**A Novel Causal Risk-Based Decision-Making Methodology: The Case of Coronavirus**

Stavros K. Stavroglou1, Bilal M. Ayyub2\*, Vasileios Kallinterakis3,

Athanasios A. Pantelous1, H. Eugene Stanley4

1Department of Econometrics and Business Statistics, Monash University, Clayton VIC 3800, Australia.

[stavros.stavroglou@monash.edu](mailto:stavros.stavroglou@monash.edu) and [Athanasios.Pantelous@monash.edu](mailto:Athanasios.Pantelous@monash.edu)

2Center for Technology and Systems Management, University of Maryland, College Park, MD 20742, USA. [ba@umd.edu](mailto:ba@umd.edu)

3University of Liverpool Management School, Chatham Building, Chatham Street, Liverpool L69 7ZH, UK. [V.Kallinterakis@liverpool.ac.uk](mailto:V.Kallinterakis@liverpool.ac.uk)

4 Department of Physics and Center for Polymer Studies, Boston University, Boston, MA 02215, USA. [hes@bu.edu](mailto:hes@bu.edu)

\*corresponding author

**Abstract**

Either in the form of nature’s wrath or a pandemic, catastrophes cause major destructions in societies, thus requiring policy and decision makers to take urgent action by evaluating a host of interdependent parameters, and possible scenarios. The primary purpose of this paper is to propose a novel risk-based, decision-making methodology capable of unveiling causal relationships between pairs of variables. Motivated by the ongoing global emergency of the coronavirus pandemic, the paper elaborates on this powerful quantitative framework drawing on data from the US at the county level aiming at assisting policy and decision makers in taking timely action amid this emergency. This methodology offers a basis for identifying potential scenarios and consequences of the ongoing 2020 pandemic by drawing on weather variables to examine the causal impact of changing weather on the trend of daily coronavirus cases.

**KEY WORDS:** Decision Making; Risk Quadruplet; Quantitative Analysis; Causality; Coronavirus; US county-level data

200-character summary:

Events, like the outbreak of a local epidemic can rapidly evolve into a looming global pandemic. Unveiling causal interactions, a novel causal risk-based decision-making framework is developed.

1. **Introduction**

Decision-making and risk analysis addresses a host of diverse real-world problems (Howard, 1988), (McGill, Ayyub & Kaminskiy, 2007), (Paté-Cornell, 2012), (Borgonovo, Cappelli, Maccheroni & Marinacci, 2018) and (Kontosakos, Hwang, Kallinterakis & Pantelous, 2020) by traditionally drawing heavily on quantitative methodologies, which rely on interdisciplinary designs via the integration of several disciplines (Aven, 2012) and (Aven & Zio, 2014).Despite the large number of risk-based decision-making frameworks developed in that respect, it is nevertheless interesting to note that they tend to only rarely incorporate causal arguments in the treatment of decisions, even though real-world practice decisions often depend heavily on causal knowledge and reasoning (Hagmayer & Fernbach, 2017). Although the scope of decisions generally involves determining potential actions (e.g., identifying the location that must go into lockdown due to steadily increasing numbers of the 2019 Coronavirus, COVID-19, cases) or outcomes requiring actions (e.g., a healthcare system under strain) (Borasio, Gamondi, Obrist & Jox, 2020) and (Hanna, Evans & Booth, 2020), it can be expanded to include risk-based criteria. For example, under the assumption of instituting and enforcing a lockdown in an area, decision and policy makers must examine a set of other actions (as they, jointly, help mitigate the COVID-19 spread); thus, allowing health systems to better cope with this disease and expedite the resumption of economic activity. The outcomes of such decisions largely depend on underlying causal mechanisms. Lockdowns, for instance, allow healthcare systems to better cope with this disease (as well as facilitating the smooth resumption of economic activities afterward); however, the outcomes depend on a complex system of factor-interactions. Lockdowns may keep people at a physical distance, resulting in flattening an area’s coronavirus curve of the number of new cases of hospitalizations or deaths; while also giving rise to increased mental health issues (thus, potentially, putting more strain onto the welfare system) (Kumar & Nayar, 2020). Thus, causal knowledge about the system under consideration assists society in making adaptive decisions, thus, enhancing risk management.

In this paper, a novel risk-based decision-making methodology is introduced for structuring the decision-making process with the COVID-19 pandemic as a case study. In this regard, the proposed methodology extends the risk triplet (Kaplan & Garrick, 1981) and (Kaplan, 1997) of scenarios, likelihoods and consequences to accommodate an explicit treatment of causality that might be implicitly included in the decision-making. Such an approach enables accounting for the performance of a complex system including likely outcomes, sensitivities, areas of importance, system interactions, and areas of uncertainty. The resulting risk quadruplet not only considers the set of the first three standard questions to define “risk” based on the standard risk triplet (*What can go wrong?*, *How likely is it?* and *What are the consequences?*), but also addresses explicitly the fourth question (*What are the causal relations?*). The latter is key to risk assessment research, since deciphering the nature of causal interactions of different variables relevant to a complex system can aid us in assessing its evolution – and, thus improve our decision and policy making abilities.

1. **Introduction to Pattern Causality**

The concept of causality permeated decision making since the antiquity, with people since then tending to interpret it across a continuum, ranging from arbitrary causality based on random observations to the identification of consistent patterns that separated random observations from truly causal observations (Cohen & Reeve, 2020), (Cornblatt, 2020), (Gerber, Hunger & Pingree, 2001) and (Crawford & Sen, 1996). In more recent times, research on causality has shifted toward extrapolating from consistent patterns in order to define governing principles of systems bearing varying degrees of complexity.

This abstraction allows us to formulate alternative scenarios and consequences of what to expect, based on the established causalities identified. This abstraction and the very principles that emerge from such a process provide the motivation to design this proposed causal risk-based decision-making methodology conceptualized in this section.

To begin with, the crux of the argument for the methodology rests upon the enhanced volatility and interconnectivity of many challenges facing humanity that, increase complexity in decision making and force us to account for multiple risk factors; this, in turn, is prompting public discourse to acknowledge the importance of unveiling the structure of complex systems in order to improve the assessment of those risk factors.

Recently, Stavroglou, Pantelous, Stanley and Zuev (2019, 2020) proposed the pattern causality (PC, hereafter) algorithm geared towards deciphering the nature of causality, drawing on theories of symbolic dynamics and attractor reconstruction. According to their treatment, each time series can be represented through patterns that account for one-step-ahead percentage changes. Symbolic dynamics allow the patterns to expose the nature of causality, yet not before a valid causal relationship is first identified. Time series belonging to the same dynamical system are parts of a common attractor, which corresponds to the states of that system. In the present paper, the PC framework’s steps are abstracted in a way that can be used for a wide spectrum of decision-making settings (hereafter, referred to abstractly as “systems”), such as in business, economy, weather, ecosystems, climate, pandemics and more. This approach breaks down decision making into a series of basic steps. It can add value to almost any situation, especially where serious or catastrophic outcomes are of concern. These steps can be used at different levels of detail and with varying degrees of formality, depending on the situation. The key to using the process is in completing each step in the most simple, practical way to provide the information a decision or policy maker needs. Some situations are so complex that detailed causal assessments are needed, but most can be addressed with simpler causal assessments.

The methodology provides guidance towards early stage or minor decisions, which at the end enables concluding or major decisions to be made in a transparent way. This is contrary to many “black box” methodologies which simply receive input and produce an output where the end product is based more on randomness, rather than causality. Thus, the purpose of the proposed methodology is to provide enough information to help inform decision or policy making practices. It focuses on organizing information for logical understanding in a highly structured fashion. However, it neither replaces the decision maker, nor does it lead into burdensome risk or inconclusive assessments involving gathering information that is irrelevant to the decision. On average, and over time, appropriate decisions made through this process should generate appropriate outcomes, while also offering logical explanations for decisions with unfavorable outcomes (Peysakhovich & Karmarkar, 2016). The PC algorithm provides a rational basis for a causal risk-based decision-making methodology as developed in the subsequent section.

1. **Causal risk-based decision-making methodology**

The purpose of developing this novel decision-making methodology is to unveil causal[[1]](#footnote-1) relationships between pairs of variables. However, as is the case with many real-world problems, the available datasets might contain well over a few thousands of variables of any type (categorical, numerical, etc.). Such complexity might lead to ill-directed analysis, although in practice, anyone interested might be centered on particular variable(s),  (see the *red* node in **Figs. 1a, b**), which can play a key role in some case-specific decision-making problem. In the proposed setting, the outcome of the decision-making process is expected to be a causality-enabled prediction of such critical variable(s). In what follows, the proposed causal decision-making methodology is presented with its five main steps, each outlined under a separate heading.

* 1. **Formulate the problem**

The first step for any data-driven decision-making methodology is to describe the raw data collected. The data typically resonate with our understanding of causal relations among variables and contain implications for subsequent decisions. However, whether data becomes information or not, depends on what the decision maker expects to extract from them. In this regard, before proceeding further with the directed analysis, as a second step, filtering out some of the key variables is necessary. In this framework, exploring dependencies throughout a time horizon[[2]](#footnote-2) retains only time series variables. Consequently, all other categories of data are removed. After completing the initial filtering, only time series variables, , depicted as nodes in **Fig. 1**[[3]](#footnote-4) are retained. These variables have some implied - yet currently unknown - causal structure.[[4]](#footnote-5)

In real-world applications, variables identified for a particular decision situation from different perspectives are not only subject to a causal network structure (see **Fig. 1a**) but also are governed by the same attractor dynamics, i.e., belong to the same manifold (see the *orange* manifold in **Fig. 1b**).[[5]](#footnote-6) These dynamics can approximately be observed by rendering scatterplots of combinations of two or three variables. For illustration purposes, **Fig. 1b** shows the Lorenz attractor which is built by taking the scatterplot of the three constituent variables that comprise the Lorenz Differential Equations (Lorenz, 2004) . In practice, it is often the case that such scatterplots, particularly when produced from real-world data, visually appear as nothing but noise, which does not help in a decision-making process. Therefore, since it is essential not to rely on visual techniques, our treatment algorithmically exploits the underlying attractor dynamics to perform causality-based predictions which can be used in decision making.

The proposed causal decision-making methodology with real time-series variables is robust to produce results in case of missing values in the data. In such cases, moving on to the next value works and is efficient for prediction purposes, especially when the accuracy is high despite the missing values. However, when interpreting causality, caution must be exercised since the higher the percentage of missing values, the greater the uncertainty of interpretation. A decision maker should also exercise judgement about whether to remove variables with missing values at the cost of possibly losing some predictive power. Thus, the decision maker’s choice on which variables to remove plays a decisive role on what the subsequent steps should be.

|  |  |
| --- | --- |
|  |  |

**Fig. 1. (a)**  A selection of time series variables with unknown interdependencies (denoted by grey links), with the focus being directed toward some critical variable(s) . (**b)** Various combinations of those selected variables , are governed by the same attractor dynamics; however, in the right part of this figure, we illustrate only in the attractor .

* 1. **Construct the dynamics**

To establish decision-enabling causality for a pair of variables, causing , first the dynamics of the governing attractor need to be retrieved. In rare cases where the governing equations of the variables are known by physics or other means, such causal dynamics should be used, e.g., as represented in **Fig. 1b** with the Lorenz attractor. Frequently, however, the causal dynamics can be extracted without knowing the governing physics or equations of real-world systems. The approximation process starts with the phase of *shadow attractor reconstruction*, during which a shadow version of the original attractor is reconstructed (**Fig. 1b**). Α shadow attractor can be plotted[[6]](#footnote-7) by assuming the scatterplot of and its lagged versions (see, **Fig. 2a** from left to right, , ).[[7]](#footnote-8)

For each pair, , the optimal combination of embedding dimension () and time lag () is identified in order to reconstruct their shadow attractors and retrieve any causal information that exists in the given pair. The optimal combination of is the one which produces the highest causality accuracy from to . However, in real-world applications, it is possible for decision makers to have to choose among different combinations of and with very comparable levels of causality accuracy. In this regard, the time scales of the variables (which might be in minutes, hours, days, years etc.) as well as the decision maker’s knowledge or understanding of the overall system (i.e., some systems involve cyclical behavior, others are more stochastic) play a pivotal role on which of the optimal combinations of and to pick. In particular, for short time series the decision maker needs to consider the *first causality point* (FCP) which reflects the minimum amount of points in the shadow attractor.[[8]](#footnote-9)

Once a decision maker identifies the optimal combination of and for each pair , then the following two questions should be answered: “Shall the optimal combination of and for each pair separately be used (namely, choice A)?” Or “shall the optimal combination over all pairs be used (namely, choice B)?” The latter is calculated by getting the average accuracy over all pairs for each combination of and . On the one hand, choice A may be better when dealing with few variables, i.e., for small datasets, where getting into an analytical narrative of the various combinations of and should not be a hurdle. However, another case for choice A would be when the accuracy variances over all pairs under the same and combinations are intolerably high, which would suggest the loss of valuable causal information and predictive capacity for some pairs. On the other hand, if a dataset comprises of many variables, i.e., big dataset, and the accuracy variance is sufficiently low for the decision maker’s demands then for simplicity, choice B is preferred. Further, using the same and combination for all pairs can enable a more viable narrative and interpretation of the results, rendering the rest of the decision-making process more straightforward.[[9]](#footnote-10)

Furthermore, for a given pair, , testing for causality at each time step, from FCP onwards is necessary. Specifically, each time step identifies the nearest neighbors of , takes their time indices denoted by for each neighbor, and finds the points on using the time indices from and adding steps ahead[[10]](#footnote-11). This process is called *neighborhood projection* and the rationale behind it is that if (causes), then topological information on should accurately predict topological information on . At this stage, two important intermediate considerations require the attention of a decision maker. Firstly, in order to identify the nearest neighbors, distances of points on the attractor need to be calculated. This can be achieved in many ways by implementing different metric distances. In general, someone may consider the Minkowski distance for some exponent (for , the Euclidean distance yields). The higher the , the more the distant points are penalized. If the decision maker decides to treat all the points equally, then should equal 1, effectively yielding the Manhattan distance[[11]](#footnote-12). Secondly, the number of nearest neighbors to keep for the next step (see **Fig. 2b**) should be set, which might vary due to the nature of the system studied. Theoretically, nearest neighbors are sufficiently enough because they create a simplex of higher dimension around the current point’s neighborhood, and thus contain all the topological information needed (Kantz & Schreiber, 2003). However, the decision maker may decide to include more nearest neighbors to transfer a more representative assembly of topological information into the next step.

|  |
| --- |
|  |
|  |

**Fig. 2. (a)**  Shadow attractor reconstruction technique: using lagged values of the same variable in order to approximate its original dynamics. Those original dynamics hold the causal information which is tapped in the next step. (**b)** Scanning for nearest neighbors in and using their projected (by ) time indices on . Those projections are used in the step of “signature extraction” in order to test for causality.

* 1. **Identify the causal profile**

For every projected nearest neighbor ,a weight is assigned, which is inversely proportional to the distance of its corresponding point on from the current point. Moreover, every point on a shadow attractor can be seen as a pattern which describes rates of change (Stavroglou et al., 2020). Patterns’ possible forms and complexity increase as the embedding dimension increases. Using the weights of the projected nearest neighbors and their corresponding patterns, their signature (see **Fig. 3a**) is calculated which is merely their weighted average pattern. If the signature coincides with the actual pattern of , then it is marked in this time step as caused . This is called in-sample prediction and is used to assess causality. Similarly, one can do out-of-sample prediction steps ahead of . The only difference is, for the projected nearest neighbors on , taking the ones with time indices , and thus the predicted signature for .

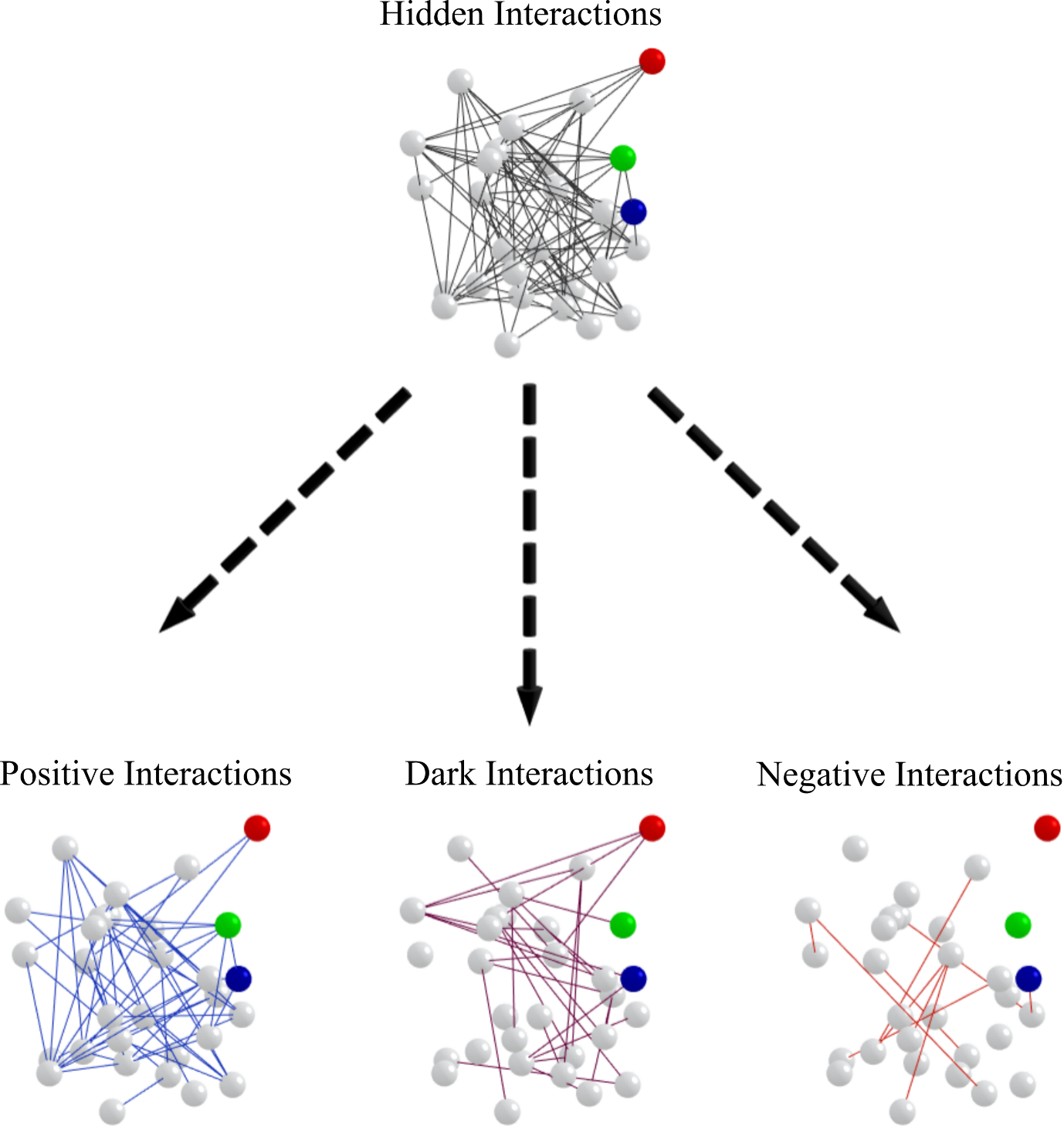
Identification of the pattern of provides a basis to assess the nature of causality by finding how it relates to the pattern of the causal variable , according to the matrix in **Fig. 3b**. This matrix charts all possible combinations of causal and affected variables at each time step. In this way, a system of three types of causal interactions is established. *Positive* causality refers to cases when ’s pattern causes the same patterns on the critical variable. *Negative* causality refers to cases of opposite pattern changes. Every other case of mixed pattern combinations is collectively termed *dark* causality (Stavroglou et al., 2019). This partition informs the decision maker about how each pair of variables of the specific problem interacts. Thus, there are variables which bolster the critical variable , others that act as inhibitors and some that influence in a more complicated way.

Repeating this process for a given pair , at each time step helps generate a certain pattern combination (depicted in slices for each in **Fig. 3c**). In order to have an overall view of the causal relationship , the frequency distribution of each pattern combination is counted, thus revealing information on how much positive, negative or dark each causal relationship is. This information is also summarized in the aggregate causality matrix which collects the aforementioned frequencies.

|  |  |
| --- | --- |
|  |  |
|  | |

**Fig. 3. (a)** Schematic representation of signature extraction; signature is the weighted average of the projected points from the previous step. The weights used are inversely proportional to the distance of the nearest neighbors. (**b)** From the predicted signature, the corresponding pattern is derived (as shown horizontally) and by seeing how it combines with the causal variable’s pattern we can derive what type of causality holds (provided that the predicted signature’s pattern is accurately predicted). (**c)** Repeating this process for each time step leads to the generation of the aggregate causality matrix which summarizes the combinations of patterns met.

Finally, once the causal profile of all relationships that are of interest is obtained (as implied in **Fig. 1a**), the initially hidden interactions of the dataset can be unveiled. Keeping for each pair their highest causality type only (positive, dark or negative) allows for a bird’s eye view of the three possible aspects of our causal network (see **Fig. 4**). Therefore, it makes sense to just focus on the single or few critical variables, and the way they are influenced by the rest of the variables.



**Fig. 4.** Having repeated the process as described in the previous steps, the hidden interactions are unveiled, thus revealing the three different aspects of the causal network.

* 1. **Calculate the causal chains frequency distributions**

**Fig. 4** provides a bird’s eye view of the three possible aspects of the causal network, and the previous section serves as a reconnaissance step, but our purpose is to exploit the causal relationships identified to perform risk analysis. In this regard, one very important perspective is the identification of causal chains (for each causal relationship). A causal chain is a successive accurate causality of the same type. Therefore, three types of causal chains are possible: positive, dark and negative (**Fig. 5a** left) in correspondence with the three types of causality identified in the previous step. The repetitive observation of a relatively long chain type (“long” depends on the problem context) can be used to probabilistically expect the future movements of the critical variable(s).

From a risk analysis perspective, the critical variable(s) needs to be understood in terms of their consequences. The first step is to identify for the critical variable(s) their possible scenarios which are entwined with them; in other words, for the critical variable(s) one needs to identify which values are: (a) *Tolerable*, (b) *Risky* and, (c) *Catastrophic* (**Fig. 5a** right). To decide the scenarios’ thresholds, the decision maker should be either an expert or seek expert knowledge. Based on these scenarios, the decision maker is able to anticipate and reduce the overall impact of any events that might ensue, to evaluate whether the potential risks can be hedged, to plan for adequate actions, to understand the financial implications, and identify the impact of and prepare for changes in the evaluation process.

To aggregate the information included in the causal chains over all the relationships that target a specific critical variable, it is important to calculate their frequency distributions for each causal type (see **Fig. 5a**). In this way, the decision maker can create a dashboard regarding the persistence of causal chains on the critical variable under consideration. Such information is very useful for a holistic risk analysis on which factors (variables) influence, and in what manner (causality type and chain lengths) according to the possible scenarios outlined beforehand. One can also build from here probabilistic models for each expected scenario.

|  |  |
| --- | --- |
|  |  |
|  |  |

**Fig. 5. (a)** Considering the type of causality at each time step we can collect the causal chains that influence the critical variable(s). The critical variable , partly influenced by the causal variables , takes roughly three ranges of values corresponding to tolerable, risky and catastrophic scenarios associated with the critical variable. (**b)** The totality of causality on the critical variable can be summarized via their respective frequency distributions in terms of their successive incidents (causal chains).

* 1. **Produce** **risk quadruplets and critical variable(s) prediction**

As in any risk-informed decision-making framework, all of the identifiable factors that affect a decision must be considered. However, the factors may have different levels of importance in the final decision. Therefore, an orderly decision analysis structure that considers more than just risk is necessary to give decision makers the information needed to make smart choices. In that context our framework, and especially the causal chains it reveals, can enhance the *risk triplet* framework (scenarios, critical variable frequency distribution, consequences (Borgonovo et al., 2018), (Kumar & Nayar, 2020) and (Kaplan & Garrick, 1981), thus generating *risk quadruplets* (i.e., risk triplet plus causal factors).

Analytically, for a given system we collect: (1) The possible scenarios () for our problem settings, (2) The frequency distribution of the critical variable(s) () which determine which scenario occurs in the first place, (3) The consequences that emerge from each scenario, both beneficial and detrimental, (4) The causal relationships that affect the critical variable(s) which can be summarized by their chains frequency distributions (as discussed in the previous section, see also **Fig. 5b**). Collectively those four components comprise the risk quadruplet, (see **Fig. 6a**), which is a risk analytic manifestation of our causal risk-based decision-making framework.

Ultimately, the crown result obtained is the assembly prediction of the critical variable(s). To do this we need to recall the out-of-sample prediction process for a single causal variable as described in a previous subsection. The assembly prediction (**Fig. 6b**) uses all the causal variables and calculates the weighted average of their individual predictions (on the critical variable). The weights used should be proportional to the in-sample prediction accuracy performed before. The out-of-sample assembly prediction then can be used to estimate what scenario to expect both in the short-term and in the long-term and thus formulate an educated decision-making.

|  |  |
| --- | --- |
|  |  |

**Fig. 6. (a)** The chief contribution of the current study is the extension of the well-established Risk Triplet framework into what the “Risk Quadruplets” with the fourth component being that of the causal information retrieved. (**b)** We can predict the future values of the critical variable(s) by using an assembly of predictions from all the causal variables considered.

1. **Making risk-based, causality-enabled decisions on the COVID-19 pandemic**

In this section, the novel causal risk-based decision-making framework developed in the previous section for the case of the COVID-19 pandemic is employed. On the one hand, the demonstrated application will be a step-by-step elaboration on how to use the risk quadruplets, from the very formulation of the problem at hand, to the risk assessment of the causal factors of the system at hand. On the other hand, our endeavor will be to provide useful implications for decision and policy makers, and thus a powerful causal-related decision tool on designing effectively, and therefore adjusting appropriately the measures to suppress the spread of COVID-19 in the country or area chosen, and simultaneously organize the actions needed to minimize the socioeconomic damage from the pandemic to several key sectors. Based then on the risk quadruplets, the appropriate decision-making path is selected, taking into account the decision maker’s tolerance to uncertainty and pursuing (but not limited to) the consideration of the actions which will produce the final decision to be made.

As it is expected for any real-world application, there are many influencing factors that can be fundamental sources of future change. In particular, examples of COVID-19 specific influencing factors include the numbers of COVID-19 cases, deaths, ages, severe travel restrictions, product suspensions, supply chain interruptions, among many others. Both epidemiological and laboratory studies have reported that weather might be an important factor in the transmission and survival of disease agents (Martens & McMichael, 2002), (Casanova, Jeon, Rutala, Weber & Sobsey, 2010) and (Gupta, Raghuwanshi & Chanda, 2020).[[12]](#footnote-13) Recently, the stability of COVID-19 has been reported to be similar to that of SARS CoV-1 on various types of inanimate surfaces in specific weather conditions (van Doremalen et al., 2020). Thus, in our case study, data only on temperature and precipitation are considered as approximate proxies for weather conditions, in order to investigate their causal interaction with the critical variable () which measures the daily number of COVID-19 cases for scenarios’ formulation.[[13]](#footnote-14) Historical data for the United States on the state and county level is collected from 23rd January till 31st May 2020.[[14]](#footnote-15) The state and county files contain FIPS[[15]](#footnote-16) codes, which is a standard geographic identifier to facilitate our analysis as it permits the combination of this data with temperature and precipitation data retrieved from the National Oceanic and Atmospheric Administration (NOAA) of the United States.[[16]](#footnote-17) The treatment proposed herein can be easily extended to include other factors once data on them become available.

In order to determine what influencing factors are likely to effect, various scenarios of the pandemic are defined as appropriate and customizable as possible. This scenario planning is used as an important basis for robust strategic decision-making, and as such to mitigate overwhelming uncertainty and create logical, consistent stories of how the pandemic situation might unfold. Further, the frequency distribution of *Y*, at the federal level is computed. As an example, our focus will be on the state-specific frequency distributions of the three most populous states, namely New York, California, and Illinois. Additionally, some expert sources are utilized in order to arrange the consequences connected to each scenario. Finally, a detailed map of how weather conditions influence the daily number of cases at the county level is presented, again focusing on these three states.

* 1. **Scenarios (Sj)**

A scenario is a succinct summary which describes what can possibly happen at a specific setup. COVID-19 cases represent our critical variable which shapes the decision-making scenarios assessed. In the current absence of vaccines and effective drug treatments, the following five scenarios will be considered, where the critical variable () is the number of cases, and are predetermined thresholds.

* **Scenario 1** (**Tolerable**): : The number of new cases is very manageable for the healthcare system, and the spread of the infection is either steady or slowing down.
* **Scenario 2** (**Risky: Warning**): : The number of new cases is manageable for the healthcare system; however, the spread of the infection is slightly increasing.
* **Scenario 3** (**Risky: Alarming**): : The number of new cases is manageable for the healthcare system; however, the spread of the infection is steadily increasing.
* **Scenario 4** (**Risky: Critical**): : The number of new cases is still manageable; however, the spread of the infection has increased significantly.
* **Scenario 5** (**Catastrophic**): : The number of new cases is not manageable, and the spread of the infection is out of control.

It is important to reiterate here that our methodology is very flexible, and can accommodate many possible scenarios with a variety of predefined thresholds which might be most suitable for the instance. Also, it does not require those thresholds to be the same across state and county levels. In this regard, the relevant team of the decision makers will have the opportunity to design the scenarios and map them with their possible consequences. The proposed framework can simultaneously be employed at the county/state/federal level.

* 1. **Frequency (Fj) distributions.**

An overview of the current problem is obtained by plotting the daily aggregate U.S. time series of COVID-19 cases (see **Fig. 7a**). As the figure illustrates, there exists an exponential growth from March to April, followed by a fluctuating pattern lasting until the end of May. Nevertheless, the picture obtained when plotting the aggregate cases over the three states of New York, California and Illinois dissolves the generic fluctuating pattern of **Fig. 7a** into distinct patterns for each state (see **Fig. 7b**).

More specifically, New York follows an exponential growth until the beginning of April and then oscillates downwards (with the exception of the last week of April, which displays a surge in COVID-19 cases similar to the top of the previous exponential rally). A different picture is plotted for California, which displays no exponential growth; rather COVID-19 cases, starting from March, display a steady and oscillating rise. No spikes, or tendencies to drop are observed. Illinois reveals a picture similar to California’s, with the difference being in May (whence COVID-19 cases seem to oscillate around a mean plateau).

The second component of the *risk quadruplet* methodology corresponds to the frequency distribution *F* of the critical variable which in our case is the daily cases of COVID-19. Calculating the frequency distribution both U.S.-wide and state-wide is a first step towards understanding the current situation and planning ahead based on the other components of the risk quadruplet.

Regarding the U.S.-wide daily COVID-19 cases (see **Fig. 7c**), it is apparent that most cases are lying between 0 and 5000 while there is also a significant “fat-tail” from 15000 to 30000 daily cases.

New York has a significant frequency of up to 1000 daily cases with the rest of the frequencies being significantly lower (see **Fig. 7d**). California displays a high frequency of up to 1000 daily cases with a significant concentration between 2000 and 5000 daily cases (see **Figure 7e**). Illinois has frequencies which decrease geometrically from up to 1000 to up to 3000 daily COVID-19 cases (see **Fig. 7f**).

|  |  |
| --- | --- |
|  | |
|  | |
|  |  |
|  |  |

**Fig. 7** (**a**)Time series of daily COVID-19 cases for all U.S. and (**b**) separately for the states of New York, California, and Illinois. Frequency distributions (*F* from *risk quadruplet*) for daily COVID-19 cases (**c**) for all U.S. and (**d**) for New York, (**e**) California and (**f**) Illinois, separately from 23rd January till 31st May 2020.

* 1. **Consequences (yj)**

The spread of COVID-19 across the United States has motivated the introduction of unprecedented measures aiming at containing the epidemic. However, apart from the consequences on citizens’ health, there are also other consequences impacting the society and the economy that need to be identified - and these need to be measured and mapped with the relevant scenarios. The consideration of the consequences, the measurement of their impact, and their mapping to the scenarios are all challenging tasks, highly related to both the decision makers and the aspects of the COVID-19 pandemic that they need to prepare to mitigate further. In our case, the following eight consequences are considered:

* Travel Restrictions
* Physical Distancing
* Face Masks Usage
* Lockdown
* Government support
* Supply chain interruption
* Reduced Consumption
* Working Capital Pressure

**Table I** maps the consequences with the five scenarios, and due to lack of relevant data, their impact is measured as neutral/low, medium and high with gray, dark yellow and red color, respectively.

* 1. **Causality (Aj)**

Each consequence is connected with every scenario, and measured accordingly (see **Table I**), and each scenario is connected with the critical variables’ frequency (in our case COVID-19 daily cases as in **Fig. 7**). However, what causes each critical variable (and, therefore what affects scenarios and, as a result, the culminating consequences) is reflected through the pattern causality analysis, from which the causal profile of the variables under study is obtained. In effect, knowledge of the causal interdependencies is the differentiating factor between risk triplets and risk quadruplets. Now, decision makers with a causal map, as in **Figs. 8** and **9,** have a top-level dashboard and can make informed decisions with deep knowledge about the bigger picture of the problem they face.

Pattern causality allows us to calculate the impact of precipitation and temperature (the main proxies of the weather conditions) on the number of daily COVID-19 cases (see **Figs. 8** and **9**, respectively) for every county in the United States. The overall striking result is that weather conditions do affect the number of daily COVID-19 cases, however in a variety of ways. The aggregate causal impacts of precipitation and temperature on daily cases for each county are illustrated in **Fig. 8** and **Fig. 9**, respectively, for the states of New York, California and Illinois.

**Table I.** Consequences per scenario.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Consequences** | **Scenario 1. Tolerable** | **Scenario 2. Warning** | **Scenario 3. Alarming** | **Scenario 4. Critical** | **Scenario 5. Catastrophic** |
| Travel Restriction |  |  |  |  |  |
| Physical Distancing |  |  |  |  |  |
| Face Masks Usage |  |  |  |  |  |
| Lockdown |  |  |  |  |  |
| Government support |  |  |  |  |  |
| Supply chain interruption |  |  |  |  |  |
| Reduced Consumption |  |  |  |  |  |
| Working Capital Pressure |  |  |  |  |  |

Neutral or low impact

Medium impact

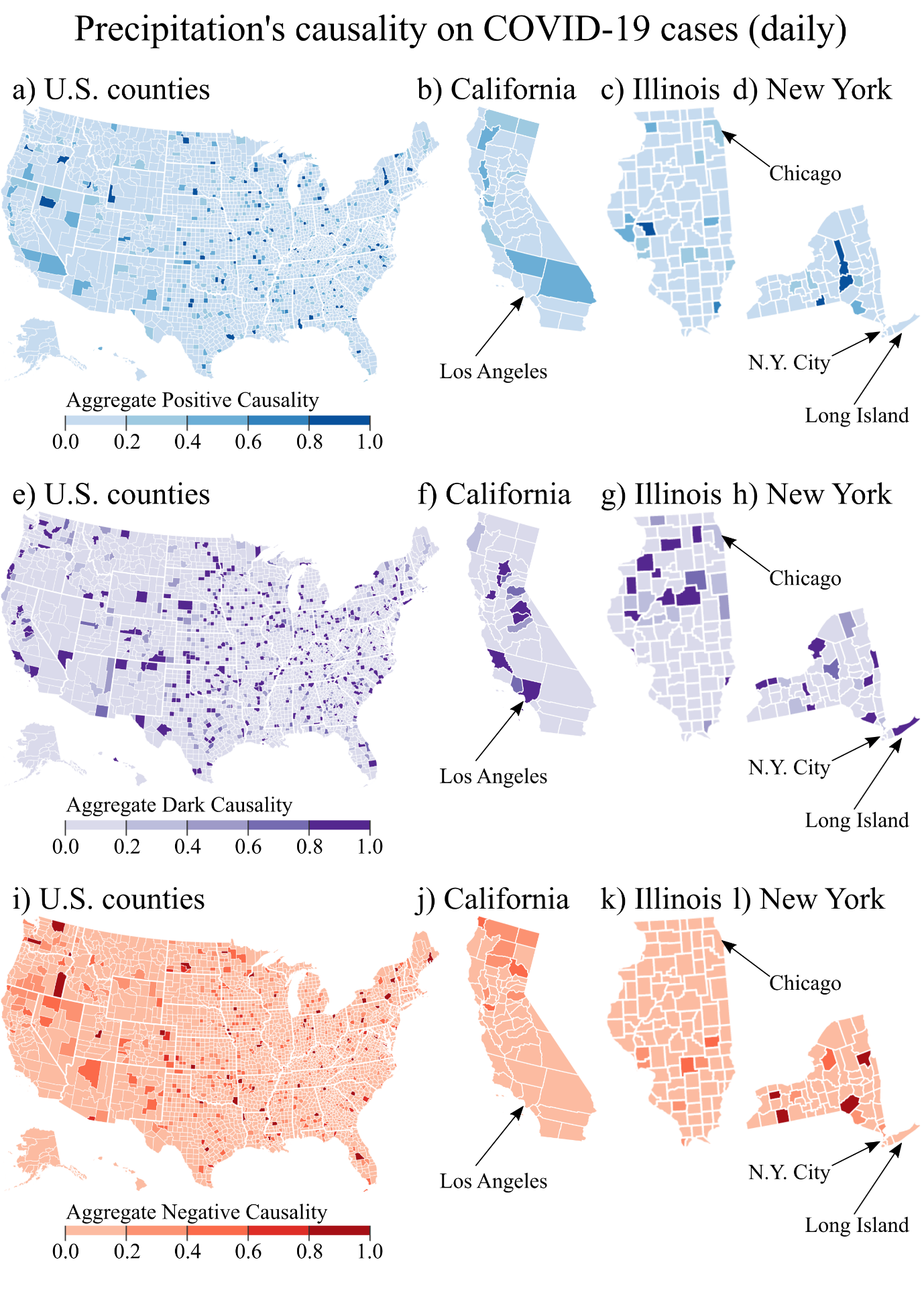
High impact

**4.4.1. Impact of Precipitation on COVID-19 cases**

The precipitation causal map of California reveals a clear structure: In northern Californian counties precipitation drives COVID-19 cases in a mostly negative way (**Fig. 8.j**), i.e., increases (decreases) in precipitation cause decreases (increases) in COVID-19 cases. Some other counties on the periphery of California display positive causality of precipitation on COVID-19 cases (**Fig. 8.b**), i.e., increases (decreases) in precipitation cause increases (decreases) in COVID-19 cases. Moreover, some counties located in central California along with the Los Angeles area expose the dark causality of precipitation on COVID-19 cases (**Fig. 8.f**). This finding suggests that changes in precipitation cause mixed changes in daily COVID-19 cases.

As far as Illinois is concerned, distinct dark causality is observed in its northern counties (**Fig. 8.g**), suggesting oscillations in COVID-19 cases which are intricately influenced by precipitation. The county of Chicago displays a dual (positive and dark) causality of precipitation on COVID-19 cases, suggesting a mixed influence with a tendency towards same-direction changes. Positive causality of precipitation is also found in central to west Illinois counties (**Fig. 8.c**). Furthermore, very few counties towards center-south Illinois exhibit some negative causality of precipitation on COVID-19 cases (**Fig. 8.k**).

Lastly, in the state of New York, precipitation displays positive causality mostly in some central counties (**Fig. 8.d**). Negative causality of precipitation in the state of New York is more eminent compared to the other two states examined here (**Fig. 8.l**), denoting that COVID-19 cases have been moving opposite to precipitation. Nevertheless, there is a significant number of counties where precipitation has dark causality as well (**Fig. 8.h**), most notably the county of Long Island.



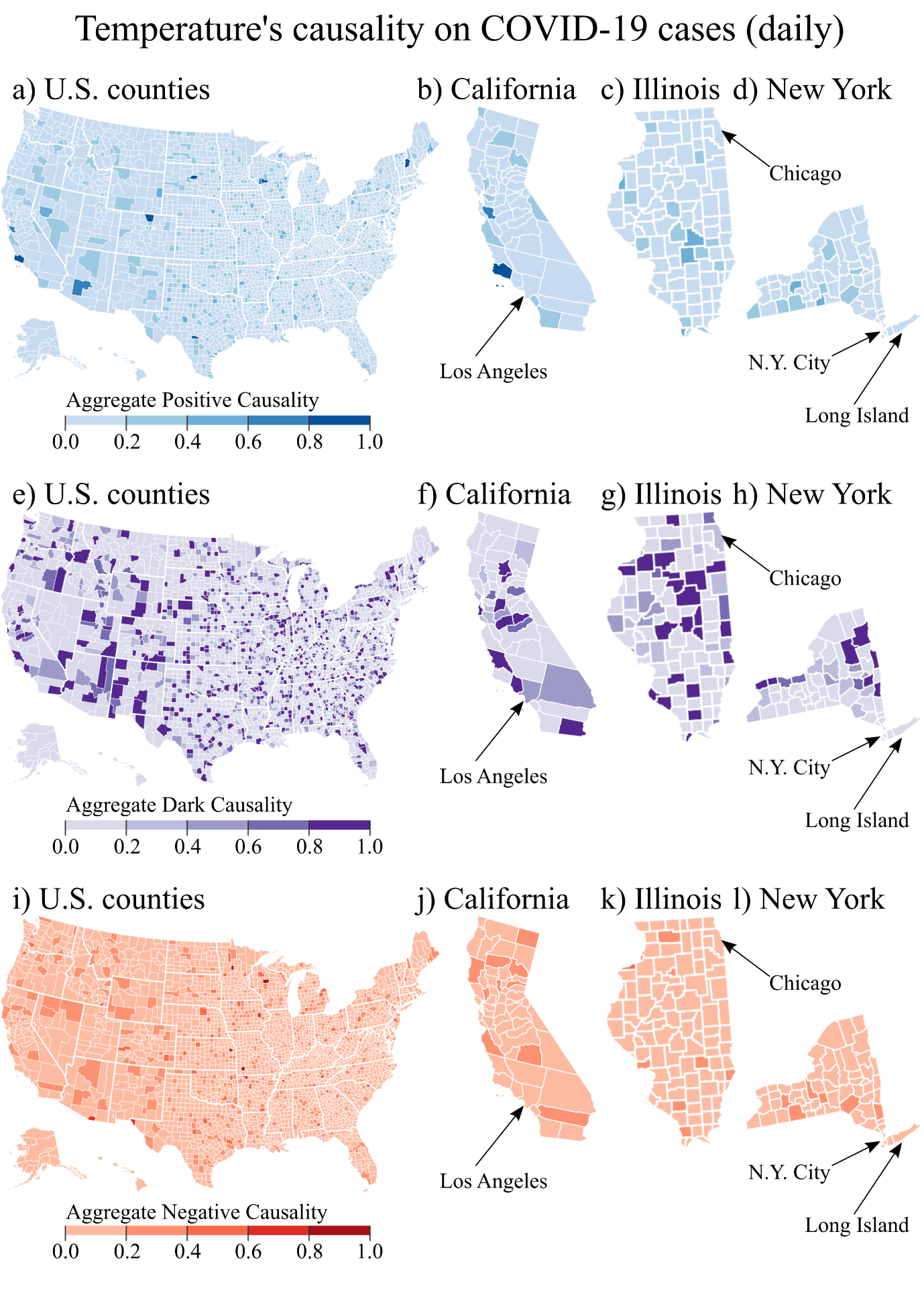
**Fig. 8.** Aggregate impact of daily precipitation on COVID-19 cases per day, over the period from 23 January 2020 to 31 May 2020. The map detail is on the county-level for the states of California, Illinois and New York, the US’ three most populous ones. To understand the way precipitation influences COVID-cases, a series of maps for each type of causality is presented. Namely, positive causality of precipitation on COVID-19 cases is presented in **a, b, c, d**; dark causality of precipitation on COVID-19 cases is presented in **e, f, g, h**; negative causality of precipitation on COVID-19 cases is presented in **i, j, k, l**.

**4.4.2. Impact of temperature on COVID-19 cases**

The temperature causal map of California is more mixed than its precipitation counterpart. Negative causality is observed in very few counties and at very low levels (**Fig. 9.j**), suggesting an insignificant opposite directional movement of cases driven by temperature. Some west-side California counties are characterized by positive causality of temperature on cases (**Fig. 9.b**). Most California counties along with Los Angeles have dark causality of temperature over cases (**Fig. 9.f**), implying a mixed oscillation.

As far as Illinois is concerned, distinct dark causality is detected mostly in central to northern counties (**Fig. 9.g**), suggesting oscillations in COVID-19 cases which are intricately influenced by temperature. The county of Chicago displays a dark causality of temperature on COVID-19 cases, suggesting mixed-direction changes. Positive causality of temperature is also found in very few (mostly central) Illinois counties (**Fig. 9.c**). Furthermore, even fewer counties exhibit negative causality of temperature on COVID-19 cases (**Fig. 9.k**) and at very low levels.

Finally, in the state of New York, temperature displays a very similar causality profile to those of California and Illinois. Positive causality is observed mostly in some south New York counties (**Fig. 9.d**). There, an increase in temperature suggests increases in COVID-19 cases. Negative causality of temperature in the state of New York is very sparse and low (**Fig. 9.l**). Nevertheless, there is a significant number of counties where precipitation has dark causality as well (**Fig. 9.h**), suggesting that temperature, overall, causes mixed oscillations in COVID-19 cases.



**Fig. 9.** Aggregate impact of daily average temperature on COVID-19 cases per day, over the period from 23 January 2020 to 31 May 2020. The map detail is on the county-level for the states of California, Illinois and New York, the US’s three most populous. To understand the way temperature influences COVID-cases, a series of maps for each type of causality is presented. Namely, positive causality of temperature on COVID-19 cases is presented in **a, b, c, d**; dark causality of temperature on COVID-19 cases is presented in **e, f, g, h**; negative causality of temperature on COVID-19 cases is presented in **i, j, k, l**.

* 1. **Risk Quadruplets (R) on the Coronavirus pandemic**

The proposed framework establishes the causal factors of the critical variable (whose frequency distribution must be known) in order to understand both the current scenario at hand (and its corresponding consequences) as well as what scenario to expect based on the causalities from the influencing factors.

**The case of California**. According to **Fig. 7**, daily COVID cases are increasing at an almost steady rate. Setting daily COVID-19 cases and daily COVID-19 cases, as a result, California is as of May 31 under Scenario 3 which is alarming. Looking at the frequency distribution of COVID-19 cases in California, reveals a significant concentration between 1000 and 2500 cases per day, a fact which renders the situation risky. Being in scenario 3, one would expect, according to **Table I**, a high impact on physical distancing, medium impact on travel restrictions, face masks usage, lockdown, government support, supply chain interruption, reduced consumption and working capital pressure. Of course, all this refers to the situation as of 31 May. However, in order to be able to better plan for the imminent future, the decision maker needs the fourth component of the risk quadruplet which is the causal impact on the critical variable of the daily COVID-19 cases. In California precipitation and temperature are influencing COVID-19 cases primarily in a dark way and then in a positive way. This means that depending on the county, local authorities should be aware of the weather forecasts and expect either a same-direction impact on the COVID-19 cases (positive causality) or most importantly mixed oscillations (dark causality).

**The case of Illinois**. According to **Fig. 7**, daily COVID cases have been increasing at an almost steady rate, before exhibiting a slowdown. Setting daily COVID-19 cases and daily COVID-19 cases, as a result, Illinois is as of May 31 under Scenario 2 (“warning”). Looking at the frequency distribution of COVID-19 cases in Illinois, reveals a significant concentration up to 3000 cases per day a fact which renders the situation risky. Being in scenario 2, this suggests, according to **Table I**, a medium impact on physical distancing, face masks usage, government support, and low impact on travel restrictions, lockdown, supply chain interruption, reduced consumption and working capital pressure. Of course, all this refers to the situation as of 31 May. However, to be able to better plan for the imminent future, the decision maker needs the fourth component of the Risk Quadruplet which is the causal impact on the critical variable of daily COVID-19 cases. In Illinois precipitation and temperature are influencing COVID-19 cases primarily in a dark way and then in a positive way, exactly like in California. This means that depending on the county, local authorities should be aware of the weather forecasts and expect either a same-direction impact on the COVID-19 cases (positive causality) or, alternatively, mixed oscillations (dark causality).

**The case of New York**. According **Fig. 7**, daily COVID cases had been increasing until April, dropping dramatically during May. Setting daily COVID-19 cases, as a result, New York is, as of May 31, under Scenario 1 (“tolerable”). Looking at the frequency distribution of COVID-19 cases in New York reveals a significant concentration between up to 1000 cases per day, a fact which renders the situation very manageable. Being in scenario 1, this suggests, according to **Table I**, a medium impact on physical distancing and low impact on travel restrictions, face masks usage, lockdown, government support, supply chain interruption, reduced consumption and working capital pressure. Of course, all this refers to the situation as of 31 May. However, in order to be able to better plan for the imminent future, the decision maker needs the fourth component of the Risk Quadruplet which is the causal impact on the critical variable of daily COVID-19 cases. In New York precipitation is influencing COVID-19 cases in all possible ways, while temperature is motivating almost exclusively dark causality. This means that depending on the county, local authorities should be mostly alert on the precipitation due to its highly uncertain influence on the daily cases of COVID-19.

1. **Conclusion**

This study develops a novel causal risk-based decision-making methodology to assist policy and decision makers to take the causal structure underlying a decision problem into account and infer the causal consequences of choosing the available options. The proposed methodology requires five main steps to be performed to allow decision and policy makers to answer four key questions with the Risk Quadruplets: (1) *What can go wrong?* (2) *How likely is it?* (3) *What are the consequences?* and (4) *What are the causal relations?*

To elaborate on the direct applicability and usefulness of our methodology, and in order to attempt to demonstrate the role of weather parameters on the formulation of the decision-making process in the US, temperature and precipitation data as representative proxies of the weather conditions were used to apply the proposed methodology and identify the causal interaction with the daily number of COVID-19 cases for the scenarios’ formulation; thereby not considering other factors where data are unavailable. The treatment proposed herein can be easily extended to including other factors once data on them become available.

This approach illustrates enabling the appropriate adjustment of measures to suppress the spread of COVID-19 at the county, state and country levels, and simultaneously identify the actions needed to minimize the socioeconomic damage from the pandemic to several key sectors. Finally, the populous states of New York, California and Illinois have been considered in some detail to formulate the risk quadruplet, and thus, arrange the consequences connected to each scenario proposed. The results of the work are actionable, and provide a basis for further research.

**References**

Aven, T. (2012). Foundational Issues in Risk Assessment and Risk Management. *Risk Analysis*, *32*(10), 1647–1656. https://doi.org/10.1111/j.1539-6924.2012.01798.x

Aven, T., & Zio, E. (2014). Foundational Issues in Risk Assessment and Risk Management. *Risk Analysis*, *34*(7), 1164–1172. https://doi.org/10.1111/risa.12132

Borasio, G. D., Gamondi, C., Obrist, M., Jox, R. (2020). COVID-19: decision making and palliative care. *Swiss Medical Weekly*. https://doi.org/10.4414/smw.2020.20233

Borgonovo, E., Cappelli, V., Maccheroni, F., & Marinacci, M. (2018). Risk analysis and decision theory: A bridge. *European Journal of Operational Research*, *264*(1), 280–293. https://doi.org/10.1016/j.ejor.2017.06.059

Casanova, L. M., Jeon, S., Rutala, W. A., Weber, D. J., & Sobsey, M. D. (2010). Effects of Air Temperature and Relative Humidity on Coronavirus Survival on Surfaces. *Applied and Environmental Microbiology*, *76*(9), 2712–2717. https://doi.org/10.1128/AEM.02291-09

Cohen, S. M. & Reeve, C. D. C. (2020). *Aristotle’s Metaphysics* (Edward N. Zalta (ed.); Winter 202). Metaphysics Research Lab, Stanford University.

Cornblatt, C. (2020). Guiguzi, China’s First Treatise on Rhetoric: A Critical Translation and Commentary. *Technical Communication Quarterly*, *29*(1), 110–113. https://doi.org/10.1080/10572252.2019.1669967

Crawford, G., & Sen, B. (1996). *Derivatives for decision makers : strategic management issues*. John Wiley & Sons.

Gerber, M., Hunger, H., & Pingree, D. (2001). Astral Sciences in Mesopotamia. *Journal of the American Oriental Society*, *121*(2), 317. https://doi.org/10.2307/606596

Gupta, S., Raghuwanshi, G. S., & Chanda, A. (2020). Effect of weather on COVID-19 spread in the US: A prediction model for India in 2020. *Science of The Total Environment*, *728*, 138860. https://doi.org/10.1016/j.scitotenv.2020.138860

Hagmayer, Y., & Fernbach, P. (2017). Causality in Decision-Making. In M. R. Waldmann (Ed.), *Oxford Handbook of Causal Reasoning* (Vol. 1, Issue March 2018, pp. 1–32). Oxford University Press. http://oxfordhandbooks.com/view/10.1093/oxfordhb/9780199399550.001.0001/oxfordhb-9780199399550-e-27

Hanna, T. P., Evans, G. A., and Booth, C. M. (2020). Cancer, COVID-19 and the precautionary principle: prioritizing treatment during a global pandemic. *Nature Reviews Clinical Oncology*, *17*(5), 268–270. https://doi.org/10.1038/s41571-020-0362-6

Howard, R.A. (1988). Decision analysis: practice and promise. Management Science, *34*(6): 679-695. <https://doi.org/10.1287/mnsc.34.6.679>

Kantz, H., & Schreiber, T. (2003). *Nonlinear Time Series Analysis*. Cambridge University Press. https://doi.org/10.1017/CBO9780511755798

Kaplan, S. (1997). The Words of Risk Analysis. *Risk Analysis*, *17*(4), 407–417. https://doi.org/10.1111/j.1539-6924.1997.tb00881.x

Kaplan, S., & Garrick, B. J. (1981). On The Quantitative Definition of Risk. *Risk Analysis*, *1*(1), 11–27. https://doi.org/10.1111/j.1539-6924.1981.tb01350.x

Kontosakos V, Hwang S, Kallinterakis V, Pantelous AA. (2020) Disappointment Aversion and Long-Term Dynamic Asset Allocation. Available at SSRN 3270686, 2020: 1-75. http://dx.doi.org/10.2139/ssrn.3270686

Kumar, A., & Nayar, K. R. (2020). COVID 19 and its mental health consequences. *Journal of Mental Health*, 1–2. https://doi.org/10.1080/09638237.2020.1757052

Lorenz, E. N. (2004). Deterministic Nonperiodic Flow. In *The Theory of Chaotic Attractors* (pp. 25–36). Springer New York. https://doi.org/10.1007/978-0-387-21830-4\_2

Martens, P. & McMichael, A. J.. (2002). *Environmental change, climate and health: issues and research methods* (1st edition). Cambridge University Press.

McGill WL, Ayyub BM, Kaminskiy M. (2007) Risk analysis for critical asset protection. *Risk Analysis*, *27*(5), 1265-1281. https://doi.org/10.1111/j.1539-6924.2007.00955.x

Paté‐Cornell E. (2012) On “Black Swans” and “Perfect Storms”: risk analysis and management when statistics are not enough. *Risk Analysis*, *32*(11): 1823-1833. https://doi.org/10.1111/j.1539-6924.2011.01787.x

Peysakhovich, A., & Karmarkar, U. R. (2016). Asymmetric Effects of Favorable and Unfavorable Information on Decision Making Under Ambiguity. *Management Science*, *62*(8), 2163–2178. https://doi.org/10.1287/mnsc.2015.2233

Stavroglou, S. K., Pantelous, A. A., Stanley, H. E., and Zuev, K. M. (2019). Hidden interactions in financial markets. *Proceedings of the National Academy of Sciences*, *116*(22), 10646–10651. https://doi.org/10.1073/pnas.1819449116

Stavroglou, S. K., Pantelous, A. A., Stanley, H. E., and Zuev, K. M. (2020). Unveiling causal interactions in complex systems. *Proceedings of the National Academy of Sciences*, *117*(14), 7599–7605. https://doi.org/10.1073/pnas.1918269117

van Doremalen, N., Bushmaker, T., Morris, D. H., Holbrook, M. G., Gamble, A., Williamson, B. N., Tamin, A., Harcourt, J. L., Thornburg, N. J., Gerber, S. I., Lloyd-Smith, J. O., de Wit, E., and Munster, V. J. (2020). Aerosol and Surface Stability of SARS-CoV-2 as Compared with SARS-CoV-1. *New England Journal of Medicine*, *382*(16), 1564–1567. https://doi.org/10.1056/NEJMc2004973

1. We specifically focus on a new measure of causality instead of the standard correlation measure as the causal relationships hold explanatory power (Stavroglou et al., 2019) and (Stavroglou et al., 2020) .On the one hand, correlation looks for statistical linear coupling between two variables, and . Thus, correlation ignores whether X drives Y or the other way around. On the other hand, causality primes in the identification of whether X drives Y or Y drives X. This is very crucial for understanding the system’s dynamics, for better formulating prediction tasks, and for better hedging against risk. The interested reader may consider (Stavroglou et al., 2019) and (Stavroglou et al., 2020). [↑](#footnote-ref-1)
2. The *time horizon* is finite and depends on the length of the available time series data. In addition, it can be measured in lower (e.g., daily, monthly, quarterly or annually) or higher (e.g., 1-minute, 5-minute or 1-hour) frequencies. The frequency preferred should be consistent for all the time series variables. [↑](#footnote-ref-2)
3. Among all the variable, in **Fig. 1b**, we illustrate only . [↑](#footnote-ref-4)
4. For instance, the *black* links in **Fig. 1a** are used only symbolically here to suggest the yet unknown causal relations between the nodes. [↑](#footnote-ref-5)
5. is the embedding dimension of the attractor, . [↑](#footnote-ref-6)
6. Limited for plotting to up to three dimensions. [↑](#footnote-ref-7)
7. Note that the shadow attractor reconstructed using time lags of and is denoted by , and , accordingly, in order to simplify the notation of the shadow attractor (time is eliminated), except where indicated. [↑](#footnote-ref-8)
8. The higher and the longer time series are needed. [↑](#footnote-ref-9)
9. Something particularly useful when the results need to be communicated to a broader audience, non-expert stakeholders, and policy makers. [↑](#footnote-ref-10)
10. when looking for contemporaneous causality (i.e. synchronous) or when looking for causality which is expressed with a time delay (i.e. the causal variable’s impact on the affected variable is manifested at some point in the future). [↑](#footnote-ref-11)
11. Different types of distances can also be considered and implemented. The proposed methodology is versatile by design, and it can be adjusted accordingly. However, further elaboration on this issue is beyond the scope of the paper. [↑](#footnote-ref-12)
12. <https://www.who.int/globalchange/summary/en/index5.html> (Retrieved at 1st May 2020) [↑](#footnote-ref-13)
13. We are not able to consider other important factors as the required data are not available (i.e., deficient or missing data) to us. We should reiterate here that the proposed causal decision-making methodology with real time-series variables is robust enough to produce results in case of missing values in the data; see Section 3.1 for details. [↑](#footnote-ref-14)
14. Data was obtained from the New York Times Github repository: <https://github.com/nytimes/covid-19-data> [↑](#footnote-ref-15)
15. FIPS: Federal Information Processing Standards [↑](#footnote-ref-16)
16. Weather data (daily precipitation and temperature per U.S. county) were retrieved from NOAA’s National Climatic Data Center: <http://www.ncdc.noaa.gov>. In order to conduct our analysis, we had to match the daily COVID-19 cases to the daily temperature and precipitation for each county separately. The common reference point was the FIPS code which was readily available for the COVID-19 data. [↑](#footnote-ref-17)