The Cognitive and Mathematical Foundations of Analytic Epidemiology

Yingxu Wang[1], Kostas N. Plataniotis[2], Jane Z. Wang[3], Ming Hou[4], Menchu Zhou[5], Newton Howard[6], Jun Peng[7], Runhe Huang[8], Shushma Patel[9], and Du Zhang[10]

[1] President, Int'l Inst. of Cognitive Informatics and Cognitive Computing
Schulich School of Engineering and Hotchkiss Brain Institute
University of Calgary, Canada
Email: yingxu@ucalgary.ca

[2] Dept. of Electrical & Computer Engineering
University of Toronto, Canada
Email: kostas@ece.utoronto.ca

[3] Dept. Electrical and Computer Engineering
Biomedical Engineering School
University of British Columbia
Email: zjanew@ece.ubc.ca

[4] Toronto Research Centre
DRDC, Canada
Email: ming.hou@drdc-rddc.gc.ca

[5] Dept. of Electrical & Computer Engineering
New Jersey Institute of Technology
Newark, NJ 07102 USA
Email: zhou@njit.edu

[6] Computational Neurosciences Lab
University of Oxford, UK
Email: newton.howard@nds.ox.ac

[7] School of Intelligent Tech. & Engineering
Chongqing University of Sci. & Tech.
Chongqing 401331, China
Email: jpeng@cqust.edu.cn

[8] Hosei University, Tokyo, Japan
Email: rhuang@hosei.ac.jp

[9] School of Engineering
London South Bank University, UK
Email: shushma@lsbu.ac.uk

[10] Faculty of Information Technology
Univ. of Science & Technology, Macau, China
Email: duzhang@must.edu.mo

Abstract – Analytic epidemiology is a transdisciplinary study on the cognitive, theoretical, and mathematical models of COVID-19 and other contagious diseases. It is recognized that analytic epidemiology may be better studied by big data explorations at the macro level rather than merely biological analyses at the micro level in order to not loss the forest for the trees. This paper presents a basic research on analytic epidemiology underpinned by sciences of cognition, computer, big data, information, AI, mathematics, epidemiology, and systems. It introduces a novel Causal Probability Theory (CPT) for explaining the Dynamic Pandemic Transmission Model (DPTM) of analytic epidemiology. It reveals how the fundamental reproductive rate (R_0) may be rigorously calibrated based on big data of COVID-19. A theoretical framework of analytic epidemiology is developed to elaborating the insights of pandemic mechanisms in general and COVID-19 in particular. Robust and accurate predictions on key attributes of COVID-19, including $R_0(t)$, forecasted infectives/resources, and the expected date of pandemic termination, are derived via rigorous experiments on worldwide big data of epidemiology.

Keywords – *Analytic epidemiology, COVID-19, cognitive pandemic models, R_0, infectious transmission models, cognitive informatics, cognitive algorithms, big data experiments*

I. INTRODUCTION

The worldwide outbreaks of COVID-19 [22] and other contemporary contagious diseases [12, 14] have triggered a wide scope of transdisciplinary studies on epidemiology towards their systematical treatments, control, prediction, prevention, management, and decision optimization [5, 8, 22, 24]. The multidisciplinary investigations into the COVID-19 pandemic have led to the emergence of analytic epidemiology underpinned not only by epidemiology, biology, and medical sciences, but also by computer, big data, information, AI, system sciences as well as mathematics, sociology, and economics.

A fundamental challenge to analytic epistemology in general and COVID-19 in particular is the lack of cognitive informatics and mathematical models for pandemic monitoring and prediction in order to support rational and optimal decision making at different levels of communities, nations, and the world. The traditional mathematical models of epidemiology have been mainly based on probability theory, statistics, Bayesian networks, and differential equations [3, 4, 7, 13, 14, 20]. There are three classical models known as the Susceptible-Infective-Susceptible (SIS) model, the Susceptible-Infective-Recovery (SIR) model, and the Susceptible-Infective-Recovery-Susceptible (SIRS) model [3, 9]. In which the populations in different epidemiological categories as variables over time are identified in the classes of *susceptible* (*S*, not yet infected), *infective* (*I*, infected and transmissive), and *recovered* (*R*, removed from both classes of *S* and *I* with immunity), respectively. However, big data of COVID-19 collected worldwide [22] do not fit the classic pandemic models very well.

The second challenge to analytic epistemology is that current pandemic models for contagious disease predication and estimation were based on classic probability theories [6, 11, 18, 19]. Hence, perceptions on the transmission mechanisms of epidemiology have been based on two biased assumption that the prior probabilities of contagious infections and transmissions are known and invariant [6, 22]. However, it is observed recently that, in general, the sample space of pandemic probability is not invariant as conventionally perceived [19, 20]. Therefore, both preconditional assumptions were untrue because none of them may be fulfilled due to the exponential

growth of the sample spaces of the affected population in COVID-19.

In epidemiology, the agents transmitting infectious diseases to the hosts (human and/or animals) may be categorized into four categories including virus, bacteria, protozoa, and helminths [14, 22]. The COVOD-19 agent is recognized as a kind of new corona virus that is highly infectious with a potentially high mortality rate among the infected hosts [22]. There is a lack of practically available and dedicated pandemic decision-making system. The popular COVID-19 Dashboard at Johns Hopkins University has no function for autonomous decision making, rapid prediction, and early alarms [10]. Further, this type of online systems is not a real-time system and therefore do not support rapid decision making. The third constraint is that the exiting pandemic information systems in other countries cannot be directly migrated to Canada because both policies and data collection formats are different.

This paper presents a basic research on the cognitive and mathematical foundations of analytic epidemiology for explaining the insights of epidemiology and COVID-19 underpinned by the causal probability theory, big data algebra [20], and causal inference algebra [17]. Section II explores the domain of analytic epidemiology and its cognitive models. Section III creates a set of mathematical models for enabling rigorous pandemic analyses and forecasts. A set of experiments on epidemiological predictions is demonstrated in Section IV based on the analytic epidemiology theory and cognitive algorithms for causal probability elicitation from worldwide pandemic big data.

II. THE COGNITIVE FOUNDATIONS OF EPIDEMIOLOGY

This section explores the domain of analytic epidemiology in order to understand its universe of discourse and essential control attributes. It leads to the cognitive models of COVID-19 and the calibration of fundamental attributes of epidemiology via big data analytics.

2.1 The Domain of Analytic Epidemiology and Control Attributes

Definition 1. *Analytic epidemiology* is a transdisciplinary field for contagious diseases and outbreaks detection, treatment, prediction, and optimal decision making underpinned by sciences of epidemiology, computer, big data, information, cognition, AI, mathematics, sociology, and systems.

The domain of analytic epidemiology encompasses a comprehensive set of pandemic attributes and variables, particularly those of COVID-19 epidemiology [1, 2, 9, 13, 22], which may be formally described as follows.

Definition 2. The *universe of discourse U* of analytic epistemology in a size *N* population is a relatively conservative (constrained) system encompassing five disjoint sets of susceptibles N_S, infectives N_I, immunized N_M, recovered N_R, dead N_D, and hospitalized N_H classes, as well as the numbers of normal birth N_B and death $N_{D'}$:

$$U \triangleq N_S(t) + N_I(t) + N_M(t) = N,\ N_B(t) \approx N_{D'}(t) \quad (1)$$

constrained by the following relations:

$$\begin{cases} N_S(t) = N - N_I(t) - N_M(t) = \sigma(N - N_M(t)) \approx \sigma N \\ N_I(t) = N_H(t) + N_R(t) + N_D(t) = \tau N_S(t) \\ N_M(t) = N_R(t) + N_0(t) = \mu N_I(t) + N_0(t) \\ N_D(t) = N_I(t) - N_R(t) - N_H(t) = \delta N_I(t) \end{cases}$$

where $N_B(t) \approx N_{D'}(t)$, in a relatively short period. Four statistical attributes are adopted as: a) λ the *average daily contact rate*; b) τ the *average daily recovery (removal) rate*; c) μ the *average death rate*; and d) σ the *average number of adequate contacts* by an infective per day.

The attributes of *U* in three sample countries, i.e., Canada, USA, and China, are listed in Table 1 with data collected from WHO [22, 24] up to July 11, 2020, which provide an overview of basic COVID-19 attributes. Data for other countries and regions may be found from the same source.

2.2 The Formal Diagnosis Model of COVID-19

On the basis of real-world COVID-19 big data as presented in Table 1, a set of statistical results is derived for the three sample countries as shown in Table 2. In Table 2, $\gamma_s(t)$ and $\gamma_S(t)$ represent the average infective rate among the susceptible class or the whole population, while $\gamma_d(t)$ and $\gamma_D(t)$ denote the average mortality rate among the infective class or the whole population, respectively.

The decision model of COVID-19 diagnoses may be formally described by a Cartesian product of the sets of symptoms [23] and test results according to Definition 3.

Table 1. Statistical Big Data of COVID-19 (partial) [23]

Country	#Infectives (N_{inf})	#Recovered (N_r)	#Deaths (N_d)	#Hospitalized (N_h)	#Tested (N_{te}= N_t)	Population (N)
Canada	107,347	71,266	8,773	27,308	3,183,516	37,751,539
USA	3,355,646	1,490,446	137,403	1,727,797	41,770,226	331,060,504
China	83,594	78, 634	4,634	326	90,410,000	1,439,323,776
World	12,848,040	7,483,451	567,760	4,796,829	-	-

Table 2. Sample Statistical Parameters of COVID-19 Pandemic

Country	Reproductive rate ($R_0(t)$: mean | max | approx.)	Average infective rate ($\gamma_S(t)$, $\gamma_S(t)$)	Test rate (γ_e)	Mortality rate ($\gamma_0(t)$, $\gamma_D(t)$)	Population (N)
Canada	[1.0921, 2.0000, 2.3513]	3.3719%, 0.2844%	8.4328%	8.1726%, 0.0232%	37,751,539
USA	[1.1043, 1.8000, 3.8930]	8.0336%, 1.0136%	12.6171%	4.0947%, 0.0415%	331,060,504
China	[1.0582, 2.9351, 7.0940]	0.0925%, 0.0006%	6.2800%	5.5435%, 0.0003%	1,439,323,776

Definition 3. Let the set of *symptoms* of COVID-19 be S = {S_1(*Fever*), S_2(*Cough*), S_3(*BreathDifficulty*), S_4(*Chills*), S_5(*ChillShaking*), S_6(*MusclePain*), S_7(*HeadAche*), S_8(*SoreThroat*), S_9(*LossOfTaste/Smell*)}, and the set of *lab tests* be L = {L_1(*NucleicAcid*), L_2(*SoreSample*), L_3(*LungImage*)}. The *diagnosis* E of COVID-19 infectives is detected by the Cartesian product between the sets of *detection symptoms* E_S and *lab confirmations* E_L as follows:

$$\begin{aligned} E \triangleq E_S \times E_L &= \mathop{R}_{i=1}^{9} S_i \times \mathop{R}_{j=1}^{3} L_j \\ &= \{S_1, S_2, S_3, S_4, S_5, S_6, S_7, S_8, S_9\} \times \{L_1, L_2, L_3\} \\ &= \begin{cases} E_S = T|\mathrm{L} \wedge E_L = T|\mathrm{L} & \text{// Symptomatic positive} \\ E_S = F|\mathrm{L} \wedge E_L = T|\mathrm{L} & \text{// Nonsymptomatic positive} \\ E_S = F|\mathrm{L} \wedge E_L = F|\mathrm{L} & \text{// Negative} \\ E_S = T|\mathrm{L} \wedge E_L = F|\mathrm{L} & \text{// Suscetibly negative} \end{cases} \end{aligned} \quad (2)$$

where T|L and F|L denotes a Boolean logical variable for True or False, respectively. The diagnosing results are classified in the categories of *symptomatic positive*, *none-symptomatic positive*, *negative*, and *susceptibly negative*; and the *big-R* notation [15] represents an iterative series of recurrent structures or embedded functions.

2.2 Fundamental Attributes of COVID-19 Pandemic and their Calibrations by Big Data Analytics

In epidemiology, the reproductive ratio R_0 of a contagious disease is adopted to estimate how fast the disease spreads in a community. The role of R_0 is formally introduced in a simplified model as a constant $\bar{R}_0$ in the following for approximate estimation. However, more rigorous analysis of $R_0(t)$ as a dynamic series will be developed in Sections III and IV.

Definition 4. The *exponential series* $N_{inf}(t)$ *of epidemical transmission* on the $t_0 + k$th day is estimated by a product of initial infectives $N_{inf}(t_o)$ and the average reproductive rate $\bar{R}_0$ raised to the kth power:

$$N_{inf}(t_0 + k) \triangleq \bar{R}_0^{\,k} N_{inf}(t_0),\ \bar{R}_0 > 1.0, k \geq 0, N_{inf}(t_0) \neq 0 \quad (3)$$

Theorem 1. The *estimated average reproductive rate* $\bar{R}_0$ of a pandemic transmission is the kth root of the average ratio between the number of infectives $N_{inf}(t_0+k)$ cumulatively infected at $t_0 + k$ by each initial infective $N_{inf}(t_0)$:

$$\bar{R}_0 \triangleq \sqrt[k]{\frac{N_{inf}(t_0 + k)}{N_{inf}(t_0)}},\ k \geq 0, N_{inf}(t_0) \neq 0 \quad (4)$$

Proof. According to Definition 4, Theorem 1 may be directly proven. ■

It is noteworthy that $\bar{R}_0 = 1.0$ when $k = 0$.

Corollary 1. The *average reproductive rate* $\bar{R}_0$ is an indicator θ for the congruous severity classified in two categories by the threshold $\bar{R}_0 = 1.0$:

$$\theta = \begin{cases} congruous, & \bar{R}_0 > 1.0 \\ incongruous, & 1.0 \geq \bar{R}_0 \geq 0 \end{cases} \quad (5)$$

The average value of $\bar{R}_0$ in COVID-19 has been estimated in a considerably inconsistent range according to different pandemic patterns and datasets in a certain period of the pandemic. For instance, WHO has empirically estimated $\bar{R}_0$ of COVID-19 in the range of 2.24 to 4.00 recently, while its preliminary estimation was from 1.40 to 2.50 on January 23, 2020 [23]. It will be explained in Section III why the WHO empirical estimations on $\bar{R}_0$ were considerably higher than those obtained in rigorous analyses with real-world big data in different periods of the COVID-19 lifecycle.

Investigating into the nature of pandemic dynamics for rigorously predict the pandemic trends, we find that in order to model more general and complex pandemic dynamics, the reproductive rate must be treated as a series of variables $R_0(t)$ over time. A formal analysis of this fundamental phenomenon of epidemiology will be elaborated in Sections III and IV based on the causal probability theory and big data analytics. It will describe how $R_0(t)$ is rigorously determined as a series of dynamic variables in epidemiology.

III. MATHEMATICAL MODELS OF ANALYTIC EPIDEMIOLOGY

The preceding section has indicated that the challenging problems in analytic epidemiology demand novel mathematical means and models. It is observed that the sample space of general probability is dynamically varying rather than static as traditionally perceived [20, 22]. This section analyzes the constraints of traditional approach to epidemiological dynamics modeling by classical probability theory. Then, a novel theory on causal probability is introduced towards rigorous epidemiological analytics.

3.1 The Causal Probability Theory (CPT) for Modeling the Dynamics of Epidemiological Processes

In order to address the instability, sensitivity, and interlocked (dependent) solutions in the SIR model and classic exponential growing sample spaces of pandemic probability, we introduce the *Causal Probability Theory* (CPT) for analytic epidemiology. CPT models the dynamics of pandemic transmissions as a causal series. Each step in the causal probability series is determined by CPT where traditional probability is a special case of it when the sample space is assumed to be invariant.

Definition 5. The *axiom of a series of causal probabilities* in CPT is based on the fundamental concepts: a) The *causal influential factor* γ_t determined by the difference between the sizes of events $\bar{e}(t)$ and $\bar{e}(t-1)$ over the current sample space $\tilde{S}(t)$, where $\bar{e}(t\text{-}1)$ is called the *cause* and $\bar{e}(t)$ the *effect*; and b) The *elemental causal probability* ρ_t of an event $\bar{e}(t)$ on a variable sample space $\tilde{S}(t-1)$ in the recursive series of a pandemic, i.e.:

$$\begin{cases} \gamma_t = \dfrac{|\bar{e}(t)| - |\bar{e}(t-1)|}{\tilde{S}(t)}, \ |\gamma_t| \geq 0, \ \tilde{S}(t) \geq 0 \\ \rho_t = \dfrac{|\bar{e}(t)|}{\tilde{S}(t-1)}, \ 0 \leq \rho_t \leq 1, \ \tilde{S}(t-1) \geq 0 \end{cases} \tag{6}$$

On the basis of Definition 5, the causal probability of a series of causes and effects in CPT may be rigorously derived as follows.

Definition 6. The *causal probability* $\vec{\rho}(t)$ of a series of *n* consecutively pairs of causal probabilities $\bar{e}(t-1) \xrightarrow{\vec{\rho}(t)} \bar{e}(t)$ in a dynamic sample space $\tilde{S}(t)$, is:

$$\begin{aligned} \mathop{\vec{R}}_{t=1}^{n} \bar{p}_t(t \mid \tilde{S}(t)) &\triangleq \mathop{\vec{R}}_{t=1}^{n} \rho_t(1+\gamma_t), \ d\tilde{S}(t)/dt \neq 0 \\ &= \mathop{\vec{R}}_{t=1}^{n} \frac{|\bar{e}(t)|}{\tilde{S}(t-1)}\left(1+\frac{|\bar{e}(t)-\bar{e}(t-1)|}{\tilde{S}(t)}\right), \ \bar{e}_0(0)=0, \tilde{S}(0) \neq 0 \end{aligned} \tag{7}$$

The causal probability model for analytic epidemiology reveals a special series of causal influences between each pair of the previous and the current events on a varying sample space $d\tilde{S}(t)/dt$. The traditional probability theory considers only a pair of conditional influence on a static sample space, which is a special case of the causal probability theory.

Theorem 2. The *sample space* $\tilde{S}(t)$ *of causal probability* in a series is not a constant due to the causal influences in the recursive series:

$$\frac{d\tilde{S}(t)}{dt} \neq 0 \tag{8}$$

Proof. According to Definition 6, Theorem 2 is proved as follows:

$$\forall \mathop{\vec{R}}_{t=1}^{n} \tilde{S}(t) = \mathop{\vec{R}}_{t=1}^{n} \bar{e}(t-1), \ \bar{e}(0)=0, \tilde{S}(0) \neq 0$$

$$\exists \bar{e}(t-1) \neq \bar{e}(t) \Rightarrow \tilde{S}(t) \neq \tilde{S}(t+1)$$

$$\text{thus } \frac{d\tilde{S}(t)}{dt} \neq 0$$

Theorem 2 indicates that the causal probability is a general probability theory that extend classic conditional probability and the Bayesian law [4, 22] to a general setting where both the sample space and events are varying influenced by the past series as that in epidemical transmissions.

Based on Definition 6 and Theorem 2, the fundamental model for explaining the dynamic behaviors of epidemical transmission may be formally perceived as follows.

Definition 7. The *number of infectives* of a pandemic on day *t* may be rigorously predicated based the causal probability $\vec{\rho}(t)$ where its prior as the cause is the cumulative number of infectives $\vec{N}_{inf}(t\text{-}1)$:

$$\begin{aligned} \mathop{\vec{R}}_{t=1}^{n} \vec{N}_{inf}(t) &\triangleq \mathop{\vec{R}}_{t=1}^{n} \vec{\rho}(t) \vec{N}_{inf}(t\text{-}1) \\ &= \mathop{\vec{R}}_{t=1}^{n} \rho_t(1+\gamma_t) \vec{N}_{inf}(t\text{-}1) \end{aligned} \tag{9}$$

The physical meaning of $\vec{N}_{inf}(t\text{-}1)$ in Definition 7, is embodied by the cumulated historical priors as the cause, and $\vec{N}_{inf}(t)$ the current event. The big-R calculus denotes the dynamics mechanisms of system updating in order to interchange roles of the effect and the cause in the recursive series of causal probability inferences.

3.2 Fundamental Theories for CODIV-19 Forecast and Control

On the basis of CPT, a dynamic transmission model of analytic epidemiology may be rigorously derived for COVID-19 prediction based on both the dynamic transmissive rates and the varying sample spaces over time.

Lemma 1. The series of dynamic *reproductive rates* $\mathop{\vec{R}}_{t=1}^{n} R_0(t-1)$ of COVID-19 is recursively determined by its causal probabilities $\rho_t = \rho_{t-1}(1+\gamma_{t-1})$ in each step of the iteration:

$$\mathop{\vec{R}}_{t=2}^{n} R_0(t-1) \triangleq \mathop{\vec{R}}_{t=2}^{n} \frac{N_{inf}(t-1)}{N_{inf}(t-2)}, \ R_0(0)=1, N_{inf}(t) \neq 0 \tag{10}$$

Proof. Let $\vec{\rho}_t = R_0(t)$ be the dynamic causal probability of a pandemic series. According to Definition 7, Lemma 1 is proved by the recursive series when of the previous values $N_{inf}(t\text{-}1)$ and $N_{inf}(t\text{-}2)$ are known in the causal series. ■

It is noteworthy that, although the average value of $R_0(t)$, $\bar{R}_0$ (Definition 4), in empirical studies on COVID-19 is assumed as a constant, it is naturally a λ-shape series due to the cumulative infective dynamics as shown in Figure 3. $R_0(t)$ may be rigorously calibrated for each step of the transmission series $\mathop{\vec{R}}_{t=1}^{n} R_0(t)$ based on Lemma 1 as follows.

Theorem 3. The *Dynamic Pandemic Transmission Model* (*DPTM*) of analytic epidemiology is a recursive series of causal probabilities driven by the reproductive rate $R_0(t)$ to determine the effect of *future number of infectives* $N_{inf}(t)$ based on prior causes $N_{inf}(t-1)$:

$$\overset{n}{\underset{t=1}{R}} N_{inf}(t) \triangleq \overset{n}{\underset{t=1}{R}} R_0(t) N_{inf}(t-1),\ R_0(0)=1, N_{inf}(0) \neq 0 \quad (11)$$

Proof. According to CPT and Lemma 1, a series of causal probabilities between adjacent events in a variant sample space $\tilde{S}(t)$ of COVIC-19 are:

$$\forall d\tilde{S}(t)/dt \neq 0,$$

$$R_0(t) = \vec{\rho}_t = \rho_t(1+\gamma_t) = \frac{\Delta N_{inf}(t-1)}{N_{inf}(t-1)}(1+\frac{\Delta N_{inf}(t)}{N_{inf}(t)}),$$

Any arbitrary signle causal influence between an adjacent pair of causal events $N_{\inf}(t-1) \xrightarrow{R_0(t)} N_{\inf}(t)$ is:

$$N_{inf}(t) = \vec{\rho}_t N_{inf}(t-1) = R_0(t) N_{inf}(t-1)$$

Thus, the causal series of pademic transmissions $\overset{n}{\underset{t=1}{R}}\{N_{\inf}(t-1) \xrightarrow{R_0(t)} N_{\inf}(t)\}$ becomes:

$$\overset{n}{\underset{t=1}{R}} N_{inf}(t) = \overset{n}{\underset{t=1}{R}} R_0(t) N_{inf}(t-1)$$

■

According to Theorem 3, the infectives in a pandemic series may be rigorously predicated at any given time as follows.

Corollary 2. The *forecasted number of infectives* $\vec{N}_{inf}(t+k)$ of a period on days $t + k$ is determined by the following sum of products of $N_{inf}(k)$ and $R_0(k)^{k-t}$:

$$N_{inf}(t+k) \triangleq \sum_{\tau=t}^{t+k} R_0(\tau) N_{inf}(\tau-1) \quad (12)$$

where if $k = 0$, Eq. 12 reduces to the simplest form $N_{inf}(t) \triangleq (R_{0-mean})^t N_{inf}(t-1)$.

Proof. According to DPTM (Theorem 3), the cumulative infectives of a subseries in $[t, t+k]$ is proven as follows:

$$\forall N_{inf}(\tau) = \overset{t+k}{\underset{\tau=t}{R}} N_{inf}(\tau),$$

$$\sum_{\tau=t}^{t+k} N_{inf}(\tau) = \sum_{\tau=t}^{t+k} R_0(\tau) N_{inf}(\tau-1) = N_{inf}(t+k)$$

■

As indicated by Corollary 2, there are two criteria to forecast the termination of a pandemic based on if the trend of $R_0(t)$ is approaching to 1.0 or if the limitation of the infective rate $dN_{inf}(t)/dt$ is approaching to 0.

Definition 8. The *forecasted endpoint* T_{max} of an epidemical lifecycle is determined at the point of t_e according to DPTM while the following conditions continuously meet for a period:

$$T_{\max}(R_0(t), \frac{dN_{\inf}(t)}{dt}) \triangleq t_e \begin{cases} \lim\limits_{t \to t_e} R_0(t) = 1.0 \\ \lim\limits_{t \to t_e} \dfrac{dN_{\inf}(t)}{dt} \leq 0 \end{cases} \quad (13)$$

where the stabilization period may be set as a week or so until either or both conditions are continuously met.

As a result, the expected maximum infectives $N_{inf\text{-}\max}(t)$ of a pandemic may be obtained according to Definition 8 given T_{max} as follows

Definition 9. The *maximum infectives* $N_{inf\text{-}max}$ towards the termination point T_{max} of a pandemic is determined by an integration or approximately a weighted sum of the incremental infectives:

$$\begin{aligned} N_{inf\text{-}max}(T_{\max}) &\triangleq \int_0^{T_{\max}} \Delta N_{inf}(t)dt \\ &= \sum_{t=0}^{T_{\max}} \Delta N_{inf}(t)h = N_{inf}(T_{\max}),\ h=1 \end{aligned} \quad (14)$$

Towards the termination point T_{max}, all other attributes of a pandemic may also be rigorously predicated.

Definition 10. The *maximum mortality* $N_{d\text{-}max}$ towards the termination point T_{max} of a pandemic is determined by the product of the total infectives and the average death rate:

$$N_{d\text{-}\max}(T_{\max}) \triangleq \gamma_d(T_{\max}) N_{inf}(T_{\max}) \quad (15)$$

Applications and big-data-based experiments of CPT and DPTM will be presented in the following section.

VI. APPLICATIONS OF CPT AND DPTM IN ANALYTIC EPIDEMIOLOGY

The theories of analytic epistemology and the Dynamic Pandemic Transmission Model (DPTM) as developed in the proceeding section have provided a rigorous foundation to reveal the insights of pandemic mechanisms. This section describes the design of a decision-making tool for analytic epidemiology. Then, a set of experimental results will be obtained to demonstrate the predictive power of the DPTM theory and the approach to rigorous forecasts of COVID-19 trends based on real-world big data.

4.1 The Architecture of the Analytic Epidemiology System

The Analytic Epidemiology System (AES) encompasses six subsystems of solutions with 15 categories of 60+ functions. AES is shown in Figure 1 for elaborating an entire picture of analytic epidemiology. Within each of the subsystems, a number of COVID-19 analytic functions and algorithms are embodied to implement the system. Details of key functions and algorithms as well as their mathematical models are based on the analytic CPT and DPTM as developed in Sections II and III.

AES provides a wide spectrum of rapid decision supports for multimodal big data gathering, analysis and visualization, event detection and alarms, situation awareness, predications, taskforce deployment, resources allocation, and machine-learning-based causal inference algorithms. As shown in Figure 1, the real-time analytic system of epidemiology encompasses the subsystems of: a) The real-time operating systems (RTOS) platform; b) The graphical user interface (GUI) of AES; c) The big data analytic engine (BDAE); d) The AI decision engine (AIDE); e) The pandemic decision making (DM) database (PDB); and f) The pandemic DM knowledge base (PKB).

The AES tool powered by the CPT and DPTM algorithms as developed in Section III is designed to address a set of key challenges to epidemiological analyses and predictions. This subsection illustrates the key functions and algorithms of the system including predictions for dynamic infectives and incrementals, maximum infectives and average reproductive rates, maximum death/average death rate, and epidemic life cycle (end time), as well as their rigorous estimations beyond those of traditional statistical methodologies.

A set of numerical and machine learning algorithms is designed for rigorously determining the important attributes of COVID-19 pandemics and their forecasts based on the mathematical models of analytic epidemiology supported by the AES tool. Rigorous requirement predications are enabled for expected infectives $N_{inf}(t)$, recovered $N_r(t)$, deaths $N_d(t)$, hospital wards allocation $N_h(t)$, and the expected termination day $N_e(t)$ of the pandemic in any region. All the key attributes of pandemic are quantitatively analyzed and reported by the AES tool for the expected values, maximums, potentials, trends, ratios, and early alarms.

4.2 Experiments on COVID-19 Trends Predication by AES

On the basis of DPTM supported by the AES tool, key attributes of analytic epidemiology, including $R_0(t)$, $N_{inf}(t)$, $T_{max,}$ and $N_{inf\text{-max}}(t)$, may be rigorously determined for COVID-19 prediction and decision-making with high accuracy and confidence.

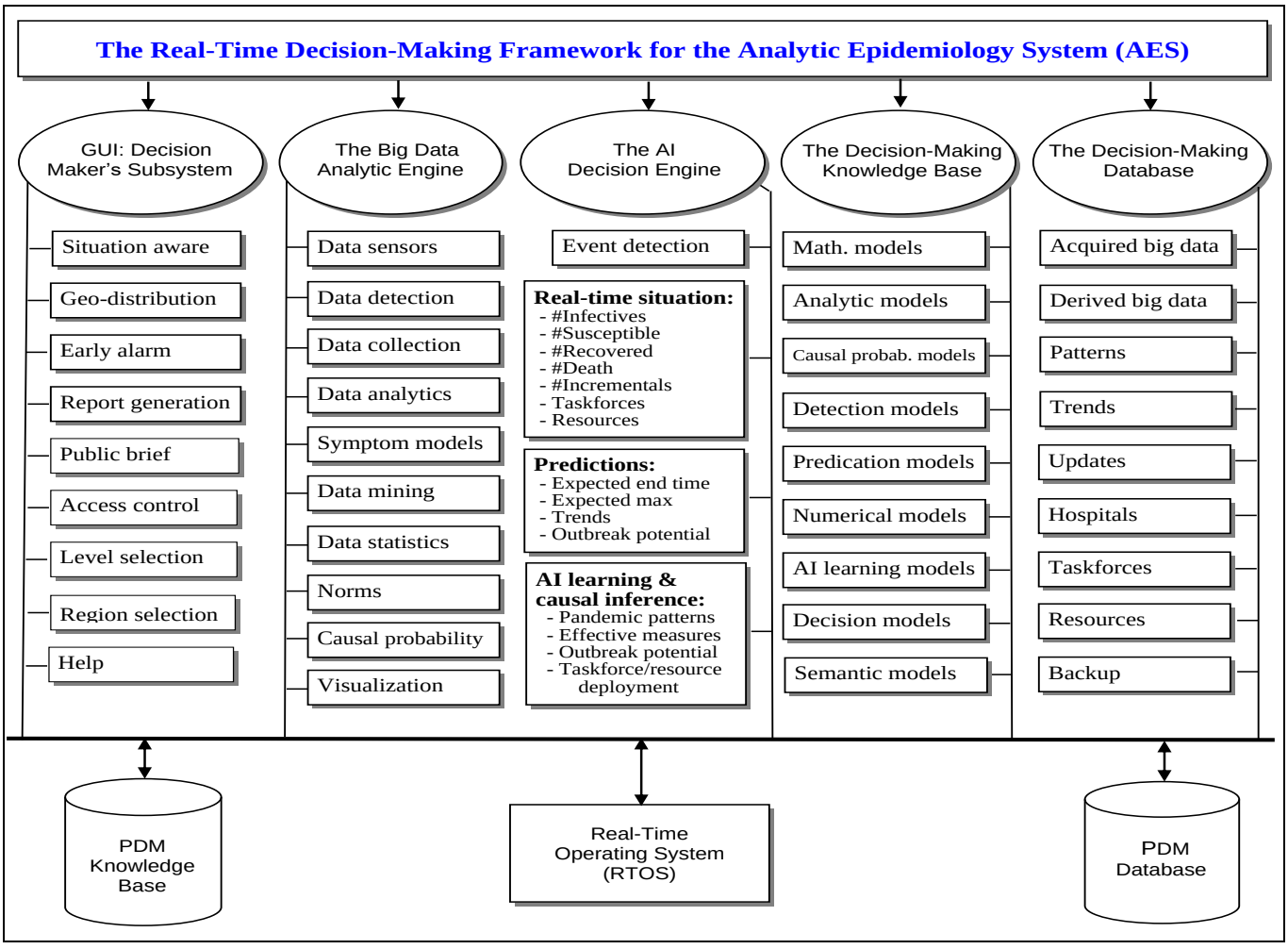
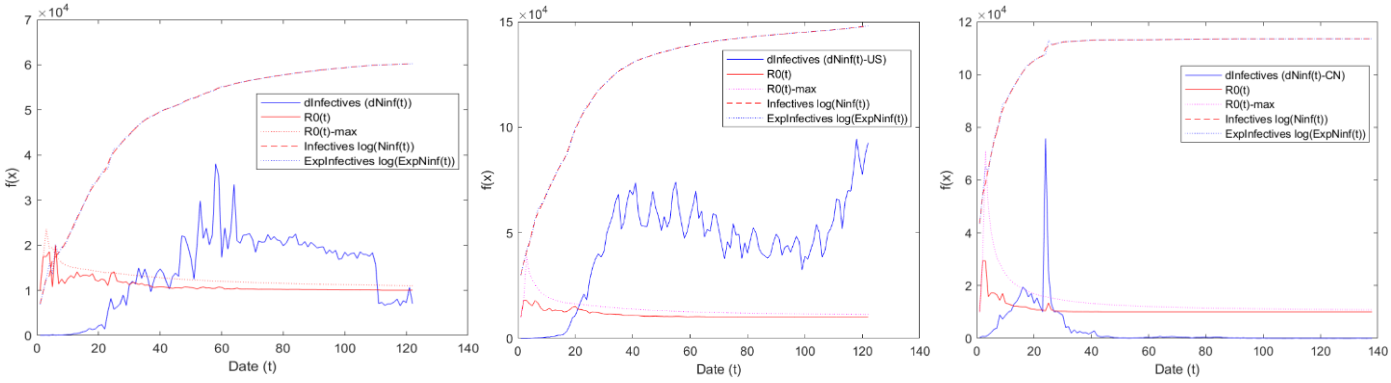


Fig. 1. The framework of pandemic decision-making (PDM) for the analytic epidemiology system (AES)



a) Canada b) USA c) China

Fig. 2 Predictive analyses of COVID-19 trends

The analysis and forecast functions of the AES tool are based on a set of raw big data adopted from WHO databases [23] during January 20 to June 30, 2020 as shown in Table 3. Key pandemic attributes are rigorously derived and/or predicated by the tool based on the DPTM algorithms, particularly CPT-driven machine learning algorithms for calibrating the dynamic trends of $R_0(t)$ and forecasting the infectives through the lifecycle of the COVID-19 pandemic in the sample countries or anywhere else.

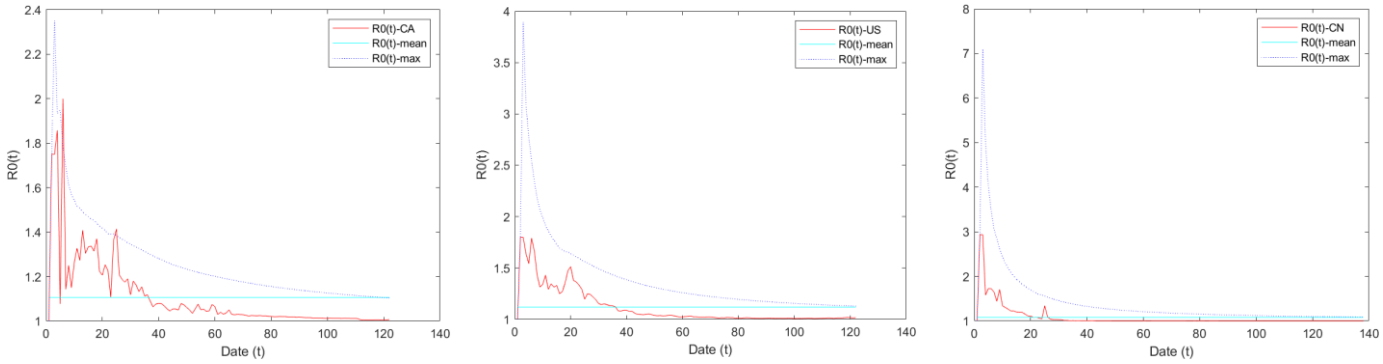
Table 3. Sample Big Data of Incremental Infectives (ΔInf(t)) of COVID-19 [WHO, 2020]

Data set	Country	Sample Duration	Days
1	Canada	March 1 to June 30, 2020	122
ΔInf_1	[4 3 6 1 14 4 8 6 12 19 21 40 42 60 81 101 156 129 146 216 242 141 527 820 582 634 702 895 665 1128 1199 1140 1497 1265 1475 1258 970 1230 1393 1475 1383 1190 1033 1300 1383 1309 2207 2187 1976 1715 1248 2027 2982 2150 2377 2007 2242 3804 3553 1732 2384 1913 2344 3346 2101 2045 2225 2229 2263 2133 1844 2096 2065 2072 2237 2122 2192 2118 2097 1989 2053 2185 2170 2250 2012 2009 2082 1962 2065 2007 1941 1959 1963 1809 1895 1827 1888 1685 1706 1845 1767 1815 1828 1708 1854 1797 1795 1756 1834 1582 706 735 667 702 701 741 805 659 770 729 1065 705 …]		
2	USA	March 1 to June 30, 2020	122
ΔInf_2	[20 16 23 32 72 108 116 122 174 294 286 438 534 739 744 1005 1674 3028 4838 5411 7156 8930 10469 9071 13863 16772 18800 20034 19249 20631 24987 27089 29019 32433 34110 25860 27667 32252 35143 33997 36886 29562 26716 26534 26485 30159 34814 30943 28773 25478 28830 26334 28158 34827 37049 31204 27315 23734 25330 28815 31462 34940 28358 30134 24059 25834 25334 29527 29204 25820 22122 18841 22872 21362 27200 24496 24639 18907 22664 20177 23321 26174 24848 21275 20055 19545 18998 19814 22748 24681 23705 19644 21531 22015 20383 22260 24176 22837 16258 19448 18746 20944 23025 26994 25082 19210 20872 24871 26033 28065 33002 32699 25932 32416 34978 34863 40170 47263 42373 38796 43449 46317 …]		
3	China	January 20 to June 5, 2020	138
ΔInf_3	[77 149 131 259 444 688 769 1771 1459 1737 1982 2102 2590 2829 3235 3887 3694 3143 3399 2656 3062 2478 2015 15152 5093 2644 2009 2053 1891 1751 825 892 399 649 416 527 411 440 329 430 573 206 128 120 143 145 103 46 45 20 31 25 11 18 27 29 39 35 84 65 46 82 102 147 99 114 118 135 128 106 98 86 93 78 19 55 75 66 86 92 56 64 113 115 99 49 52 27 31 21 36 13 37 15 9 12 25 3 6 22 4 12 4 5 7 4 2 3 6 1 14 17 1 7 6 5 10 5 7 6 5 2 13 0 3 11 7 1 3 0 4 5 18 9 7 1 11 6]		

Case 1. The analysis and predictive results of Canada by the AES tool is shown in Figure 2(a) based on the daily sample dataset of incremental infectives *ΔInf(t)* during March 1 to June 30, 2020 with the first 122 days of a partial COVID-19 lifecycle as presented in Table 3. The pandemic attributes derived by the AES tool including the reproductive rate $R_0(t)$, expected infectives $N_{\text{expinf}}(t)$, and real infectives $N_{\text{inf}}(t)$. Figure 2(a) demonstrates a fairly high accuracy in prediction and the insights of the dynamic $R_0(t)$ in COVID-19 pandemic. It is noteworthy that the maximums of the cumulated infectives $N_{inf\text{-max}}(t)$ and expected infectives $N_{inf}(t)$ are upto 169,961 vs. 170,330, respectively, which are shown in logarithmic scale in order to highlight the other attributes of COVID-19 dynamics in Canada in the given period.

Cases 2 and 3. Analyses and predications for COVID-19 in USA and China are shown in Figures 2(b) and 2(c), respectively. The big data of the former [10] are sampled in the same period as that of Canada. While the dataset of the latter is obtained from January 20 to June 5, 2020 with 138 days for a complete pandemic lifecycle [23]. The pandemic systems of the three sample countries in Figure 2 show that although the absolute values of pandemic trends may be widely different in the world or local communities, the basic pandemic patterns across them are common, that may be rigorously predicted based on the DPTM theory for COVID-19. The analytic results and simulations provide empirical support for the theories of DPTM and CPT as the general pandemic model for any other countries, regions, cities, or communities. Rational decision models may be generated by the AES tool for supporting rapid reactions and rational policy making based on the theories proven in Section III.

Case 4. Extrapolative forecasts for the key dynamic trends of $R_0(t)$ and its average value have been derived as illustrated in Figure 3, where the sample countries are comparatively studied. Figure 3 indicates that the driving initial $R_0(t_0)$ for triggering a pandemic is much greater than 1.0. However, $R_0(t_0)$ is decreasing through the lifecycle of the pandemic until it reaches 1.0. In the visualized results of Figure 3, the CPT and DPTM models as obtained in Definition 6 and Theorem 3 have successfully applied to accurately predict the key attributes and expectations of COVID-19 trends. Comparative analyses and calibrations of $R_0(t)$ by the AES tool are summarized in Table 4 where the approximate $\bar{R}_0(t)$ is determined by Eq. 4.



a) Canada b) USA c) China

Fig. 3. Forecasts of $R_0(t)$ trends in the COVID-19 dynamics

Table 4. COVID-19 $R_0(t)$ Calibrations

Country	**Dynamic $R_0(t)$**			**Approximate $\bar{R}_0(t)$**		
	Mean $\bar{R}_0(t)$	**Max** $R_{0\text{-}max}(t)$	**Min** $R_{0\text{-}min}(t)$	**Mean** $\bar{R}_{0x}(t)$	**Max** $R_{0x\text{-}max}(t)$	**Min** $R_{0x\text{-}max}(t)$
Canada	1.0921	2.0000	1.0000	1.2638	2.3513	1.1038
USA	1.1043	1.8000	1.0000	1.3953	3.8930	1.1290
China	1.0582	2.9351	1.0000	1.4221	7.0940	1.0857

Figure 3 reveals that the continuous increments of infectives in a COVID-19 series is not caused by an increasingly higher transmission rate $R_0(t)$ in the sample countries as traditional empirical explanations suggested. However, big dada analyzing results derived from the DPTM theory indicate that the main factor of a COVID-19 pandemic is driven by the exponential magnitude of cumulatively growing base of total infectives $\sum_{t=1}^{n} N_{inf}(t)$.

The experimental results as reported in Case Studies 1-4 as well as Figures 2 and 3 have demonstrated the strengths of the AES tool in real-world COVID-19 applications powered by the robust and rational CPT and DPTM theories. This basic research s a way to rigorously explain the myths of COVID-19 by an explainable and forecastable causal probability theory, the DPTM methodology, and the AES tool with associated cognitive and analytic algorithms.

V. CONCLUSION

This work has revealed a broader picture and deep insights of analytic epidemiology by the dynamic pandemic transmission model (DPTM). It has explored the cognitive, mathematical, and predicative foundations of analytic epidemiology. The causal probability theory (CPT) has been created for rigorously explaining how the fundamental reproductive rate $R_0(t)$ is rigorously defined and calibrated based on the big data of COVID-19. A theoretical framework of analytic epidemiology has been designed to elaborate the DPTM of epidemiology in general and COVID-19 in particular. Robust and accurate predictions on key attributes of COVID-19, including the calibrated transmissive rate $R_0(t)$, the predicated infectives at any day of the pandemic lifecycle, and the expected end of the pandemic, have been derived and demonstrated via four case studies with epidemical big data.

ACKNOWLEDGEMENT

This work is supported in part by the Department of National Defence's Innovation for Defence Excellence and Security (IDEaS) program, Canada, through the project of *AutoDefence: Towards Trustworthy Technologies for Autonomous Human-Machine Systems*, NSERC, and IEEE TC-BCS. The authors would like to thank the anonymous reviewers for their valuable suggestions and comments.

REFERENCES

[1] Anderson, R.M., May, R.M. (1979) Population biology of infectious diseases I, *Nature* 280, 361- 367.

[2] Anderson, R.M. (1982), ed. *Population Dynamics of infections Diseases: Theory and Applications*, Chapman and Hall, NewYork.

[3] Bailey, N.T.J. (1975), *The Mathematical Theory of Infectious Diseases*, 2nd ed. Harner, New York.

[4] Bender, E.A. (2000), *Mathematical Methods in Artificial Intelligence*, IEEE CS Press, Los Alamitos, CA.

[5] CDC (2020), *Coronavirus (COVID-19), Cases and Data*, https://www.cdc.gov/coronavirus/2019-ncov/covid-data/covidview/index.html.

[6] di Finetti, B. (1970), *Theory of Probability*, John Wiley & Sons, New York.

[7] Dietz, K. (1975), Transmission and Control of Arbovirus Diseases, *Epidemiology*, Philadelphia, USA, pp. 104-121.

[8] Government of Canada (2020), Coronavirus Disease (COVID-19), https://www.canada.ca/en/public-health/services/diseases/2019-novel-coronavirus-infection.html.

[9] Hethcote, H.W. (2008), Three Basic Epidemiological Models, Mathematical Understanding of Infectious Disease Dynamics, in *Models, Expressions for R0, Parameter Estimation, and Applications*. World Scientific, pp. 1-61.

[10] JHU (2020), *OVID-19 Map* - Johns Hopkins Coronavirus Resource Center, https://coronavirus.jhu.edu/map.html.

[11] Johnson, R.A. and G.K. Bhattacharyya (1996), *Statistics: Principles and Methods*, 3rd ed., John Wiley & Sons, Inc., NY.
[12] Kamps, B.S., C. Hoffmann (2003), eds. *SARS Reference*, 3rd ed. Flying Publisher, USA.
[13] Kucharski, A.J., T.W. Russell, C. Diamond, Y. Liu, J. Edmunds, S. Funk, and R.M. Eggo (2020), Early Dynamics of Transmission and Control of COVID-19: A Mathematical Modelling Study, *The Lancet Infectious Diseases*, 20(5), May, pp. 553-558.
[14] Last, J.M. (2001), ed. *Dictionary of Epidemiology*, 4th ed., Oxford University Press, New York, USA.
[15] Wang, Y. (2007), *Software Engineering Foundations: A Software Science Perspective*, Auerbach Publications (CRC), NY, USA.
[16] Wang, Y. (2011), On Cognitive Models of Causal Inferences and Causation Networks, *International Journal of Software Science and Computational Intelligence*, 3(1), 50-60.
[17] Wang, Y. (2011), Inference Algebra (IA): A Denotational Mathematics for Cognitive Computing and Machine Reasoning (I), *International Journal of Cognitive Informatics and Natural Intelligence*, 5(4), 61-82.
[18] Wang, Y. (2014), Fuzzy Causal Inferences based on Fuzzy Semantics of Fuzzy Concepts in Cognitive Computing, *WSEAS Transactions on Computers*, 13, 430-441.
[19] Wang, Y. (2015), Fuzzy Probability Algebra (FPA): A Theory of Fuzzy Probability for Fuzzy Inference and Computational Intelligence, *Journal of Advanced Mathematics and Applications*, 4(1), 38-55.
[20] Wang, Y. (2016), On Probability Algebra: Classic Theory of Probability Revisited, WSEAS *Trans. on Mathematics*, 15, 550-565.
[21] Wang, Y. (2016), Big Data Algebra: A Denotational Mathematics for Big Data Science and Engineering, *Journal of Advanced Mathematics and Applications*, 5(1), 3-25.
[22] Wang, Y. (2020), Keynote: Intelligent Mathematics: A Basic Research on Foundations of Autonomous Systems, General AI, Machine Learning, and Intelligence Science, IEEE 19th Int'l Conf. on Cognitive Informatics and Cognitive Computing (ICCI*CC'20), Tsinghua Univ., Beijing, China, Sept., pp. 4.
[23] WHO (2020), *COVID-19 Cases by Countries and Territories*, Geneva: World Health Organization, https://portal.who.int/report/eios-covid19-counts/.
[24] Worldometer (2020), *COVID-19 Coronavirus Pandemic*, https://www.worldometers.info/.