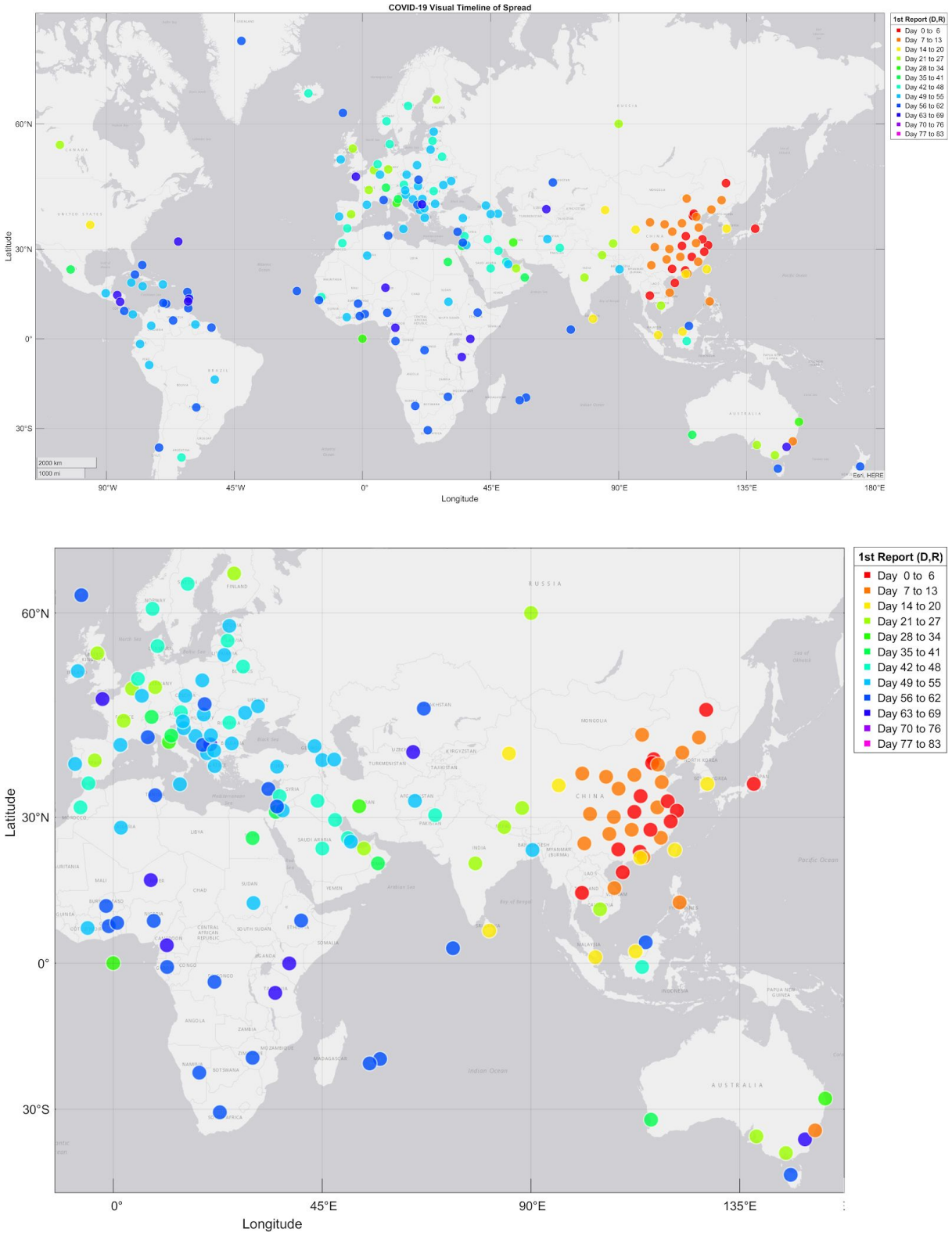

Figure 4 - (top) Visual timeline of global spread of COVID-19
(bottom) magnified section showing Europe, Africa, Asia and Oceania.

Severity by Location - ESI Analysis

We calculated the Epidemic Severity Index for all 239 reporting locations. We conditioned the ESI to judge severity against a reference baseline survival ratio, *S = 6.5,* which can be interpreted as meaning any outbreak that is worse than approximately “twice as severe as US seasonal flu”.

Figure 5 shows peak ESI scores (linearised form[1]) for all locations up to the end of March 2020 marking the end of winter and the start of spring. Nine locations experienced severe outbreaks over the period (ESI > 2.5), all of them above the 30°N latitude line. Considering only locations above 30°N, the timeline of spread (figure 4) is broadly consistent with outbreak severity; locations infected early generally experienced more severe outbreaks. South of this line, this is not the case. Here, multiple locations reported infections early yet none progressed into severe outbreaks. Figure 6 shows the log scale version of the same data to show low severity outbreaks more clearly - note that some location’s ESI scores went negative immediately and therefore do not show since their peak ESI is 0.

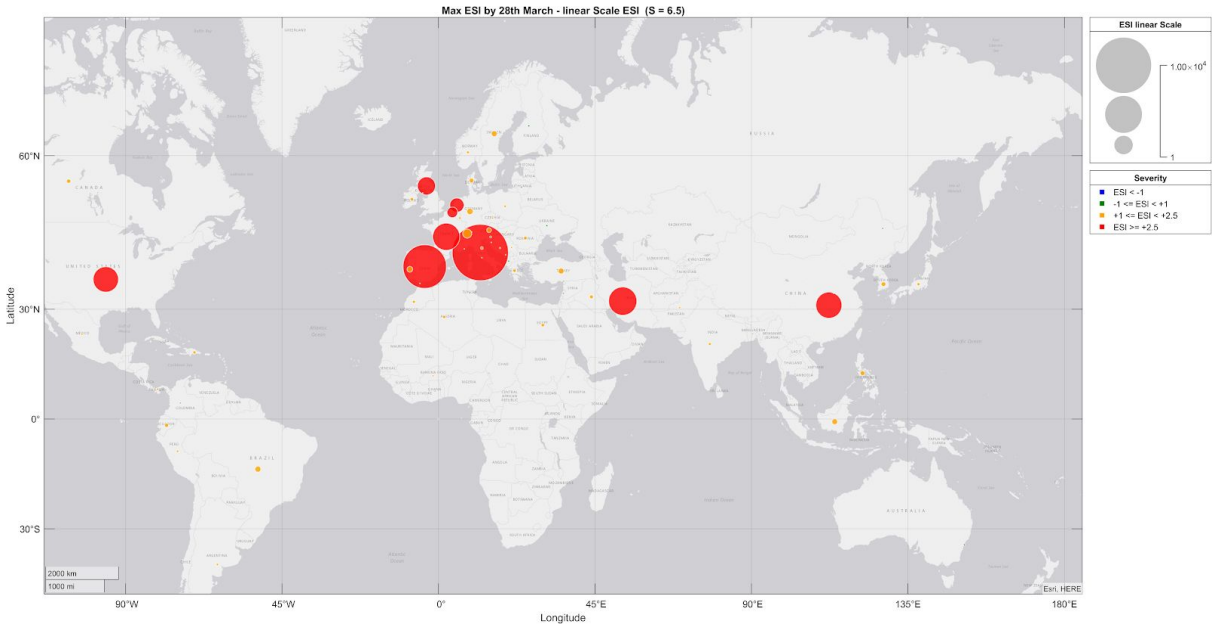

Figure 5 - peak ESI scores up to 28th March 2020 (S = 6.5; linearised scale)
Colour Key: Yellow = ESI < +2.5 (mild); Red = ESI > +2.5 (severe)

[1] Linearised form of ESI raises ten to the absolute ESI value so that geobubble areas are directly comparable with the number of deaths.

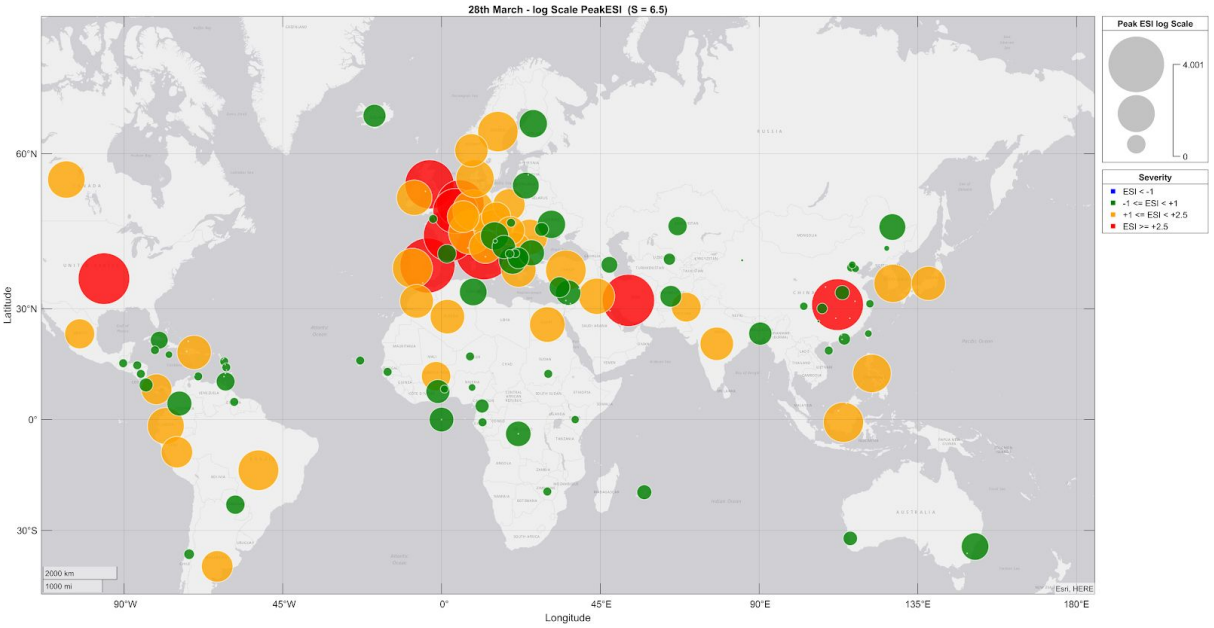

Figure 6 - peak ESI scores for all locations up to 28th March 2020 (log scale)
Colour Key: Green = ESI < +1 (very mild); Yellow = +1 < ESI < +2.5 (mild); Red = ESI > +2.5 (severe)

Latitude	Location	Comment	1st Report Date	Relative Day	Peak ESI	ESI 28-Mar	Deaths 28-Mar	Recovs 28-Mar	Total D+R	Hosp Survival Ratio 28-Mar
64.0	Finland		12-Feb	21	0.983	0.983	9	10	19	too small
60.0	Russia	Report accuracy questioned	12-Feb	21	-	-0.554	4	49	53	too small
56.1	Canada	6% <25nmol/L in 60 - 79 age	12-Feb	21	1.737	-1.739	61	466	527	7.6
55.4	UK		12-Feb	21	3.009	3.009	1,019	135	1154	0.1
51.0	Germany		13-Feb	22	2.093	-2.637	433	8481	8914	19.6
50.8	Belgium		17-Feb	26	2.549	2.549	353	1063	1416	3.0
46.8	Switzerland		03-Mar	41	2.408	2.408	264	1530	1794	5.8
46.2	France		12-Feb	21	3.365	3.365	2,314	5700	8014	2.5
43.9	San Marino	Mountain microstate of Italy	03-Mar	41	1.362	1.362	22	6	28	too small
43.0	Italy		21-Feb	30	4.001	4.001	10,023	12384	22407	1.2
40.0	Spain		15-Feb	24	3.777	3.777	5,982	12285	18267	2.1
37.1	US		09-Feb	18	3.307	3.307	2,026	1072	3098	0.5
36.0	Japan	High fish diet	26-Jan	4	1.450	-1.674	52	404	456	7.8
36.0	Korea, South	8-10% <25nmol/L in 65+ ages	07-Feb	16	1.820	-2.161	144	4811	4955	33.4
32.0	Iran	62.0% <50nmol/L (c.f. Israel)	19-Feb	28	3.401	3.401	2,517	11679	14196	4.6
31.0	Israel	14.4% <25nmol/L 2012 (c.f. Iran)	27-Feb	36	-	-0.505	12	89	101	7.4
28.2	Nepal		12-Feb	21	-	-	-	1	1	too small
26.0	Egypt		28-Feb	37	1.568	1.568	36	121	157	3.4
24.0	United Arab Emirates		12-Feb	21	-	-0.453	2	52	54	too small
23.7	Taiwan		06-Feb	15	0.066	-0.321	2	30	32	too small
23.6	Mexico		03-Mar	41	1.112	1.112	12	4	16	too small
21.0	India		16-Feb	25	1.393	1.393	24	84	108	3.5
21.0	Oman		29-Feb	38	-	-	-	23	23	too small
16.0	Vietnam		01-Feb	10	-	-	-	21	21	too small
15.0	Thailand		26-Jan	4	-	-0.835	6	97	103	16.2
13.0	Philippines	3 out of 5 Vit-D deficient	02-Feb	11	1.839	1.839	68	35	103	0.5
11.6	Cambodia		12-Feb	21	-	-	-	13	13	too small
7.0	Sri Lanka		08-Feb	17	-	-0.037	1	9	10	too small
2.5	Malaysia		07-Feb	16	-	-1.447	27	320	347	11.9
1.3	Singapore		08-Feb	17	-	-0.477	2	198	200	99.0
-28.0	Queensland, Aust.		21-Feb	30	-	-0.022	1	8	9	too small
-32.0	Western Australia, Aust.		01-Mar	39	0.255	-0.295	2	28	30	too small
-33.9	New South Wales, Aust.		30-Jan	8	0.938	0.938	8	4	12	too small
-34.9	South Australia, Aust.		17-Feb	26	-	-	-	6	6	too small
-37.8	Victoria, Aust.		13-Feb	22	-	-0.602	3	191	194	63.7

Table 1 - Detailed data for locations outside China reporting infections in the first 6 weeks sorted by latitude. Pattern outliers marked in orange. Peak ESI and latest ESI up to 28th March show severity. Death and Recoveries and Hospital Survival Ratios included.

Table 1 lists locations outside China that reported at least one death or recovery within 6 weeks (42 days) of 22nd January along with severity measures. Peak and final ESI scores are shown with details of total deaths and recoveries by 28th March (66 days, 9.5 weeks). Where the total exceeds 100 the hospital survival ratio is also shown. Outliers breaking the latitude dependency are marked in orange.

The latitude dependency of outbreaks is shown by figure 7 and figure 8 and which show the Epidemic Severity Index over time for all locations in table 1 that reported a death or recovery within the first six weeks. Locations have been grouped by descending order of latitude into six groups so there are no more than six locations per plot for readability. Ffigure 7 shows ESI by actual date and figure 8 shows ESI by day relative to each location's "day zero" - the first reported death or recovery.

Risk factors for Vitamin D deficiency

During winter months in locations outside the tropics, the UV Index (UVI)[2] reaches a maximum of 3 (see table 2) which is insufficient for the skin to produce Vitamin D [9].

		winter					summer						winter
Country (City)	**Latitude**	**Jan**	**Feb**	**Mar**	**Apr**	**May**	**Jun**	**Jul**	**Aug**	**Sep**	**Oct**	**Nov**	**Dec**
Russia (St Petersburg)	60°N	0	0	1	3	4	5	5	4	2	1	0	0
USA (New York)	41°N	2	3	4	6	7	8	9	8	6	3	2	1
Spain (Palma de Mallorca)	39°N	2	3	4	6	8	9	9	8	6	4	2	1
Vietnam (Hanoi)	21°N	6	8	10	11	11	11	12	12	10	8	6	6
Singapore (Singapore)	1°N	11	12	13	13	11	11	11	11	12	12	11	10
Madagascar (Tananarive)	19°S	12	12	11	9	7	6	6	8	11	11	12	12
Australia (Melbourne)	37°S	8	8	6	4	2	2	2	3	5	6	8	9
New Zealand (Wellington)	42°S	7	7	5	3	1	1	1	2	4	6	7	8
Falkland-Islands (Port Stanley)	58°S	5	4	2	1	0	0	0	1	2	3	5	5
		Jan	**Feb**	**Mar**	**Apr**	**May**	**Jun**	**Jul**	**Aug**	**Sep**	**Oct**	**Nov**	**Dec**
		summer					winter						summer

Table 2 - Maximum UV Index by month and latitude [10].

The most severe outbreak in Europe has been Italy where it is noted vitamin D deficiency is one of the highest in Europe [11], 76% of women aged 60-80 were found to have 25(OH)D levels below 12 ng/ml [5] and hypovitaminosis D was prevalent in 82% of patients who engaged in long-term rehabilitation programmes because of various neurological disorders [5,12].

Old age, pregnancy and infancy are typically high risk for disease due to compromised or low immunity and hypovitaminosis D [13,14][15]. Darker skin pigmentation , less sunlight exposure, dietary habits and absence of vitamin D fortification are also risk factors[4].

[2] The UV Index is a convenient standard for expressing UV radiation levels.

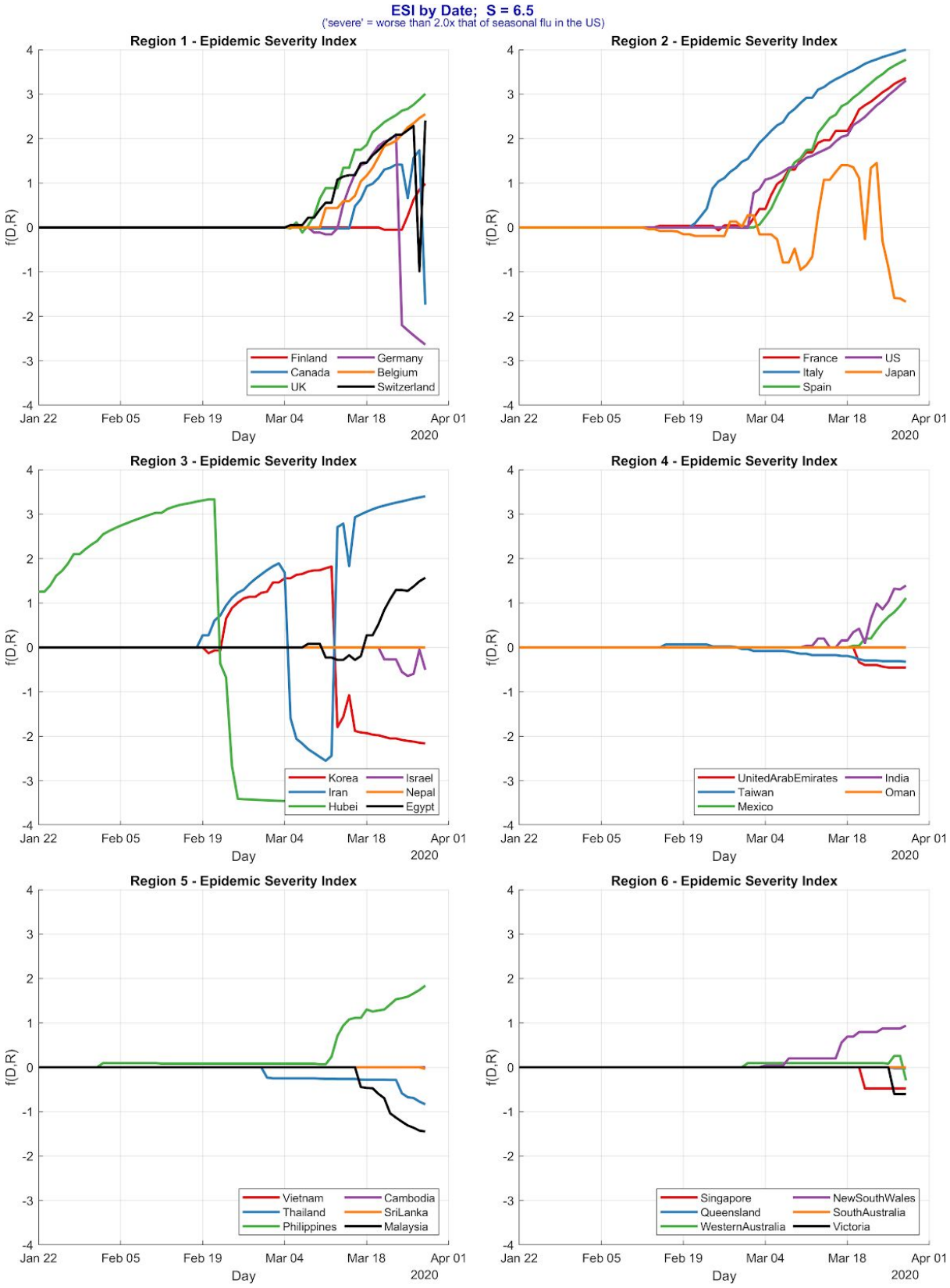

Figure 7 - Epidemic Severity Index (*S = 6.5*) by date for all locations reporting within 6 weeks of Hubei, grouped into six regions by descending order of latitude.

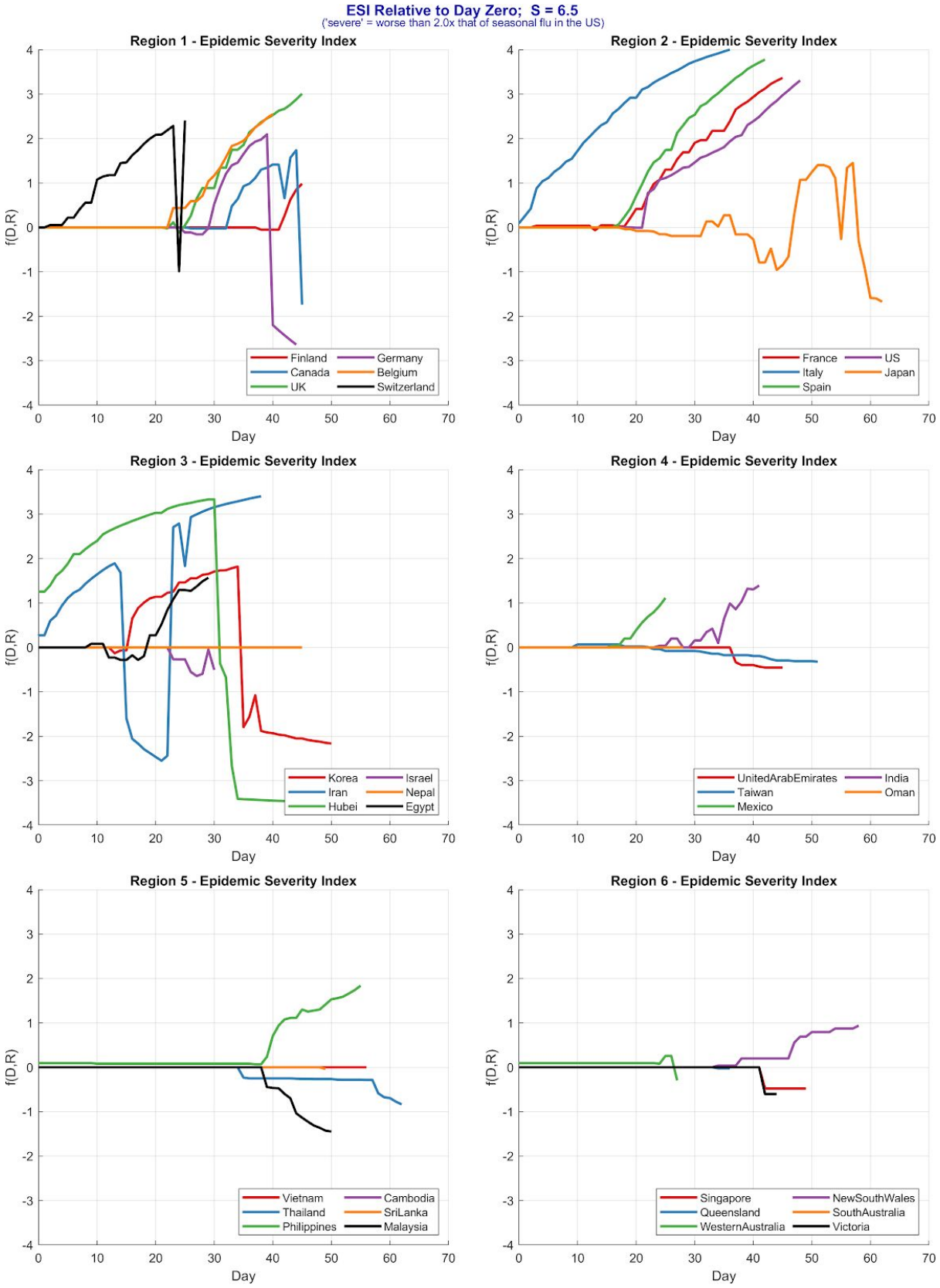

Figure 8 - ESI (*S = 6.5*) relative to location's first reported death or recovery for all locations reporting within 6 weeks of Hubei, grouped into six regions by descending order of latitude.

Outliers infected within 6 weeks

Four locations above 35°N appear as pattern outliers with relatively mild outbreak severities: Canada, Germany, Japan and South Korea. Available data for all four countries suggest that a common link is relatively low Vitamin D deficiency levels [4][16][17][18][19][20]. Canada may also benefit from a low population density. Japan has a very high population density but has the lowest incidence of Vitamin D deficiency which is attributed to its high fish-content diet. The reason behind the other countries is less clear but presumed to be lifestyle, diet and nationwide policy on supplementation. A lack of more recent data and issues with inconsistent testing methods and definitions make it difficult to assert this with certainty. Nonetheless, the broad pattern of serum 25(OH)D correlations with outbreak severity appears to hold consistently and account for outliers across the full latitude range.

Two locations worthy of comparison lie in the relatively sunny latitude range 31-32°N are Iran and Israel. Iran suffered a severe outbreak despite its lower latitude with an epidemic severity index of 3.4 and more than 2,500 deaths. In stark contrast, Israel reported just 12 deaths and 101 recoveries. Studies of vitamin D levels for the two countries reveal that Iran has a very high prevalence of deficiency [21] whereas Israel is relatively low [22]. High deficiency levels in Iran have been attributed to skin colour, religious full-body clothing and lifestyle.

Of the locations in the region below 30°N, two countries break the latitude pattern. The Phillipines appears in the first 6 weeks infections list as outlier. The country has noted serious concerns about Vitamin D levels after a small study found that 58% of randomly tested office workers in Manila were deficient [23]. Phillipine Dr Mark Alipio recently published a (preprint) retrospective multicentre study of 212 cases with laboratory-confirmed infection of SARS-CoV-2 from three hospitals in Southern Asian countries. Data pertaining to clinical features and serum 25(OH)D levels were extracted from the medical records. Alipio concludes, “Vitamin D status is significantly associated with clinical outcomes ($p<0.001$). For each standard deviation increase in serum 25(OH)D, the odds of having a mild clinical outcome rather than a severe outcome were increased approximately 7.94 times; the odds of having a mild clinical outcome rather than a critical outcome were increased approximately 19.61 times,” indicating that, in COVID-19 patients, increased serum 25(OH)D level could improve clinical outcomes, and/or mitigate the worst (severe to critical) outcomes. Conversely, decreased serum 25(OH)D levels could worsen clinical outcomes.” [24].

Brazil also appears as an outlier in the southern hemisphere, though it reported its first deaths and recoveries outside the first six weeks. Vitamin D deficiency data for countries in the southern hemisphere is less complete than in the north, but available data suggest that Brazil has a very high prevalence of vitamin D deficiency in elderly populations. Studies of two nursing homes in São Paulo found that residents had very low mean serum levels in winter which did not improve over summer. Mean ± standard deviation was reported as 37.6 ± 29.9 nmol/L, compared to the recommended value >75.0 nmol/L [25]. It was noted

that many of these individuals are from low income families that have little access to health services. Skin colour is likely to contribute significantly to vitamin D deficiency in Brazil, since it has a large population of black (8%) and mixed race (43%) ethnic groups [26].

Mexico's initial trajectory appears concerning but on 30th March it had reported 11,423 recoveries and only 1,859 deaths, indicating a hospital survival rate of about 14%, just under twice that of seasonal US flu. Similarly, India reported 9,068 recoveries and 1,154 deaths by 30th March equivalent to 11.2% hospital survival rate.

Causal Model Analysis

Confounder Bias and Special Cases

In general, any "upstream" causes of forks may be sources of confounding in observational data. Given the broad nature of causes of morbidities, this can cause complexity (figure 9). This can be handled by considering specific morbidities and controlling for common causes, or employing alternative decounding mechanisms. Modern techniques, such as the "front door criterion" enabled deconfounding of even unknown or unmeasured variables through the use of mediator variables.

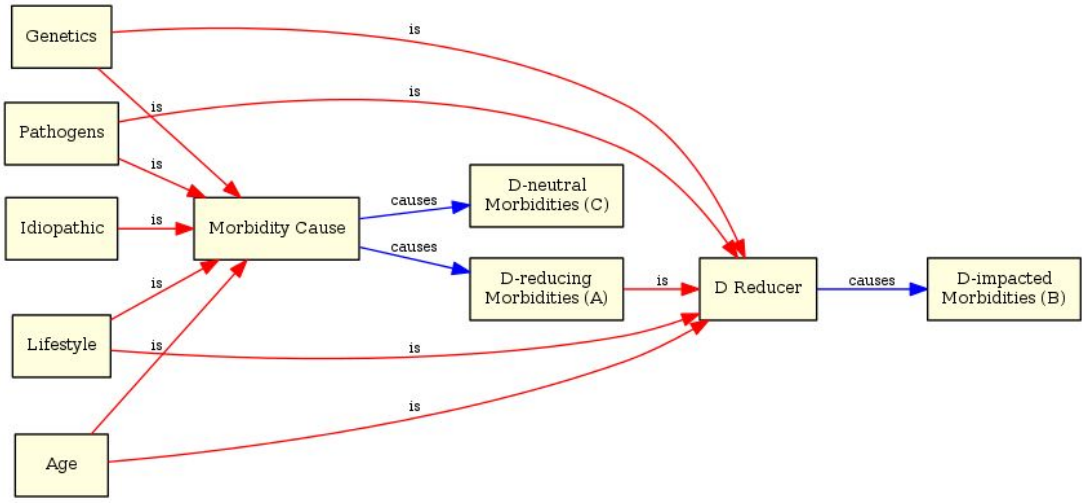

Figure 9 - upstream causes must be controlled for when analysing morbidity statistics

Another way to reduce bias is triangulation. Triangulation integrates results from different data where each set has different sources of potential bias that are unrelated. If these different data all point to the same conclusion, this strengthens confidence in the finding [27].

Special Cases: Senescence, Pregnancy and Infancy

In developed countries, pregnant women and infants are routinely supplemented with Vitamin D [28], so in Model A we consider these as belonging to the Supplementation "D increaser" root variable category.

In model B, where Vitamin D is not causal, we consider them as part of "D-neutral Morbidities", expanding the condition of "morbidity" to include the immunosuppressed state of pregnancy and still-developing states of infant immune systems.

Other modelled root causes that we shall consider are, to all intents and purposes, independent[3]. I.e. we expect that taking supplements has zero effect on skin colour genes, and does not change the orientation of the Earth with respect to the sun *etc*. The special case of season impacting Vitamin D supplementation is *not* confounding since the effects compensate for each other in terms of impact on outcome (which is the point of taking supplements during winter). However, since we will be considering specific cases of season in our qualitative analysis, this potential confounder will be "controlled for" anyway.

Cyclic Dependencies

Pathogen-caused disease spread has a cyclic nature: pathogens cause infections which cause disease which results in more pathogens (figure 10). Cyclic dependencies can't be handled by DAG causal models which are "acyclic" by definition. The acyclic restriction is necessary for quantitative analysis because cyclic dependencies (positive or negative feedback loops) are mathematically very complex to model[4]. However they can be analysed qualitatively and the relationship is worthy of discussion.

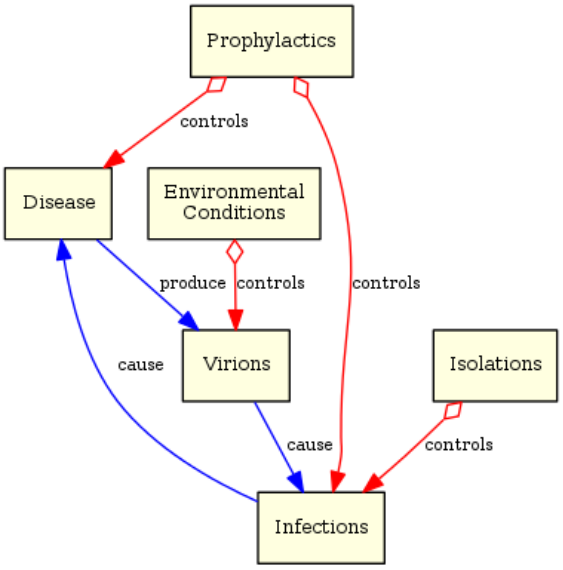

Figure 10 - probability of spreading and catching viral diseases have cyclic dependencies

It's clear from the viral cycle that anything that controls disease severity, reduces the number of viral particles, or lowers the probability of infection will suppress the feedback loop. Environmental conditions hostile to virus particle survival will reduce transmission rates; Isolation, quarantine and social distancing measures similarly. Neither is expected to reduce

[3] Any challenges to the assertion of independence can be handling by modifying the diagram appropriately or similarly considering the expected nature of the impacts.
[4] Sophisticated techniques using DCGs have been developed recently that can handle cyclic dependencies.

disease severity or case fatalities rates. However, anything that directly controls disease severity will reduce case fatalities rates as well as reducing transmission rates since fewer viral particles are produced and this further reduces the possibility of infection. Prophylaxis will lower infection rates as well as CFRs. This is an important distinction.

Informed by these considerations and armed with two contrasting models we can now make predictions for each.

Causal Model Predictions

Model A and Model B predict multiple different outcomes.

Season is a function of latitude and can only take on one value in any given location at a given time. Since D reducers and D increasers act in opposition, we consider the cases of D increasers counteracting the D-reducing effects of weak sunlight (winter latitudes); and D reducers counteracting the D-increasing effects of strong sunlight (summer and tropical latitudes).

Causal Model A Predicts

1. Latitude will correlate strongly with case fatality rates, with winter locations showing worse CFRs than tropical and summer locations.
2. Winter locations with atypically low prevalence of D deficiency will appear as outliers with unusually low transmission and case fatality rates - even in the absence of the absence of quarantine & isolation measures.
3. In winter locations, typically high-risk subpopulations that are routinely supplemented with vitamin D will stand out as unexpectedly low-risk: pregnancy and infancy.
4. in winter locations, subpopulations with dark skin will correlate with higher CFRs.
5. Tropical and summer hemisphere locations will in general display lower transmission and case fatality rates - even in the absence of quarantine & isolation measures.
6. Tropical and summer hemisphere locations with atypically high prevalence of D deficiency - especially in elderly populations - will appear as outliers experiencing higher levels of transmission and case fatalities.
7. All D reducers will correlate with higher CFRs even in the absence of any obvious comorbidities.
8. D-reducing Morbidities (A) will correlate with CFRs.
9. D-impacted Morbidities (B) will correlate with CFRs even when controlled for confounder bias.
10. In tropical and summer locations, populations with high prevalence of D deficiency from e.g. extreme sun-blocking or avoidance behaviours - such as Islamic women wearing full body-covering clothing - will correlate with higher CFRs.

Bystander Model B Predicts

1. Latitude will not correlate with case fatality rates.
2. In winter locations, Vitamin D sufficiency, and high-fish diets will not correlate with lower CFRs.

3. Routinely supplemented subpopulations will not correlate with CFRs: pregnancy and infancy will show up as typically high-risk.
4. D reducers will not correlate with CFRs in the absence of illness.
5. In winter locations, only quarantine, isolation and social distancing measures will correlate with lower transmission rates.
6. Only D-neutral Morbidities (C) will correlate with CFRs once disease confounders are controlled for.

Predictions Compared to Available Data

We use total hospital cases per location (Deaths + Recoveries) (see table 1) as a proxy indicator of transmission rates, and Hospital Recovery Rates (R/D) as a proxy indicator of CFR. Table 3 summarises the predictions from each model, grouping broadly by latitude to consider counterations and lastly a general predictions group and compares predictions (P) with observations (O). Table cells have been coloured to clearly indicate where predictions match observations: green = strong match; orange = mismatch; red = strong mismatch. Predictions that are the same in both models are coloured grey as these are not informative. Cells where predictions cannot be validated against observations are left white.

Variables	**Causal Model A**		**Bystander Model B**	
Winter Latitudes > 40ºN	**Transmission**	**CFRs**	**Transmission**	**CFRs**
General Winter Populations	P: higher O: higher	P: higher O: higher	P: higher O: higher	P: baseline O: higher
Quarantine, Distancing Measures	lower	P: same as general (higher) O: higher	lower	P: same as general (baseline) O: higher
High Prevalence Hypo-D Countries	P: much higher O: much higher	P: much higher O: much higher	P: higher O: much higher	P: baseline O: much higher
Low Prevalence Hypo-D Countries	P: lower O: lower	P: lower O: lower	P: higher O: lower	P: baseline O: lower
High Fish Diet Countries	P: lower O: lower	P: lower O: lower	P: higher O: lower	P: baseline O: lower
Dark-skinned Sub-populations	-	P: much higher O: much higher	-	P: baseline O: much higher
Tropic/Summer Lat < 40ºN	**Transmission**	**CFRs**	**Transmission**	**CFRs**
General Summer Populations	P: lower O: lower	P: lower O: lower	P: lower O: lower	P: baseline O: lower
Quarantine, Distancing Measures	P: lower O: lower	same as general (lower)	P: lower O: lower	same as general (baseline)

High Prevalence Hypo-D Countries	P: higher O: higher	P: higher O: higher	P: baseline O: higher	P: baseline O: higher
Sun-Blocking Populations	-	P: higher O: higher	-	P: baseline O: higher
General Predictions	**Transmission**	**CFRs**	**Transmission**	**CFRs**
Elderly Sub-populations	-	P: much higher risk O: much higher risk	-	P: higher risk O: much higher risk
Pregnancy, Infancy Sub-populations	-	P: lower risk O: lower risk	-	P: high risk O: lower risk
D-reducing morbidities (A)	-	higher risk	-	typical risk
D-impacted morbidities (B)	-	higher risk	-	typical risk
D-neutral morbidities (C)	-	typical risk	-	typical risk

Table 3 - Predictions from models A and B. Cases where the models predict the same outcome have been coloured grey.

UK COVID-19 deaths in BAME communities have been alarmingly high with data from intensive care units indicating that over a third of patients are from BME backgrounds [29,30] and 63% of the first 106 health and social care staff known to have died from the virus were black or Asian [31]. The New York Department of Health and Mental Hygiene reports that Black/African American and Hispanic/Latino populations are suffering up to twice the case fatality rates of white New Yorkers [32].

Summary of Predictions versus Observations

Model A: 16 predictions match observed data; 3 predictions cannot be determined.
Model B: 14 predictions strongly contradict observed data; 2 may contradict data; 3 cannot be determined.

Discussion

Qualitative Evidence of Causation from Triangulation

The COVID-19 pandemic is a global phenomenon that has affected millions worldwide providing observational data with statistical power many orders of magnitude greater than a devised RCT or observational trial that could be conducted even at national level. Striking patterns emerge directly from this statistical power that are so large they are evident without the need for sophisticated regression analyses.

Contrastive causal models allow us to distinguish a large set of predicted features from a multitude of predominantly *independent* exogenous variables driving data features. This greatly increases the causal inference power available because of triangulation: the more independent root causes whose natural variation concurs with the model predictions, the narrower the the possible range of explanations, and the less likely the model is to be wrong. This holds true even for hidden or unknown biases so long as the source of bias is different for each root variable.

Even if we discount possible confounders Age and Underlying Conditions, a large set of predictions remain that match those of the causal model and that do not match predictions made by the acausal “bystander” model. Clear confirmation of the causal model is evident in:

- Severe outbreaks with high fatality rates happened in general only in the northern hemisphere post-winter locations
- In general, outbreaks in tropical and southern post-summer locations were mild
- Northern outliers Canada, Germany, Japan and South Korea all correlate with known low prevalence of hypovitaminosis D relative to countries with severe outbreaks, presumed to be due to either high-fish containing diets or supplementation (actual cause is immaterial)
- Southern outliers: Philippines and Brazil suffered the worst outbreak and have known high prevalence of hypovitaminosis D.
- Pregnancy and infancy are both highly-unexpected exceptions to the typical at-risk groups most diseases include.
- Statistics for communities with genetically dark skin in the UK and US have confirmed case fatality rates twice the average rate.

Further Historical Supporting Evidence

Vitamin D was hypothesised in 1922 and confirmed to exist in 1925 [33] five years after the end of the 20th centuries deadliest pandemic, Spanish Influenza, caused by the avian virus H1N1. The pandemic is estimated to have infected 500 million people which was one third of the world’s population at the time. Its mortality rate is estimated to be between 10% and 20%, with a death toll of 50 to 100 million people. By the turn of the 20th century, signs of hypovitaminosis D and rickets were present in upwards of 80 to 90% of children living in northern Europe and north eastern United States [34]. This was an era before the discovery of antibiotics and many died of secondary bacterial infections [35]. Sanatoria promoted fresh air and sunshine as the only treatments that appeared to be effective. Explanations for their efficacy were lacking.

After the discovery of Vitamin D as a cure for rickets it became common practice to add it to many different food substances, and for two decades between 1930 and 1950, Vitamin D food additives eradicated hypovitaminosis D but at the expense of causing cases of hypercalcemia which eventually led to the practice being banned. Hypovitaminosis D returned and was exacerbated in the 1970s by the introduction of sunscreen creams containing UVB absorbing chemicals [34].

We note with interest, that when specific strains of influenza are plotted on an epidemic and pandemic timeline, there is a 37 year period from 1920 to 1957 where no new flu strains seem to appear and no epidemics occurred (figure 11) which coincides with the only known period during which the population at large was routinely supplemented with Vitamin D.

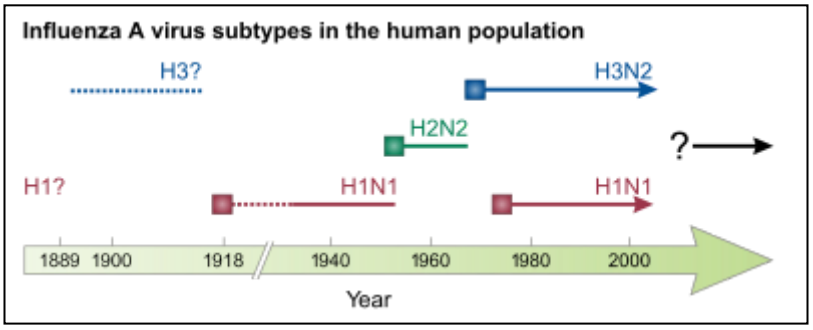

Figure 11 - Specific strains of influenza infection throughout the 20th century[5].

Discussion of Biochemical Mechanisms

SARS-COV-2 and COVID-19

It is has long been known that the renin-angiotensin system (RAS) is critically involved in normal cardiovascular and renal function as well as and in disease conditions.

Angiotensin-converting enzyme 2 (ACE2) is a membrane-bound aminopeptidase that has a vital role in the cardiovascular and immune systems and is abundantly present in humans in the epithelia of the lung and small intestine [37]. ACE2 is involved in heart function and the development of hypertension and diabetes mellitus.

ACE2 negatively regulates the renin-angiotensin system (RAS) [38] and is a balancing mechanism for ACE which positively regulates the RAS [39].

The respiratory tract is a major site of coronavirus infection and disease morbidity: studies of related coronaviruses SARS-Cov and NL63 showed that undifferentiated cells expressing little ACE2 are poorly infected with SARS-CoV, while well-differentiated cells expressing more ACE2 are readily infected. [40]. Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) infects host cells through ACE2 receptors, leading to coronavirus disease (COVID-19)-related pneumonia, while also causing acute myocardial injury and chronic damage to the cardiovascular system. SARS-Cov-2 virus enters cells via ACE2 receptors in cells [41] in epithelial cells in the lungs. Studies of related coronaviruses show that viral replication downregulates ACE2 [42]. This dysregulates the RAS since it depends on the counterbalancing effects of ACE and ACE2: as ACE2 is depleted the RAS becomes over-activate, resulting in a cytokine storm in the host which causes ARDS [43][44].

[5] Influenza: old and new threats. [36]

Vitamin D deficiency is strongly associated with ARDS [45], occurring in 90% of patients with ARDS and correlates with poor mortality outcomes [46]. Although the precise nature of COVID-19 respiratory distress differs from ARDS, this association is notable. Chronic vitamin D deficiency induces lung fibrosis through activation of the RAS [47].

Vitamin D is also strongly associated with many comorbidities that correlate with COVID-19 case fatalities, such as cardiovascular disease [48], diabetes [49] and dementia [50]. Vitamin D deficiency is associated with a substantially increased risk of all-cause dementia and Alzheimer disease [51].

Since SARS-Cov-2 infections cause acute myocardial injury and chronic damage to the cardiovascular system [38,52] it is unsurprising that these comorbidities correlate strongly with COVID-19 case fatalities.

Several large studies and meta-analyses have found that vitamin D deficiency is associated with greater illness severity, morbidity and mortality in both critically ill adult and geriatric patients, but there is a lack of clinical data demonstrating the efficacy of normalising levels as an effective therapy and trials are being designed to this end [53].

Vitamin D acts via multiple mechanisms

It has been shown that Vitamin D acts to rebalance the RAS: it alleviates and attenuates lung injury by regulating the RAS and blocking the Ang-2-Tie-2-MLC kinase cascade [54][55] and suppresses renin gene expression [56].

HIV studies have illuminated in detail the mechanisms by which Vitamin D supplementation increases immunity and reduces inflammatory responses: vitamin D promotes an anti-inflammatory response by inhibiting the maturation of dendritic cells (DCs), resulting in a phenotype characterized by the downregulation of antigen presenting molecules (MHC-class II), costimulatory molecules (e.g., CD40, CD80, and CD86), and pro-inflammatory cytokines (e.g., IL-12 and IL-23) and Vitamin D enhances anti-inflammatory cytokine (IL-10) and T-cell inhibitory molecule (PD-1) [57].

Vitamin D reduces the risk of acute respiratory tract infection [58]. High dose oral Vitamin D has been shown to improve mortality in critically ill patients with severe vitamin D deficiency [59].

Causal evidence for Vitamin D

Vitamin D associations with all cause mortality and all cause morbidity appear in research spanning decades. A small number of recent studies have used causal inference to demonstrate a causal link between supplementation and outcomes. A combined sample from three European cohort studies used Mendelian Randomisation to provide further support for a causal relationship between vitamin D deficiency and increased all-cause mortality. However, the study size, (~10,000 participants), was underpowered and larger studies on genetics and mortality were called for [60]. Meta-analyses showing causal relationships between Vitamin D supplementation and morbidity and mortality exist. Intake of

ordinary doses of vitamin D supplements seems to be associated with decreases in total mortality rates [61]. A more recent 2019 study of higher dose random control trials showed that while vitamin D supplementation did not reduce total cancer incidence, it significantly reduced total cancer mortality [62]. There is evidence that critically ill patients with very low 25(OH)D concentrations have blunted responses to vitamin D replacement which may also explain the apparently negative results of earlier trials and meta-analyses [63].

We note that the mechanisms by which Vitamin D has been proven in HIV patients to improve immunity and reduce inflammatory responses appear to be slow and may take weeks to develop [57]. Since supplementation with D3 increases serum Vitamin D levels over many days, this may additionally contribute to negative results in critical care trials attempting treatment with Vitamin D, especially at low doses. More clinical data is needed to prove which, if any, diseases respond, at which doses and over what time frames.

Conclusion

Recent and historical data are highly consistent with a causal protective role for Vitamin D in respiratory disease risk and especially in the case of COVID-19. By contrast, the same evidence conflicts with predictions made by an acausal/bystander model. Since data come from multiple independent exogenous variables this provides “triangulation” consistency and the added global scale of the data suggest a high level of confidence in this conclusion is merited. The exact size of effect may be calculated by future quantitative analyses. Our analysis demonstrates it is large enough to be seen qualitatively in high-level data. A full quantitative analysis is called for and require large reliable and reliable data sets in order to more precisely quantify the size of the effect.

Vitamin D supplementation is an effective, safe and cheap method to protect against seasonal respiratory diseases and can play a key role in combatting the COVID-19 pandemic. Cholecalciferol (D3) oral supplements in doses up to a maximum of 4,000iu/d for short periods are considered safe [64]. NICE recommends daily supplements for all UK adults all year [65] [66].

Dr Gareth Davies (PhD)
Dr Attila R Garami (MD, PhD)
Dr Joanna Byers (MBChB)

Footnotes

Acknowledgements

We would like to thank Professor Andrea Giustina and Dr Thomas Hiemstra for advice and contributions during the process of writing this paper.

Funding and Competing Interests

This work was conducted *pro bono* as part of an unfunded, independent international response to the pandemic crisis.

The authors declares no competing interests.

About the authors

Dr Gareth Davies has a BSc in Physics and a PhD in Medical Physics from Imperial College, London, though he is not currently affiliated with Imperial College as a research scientist. He has more than three decades of experience of complex data analysis, systems modelling, software engineering and machine learning. In 2019, he was named as one of Codex World's Top 50 Innovators.

Dr Attila R Garami has a medical degree from Hungary and a PhD in Multidisciplinary Medical Sciences. He has a highly multidisciplinary background including tropical diseases, immunology, biochemistry and molecular biology from Max-Planck-Institutes, Tübingen in Germany - where he earned a reputation as an outstanding research fellow. He specialised in growth- and cancer-related signal transduction pathways from Friedrich-Miescher-Institute for Biomedical Research, Basel in Switzerland. He works as a senior biomarker consultant for innovative biomedical approaches in Switzerland.

Dr Joanna Byers is a British medical doctor (MBChB, University of Birmingham) with years of clinical medical experience across both primary and secondary care sectors. She has an interest in integrated systems design and preventative healthcare and holds a diploma in Healthcare Management, Leadership and Innovation from the University of Plymouth and is currently undertaking an MSc in Occupational Therapy with the University of Essex. She teaches Public Health at the University of Cambridge

Background to this paper

In response to the urgency of the unfolding crisis, on 23rd March, 2020, we published an early online report [67] outlining evidence supporting a beneficial role for Vitamin D in the prevention and treatment of COVID-19, calling for hospitals to test serum levels, treat deficiency and report data and outcomes. Response to the report was highly polarised. Those critical of it were quick to dismiss citing "correlation isn't causation". The report contained direct and indirect evidence for cause but lacked a formal framework justifying causal inference which has slowed the adoption of this critical information. A causal inference framework has a great deal of value to add beyond the current pandemic. Vitamin D has received a great deal of renewed attention in recent years, but a beneficial role for it in disease states has been controversial [68]. Vitamin D deficiency is association with many serious diseases has raised the question of whether it is a bystander, simply a general marker of illheath, or plays an active role in the cause and progression of diseases. The lack of standardised serum level testing and a long history of research and clinical trials with significant design flaws adds significant noise and confusion to the body of research making

it difficult to find clear answers even when present. We hope that the causal model framework presented in this paper will help to resolve some of these controversies.

Appendix - Background on Causal Inference

A new science of “causal inference” (CI) emerged in recent years from the field of artificial intelligence and was introduced to health research around the turn of the millenium [69]. CI is an extension to statistics that defines causation mathematically and allows it to be determined if causation can be inferred from observational data. If so, the size of the causal effect can be calculated.

CI employs causal diagrams (models) which are mathematically and graphically represented by “directed acyclic graphs” (DAGs). These network-like diagrams of connected nodes declare the flow of causes-to-effects in the model, and CI’s new symbolic language makes it possible to mathematically express and answer the causal questions we wish to pose. Formal rules dictate how DAGs are created, manipulated and interpreted soundly, and CI gives ways to test models for correctness and identify potential flaws. CI formally distinguishes *seeing* from *doing*, (adding a “do-operator”) which allows for a new, highly-succinct definition of “confounding” as: the difference between *seeing* X and *doing* X on an outcome (Y). If *seeing* and *doing* are equal in outcomes, there is no confounding, otherwise there is. Prior to CI, statistics has been unsuccessful in clearly, succinctly and correctly defining confounding. The nature of confounding was ironically itself confounding.

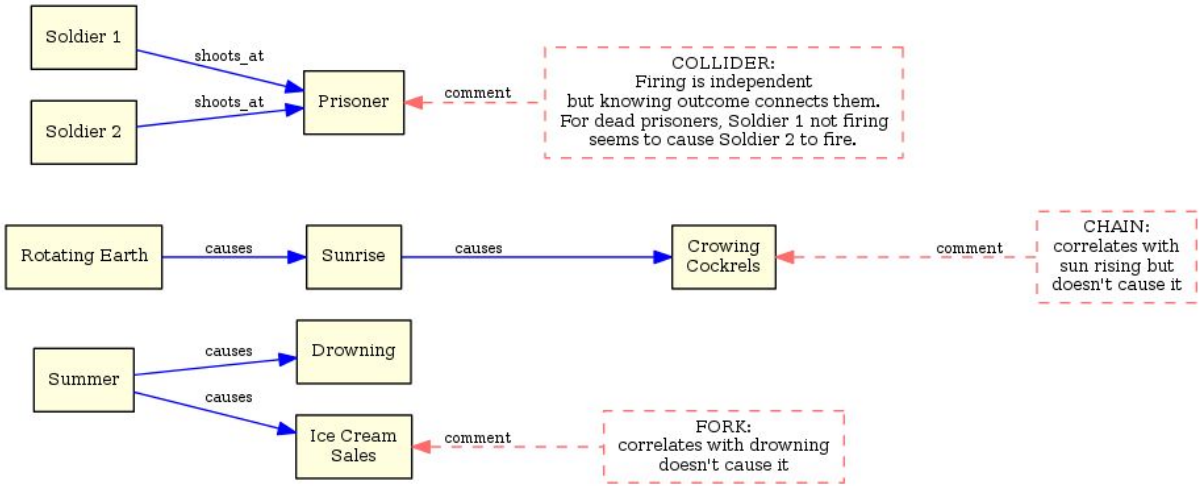

Figure 12 - examples of sources of confounding for the three types of causal relationships in DAGs: “collider” (top), “chain” (middle) and “fork” (bottom).

The purpose of a causal model is to *explain* the underlying process generating the observed data. The process of drawing and iterating DAGs brings clarity and interpretability to what can otherwise be highly-confusing data. Most importantly, models generate predictions which are testable against observed data and can be verified or falsified. Model relationships can be expressed mathematically to calculate the expected size of a causal effect purely from observed data with remarkable accuracy [70]. Models help to identify spurious

correlations and also show when inference is not possible. A credible causal model must first be defined then analysed to identify potential confounding variables, and any variables that may appear to be confounders but are not. Confounding variables can be "controlled for"[6] enabling the deconfounding of statistics. Models help to identify variables that should not be controlled for (colliders), since doing so may inadvertently introduce selection bias. CI gives us the power to interpret the size of causal effects from correlations in observational data.

References

1 Gaunt ER, Hardie A, Claas ECJ, *et al.* Epidemiology and clinical presentations of the four human coronaviruses 229E, HKU1, NL63, and OC43 detected over 3 years using a novel multiplex real-time PCR method. *J Clin Microbiol* 2010;**48**:2940–7. doi:10.1128/JCM.00636-10

2 J W T. The effect of environmental parameters on the survival of airborne infectious agents. *Journal of the Royal Society* Published Online First: 22 September 2009. doi:10.1098/rsif.2009.0227.focus

3 Re: Preventing a covid-19 pandemic Can high prevalence of severe hypovitaminosis D play a role in the high impact of Covid infection in Italy? Published Online First: 29 March 2020.https://www.bmj.com/content/368/bmj.m810/rr-36 (accessed 30 Mar 2020).

4 Mithal A, on behalf of the IOF Committee of Scientific Advisors (CSA) Nutrition Working Group, Wahl DA, *et al.* Global vitamin D status and determinants of hypovitaminosis D. Osteoporosis International. 2009;**20**:1821–1821. doi:10.1007/s00198-009-1030-y

5 Isaia G, Giorgino R, Rini GB, *et al.* Prevalence of hypovitaminosis D in elderly women in Italy: clinical consequences and risk factors. *Osteoporos Int* 2003;**14**:577–82. doi:10.1007/s00198-003-1390-7

6 Johns Hopkins CSSE. 2019 Novel Coronavirus COVID-19 (2019-nCoV) Data Repository by Johns Hopkins CSSE. https://github.com/CSSEGISandData/COVID-19 (accessed 26 Apr 2020).

7 CIESIN, Columbia University. Population Count, v4.11: Gridded Population of the World (GPW), v4 | SEDAC. 2020. doi:10.7927/H4JW8BX5

8 Davies G. The Epidemic Severity Index: Estimating Relative Local Severity of Novel Disease Outbreaks. *medRxiv* 2020;:2020.04.23.20077685. doi:10.1101/2020.04.23.20077685

9 Mckensie R L, Liley J B, Björn L O. UV Radiation: Balancing Risks and Benefits. Wiley Online Library. 2009.https://onlinelibrary.wiley.com/doi/full/10.1111/j.1751-1097.2008.00400.x (accessed 29 Mar 2020).

[6] To "control for a variable" is to hold it constant, or within a range, in order to remove it as a possible cause for some outcome further down the causal chain. This can clarify if another variable is truly at cause. Care must be exercised choosing control variables: controlling inappropriate variables may generate statistical bias incorrectly implying cause where there is none.

10 WHO | UV Index. Published Online First: 16 October 2017.http://www.who.int/uv/intersunprogramme/activities/uv_index/en/ (accessed 29 Mar 2020).

11 Watkins J. Preventing a covid-19 pandemic. *BMJ* 2020;**368**. doi:10.1136/bmj.m810

12 Romagnoli E, Caravella P, Scarnecchia L, *et al.* Hypovitaminosis D in an Italian population of healthy subjects and hospitalized patients. British Journal of Nutrition. 1999;**81**:133–7. doi:10.1017/s0007114599000264

13 Katharina Simon A, Hollander GA, McMichael A. Evolution of the immune system in humans from infancy to old age. *Proceedings of the Royal Society B: Biological Sciences* 2015;**282**. doi:10.1098/rspb.2014.3085

14 Gil Mor IC. The Immune System in Pregnancy: A Unique Complexity. *Am J Reprod Immunol* 2010;**63**:425. doi:10.1111/j.1600-0897.2010.00836.x

15 Montecino-Rodriguez E, Berent-Maoz B, Dorshkind K. Causes, consequences, and reversal of immune system aging. *J Clin Invest* 2013;**123**:958. doi:10.1172/JCI64096

16 Government of Canada, Statistics Canada. Vitamin D levels of Canadians, 2012 to 2013. 2014.https://www150.statcan.gc.ca/n1/pub/82-625-x/2014001/article/14125-eng.htm (accessed 26 Apr 2020).

17 Park J-H, Hong I, Chung J, *et al.* Vitamin D status in South Korean population: Seven-year trend from the KNHANES. *Medicine* 2018;**97**. doi:10.1097/MD.0000000000011032

18 Rabenberg M, Scheidt-Nave C, Busch MA, *et al.* Vitamin D status among adults in Germany – results from the German Health Interview and Examination Survey for Adults (DEGS1). *BMC Public Health* 2015;**15**:1–15. doi:10.1186/s12889-015-2016-7

19 Cristina Palacios LG. Is vitamin D deficiency a major global public health problem? *J Steroid Biochem Mol Biol* 2014;**144PA**:138. doi:10.1016/j.jsbmb.2013.11.003

20 Rabenberg M, Scheidt-Nave C, Busch MA, *et al.* Implications of standardization of serum 25-hydroxyvitamin D data for the evaluation of vitamin D status in Germany, including a temporal analysis. *BMC Public Health* 2018;**18**. doi:10.1186/s12889-018-5769-y

21 Tabrizi R, Moosazadeh M, Akbari M, *et al.* High Prevalence of Vitamin D Deficiency among Iranian Population: A Systematic Review and Meta-Analysis. *Iran J Med Sci* 2018;**43**:125.https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5936844/ (accessed 26 Apr 2020).

22 Saliba W, Rennert HS, Kershenbaum A, *et al.* Serum 25(OH)D concentrations in sunny Israel. *Osteoporos Int* 2011;**23**:687–94. doi:10.1007/s00198-011-1597-y

23 Yu MG, Castillo-Carandang N, Sison MEG, *et al.* Development and validation of a sunlight exposure questionnaire for urban adult Filipinos. *Epidemiol Health* 2018;**40**. doi:10.4178/epih.e2018050

24 Alipio M. Vitamin D Supplementation Could Possibly Improve Clinical Outcomes of

Patients Infected with Coronavirus-2019 (COVID-2019). Published Online First: 9 April 2020. doi:10.2139/ssrn.3571484

25 Maeda SS, Saraiva GL, Kunii IS, *et al.* Factors affecting vitamin D status in different populations in the city of São Paulo, Brazil: the São PAulo vitamin D Evaluation Study (SPADES). *BMC Endocr Disord* 2013;**13**:14. doi:10.1186/1472-6823-13-14

26 CIA. Ethnic groups. CIA World Factbook. 2010.https://www.cia.gov/library/publications/the-world-factbook/fields/400.html (accessed 26 Apr 2020).

27 Lawlor DA E al. Triangulation in aetiological epidemiology. - PubMed - NCBI. *International Journal of Epidemiology* Published Online First: 1 December 2016.https://www.ncbi.nlm.nih.gov/pubmed/28108528 (accessed 18 Apr 2020).

28 Website NHS. Pregnancy supplements 'don't help, just take vit D and folic acid'. nhs.uk. https://www.nhs.uk/news/pregnancy-and-child/pregnancy-supplements-dont-help-just-take-vit-d-and-folic-acid/ (accessed 1 May 2020).

29 Are more black, Asian and minority ethnic people dying with Covid-19 than might be expected? The Nuffield Trust. 2020.https://www.nuffieldtrust.org.uk/news-item/are-more-black-asian-and-minority-ethnic-people-dying-with-covid-19-than-might-be-expected (accessed 1 May 2020).

30 Website. https://www.thelancet.com/pdfs/journals/lancet/PIIS0140-6736(20)30922-3.pdf (accessed 1 May 2020).

31 Cook T, Kursumovic E, Lennane S. Exclusive: deaths of NHS staff from covid-19 analysed. Health Service Journal. https://www.hsj.co.uk/exclusive-deaths-of-nhs-staff-from-covid-19-analysed/7027471.article (accessed 1 May 2020).

32 New York Department of Health and Mental Hygiene. Age-adjusted rates of lab confirmed COVID-19 nonhospitalized cases, estimated non-fatal hospitalized cases, and patients known to have died as of April16,2020. NYC.gov. https://www1.nyc.gov/assets/doh/downloads/pdf/imm/covid-19-deaths-race-ethnicity-04162020-1.pdf (accessed 1 May 2020).

33 McCollum EV, Pitz W, Simmonds N, *et al.* The effect of additions of fluorine to the diet of the rat on the quality of the teeth. 1925. Studies on experimental rickets. XXI. An experimental demonstration of the existence of a vitamin which promotes calcium deposition. 1922. The effect of additions of fluorine to the diet of the rat on the quality of the teeth. 1925. *J Biol Chem* 2002;**277**:E8.https://www.ncbi.nlm.nih.gov/pubmed/11991957

34 Matthias Wacker MFH. Sunlight and Vitamin D: A global perspective for health. *Dermatoendocrinol* 2013;**5**:51. doi:10.4161/derm.24494

35 Kilbourne ED. Influenza Pandemics of the 20th Century. *Emerg Infect Dis* 2006;**12**:9. doi:10.3201/eid1201.051254

36 Palese P. Influenza: old and new threats. *Nat Med* 2004;**10**:S82–7. doi:10.1038/nm1141

37 Hamming I, Timens W, Bulthuis ML, *et al.* Tissue Distribution of ACE2 Protein, the

Functional Receptor for SARS Coronavirus. A First Step in Understanding SARS Pathogenesis. *J Pathol* 2004;**203**. doi:10.1002/path.1570

38 Oudit GY E al. The role of ACE2 in cardiovascular physiology. - PubMed - NCBI. 2003.https://www.ncbi.nlm.nih.gov/pubmed/12691672 (accessed 29 Mar 2020).

39 Chamsi-Pasha MAR, Shao Z, Wilson Tang WH. Angiotensin-Converting Enzyme 2 as a Therapeutic Target for Heart Failure. *Curr Heart Fail Rep* 2014;**11**:58. doi:10.1007/s11897-013-0178-0

40 Jia HP, Look DC, Shi L, *et al.* ACE2 Receptor Expression and Severe Acute Respiratory Syndrome Coronavirus Infection Depend on Differentiation of Human Airway Epithelia. *J Virol* 2005;**79**:14614. doi:10.1128/JVI.79.23.14614-14621.2005

41 Hoffmann M, Kleine-Weber H, Krüger N, *et al.* The novel coronavirus 2019 (2019-nCoV) uses the SARS-coronavirus receptor ACE2 and the cellular protease TMPRSS2 for entry into target cells. doi:10.1101/2020.01.31.929042

42 Dijkman R, Jebbink MF, Deijs M, *et al.* Replication-dependent downregulation of cellular angiotensin-converting enzyme 2 protein expression by human coronavirus NL63. *J Gen Virol* 2012;**93**:1924–9. doi:10.1099/vir.0.043919-0

43 Ji X, Zhang C, Zhai Y, *et al.* TWIRLS, an automated topic-wise inference method based on massive literature, suggests a possible mechanism via ACE2 for the pathological changes in the human host after coronavirus infection. *medRxiv* 2020;:2020.02.24.20025437. doi:10.1101/2020.02.24.20025437

44 Chen I-Y, Chang SC, Wu H-Y, *et al.* Upregulation of the Chemokine (C-C Motif) Ligand 2 via a Severe Acute Respiratory Syndrome Coronavirus Spike-ACE2 Signaling Pathway. *J Virol* 2010;**84**:7703. doi:10.1128/JVI.02560-09

45 Dancer RCA, Parekh D, Lax S, *et al.* Vitamin D deficiency contributes directly to the acute respiratory distress syndrome (ARDS). *Thorax* 2015;**70**:617–24. doi:10.1136/thoraxjnl-2014-206680

46 Ednan K. Bajwa , Ishir Bhan , Sadeq Quraishi , Michael Matthay , B. T. Thompson. Low Vitamin D Status Occurs in 90% of Patients with ARDS and Is Associated with Longer Duration of Mechanical Ventilation. 2016.https://www.atsjournals.org/doi/abs/10.1164/ajrccm-conference.2016.193.1_MeetingAbstracts.A1846 (accessed 24 Mar 2020).

47 Et al SY. Chronic vitamin D deficiency induces lung fibrosis through activation of the renin-angiotensin system. - PubMed - NCBI. 2017. https://www.ncbi.nlm.nih.gov/pubmed/28607392# (accessed 24 Mar 2020).

48 Ivanov VI, Dorofeykov VV, Kaystrya IV, *et al.* [Vitamin D and laboratory indicators of cardiovascular risk in the elderly.]. *Klin Lab Diagn* 2020;**65**:11–5. doi:10.18821/0869-2084-2020-65-1-11-15

49 Ganji V, Tangpricha V, Zhang X. Serum Vitamin D Concentration ≥75 nmol/L Is Related to Decreased Cardiometabolic and Inflammatory Biomarkers, Metabolic Syndrome, and Diabetes; and Increased Cardiorespiratory Fitness in US Adults. *Nutrients* 2020;**12**. doi:10.3390/nu12030730

50 Feart C, Helmer C, Merle B, *et al.* Associations of lower vitamin D concentrations with cognitive decline and long-term risk of dementia and Alzheimer's disease in older adults. *Alzheimers Dement* 2017;**13**:1207–16. doi:10.1016/j.jalz.2017.03.003

51 Littlejohns TJ, Henley WE, Lang IA, *et al.* Vitamin D and the risk of dementia and Alzheimer disease. *Neurology* 2014;**83**:920–8. doi:10.1212/WNL.0000000000000755

52 Zheng Y-Y, Ma Y-T, Zhang J-Y, *et al.* COVID-19 and the cardiovascular system. *Nat Rev Cardiol* 2020;:1–2. doi:10.1038/s41569-020-0360-5

53 Martucci G, McNally D, Parekh D, *et al.* Trying to identify who may benefit most from future vitamin D intervention trials: a post hoc analysis from the VITDAL-ICU study excluding the early deaths. *Crit Care* 2019;**23**:1–11. doi:10.1186/s13054-019-2472-z

54 Kong J, Zhu X, Shi Y, *et al.* VDR Attenuates Acute Lung Injury by Blocking Ang-2-Tie-2 Pathway and Renin-Angiotensin System. *Mol Endocrinol* 2013;**27**:2116. doi:10.1210/me.2013-1146

55 Xu J, Yang J, Chen J, *et al.* Vitamin D alleviates lipopolysaccharide‐induced acute lung injury via regulation of the renin‐angiotensin system. *Mol Med Rep* 2017;**16**:7432–8. doi:10.3892/mmr.2017.7546

56 Yuan W E al. 1,25-dihydroxyvitamin D3 suppresses renin gene transcription by blocking the activity of the cyclic AMP response element in the renin gene promoter. - PubMed - NCBI. https://www.ncbi.nlm.nih.gov/pubmed/17690094 (accessed 24 Mar 2020).

57 Jiménez-Sousa MÁ, Martínez I, Medrano LM, *et al.* Vitamin D in Human Immunodeficiency Virus Infection: Influence on Immunity and Disease. *Front Immunol* 2018;**9**. doi:10.3389/fimmu.2018.00458

58 Martineau AR, Jolliffe DA, Hooper RL, *et al.* Vitamin D supplementation to prevent acute respiratory tract infections: systematic review and meta-analysis of individual participant data. *BMJ* 2017;**356**. doi:10.1136/bmj.i6583

59 Christopher KB. Vitamin D and critical illness outcomes. *Curr Opin Crit Care* 2016;**22**:332–8. doi:10.1097/MCC.0000000000000328

60 Aspelund T, Grübler MR, Smith AV, *et al.* Effect of Genetically Low 25-Hydroxyvitamin D on Mortality Risk: Mendelian Randomization Analysis in 3 Large European Cohorts. *Nutrients* 2019;**11**. doi:10.3390/nu11010074

61 And Gandini AP. Vitamin D supplementation and total mortality: a meta-analysis of randomized controlled trials. - PubMed - NCBI. https://www.ncbi.nlm.nih.gov/pubmed/17846391 (accessed 29 Mar 2020).

62 Keum N E al. Vitamin D supplementation and total cancer incidence and mortality: a meta-analysis of randomized controlled trials. - PubMed - NCBI. https://www.ncbi.nlm.nih.gov/pubmed/30796437 (accessed 29 Mar 2020).

63 Amrein K, Papinutti A, Mathew E, *et al.* Vitamin D and critical illness: what endocrinology can learn from intensive care and vice versa. *Endocrine Connections* 2018;**7**:R304–15. doi:10.1530/EC-18-0184

64 Vitamin D and Health - SACN.

2016.https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/537616/SACN_Vitamin_D_and_Health_report.pdf (accessed 24 Mar 2020).

65 Colecalciferol - NICE, BNF. https://bnf.nice.org.uk/drug/colecalciferol.html (accessed 24 Mar 2020).

66 Scenario: Prevention of vitamin D deficiency in adults - NICE. NICE. 2018.https://cks.nice.org.uk/vitamin-d-deficiency-in-adults-treatment-and-prevention#!scenario:1

67 Davies G, Garami AR, Byers J. COVID-19 and Vitamin D Information - 2 Page Summary. 2020.https://bit.ly/VitDCovid19SummaryShort

68 Giustina A, Adler RA, Binkley N, *et al.* Controversies in Vitamin D: Summary Statement From an International Conference. *J Clin Endocrinol Metab* 2019;**104**:234–40. doi:10.1210/jc.2018-01414

69 Pearl J. Causal Inference in the Health Sciences: A Conceptual Introduction. *Health Serv Outcomes Res Methodol* 2001;**2**:189–220. doi:10.1023/A:1020315127304

70 Pearl J. *Causality*. Cambridge University Press 2009. https://play.google.com/store/books/details?id=LLkhAwAAQBAJ