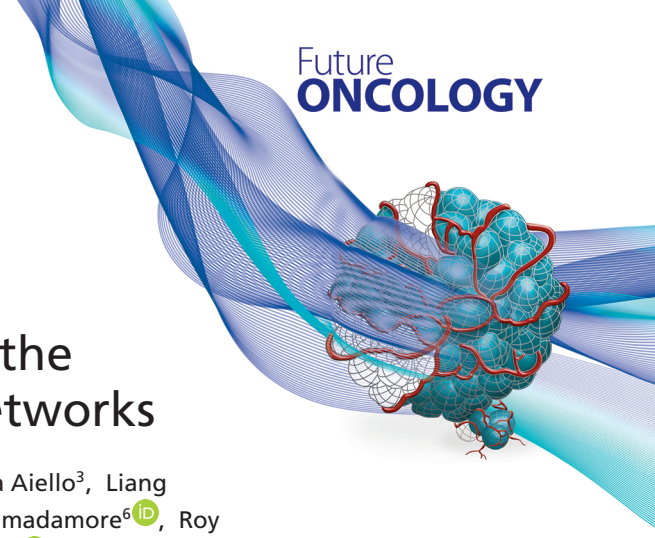
Research Article

For reprint orders, please contact: reprints@futuremedicine.com



Predicting future cancer burden in the United States by artificial neural networks

Francesco Piva[‡,1], Francesca Tartari[‡,2], Matteo Giulietti[1], Marco Maria Aiello[3], Liang Cheng[4], Antonio Lopez-Beltran[5], Roberta Mazzucchelli[6], Alessia Cimadamore[6], Roy Cerqueti[7,8], Nicola Battelli[9], Rodolfo Montironi*[,6] & Matteo Santoni[9]

[1]Department of Specialistic Clinical & Odontostomatological Sciences, Polytechnic University of Marche, 60126, Ancona, Italy
[2]Department of Economics & Law, University of Macerata, via Crescimbeni, 20, 62100, Macerata, Italy
[3]Oncology Unit, Policlinico Hospital, 95123, Catania, Italy
[4]Department of Pathology & Laboratory Medicine, Indiana University School of Medicine, Indianapolis, IN 46202, USA
[5]Department of Surgery, Cordoba University Medical School, 14071, Cordoba, Spain
[6]Section of Pathological Anatomy, Polytechnic University of the Marche Region, School of Medicine, United Hospitals, 60126, Ancona, Italy
[7]Department of Social & Economic Sciences, Sapienza University of Rome, Piazzale Aldo Moro, 5 - I-00185, Rome, Italy
[8]School of Business, London South Bank University, London, SE1 0AA, UK
[9]Oncology Unit, Macerata Hospital, 62012, Macerata, Italy
*Author for correspondence: Tel.: +39 071 596 4830; r.montironi@univpm.it
‡These authors contributed equally to this work

Aims: To capture the complex relationships between risk factors and cancer incidences in the US and predict future cancer burden. **Materials & methods:** Two artificial neural network (ANN) algorithms were adopted: a multilayer feed-forward network (MLFFNN) and a nonlinear autoregressive network with eXogenous inputs (NARX). Data on the incidence of the four most common tumors (breast, colorectal, lung and prostate) from 1992 to 2016 (available from National Cancer Institute online datasets) were used for training and validation, and data until 2050 were predicted. **Results:** The rapid decreasing trend of prostate cancer incidence started in 2010 will continue until 2018–2019; it will then slow down and reach a plateau after 2050, with several differences among ethnicities. The incidence of breast cancer will reach a plateau in 2030, whereas colorectal cancer incidence will reach a minimum value of 35 per 100,000 in 2030. As for lung cancer, the incidence will decrease from 50 per 100,000 (2017) to 31 per 100,000 in 2030 and 26 per 100,000 in 2050. **Conclusion:** This up-to-date prediction of cancer burden in the US could be a crucial resource for planning and evaluation of cancer-control programs.

First draft submitted: 16 April 2020; Accepted for publication: 24 August 2020; Published online: 11 December 2020

Keywords: artificial neural network • breast cancer • colorectal cancer • future tumor burden • lung cancer • prostate cancer

Cancer represents the second most lethal disease condition in the US, with more than 1600 deaths per day and an estimated total that will exceed 600,000 Americans in 2020 [1]. The number of new diagnoses has progressively increased in recent decades, with more than 1,800,000 new cases estimated in the US in 2020 [1]. However, it is important to note that in the past 25 years, the rate of cancer-related deaths has declined for the four types of tumor with the highest incidences (prostate, breast, colorectal and lung) [1]. This can be explained by advances achieved through the introduction of molecular targeted drugs and novel immunotherapies as well as the development of more effective diagnostic techniques and the downward trend of cancer-related risk factors, particularly smoking.

The predictions of incidence cancer rates could be useful to optimize the allocation of finite resources, the key elements of cancer control in the coming years and the future planning of cancer control programs. This has rapidly become fundamental due to the rising costs of oncological treatments approved by the US FDA in the recent decades [2–5]. The trends of population growth and the aging represent crucial factors for predicting the future burden of cancer, as the majority of tumors are age dependent [6]. Furthermore, we might attempt to estimate the future incidence of different tumor types based on the plausible changes of risk patterns over time.

Previously, several prediction techniques have been employed to estimate future cancer trends. These methods include linear extrapolation of trends [7,8], simple linear Poisson models [9,10] and classical [6,11] or Bayesian age-period-cohort models [12]. However, the limits of these techniques [13] have led to research to determine more accurate prediction models. To capture the complex relationships between input and target variables, we adopted two artificial neural network (ANN) algorithms: the classical multilayer feed-forward network (MLFFNN) and the nonlinear autoregressive network with exogenous input (NARX). ANN imitates the human brain's strategy of solving problems and is able to extract knowledge directly from the raw data. These nonparametric modeling algorithms are flexible and can perform complex function mapping. Because ANN algorithms can be trained by inputting the data from various circumstances, they are used in numerous fields, from finance to medicine [14–17]. In this study, we estimated the incidence of the four most frequent cancer diseases (prostate, breast, colorectal and lung) by developing two ANN algorithms to predict the number of new cases in the US through 2050.

Materials & methods

Data sets

The US population and life expectancy data were derived from Gapminder (www.gapminder.org). Observed and projected obesity data in the US for males and females come from the work of Wang *et al.* [18]. The number of cancer cases from 1975 to 2016 in the US was obtained from the National Cancer Institute (https://progressreport.cancer.gov/diagnosis/incidence). We extracted data about tobacco consumption in developed countries derived from the work of Ng *et al.* [19]. Note that the best fitting polynomial to predict and interpolate missing data was $Y = -0.003737 \times X^2 + 14.76056 \times X - 14551.75$, where X is the year and Y the tobacco consumption expressed in percentage of smokers.

Implementation of ANNs

Four MLFFNN ANNs, one for each type of cancer, were built. To predict future incidence of prostate cancer in the US, we considered the three main risk factors internationally recognized: demographic trends, life expectancy data and race/ethnicity [20–22]. Because the output also depends on race/ethnicity, we retrained this ANN with prostate cancer cases of White, Black, Hispanic, Asian/Pacific islander and American–Indian/Alaska native races. Our predictions did not include prostate cancer with positive family history due to the lack of a historical series in this setting. Data from 1992 to 2016 were used for training and validation, and data from 2017 to 2050 were predicted.

Age is a commonly known risk factor associated with breast cancer [23]. It is estimated that females who survive to 85 years will present a lifetime rate of approximately 11% of developing breast cancer [23]. Similarly, obesity is associated with an increased risk of both premenopausal [24] and postmenopausal breast cancer [25] and is also associated with worse cancer-related survival in all breast cancer subtypes. To predict the incidence of breast cancer in the US, our ANN was based on demographic trends, life expectancy data and (3) obesity. Data from 1992 to 2016 were used for training and validation and data from 2017 to 2050 were predicted. As for prostate cancer, here we considered total population as an input variable.

The risk of developing colorectal cancer increases with age, with more than 90% of cases occurring in patients aged 50 years or older [26]. It is estimated that 29.5% of colorectal cancers are attributable to a body mass index >22.5 [27], with an incidence that results different between men and women [28]. To predict the future incidence of colorectal cancer in the US, we included demographic trends, life expectancy data and obesity as inputs of the ANN algorithm. Data from 1975 to 2016 were used for training and validation and data from 2017 to 2050 were predicted.

The vast majority of lung cancers can be attributable to tobacco use, with more than 80% of lung tumors related to smoking [29]. As an indirect measure of exposure to pulmonary carcinogens (ionizing radiation, radon gas, etc.), age can be considered as a predictive factor of lung cancer in both smokers and never-smokers patients [30]. Indeed, the relationship between age and lung cancer in nonsmokers has been widely reported [31–33]. To predict the future incidence of lung cancer in the US, we considered demographic trends, life expectancy data and smoking prevalence as input variables. Data from 1975 to 2016 were used for training and validation and data from 2017 to 2050 were predicted.

As a forecasting model, we used multilayer perceptron, a layered feed-forward network that is trained by back propagation algorithm. Generally, a perceptron has several inputs and one output, and the function connecting inputs and output is nonlinear. A back propagation training algorithm adjusts the connection strength between

adjacent nodes. This method is easy to use, and it can model any kind of data, although for training it takes longer time than other methods, such as support vector machine (SVM), random forest (RF) and maximum likelihood (ML), and it requires large amounts of training data. The inputs were used without taking into account that they constituted time series to capture instantaneous relationships between inputs and output. Inputs were scaled by 'mapminmax' function of Matlab, to fall in between -1 and 1 to account for different variations and degrees of magnitude of input variables and achieve a better training. We used 70% of the data for training and 30% for validation, and these data sets were chosen randomly. Our MLFFNNs are constructed with three layers: the input layer has as many neurons as inputs and the output layer has one neuron. There were 10 neurons that implemented the 'tansig' function in the hidden layer.

To obtain the best scores, the network structure and its internal parameters, hidden neuron number and transfer functions, the learning rate and learning momentum were determined by many trials according to the trial-and-error method. Performance of each topology was assessed by the mean square error and the regression values. We sought to reach the best results with a minimum number of nodes to avoid that the ANN memorized data rather than learning them for generalization.

ANN algorithms can achieve excellent performance on the training data but poor performance on data used for validation or testing. This problem is called overfitting and it occurs when the algorithm takes false relationships among data that follows the noise and loses generality. We used early topping of the training process and cross-validation against overfitting. In Supplementary figures S6–S9 we show the performance of each network with regard to the training, validation and test sets.

We next implemented a Nonlinear AutoregRessive network with eXogenous inputs (NARX) algorithm, a recurrent dynamic network that should be more suitable in time-series modeling. In fact, it predicts future values of a time series (output) from past values of the same time series and past values of other time series (input). For example, in the case of colorectal cancer, the nonlinear function implemented by NARX algorithm could be

$$y(t) = f[y(t-1), \ldots, y(t-d),\ x(t-1), \ldots, x(t-d),\ z(t-1), \ldots, z(t-d),\ w(t-1), \ldots, w(t-d)]$$

where y(t) is the number of cancers, x(t) is the population number, z(t) is the life expectancy, w(t) is the obesity and d is the delay. Inputs and targets were scaled by the 'mapminmax' function, and the 'preparets' function was used to prepare the data with different numbers of delay. Delay from 1 to 4 were used. The NARX algorithm had 10 hidden neurons, and training was performed in an open loop; it was then transformed to closed loop for multistep-ahead prediction.

Software package Matlab R2016b (Mathworks Inc.) and its Neural Network Toolbox Version 9.1 were used in this study.

Results

Prostate cancer

The incidence of prostate cancer has decreased from an incidence of 200 cases per 100,000 people registered in the 1990s in the US to less than 150 per 100,000 in 2010 and, according to our predictions, will fall under 90 per 100,000 in 2025 (Figure 1A). As for almost all the different races/ethnicities, our algorithm predicts that the rapid decreasing trend starting in 2010 will continue until 2018–19 and then will slow down and reach a plateau after 2050 (Figure 1A). In white patients, the incidence of <200 per 100,000 cases in the 1990s, with a drop to 100 per 100,000 in the 2020s (Figure 1B). The incidence is similar in black patients (Figure 1C), who have had an incidence of <200 per 100,000 since 2012 and will reach a value of approximately 120 per 100,000 more than 10 years later (Figure 1C). Otherwise, patients with Asian/Pacific ethnicity as well as American–Indian and Alaska natives are characterized by a lower incidence (Figure 1D–F). Only for American–Indian/Alaska native races does the decreasing trend seem not to reach a plateau (Figure 1F).

The reduction in prostate cancer decline from 2018 could reflect reduced life expectancy and increasing population. The racial disparities are caused by behavioral differences and unequal access to high-quality health care, but this gap is declining.

Breast cancer

The incidence of breast cancer has decreased slightly since the 1990s (133 cases per 100,000 people), registering a drop to 125 per 100,000 in 2002 to 124 per 100,000 in 2015 (Figure 2). Based on our prediction algorithm,

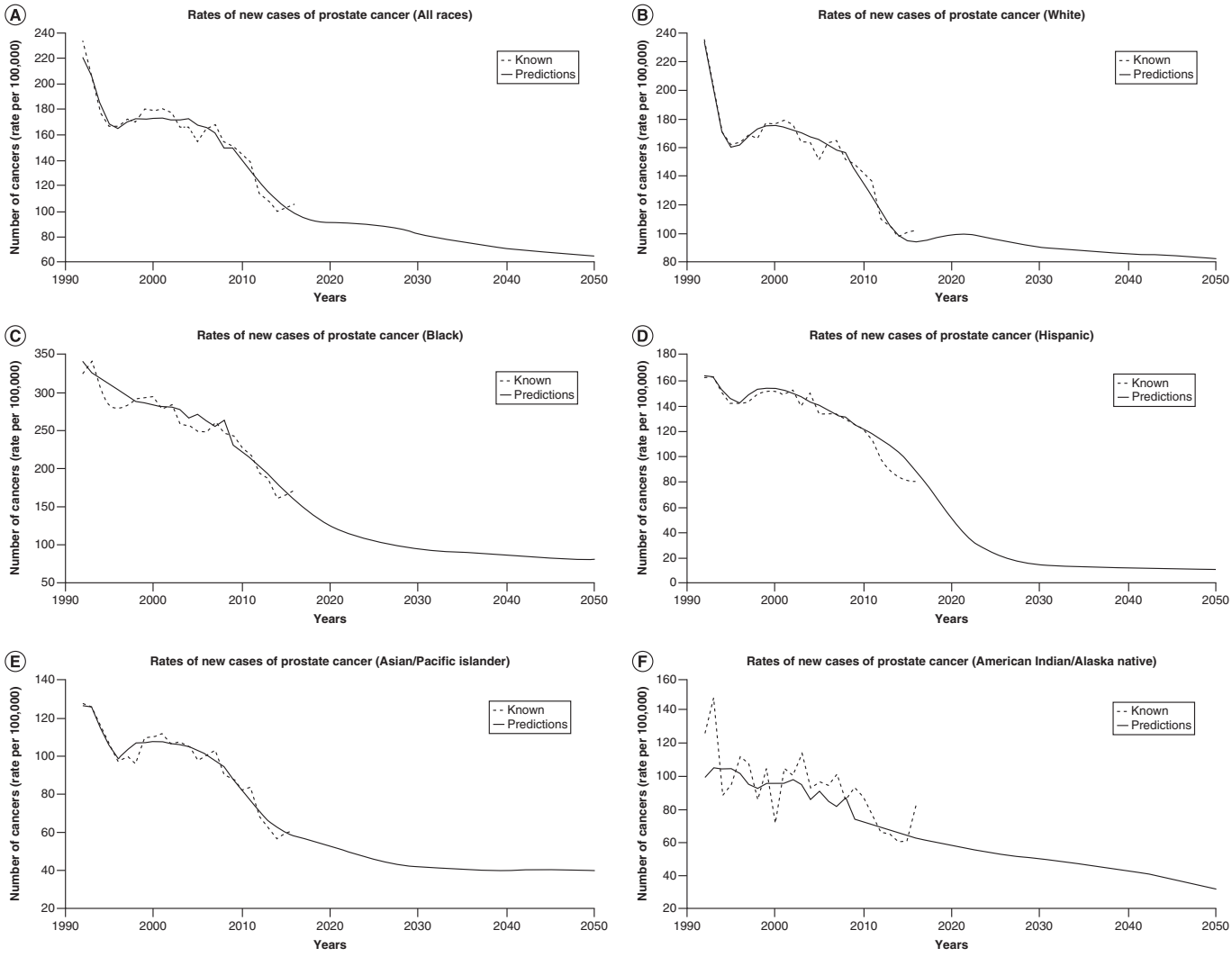


Figure 1. Trend and predicted new cases of prostate cancer overall (A) and by ethnicity: (B) white; (C) black; (D) Hispanic; (E) Asian/Pacific Islander; (F) American–Indian/Alaska Native.

the incidence will be almost constant; in particular, it will slowly increase up to approximately 128 per 100,000 in 2033, then slowly decrease (Figure 2). Performance of test phases (Supplementary Figure 2) showed that in this setting, the ANN algorithm obtained worse predictive results (test phases = 0.888) compared with prostate, colorectal, and lung cancer. The reasons for this difference is discussed in the next sections.

Colorectal cancer

The incidence of colorectal cancer has progressively increased since the 1970s (60 cases per 100,000 people), reaching its maximum in 1985 (66 per 100,000) (Figure 3A). From the late 1980s, a downward trend led to an incidence of 55 per 100,000 in 2000, with a progressive reduction to 35 per 100,000 in 2015 (Figure 3A). On the basis of our prediction algorithm, the incidence will reach a minimum value of 35 per 100,000 in 2030, followed by a plateau until 2050 (Figure 3A).

Due to the influence of gender on colorectal cancer incidence [33], we further predicted the incidence in males (Figure 3B) and females (Figure 3C). We found that the incidence was greatest for men in 1985 (79 per 100,000) and for women in 1985 (57 per 100,000), respectively (Figure 3B & C). Interestingly, we predicted that the reduction of the incidence trend will lead to a plateau of about 41 per 100,000 in 2030 in men, whereas women have a lower incidence, dropping to below 30 per 100,000 in 2050 (Figure 3B & C).

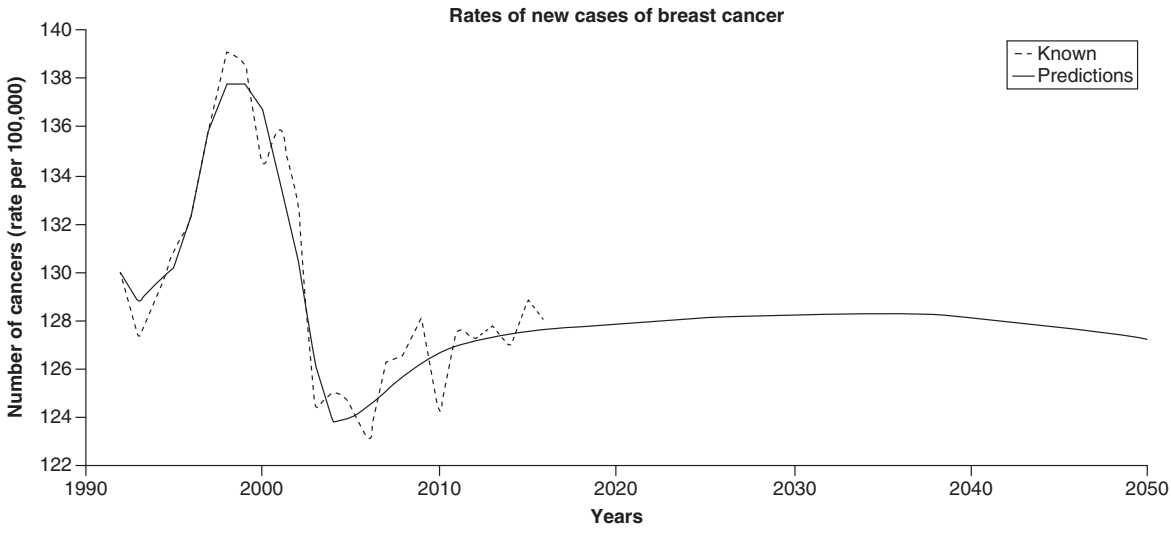


Figure 2. Trend and predicted new cases of breast cancer. Our calculations are based on population, life expectancy and obesity data for females.

Lung cancer

The incidence of lung cancer has progressively increased since 1970s (53 cases per 100,000 habitants), registering its maximum in 1992, characterized by 69 per 100,000 (Figure 4A). From 1990s, the downward trend in the smoking attitude has led to a gradual reduction of new cases of lung cancer, with an incidence in 2015 copmpared with 1970s (53 per 100,000) (Figure 4A). On the basis of this prediction algorithm, the incidence will decrease to 31 per 100,000 in 2030 and fall to 26 per 100,000 in 2050 (Figure 4A).

Due to the different time trends in tobacco consumption, we further predicted the incidence in males (Figure 4B) and females (Figure 4C). Whereas the highest incidence was registered in 1984 for males (102 per 100,000), the highest value (54 per 100,000) for women was reported in 2005 due to the rapid increase of smoking among women about 20 years later than men (Figure 4B & C). Interestingly, the drop of the incidence trend appears slower in males than in females, reaching a plateau beyond 2050 (32/100.000), 15 years after the plateau predicted for females (27/100.000) (Figure 4B & C).

Regression Analysis

The ANN outputs with respect to targets for training and validation sets are shown in scatter plots (Supplementary Figures 1–4). The dashed line represents the perfect result, that is, outputs = targets. The solid line depicts linear best fit between the outputs and the targets. The R-value summarizes the relationship between the outputs and targets. In particular, if R = 1 there is an exact linear relationship between outputs and targets. If R is close to zero, then no linear relationship links outputs to targets. In most of our models, apart for breast cancer and prostate cancer in American–Indian/Alaska Native Races, the R-values approximating $\geq$0.9 indicate a reasonably good fit for a data set. Analogously, a very small value of mean square error suggests the goodness of the models.

NARX algorithm results

The results of NARX algorithm were very similar to those of the previous MLFFNN algorithm and are shown in the Supplementary Figures 10–13.

Discussion

The lifetime risk of developing cancer is related, in general, to a longer life expectancy. However, a range of influences, from environmental and lifestyle (e.g., tobacco use) changes to prevention campaigns, screening programs and innovation technologies, should be taken into account to increase the accuracy of predicting models for cancer incidence. Our forecasts go until 2050 because beyond that date they would be unreliable.

Among the four most frequent tumor types, we showed a general decline in the incidences in the United States. The changes in prostate cancer incidence observed in the late 1980s and early 1990s (Figure 1A) were probably due

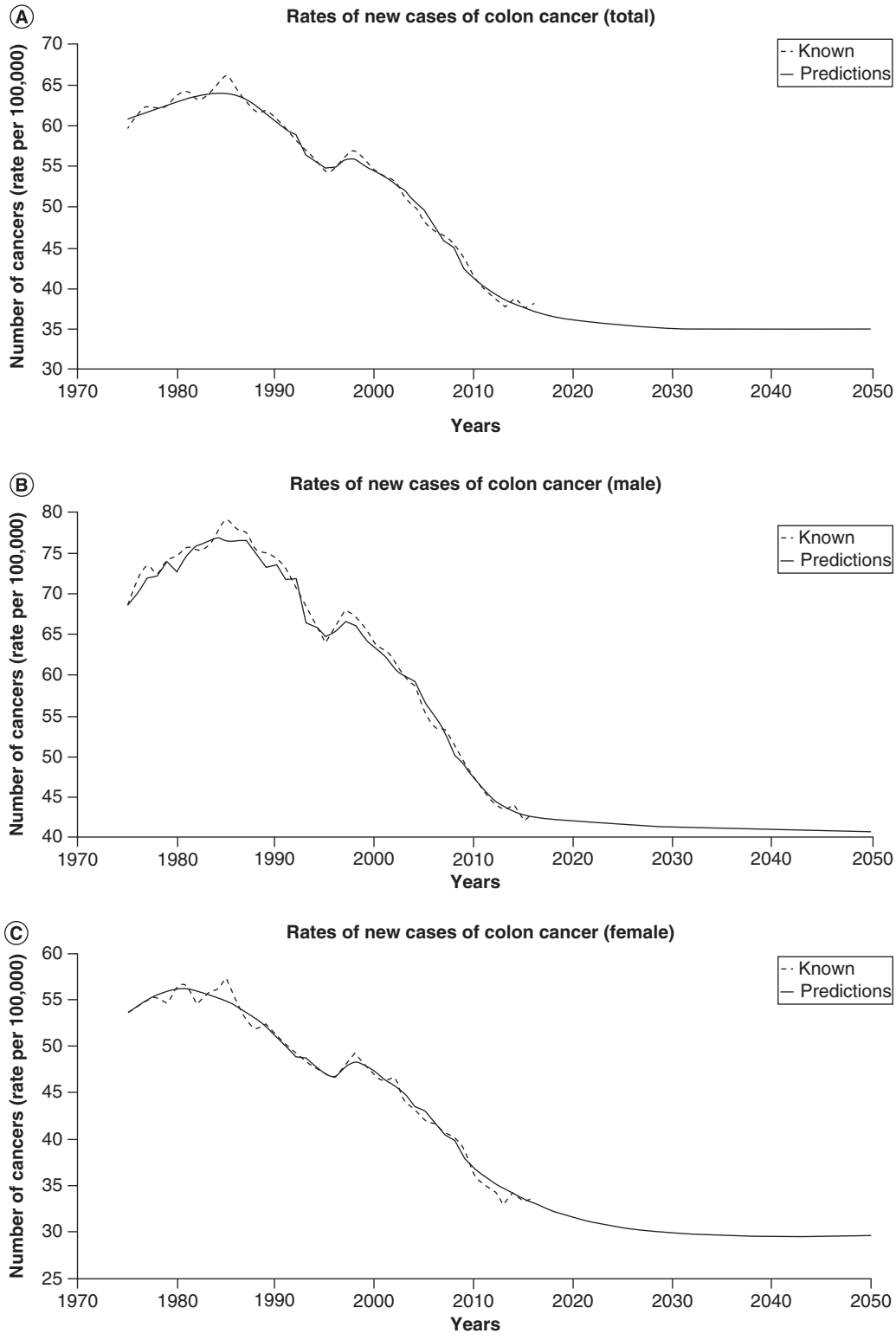


Figure 3. Trend and predicted new cases of colon cancer overall (A) and by gender (B) males; (C) females. Our calculations are based on population, life expectancy and obesity for males and females.

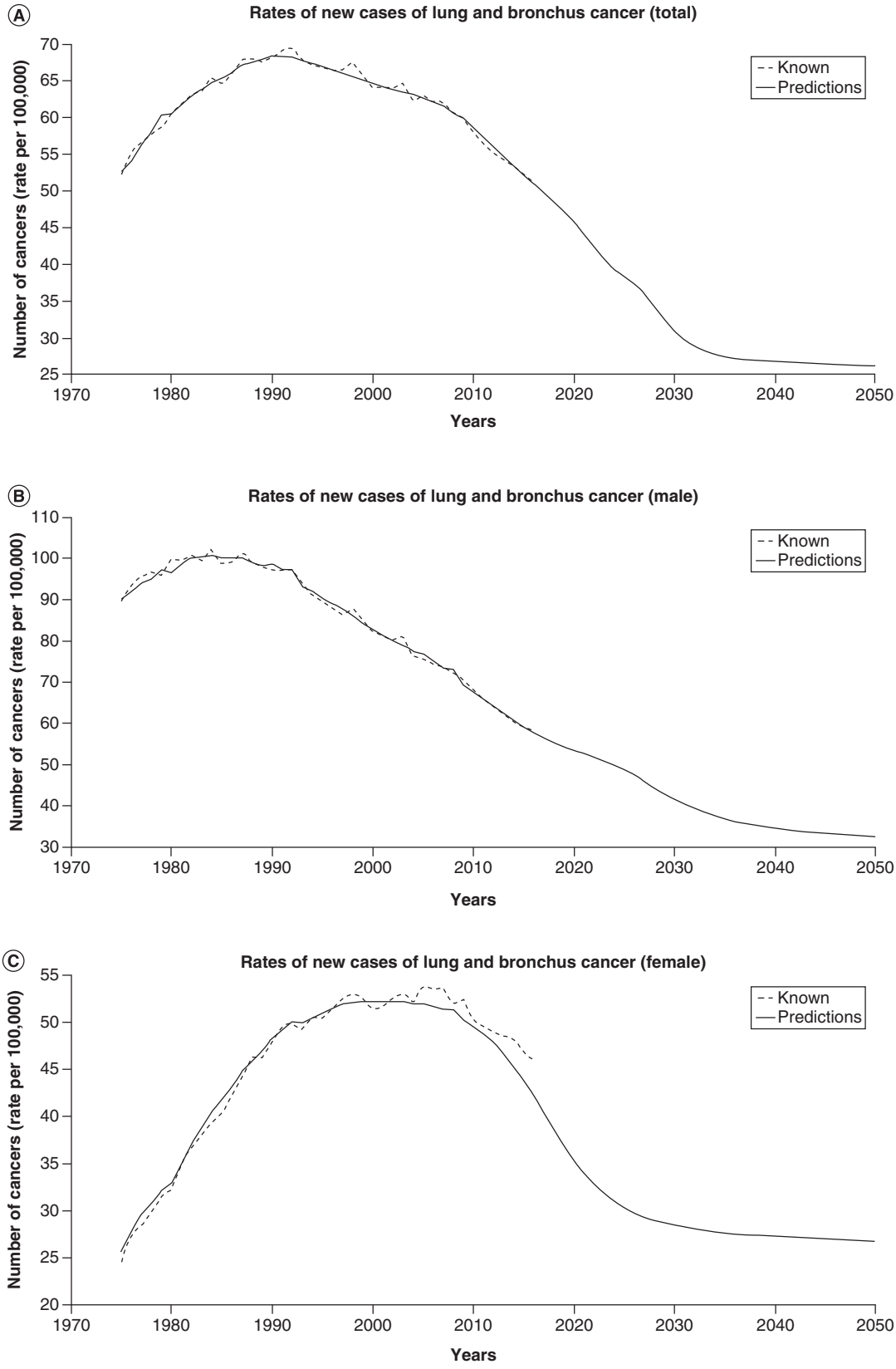


Figure 4. Trend and predicted new cases of lung cancer overall (A) and by gender (B) males; (C) females. Our calculations are based on population, life expectancy and smoking data for males and females.

10.2217/fon-2020-0359

to the introduction of widespread prostate-specific antigen (PSA) testing that allowed the detection of asymptomatic disease [34]. The reduction in the incidence of prostate cancer from 2010 to 2013 can be attributed to decreased PSA testing. In fact, the US Preventive Services Task Force recommended the use of PSA as a screening method for prostate cancer. The task force, basing on data from Prostate, Lung, Colorectal and Ovary cancer screening study (PLCO) and the European Randomized Study of Screening for Prostate Cancer (ERSPC) trial, determined that the potential harms of testing and associated treatment (erectile dysfunction, incontinence and serious surgical complications) outweighed the benefits (PSA screening reduced cancer-related mortality by 4 men for every 1000 men, after 14 years of follow-up) [35].

In breast cancer, the ANN did not reveal promising performances. The small variability of incidence data from 1990 to 2050, together with the high multifactorial presentation of this tumor, may partially explain the poorer performance of the ANN in this disease (see the regression curves in Supplementary Figure 3). This evidence suggests that the number of new cases does not strictly follow the trends registered for age and obesity in the US (i.e., the peak from 1995 to 2002 in cancer incidence in a time-interval characterized by the reduction of both risk factors) and underlines the necessity of identifying more effective input variables beyond the most commonly recognized risk factors.

As for colorectal cancer, the drop in incidence rates before 2000 can be explained by the changes in risk factors and the introduction of screening (fecal occult blood testing and endoscopy) [36]. The prevalence results are clearly different between men and women, due to underlying mechanisms that include estrogen exposure, menopausal status, insulin resistance, chronic inflammation and steroid hormones [37,38].

Lung cancer is among the most deadly cancers in both men and women [39]. Reducing its incidence represents a major goal for cancer researchers, and our findings reenforce worldwide prevention campaigns to decrease tobacco use (Figure 4A). Progressive reduction will ideally be more rapid as the effects of global action toward the 2040 tobacco-free world goal become evident [40]. As for gender differences (Figure 4B & C), they reflect the historical changes in tobacco use, with more women smoking and at older ages than men in recent decades.

Our study presents several limitations. As with other prediction systems, ANN algorithms are affected by errors and biases compared with real-world data. However, ANNs provide for training, performance and validation phases that may help to reduce system biases and increase the accuracy of predictions.

Conclusions

This up-to-date prediction of cancer burden in the United States could be a crucial resource for planning and evaluation of cancer-control programs. Urgent global actions toward a dramatic reduction of cancer-related risk factors are necessary to reduce the incidence of cancer.

Summary points

- Predictions of cancer incidence rates may be useful to optimize allocation of finite resources, determine the key elements of future cancer control and allow for planning of cancer control programs.
- Artificial neural network (ANN) algorithms such as the classical multilayer feed-forward network (MLFFNN) and the nonlinear autoregressive network with exogenous input (NARX) can be used to estimate the incidences of cancer and to predict the number of new cases in the US through 2050.
- To predict future incidence of the four most common tumors (breast, colorectal, lung and prostate) in the US, we considered the main risk factors as input variables: demographic trends, life-expectancy data and race/ethnicity for prostate cancer; obesity for breast and colon cancer; and smoking prevalence for lung cancer.
- A rapid decreasing trend of prostate cancer incidence was seen in 2010 and continued into 2018-19; it will then slow and reach a plateau after 2050, with several differences among races and ethnicities.
- The incidence of breast cancer will be almost constant; in particular, it will slowly increase to about 128 per 100,000 in 2033, then slowly decrease.
- For colorectal cancer, the incidence will reach a minimum value of 35 per 100,000 in 2030, followed by a plateau through 2050.
- For lung cancer, the incidence will decrease to 31 per 100,000 in 2030 and fall to 26 per 100,000 in 2050, with a trend that appears slower in males than females.
- To increase the accuracy of predicting models for cancer incidence, a range of influences, from environmental to lifestyle changes to prevention campaigns, screening programs and innovation technologies, should be taken into account.

Supplementary data

To view the supplementary data that accompany this paper please visit the journal website at: www.futuremedicine.com/doi/suppl/10.2217/fon-2020-0359

Author contributions

Conceptualization: M Santoni and F Tartari; methodology: F Piva, M Giulietta, MM Aiello; data curation: F Piva, M Santoni, R Mazzucchelli, R Cerqueti; writing (original draft preparation): F Tartari, M Santoni, F Piva, N Battelli; writing (final editing): A Cimadamore, L Cheng, A Lopez-Betran, R Montironi.

Financial & competing interests disclosure

The authors have no relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript. This includes employment, consultancies, honoraria, stock ownership or options, expert testimony, grants or patents received or pending, or royalties.

No writing assistance was utilized in the production of this manuscript.

References

Papers of special note have been highlighted as: • of interest; •• of considerable interest

1. American Cancer Society. Cancer Facts & Figures 2020. https://www.cancer.org/research/cancer-facts-statistics/all-cancer-facts-figures/cancer-facts-figures-2020.html (Accessed 10 August 2020).
2. Tartari F, Santoni M, Burattini L, Mazzanti P *et al.* Economic sustainability of anti-PD-1 agents nivolumab and pembrolizumab in cancer patients: Recent insights and future challenge. *Cancer Treat. Rev.* 48, 20–24 (2016).
3. Montironi R, Santoni M, Tartari F *et al.* Testing PD-1/PD-L1 expression in cancer therapy: pathologic insights and economic sustainability. *Arch. Pathol. Lab. Med.* 140(6), 501–502 (2016).
4. Tartari F, Santoni M, Pistelli M, Berardi R. Healthcare cost of HER2-positive and negative breast tumors in the United States (2012–2035). *Cancer Treat. Rev.* 60, 12–17 (2017).
5. Tartari F, Conti A, Cerqueti R. Assessing the relationship between toxicity and economic cost of oncological target agents: a systematic review of clinical trials. *PLoS One.* 12(8), e0183639 (2017).

•• **This paper deals with the relationship between economic costs and toxicity of target agents. The analysis of the toxicity is based on the severe adverse events and discontinuation rates of each target agent.**

6. Møller B, Fekjaer H, Hakulinen T *et al.* Prediction of cancer incidence in the Nordic countries up to the year 2020. *Eur. J. Cancer Prev.* 11(Suppl.), S1–S96 (2002).
7. Teppo L, Hakulinen T, Saxen E. The prediction of cancer incidence in Finland for the year 1980 by means of cancer registry material. *Ann. Clin. Res.* 6, 122–125 (1974).
8. Hakulinen T, Teppo L, Saxe E. Do the predictions for cancer incidence come true? Experience from Finland. *Cancer* 57, 2454–2458 (1986).
9. Hakulinen T, Dyba T. Precision of incidence predictions based on poisson distributed observations. *Statist. Med.* 13, 1513–1523 (1994).
10. Dyba T, Hakulinen T, Paivarinta L. A simple non-linear model in incidence prediction. *Statist. Med.* 16, 2297–2309 (1997).
11. Engeland A, Haldorsen T, Tretli S *et al.* Prediction of cancer incidence in the Nordic countries up to the years 2000 and 2010. A collaborative study of the five Nordic Cancer Registries. *APMIS Suppl.* 38, 1–124 (1993).
12. Coory M, Armstrong BK. Cancer Incidence Projections for Area and Rural Health Services in New South Wales. NSW Cancer Council, Sydney (1998). https://trove.nla.gov.au/version/45739318
13. Bray F, Møller B. Predicting the future burden of cancer. *Nat. Rev. Cancer.* 6(1), 63–74 (2006).

• **In this review, the authors illustrate the science of making 'good predictions' of cancer burden, how predictions are made, what predictions can tell us and the challenges that remain; they further discuss the practical interpretations of this approach, the drawbacks that motivate more complex modeling and the use of data collected for specific purposes.**

14. Malik F, Nasereddin M. Forecasting output using oil prices: a cascaded artificial neural network approach. *J. Econom. Business.* 58(2), 168–180 (2006).
15. Kim D. Neural networks theory and applications. Jinhan M&B, Seoul (2005).
16. Cooper JCB. Artificial neural networks versus multivariate statistics: An application from economics. *J. Appl. Statist.* 26(8), 909–921 (1999).
17. Zhang G, Patuwo EB, Hu MY. Forecasting with artificial neural networks: The state of the art. *Int. J. Forecast.* 14(1), 35–62 (1998).

- **This paper presents a state-of-the-art survey of ANN applications in forecasting and provide a synthesis of published research in this area, insights on ANN modeling issues, and the future research directions.**

18. Wang Y, Beydoun MA, Liang L, Caballero B, Kumanyika SK. Will all Americans become overweight or obese? Estimating the progression and cost of the US obesity epidemic. *Obesity (Silver Spring).* 16, 2323–2330 (2008).
19. Ng M, Freeman MK, Fleming TD *et al.* Smoking prevalence and cigarette consumption in 187 countries, 1980–2012. *JAMA.* 311, 183–192 (2014).
20. Gann PH. Risk factors for prostate cancer. *Rev Urol.* 4(Suppl. 5), S3–S10 (2002).
21. Leitzmann MF, Rohrmann S. Risk factors for the onset of prostatic cancer: age, location, and behavioral correlates. *Clin. Epidemiol.* 4, 1–11 (2012).
22. Kheirandish P, Chinegwundoh F. Ethnic differences in prostate cancer. *Br J Cancer.* 105(4), 481–485 (2011).
23. Singletary SE. Rating the Risk Factors for Breast Cancer. *Ann Surg.* 237(4), 474–482 (2003).
24. Laudisio D, Muscogiuri G, Barrea L, Savastano S, Colao A. Obesity and breast cancer in premenopausal women: current evidence and future perspectives. *Eur. J. Obstet. Gynecol. Reprod. Biol.* 230, 217–221 (2018).
25. Argolo DF, Hudis CA, Iyengar NM. The impact of obesity on breast cancer. *Curr. Oncol. Rep.* 20(6), 47 (2018).
26. Amersi F, Agustin M, Ko CY. Colorectal cancer: epidemiology, risk factors, and health services. *Clin. Colon. Rectal Surg.* 18(3), 133–140 (2005).
27. Larsson SC, Bergkvist L, Wolk A. High-fat dairy food and conjugated linoleic acid intakes in relation to colorectal cancer incidence in the Swedish Mammography Cohort. *Am. J. Clin. Nutr.* 82, 894–900 (2005).
28. Kim SE, Shim KN, Jung SA *et al.* An association between obesity and the prevalence of colonic adenoma according to age and gender. *J. Gastroenterol.* 42, 616–623 (2007).
29. Centers for Disease Control and Prevention. Annual smoking-attributable mortality, years of potential life lost, and productivity losses – United States 1997–2001. *MMWR Morb. Mortal. Wkly. Rep.* 54(25), 625–628 (2005).
30. Doll R. The age distribution of cancer: implications for models of carcinogenesis. *J. Royal Statisti. Soc. A.* 134, 133–155 (1971).
31. Doll R, Peto R. Cigarette smoking and bronchial carcinoma: dose and time relationships among regular smokers and lifelong non-smokers. *J. Epidemiol. Commun. Health.* 32(4), 303–313 (1978).
32. Meza R, Hazelton WD, Colditz GA, Moolgavkar SH. Analysis of lung cancer incidence in the Nurses' Health and the Health Professionals' Follow-Up Studies using a multistage carcinogenesis model. *Cancer Causes Control.* 19(3), 317–328 (2008).
33. Thun MJ, Henley SJ, Burns D *et al.* Lung cancer death rates in lifelong nonsmokers. *J. Natl. Cancer Inst.* 98(10), 691–699 (2006).
34. Potosky AL, Miller BA, Albertsen PC, Kramer BS. The role of increasing detection in the rising incidence of prostate cancer. *JAMA.* 273, 548–552 (1995).
35. Moyer VA. U.S. Preventive Services Task Force recommendation statement. *Ann. Intern. Med.* 157, 120–134 (2012).
36. Edwards BK, Ward E, Kohler BA *et al.* Annual report to the nation on the status of cancer, 1975–2006, featuring colorectal cancer trends and impact of interventions (risk factors, screening, and treatment) to reduce future rates. *Cancer.* 116, 544–573 (2010).

•• **This article reports the use of microsimulation modeling as a tool for interpreting past trends and projecting future trends to assist in cancer control planning and policy decisions.**

37. Slattery ML, Ballard-Barbash R, Edwards S *et al.* Body mass index and colon cancer: an evaluation of the modifying effects of estrogen (United States). *Cancer Causes Control.* 14(1), 75–84 (2003).
38. Jochem C, Leitzmann M. Obesity and Colorectal Cancer. *Recent Results Cancer Res.* 208, 17–41 (2016).
39. Siegel R, Ma J, Zou Z *et al.* Cancer statistics, 2014. *CA Cancer J. Clin.* 64, 9–29 (2014).
40. Beaglehole R, Bonita R, Yach D *et al.* A tobacco-free world: a call to action to phase out the sale of tobacco products by 2040. *Lancet.* 385(9972), 1011–1018 (2015).