ORIGINAL ARTICLE

CANCER INNOVATION

Prediction of death rates for cardiovascular diseases and cancers

Oleg Gaidai[1] | Yihan Xing[2] | Rajiv Balakrishna[2] | Jiayao Sun[3] | Xiaolong Bai[3]

[1]Shanghai Engineering Research Center of Marine Renewable Energy, College of Engineering Science and Technology, Shanghai Ocean University, Shanghai, China

[2]Department of Mechanical and Structural Engineering and Materials Science, University of Stavanger, Stavanger, Norway

[3]School of Naval Architecture & Ocean Engineering, Jiangsu University of Science and Technology, Zhenjiang, China

Correspondence
Yihan Xing, Department of Mechanical and Structural Engineering and Materials Science, University of Stavanger, Kjell Arholms gate 41, Stavanger 4021, Norway.
Email: yihan.xing@uis.no

Funding information
None

Abstract

Background: To estimate cardiovascular and cancer death rates by regions and time periods.

Design: Novel statistical methods were used to analyze clinical surveillance data.

Methods: A multicenter, population-based medical survey was performed. Annual recorded deaths from cardiovascular diseases were analyzed for all 195 countries of the world. It is challenging to model such data; few mathematical models can be applied because cardiovascular disease and cancer data are generally not normally distributed.

Results: A novel approach to assessing the biosystem reliability is introduced and has been found to be particularly suitable for analyzing multiregion environmental and healthcare systems. While traditional methods for analyzing temporal observations of multiregion processes do not deal with dimensionality efficiently, our methodology has been shown to be able to cope with this challenge.

Conclusions: Our novel methodology can be applied to public health and clinical survey data.

KEYWORDS
cardiovascular disease, cancer, probability forecast, public health, mathematical biology

1 | BACKGROUND

Cardiovascular disease (CVD) refers to a range of diseases affecting the heart and blood vessels including hypertension (high blood pressure), coronary heart disease and heart attacks, cerebrovascular diseases (e.g., stroke and heart failure), and various other heart diseases. Cancers are defined by the National Cancer Institute as diseases in which abnormal cells can divide and infiltrate nearby tissues. Cancers can arise in many parts of the body; thus, there is a wide range of cancer types, as shown below, some of which spread to other parts of the body through the blood and lymph systems. CVD and cancer are the leading causes of death worldwide, therefore analyzing bivariate statistics is important. This study is concerned with public health systems rather than health at the level of the individual. The research is not clinical in nature; the goal is to

---

estimate the burden imposed by CVD and cancer on public health systems in different countries at any given time. We analyze mortality literature data for both CVDs [1–8] and cancer [9–29].

Assessing the reliability of healthcare systems and estimating excess mortality from CVDs using conventional statistical methods are challenging [30–35]. To achieve the latter goal over large areas, degrees of freedom are typically calculated for random variables governing dynamic biological systems. In principle, the reliability of a complex biological system can be accurately estimated if there are sufficient measurements or by using Monte Carlo simulations. For CVDs and cancers, however, data are scarce before 1990 [30]. Against this background, we introduce a novel method for assessing the reliability of biological and healthcare systems, to aid prediction and management of excess mortality from CVD. This study focused on cross-correlations in CVD and cancer deaths among countries within the same climatic zone. Worldwide health data and related research are readily available online [30].

Lifetime data analysis with the application of extreme value theory is widespread in the fields of medicine and engineering, [30]. A recent paper presented the arguments for and against using the upper distribution of life expectancy data [1]. A bivariate lifetime distribution is often assumed when analyzing statistical data [3]. A new approach that uses Clayton, Gumbel, and inverse Gaussian power variance functions, as well as conditional sampling and numerical approximation, was applied for survival analysis [2]. However, few studies have aimed to predict excess CVD and cancer mortality; this paper aimed to address this deficit.

In this paper, excess mortality from CVD is viewed as an unexpected event that may occur in any country at any time. The nondimensional factor λ is used to predict CVD risk. Biological systems are influenced by environmental parameters that can be modeled as ergodic processes. The CVD and cancer incidence data for 195 countries during the period 1990–2019 were retrieved [30]. The biological system under consideration herein can be regarded as a multidegree of freedom (MDOF) dynamic system with highly interrelated regional components/dimensions. This study focused on predicting excess mortality rather than symptoms.

2 | METHODS

Consider an MDOF biosystem subjected to random ergodic environmental influences. The other alternative is to view the process as being dependent on specific

environmental parameters whose variation in time may be modeled as an ergodic process on its own. The MDOF biomedical response vector process $\vec{R} \equiv \mathbf{R}(t)$ is measured and/or simulated over a sufficiently long time interval $(0, T)$. Unidimensional global maxima over the entire time span $(0, T)$ are denoted as $X_T^{\max} = \max_{0 \leq t \leq T} X(t)$, $Y_T^{\max} = \max_{0 \leq t \leq T} Y(t)$, $Z_T^{\max} = \max_{0 \leq t \leq T} Z(t)$, ... By sufficiently long time T, one primarily means a large value of T with respect to the dynamic system autocorrelation time.

Let $X_1, ..., X_{N_X}$ be consequent in the time local maxima of the bioprocess $X(t)$ at monotonously increasing discrete time instants $t_1^X < ... < t_{N_X}^X$ in $(0, T)$. The analogous definition follows for other MDOF biological system response components $Y(t)$, $Z(t)$, ... with $Y_1, ..., Y_{N_Y}$; $Z_1, ..., Z_{N_Z}$, and so on. For simplicity, all $\mathbf{R}(t)$ components, and therefore, its maxima are assumed to be nonnegative. The aim is to estimate system failure probability

$$
1 - P = \text{Prob}\Big(X_T^{\max} > \eta_X \cup Y_T^{\max} \\ > \eta_Y \cup Z_T^{\max} > \eta_Z \cup...\Big),
$$
(1)

with

$$
P = \int\!\!\!\int\!\!\!\int_{(0,0,0,...)}^{(\eta_X, \eta_Y, \eta_Z, ...)} p_{X_T^{\max}, Y_T^{\max}, Z_T^{\max}, ...} \\ \Big(X_T^{\max}, Y_T^{\max}, Z_T^{\max}, ...\Big) dX_T^{\max} dY_{N_Y}^{\max} dZ_{N_z}^{\max} ...,
$$
(2)

being the probability of nonexceedance for response components η_X, η_Y, η_Z, ... critical values; $\cup$ denotes logical unity operation «or»; and $p_{X_T^{\max}, Y_T^{\max}, Z_T^{\max}, ...}$ being joint probability density of the global maxima over the entire time span $(0, T)$.

In practice, however, it is not feasible to estimate the latter joint probability distribution directly $p_{X_T^{\max}, Y_T^{\max}, Z_T^{\max}, ...}$ due to its high dimensionality and available data set limitations. In other words, the time instant when either X exceeds, Y exceeds, Z exceeds, and so on, the system is regarded as immediately failed. Fixed failure levels η_X, η_Y, η_Z, ... are, of course, individual for each unidimensional response component of $\mathbf{R}(t)$. $X_{N_X}^{\max} = \max\{X_j; j = 1, ..., N_X\}$ $= X_T^{\max}$, $Y_{N_Y}^{\max} = \max\{Y_j; j = 1, ..., N_Y\} = Y_T^{\max}$, $Z_{N_z}^{\max} = \max\{Z_j; j = 1, ..., N_Z\} = Z_T^{\max}$, and so on, see Naess and Gaidai [32] and Naess and Moan [49].

Next, the local maxima temporal instants $[t_1^X < ... < t_{N_X}^X; t_1^Y < ... < t_{N_Y}^Y; t_1^Z < ... < t_{N_z}^Z]$ in monotonously nondecreasing order being sorted into one single merged synthetic time vector $t_1 \leq ... \leq t_N$. Note that $t_N = \max$

$\{t_{N_X}^X, t_{N_Y}^Y, t_{N_Z}^Z, ...\}$, $N = N_X + N_Y + N_Z + ...$. In this case, t_j represents the local maxima of one of the MDOF biosystem response components either $X(t)$, $Y(t)$, or $Z(t)$, and so on. That means that having $\boldsymbol{R}(t)$ time record, one just needs to continuously and simultaneously screen for unidimensional response component local maxima and record its exceedance of the MDOF limit vector $(\eta_X, \eta_Y, \eta_Z, ...)$ in any of its components $X, Y, Z, ...$. The local unidimensional response component maxima are merged into one temporal nondecreasing vector $\vec{R} = (R_1, R_2, ..., R_N)$ in accordance with the merged time vector $t_1 \leq ... \leq t_N$. That is to say, each local maxima R_j is the actual encountered local maxima corresponding to either $X(t)$, $Y(t)$, or $Z(t)$, and so on. Finally, the unified limit vector $(\eta_1, ..., \eta_N)$ is introduced with each component η_j is either η_X, η_Y, or η_Z and so on, depending on which of $X(t)$ or $Y(t)$ or $Z(t)$, and so forth, corresponds to the current local maxima with the running index j.

Next, a scaling parameter $0 < \lambda \leq 1$ is introduced to artificially simultaneously decreases limit values for all biosystem response components, namely, the new MDOF limit vector $(\eta_X^\lambda, \eta_Y^\lambda, \eta_z^\lambda, ...)$ with $\eta_X^\lambda \equiv \lambda \cdot \eta_X$, $\equiv \lambda \cdot \eta_Y$, $\eta_z^\lambda \equiv \lambda \cdot \eta_Z$, ... is introduced. The unified limit vector $(\eta_1^\lambda, ..., \eta_N^\lambda)$ introduced with each component η_j^λ is either η_X^λ, η_Y^λ, or η_z^λ and so on. The latter automatically defines probability $P(\lambda)$ as a function of λ; note that $P \equiv P(1)$ from Equation (1). Nonexceedance probability $P(\lambda)$ can be now estimated as follows:

$$
\begin{aligned}
P(\lambda) &= \mathrm{Prob}\left\{R_N \leq \eta_N^\lambda, ..., R_1 \leq \eta_1^\lambda\right\} \\
&= \mathrm{Prob}\left\{R_N \leq \eta_N^\lambda \middle| R_{N-1} \leq \eta_{N-1}^\lambda, ..., R_1 \leq \eta_1^\lambda\right\} \\
&\quad \cdot \mathrm{Prob}\left\{R_{N-1} \leq \eta_{N-1}^\lambda, ..., R_1 \leq \eta_1^\lambda\right\} \\
&= \prod_{j=2}^N \mathrm{Prob}\left\{R_j \leq \eta_j^\lambda \middle| R_{j-1} \leq \eta_{1j-}^\lambda, ..., R_1 \leq \eta_1^\lambda\right\} \\
&\quad \cdot \mathrm{Prob}\left(R_1 \leq \eta_1^\lambda\right).
\end{aligned}
\tag{3}
$$

In practice, the dependency between neighboring R_j values is not always negligible; thus, the following one-step (i.e., "conditioning level"; $k = 1$) memory approximation is introduced

$$
\begin{aligned}
&\mathrm{Prob}\left\{R_j \leq \eta_j^\lambda \middle| R_{j-1} \leq \eta_{j-1}^\lambda, ..., R_1 \leq \eta_1^\lambda\right\} \\
&\approx \mathrm{Prob}\left\{R_j \leq \eta_j^\lambda \middle| R_{j-1} \leq \eta_{j-1}^\lambda\right\},
\end{aligned}
\tag{4}
$$

for $2 \leq j \leq N$ (called here conditioning level $k = 2$). Approximation being introduced by Equation (4) may be further expressed as

$$
\begin{aligned}
&\mathrm{Prob}\left\{R_j \leq \eta_j^\lambda \middle| R_{j-1} \leq \eta_{j-1}^\lambda, ..., R_1 \leq \eta_1^\lambda\right\} \\
&\approx \mathrm{Prob}\left\{R_j \leq \eta_j^\lambda \middle| R_{j-1} \leq \eta_{j-1}^\lambda, R_{j-2} \leq \eta_{j-2}^\lambda\right\},
\end{aligned}
\tag{5}
$$

where $3 \leq j \leq N$ (will be called conditioning level $k = 3$) and so on. The motivation is to monitor each independent failure that happened locally first in time, thus avoiding cascading local intercorrelated exceedances [36–48].

Equation (5) presents subsequent refinements of the statistical independence assumption. The latter type of approximations enables capturing the statistical dependence effect between neighboring maxima with increased accuracy. Since the original MDOF bioprocess $\boldsymbol{R}(t)$ was assumed ergodic and therefore stationary, probability $p_k(\lambda) := \mathrm{Prob}\{R_j > \eta_j^\lambda | R_{j-1} \leq \eta_{j-1}^\lambda, R_{j-k+1} \leq \eta_{j-k+1}^\lambda\}$ for $j \geq k$ will be independent of j but only dependent on conditioning level k. Thus, the nonexceedance probability can be approximated as in the Naess–Gaidai method, see [32, 49], where:

$$
P_k(\lambda) \approx \exp(-N \cdot p_k(\lambda)), \ k \geq 1.
\tag{6}
$$

Note that Equation (6) follows from Equation (1) by neglecting $\mathrm{Prob}(R_1 \leq \eta_1^\lambda) \approx 1$, as the design failure probability is usually very small. Further, it is assumed that $N \gg k$. Note that Equation (5) is similar to the well-known mean up-crossing rate equation for the probability of exceedance [32, 49]. There is observed convergence with respect to conditioning parameter k

$$
P = \lim_{k \to \infty} P_k(1); \ p(\lambda) = \lim_{k \to \infty} p_k(\lambda).
\tag{7}
$$

Note that Equation (6) for $k = 1$ turns into the quite well-known nonexceedance probability relationship with the mean up-crossing rate function

$$
P(\lambda) \approx \exp(-\nu^+(\lambda)T); \ (\lambda) = \int_0^\infty \zeta p_{R\dot{R}}(\lambda, \zeta)d\zeta,
\tag{8}
$$

where $\nu^+(\lambda)$ is the mean up-crossing rate of the response level λ for the above assembled nondimensional vector $R(t)$ assembled from scaled MDOF biosystem response $\left(\frac{X}{\eta_X}, \frac{Y}{\eta_Y}, \frac{Z}{\eta_Z}, ...\right)$. The proposed methodology can also

treat nonstationary cases. An illustration of how the methodology can be used to treat nonstationary cases is provided as follows. Consider a scattered diagram of $m = 1, ..., M$ bioenvironmental states, with each short-term bioenvironmental state having probability q_m so that $\sum_{m=1}^{M} q_m = 1$. The corresponding long-term equation is then

$$p_k(\lambda) \equiv \sum_{m=1}^{M} p_k(\lambda, m) q_m, \qquad (9)$$

with $p_k(\lambda, m)$ being the same function as in Equation (7) but corresponding to a specific short-term environmental state with the number m. Note that this statistical model has already been validated [47, 50–52].

3 | RESULTS

Prediction of CVD and cancer has long been a target in the fields of epidemiology and mathematical biology. Public health systems are dynamic, highly nonlinear, multidimensional, and spatially diverse systems that are challenging to analyze. Previous studies have used a variety of approaches to predict CVD and cancer cases. In this section, the above-described methodology is applied to real-world CVD data sets for all countries of the world.

The statistical data in the present section are from the "Our World in Data" website [30], which provides annual CVD death rates for all countries for the period 1990–2019. The death rates for the 195 countries (components $X, Y, Z...$) constitute 195 dimensional (195D) data for a dynamic biological system.

General failure limits $(\eta_X, \eta_Y, \eta_Z, ...)$, that is, CVD thresholds, are less intuitive than setting failure limits for each individual country according to its population, such that $X, Y, Z, ...$ are equal to the annual death rate of a given country. The death rate for cancer is lower than that for CVD, but it is typically more painful to die from cancer. In this paper, the "failure limit" for cancer is lowered fourfold to match that for CVD.

Next, the local maxima from all nondimensionalized time series data are merged into a single time series using Equation (5):

$$\vec{R} = \left(\left\{ \max \left\{ X_1^{\text{cardio}}, X_1^{\text{cancer}} \right\}, \max \left\{ Y_1^{\text{cardio}}, Y_1^{\text{cancer}} \right\}, \right. \right.$$
$$\max \left\{ Z_1^{\text{cardio}}, Z_1^{\text{cancer}} \right\}, ... \right\}, ..., \left\{ \max \left\{ X_N^{\text{cardio}}, X_N^{\text{cancer}} \right\}, \right.$$
$$\left. \left. \max \left\{ Y_N^{\text{cardio}}, Y_N^{\text{cancer}} \right\}, \max \left\{ Z_N^{\text{cardio}}, Z_N^{\text{cancer}} \right\}, ... \right\} \right).$$
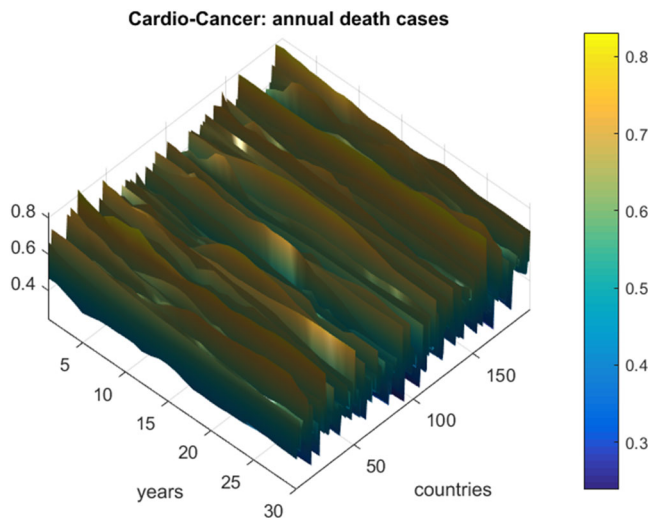$$(10)$$

FIGURE 1 Annual deaths from cardiovascular disease and cancer as a percentage of the population for 195 countries.

Each maximum, such as $\max\{X_j^{\text{cardio}}, X_j^{\text{cancer}}\}$, is inserted into single time series according to its temporal occurrence (denoted by subscript j).

Figure 1 presents the annual deaths from CVD and cancer by country and year. Figure 2 presents the number of new deaths as a 195D vector $\vec{R}$. Data for Uzbekistan were excluded from the analysis because they were regarded as outliers. $\vec{R}$ was assembled from different regional components, that is, CVD data sets. Index j is a running index of local maxima encountered in the "non-decreasing" time series.

Overall, there is a clear East–West divide in the CVD death rates. Rates across North America and Western/Northern Europe tended to be lower than those across Eastern Europe, Asia, and Africa. For most of Latin America, the rates were moderate. As an example, in France, the age-standardized CVD death rate was around 86 per 100,000 in 2017, while across Eastern Europe, it was around five times higher (400–500 per 100,000). Uzbekistan had the highest rate of 724 per 100,000.

Figure 3 presents the predicted annual CVD death rates (percentage relative to the entire population of a given country) over 100 years, extrapolated from Equation (10). $\lambda = 0.6\%$ was used as a cut-off value. The 95% confidence intervals (CIs) were calculated. According to Equation (5), $p(\lambda)$ is directly related to the target failure probability $(1 - P)$ derived from Equation (1). Therefore, system failure probability can be estimated as $1 - P \approx 1 - P_k(1)$. Note that, in Equation (6), N corresponds to the total number of local maxima in response vector $\vec{R}$. Conditioning parameter $k = 3$ was
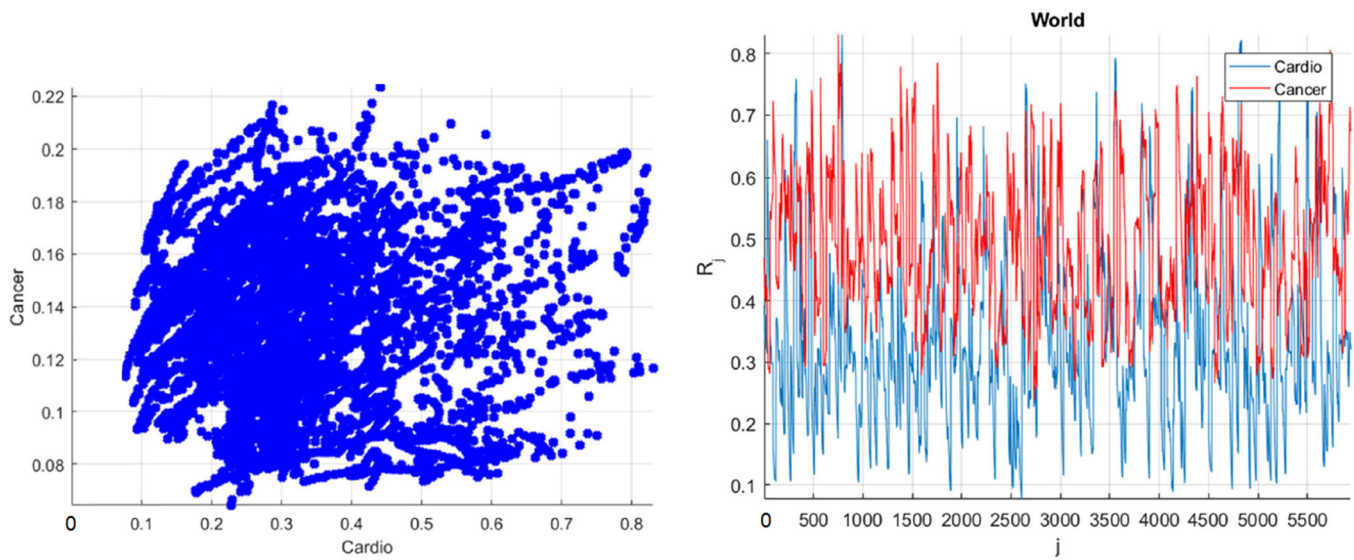
FIGURE 2   Left: Cross-correlations between cardiovascular disease (CVD) and cancer cases as a percentage of the population. Right: Annual death rates as a 195-dimensional vector $\vec{R}$, as a percentage of the population of the corresponding country. The cancer rate was increased fourfold to match that of CVD.
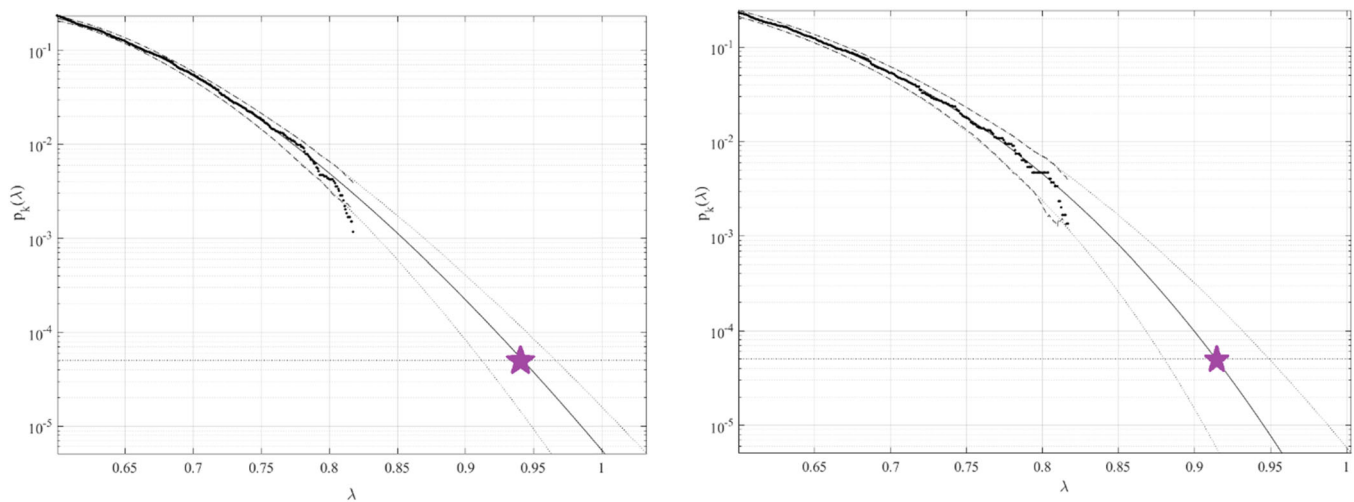

FIGURE 3   Death rate predictions over 100 years extrapolated from $p_k(\lambda)$. The critical level is indicated by a star. The 95% confidence intervals are indicated by dotted lines. The percentage of the population is represented by the horizontal axis. Left: Predictions based on 30 years of data; Right: predictions based on 15 years of data.

TABLE 1   Predicted cardiovascular disease death rates over 100 years based on 30- and 15-year data sets.

	Predicted death rate (%)	95% CI, lower bound	95% CI, upper bound
30-year data set	0.942	0.909	0.966
15-year data set	0.914	0.879	0.949

Abbreviation: CI, confidence interval.

found to be sufficient because of the convergence of k (see Equation 6). In Figure 3, the 95% CIs are relatively narrow, which represents an advantage of the proposed method. Table 1 compares 100-year predictions based on data for 15- and 30-year periods. The 15-year data set was derived from the full 30-year data set by omitting odd years. The 95% CIs were wider for the truncated data set, as expected.

The predicted average annual CVDs over the next 100 years, among all years and countries, were found below 1%. Our methodology uses available data efficiently by assuming that healthcare system data sets are multidimensional and extrapolates death rates even when the data set is relatively limited. The predicted nondimensional factor λ, indicated by the star in Figure 3, represents the probability of excess CVD mortality for any given country. Our method could be applied to predict cancer clusters, rather than merely death rates over time, which would be of high practical importance.

4 | CONCLUSIONS

Traditional methods for assessing the reliability of healthcare systems on the basis of time series data do not efficiently deal with systems characterized by high dimensionality and cross-correlations. The main advantage of our methodology is its ability to assess the reliability of high-dimensional nonlinear dynamic systems. Despite its simplicity, the novel multidimensional modeling strategy introduced herein can be used for accurate forecasting of CVD death rates in individual countries.

We analyzed 195D data, that is, CVD and cancer death rates for 195 countries worldwide, for the period 1990–2019. A novel method for analyzing the reliability of a multidimensional biosystem was applied and the mechanisms of the proposed method were described in detail. Direct measurements and Monte Carlo simulations are both suitable for assessing the reliability of dynamic biological systems; however, the complexity and high dimensionality of such systems necessitate the further development of robust and accurate techniques that can use limited data sets in an efficient manner.

This study predicted an average annual death rate for CVD over a 100-year period of about 1% across countries and years. Under current national health management approaches, CVDs will continue to represent a threat to the health of the world population.

This study introduced a general-purpose, robust, and easy-to-apply method for analyzing the reliability of multidimensional systems. The method has previously been validated by application to a wide range of simulation models but only in the context of one-dimensional systems; in general, highly accurate predictions were obtained. Both measurement and numerically simulated time series data can be analyzed. Applying the method to the data set used in this study yielded reasonable confidence intervals, indicating that it could serve as a useful tool for reliability studies of various nonlinear dynamic biological systems. Finally, the suggested methodology has many potential public health applications beyond the prediction of CVD death rates.

AUTHOR CONTRIBUTIONS
Oleg Gaidai: Conceptualization (equal). **Yihan Xing**: Validation (equal). **Rajiv Balakrishna**: Investigation (equal). **Jiayao Sun**: Methodology (equal). **Xiaolong Bai**: Validation (equal).

CONFLICT OF INTEREST STATEMENT
The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT
Data sets analyzed during the current study are available online at https://ourworldindata.org/causes-of-death ("Our World in Data" [30]).

ETHICS STATEMENT
Not applicable.

INFORMED CONSENT
Not applicable.

ORCID
Oleg Gaidai http://orcid.org/0000-0002-3196-8562

REFERENCES

1. Cox B, Vangronsveld J, Nawrot TS. Impact of stepwise introduction of smokefree legislation on population rates of acute myocardial infarction deaths in Flanders, Belgium. Heart. 2014; 100(18):1430–5. https://doi.org/10.1136/heartjnl-2014-305613

2. Tsao CW, Aday AW, Almarzooq ZI, Alonso A, Beaton AZ, Bittencourt MS, et al. Heart disease and stroke statistics—2022 update: a report from the American Heart Association. Circulation. 2022;145(8):e153–639. https://doi.org/10.1161/CIR.0000000000001052

3. Smolina K, Wright FL, Rayner M, Goldacre MJ. Determinants of the decline in mortality from acute myocardial infarction in England between 2002 and 2010: linked national database study. BMJ. 2012;344:d8059. https://doi.org/10.1136/bmj.d8059

4. Balakrishna R, Bjørnerud T, Bemanian M, Aune D, Fadnes LT. Consumption of nuts and seeds and health outcomes including cardiovascular disease, diabetes and metabolic disease, cancer, and mortality: an umbrella review. Adv Nutr. 2022;13(6):2136–48. https://doi.org/10.1093/advances/nmac077

5. Mackay DF, Irfan MO, Haw S, Pell JP. Meta-analysis of the effect of comprehensive smoke-free legislation on acute coronary events. Heart. 2010;96(19):1525–30. https://doi.org/10.1136/hrt.2010.199026

6. Alzuhairi KS, Søgaard P, Ravkilde J, Gislason G, Køber L, Torp-Pedersen C. Incidence and outcome of first myocardial infarction according to gender and age in Denmark over a 35-year period

(1978–2012). Eur Heart J Qual Care Clin Outcomes. 2015;1(2):72–8. https://doi.org/10.1093/ehjqcco/qcv016

7. Mirzaei M, Truswell AS, Taylor R, Leeder SR. Coronary heart disease epidemics: not all the same. Heart. 2009;95:740–6. https://doi.org/10.1136/hrt.2008.154856

8. NCDRF Collaboration. Trends in adult body-mass index in 200 countries from 1975 to 2014: a pooled analysis of 1698 population-based measurement studies with 19.2 million participants. Lancet. 2016;387(10026):1377–96. https://doi.org/10.1016/S0140-6736(16)30054-X

9. Siegel R, Miller K, Fuchs H, Jemal A. Cancer statistics. CA Cancer J Clin. 2022;72(1):7–33. https://doi.org/10.3322/caac.21708

10. Surveillance, Epidemiology, and End Results (SEER) Program. SEER*Stat Database: North American Association of Central Cancer Registries (NAACCR) Incidence Data—Cancer in North America Analytic File, 1995–2018, With Race/Ethnicity, Custom File With County, American Cancer Society Facts and Figures Projection Project (which includes data from the Center for Disease Control and Prevention's National Program of Cancer Registries, the Canadian Council of Cancer Registries' Provincial and Territorial Registries, and the National Cancer Institute's SEER Registries, certified by the NAACCR as meeting high-quality incidence data standards for the specified time periods). National Cancer Institute, Division of Cancer Control and Population Sciences, Surveillance Research Program; Atlanta, Georgia, 2021.

11. Sherman R, Firth R, Charlton M, De P, Prithwish D, Green D, et al., editors. Cancer in North America: 2014–2018. Volume two: registry-specific cancer incidence in the United States and Canada. North American Association of Central Cancer Registries Inc.; Springfield, Illinois, 2021.

12. Surveillance, Epidemiology, and End Results (SEER) Program. SEER*Stat Database: mortality—all causes of death, total U.S. (1969–2019)—Katrina/Rita population adjustment—linked to county attributes—total U.S., 1969–2019 Counties (underlying mortality data provided by the National Center for Health Statistics). Springfield, Illinois: National Cancer Institute, Division of Cancer Control and Population Sciences, Surveillance Research Program; 2021.

13. Wingo PA, Cardinez CJ, Landis SH, Greenlee RT, Ries LAG, Anderson RN, et al. Long-term trends in cancer mortality in the United States, 1930–1998. Cancer. 2003;97(12 suppl):3133–275. https://doi.org/10.1002/cncr.11380

14. Fritz A, Percy C, Jack A, Shanmugaratnam K, Sobin L, Parkin DM, et al., editors. International classification of diseases for oncology. 3rd ed. Geneva, Switzerland: World Health Organization; 2000.

15. World Health Organization (WHO). International statistical classification of diseases and related health problems, 10th revision. Vol. I-III. Geneva, Switzerland: WHO; 2011.

16. Surveillance Research Program. SEER*Stat software, version 8.3.8. Bethesda, Maryland: National Cancer Institute; 2020.

17. Surveillance Research Program. Joinpoint Regression Program version 4.9.0.1. Bethesda, Maryland: National Cancer Institute, Statistical Research and Applications Branch; 2021.

18. Mariotto AB, Zou Z, Johnson CJ, Scoppa S, Weir HK, Huang B. Geographical, racial and socio-economic variation in life expectancy in the US and their impact on cancer relative survival. PLoS One. 2018;13(7):e0201034. https://doi.org/10.1371/journal.pone.0201034

19. Clegg LX, Feuer EJ, Midthune DN, Fay MP, Hankey BF. Impact of reporting delay and reporting error on cancer incidence rates and trends. J Natl Cancer Inst. 2002;94(20):1537–45. https://doi.org/10.1093/jnci/94.20.1537

20. Yabroff KR, Wu XC, Negoita S, Stevens J, Coyle L, Zhao J, et al. Association of the COVID-19 pandemic with patterns of statewide cancer services. J Natl Cancer Inst. 2021;114(6):907–9.

21. Surveillance, Epidemiology, and End Results (SEER) Program. SEER*Stat Database: incidence—SEER 9 registries research data with delay—adjustment, malignant only, November 2020 submission (1975–2018)—Katrina/Rita population adjustment—linked to county attributes—total U.S., 1969–2018 counties. Bethesda, Maryland: National Cancer Institute, Division of Cancer Control and Population Sciences, Surveillance Research Program, Surveillance Systems Branch; 2021.

22. Surveillance, Epidemiology, and End Results (SEER) Program. SEER*Stat Database: incidence—SEER 18 registries research data + Hurricane Katrina impacted Louisiana cases, November 2020 submission (2000–2018)—Katrina/Rita population adjustment—linked to county attributes—total U.S., 1969–2018 counties. Bethesda, Maryland: National Cancer Institute, Division of Cancer Control and Population Sciences, Surveillance Research Program, Surveillance Systems Branch; 2021.

23. Surveillance Research Program. SEER*Explorer: an interactive website for SEER cancer statistics. Bethesda, Maryland: National Cancer Institute; 2021. https://seer.cancer.gov/explorer/. Accessed 15 April 2021.

24. Surveillance, Epidemiology, and End Results (SEER) Program. SEER*Stat Database: incidence—SEER research limited—field data with delay—adjustment, 21 registries, malignant only, November 2020 submission (2000–2018)—linked to county attributes—time dependent (1990–2018) income/rurality, 1969–2019 counties. Bethesda, Maryland: National Cancer Institute, Division of Cancer Control and Population Sciences, Surveillance Research Program; 2021.

25. Surveillance Research Program, Statistic Methodology and Applications. DevCan: probability of developing or dying of cancer software. Version 6.7.9. Bethesda, Maryland: National Cancer Institute; 2021.

26. Murphy SL, Kochanek KD, Xu J, Heron M. Deaths: final data for 2012. National Vital Statistics Reports. Vol. 63, No. 9. Hyattsville, Maryland: National Center for Health Statistics; 2015.

27. Steliarova-Foucher E, Stiller C, Lacour B, Kaatsch P. International classification of childhood cancer, third edition. Cancer. 2005;103(7):1457–67. https://doi.org/10.1002/cncr.20910

28. Gumbel E. Statistics of extremes. New York: Columbia University Press; 1958.

29. Sherman R, Firth R, Charlton M, De P, Green D, Hofer B, et al., editors. Cancer in North America: 2014–2018. Volume one: combined cancer incidence for the United States, Canada and North America. North American Association of Central Cancer Registries Inc.; 2021.

30. Ritchie H, Spooner F, Roser M. Causes of death. https://ourworldindata.org/causes-of-death

31. Choi S-K, Grandhi RV, Canfield RA. Reliability-based structural design. London: Springer-Verlag; 2007.

32. Naess A, Gaidai O. Estimation of extreme values from sampled time series. Struct Saf. 2009;31(4):325–34. https://doi.org/10.1016/j.strusafe.2008.06.021

33. Madsen HO, Krenk S, Lind NC. Methods of structural safety. Englewood Cliffs: Prentice-Hall Inc.; 1986.

34. Ditlevsen O, Madsen HO. Structural reliability methods. Chichester (World): John Wiley & Sons Inc.; 1996.

35. Melchers RE. Structural reliability analysis and prediction. New York: John Wiley & Sons Inc.; 1999.

36. Xing Y, Gaidai O, Ma Y, Naess A, Wang F. A novel design approach for estimation of extreme responses of a subsea shuttle tanker hovering in ocean current considering aft thruster failure. Appl Ocean Res. 2022;123:103179. https://doi.org/10.1016/j.apor.2022.103179

37. Gaidai O, Wang F, Wu Y, Xing Y, Medina AR, Wang J. Offshore renewable energy site correlated wind-wave statistics. Probabilistic Eng Mech. 2022;68:103207. https://doi.org/10.1016/j.probengmech.2022.103207

38. Xu X, Xing Y, Gaidai O, Wang K, Sandipkumar Patel K, Dou P, et al. A novel multi-dimensional reliability approach for floating wind turbines under power production conditions. Front Marine Sci. 2022;9:970081. https://doi.org/10.3389/fmars.2022.970081

39. Sun J, Gaidai O, Wang F, Naess A, Wu Y, Xing Y, et al. Extreme riser experimental loads caused by sea currents in the Gulf of Eilat. Probabilistic Eng Mech. 2022;68:103243. https://doi.org/10.1016/j.probengmech.2022.103243

40. Xu X, Wang F, Gaidai O, Naess A, Xing Y, Wang J. Bivariate statistics of floating offshore wind turbine dynamic response under operational conditions. Ocean Eng. 2022;257:111657. https://doi.org/10.1016/j.oceaneng.2022.111657

41. Gaidai O, Xing Y, Wang F, Wang S, Yan P, Naess A. Improving extreme anchor tension prediction of a 10-MW floating semi-submersible type wind turbine, using highly correlated surge motion record. Front Mech Eng. 2022;8:888497. https://doi.org/10.3389/fmech.2022.888497

42. Gaidai O, Xing Y, Xu X. COVID-19 epidemic forecast in USA East coast by novel reliability approach. Res Sq. 2022. https://doi.org/10.21203/rs.3.rs-1573862/v1

43. Gaidai O, Xing Y, Balakrishna R. Improving extreme response prediction of a subsea shuttle tanker hovering in ocean current using an alternative highly correlated response signal. Results Eng. 2022;15:100593. https://doi.org/10.1016/j.rineng.2022.100593

44. Cheng Y, Gaidai O, Yurchenko D, Xu X, Gao S. The 32nd International Ocean and Polar Engineering Conference, paper number: ISOPE-I-22-342, Shanghai, China. 2022.

45. Gaidai O, Storhaug G, Wang F, Yan P, Naess A, Wu Y, et al. On-Board Trend Analysis for Cargo Vessel Hull Monitoring Systems. The 32nd International Ocean and Polar Engineering Conference, paper number: ISOPE-I-22-541, Shanghai, China. 2022.

46. Gaidai O, Xu X, Naess A, Cheng Y, Ye R, Wang J. Bivariate statistics of wind farm support vessel motions while docking. Sh Offshore Struct. 2020;16(2):135–43. https://doi.org/10.1080/17445302.2019.1710936

47. Gaidai O, Fu S, Xing Y. Novel reliability method for multi-dimensional nonlinear dynamic systems. Mar Struct. 2022;86:103278. https://doi.org/10.1016/j.marstruc.2022.103278

48. Gaidai O, Xu J, Yan P, Xing Y, Wu Y, Zhang F. Novel methods for wind speeds prediction across multiple locations. Sci Rep. 2022;12:19614. https://doi.org/10.1038/s41598-022-24061-4

49. Naess A, Moan T. Stochastic dynamics of marine structures. New York: Cambridge University Press; 2013.

50. Gaidai O, Xing Y. A novel multi regional reliability method for COVID-19 death forecast. Eng Sci. 2022. https://doi.org/10.30919/es8d799

51. Gaidai O, Yihan Y. A novel bio-system reliability approach for multi-state COVID-19 epidemic forecast. Eng Sci. 2022. https://doi.org/10.30919/es8d797

52. Gaidai O, Yan P, Xing Y, Xu J, Wu Y. A novel statistical method for long-term coronavirus modelling. F1000 Res. 2022;11:1282.