Exposure and Health (2026) 18:38
https://doi.org/10.1007/s12403-026-00774-z

ORIGINAL PAPER

Estimating Cancer Mortality Risk Associated with Arsenic Concentration in Topsoil

George Tsirogiannis[1] · Michalis Leotsinidis[2] · Eleni Sazakli[2]

Received: 31 October 2025 / Revised: 18 March 2026 / Accepted: 22 March 2026

Abstract

One major pathway of human exposure to arsenic (As) is through direct or indirect contact with contaminated topsoil. This study utilizes arsenic concentration data measured in the topsoil of rural regions in Greece to identify critical threshold levels associated with significant increases in cancer mortality rates. Topsoil element data were obtained from the GEMAS (Geochemical Mapping of Agricultural and Grazing Land Soil) database. Age-stratified cancer mortality rates were extracted for 60 small municipalities and communities with populations under 20,000, covering a 15-year period (2000–2015). Poisson and nonlinear growth models were employed to fit and predict mortality across all age groups using 53 different topsoil elements. Statistical analyses of model performance, uncertainty quantification, and sensitivity analysis reveal a significant association between topsoil As concentrations and cancer mortality among younger age groups (0–29 and 30–39 years). Specifically, As concentration of 12.7 mg kg^{-1} is associated with a 50% increase relative to median mortality, while 19.5 mg kg^{-1} corresponds to a doubling of cancer mortality rate in the 0–29 age group. For the 30–39 age group, 18.2 mg kg^{-1} is related to a 50% increase, and 29.2 mg kg^{-1} to a 100% increase in cancer mortality. The findings underscore the public health significance of As concentration in topsoil and highlight the need for targeted environmental monitoring and preventive interventions.

Keywords Arsenic · Topsoil · Environmental epidemiology · Geochemical association · Cancer mortality · Health risk

Introduction

Arsenic (As) is a pervasive metalloid and potent toxicant classified among the ten most hazardous substances to human health and designated as a Group I carcinogen by the World Health Organization (IARC 2012; WHO 2022). Natural geological processes such as weathering of As-bearing minerals in combination with anthropogenic activities, including mining, industrial emissions and intensive groundwater extraction, contribute to its widespread distribution in soils and aquifers (Mandal and Suzuki 2002; Masuda 2018; Poghosyan et al. 2025). Concentrations of As vary across lithological matrices (Bradl 2005), while its environmental mobility and bioavailability are dictated by soil physicochemical properties (Huang et al. 2024).

Mounting epidemiological evidence implicates As as a potential etiological factor in various malignancies, including lung, bladder, skin and digestive tract (Cogliano et al. 2011; Naujokas et al. 2013). The mechanisms underlying its carcinogenicity involve epigenetic dysregulation, oxidative damage and disruption of DNA repair pathways (Hughes 2002; Rossman 2003).

Human exposure to As primarily occurs through the consumption of contaminated water, food, and soil (EFSA 2024; Spaur et al. 2024; Hou et al. 2025). Recently, the European Food Safety Authority (EFSA) has updated its risk assessment for inorganic As in food, recommending a reference point of 0.06 µg kg^{-1} body weight per day to minimize both cancer (skin, lung, and bladder), and non-cancer health effects (EFSA 2024). Given the epidemiological evidence of low-dose effects, which has led the EPA to set the Maximum Contaminant Level Goal (MCLG) for As in drinking water at zero, and the ongoing trend toward biologically safe thresholds, the importance of controlling

✉ Eleni Sazakli
elsazak@upatras.gr

[1] Department of Food Science and Technology, University of Patras, 26504 Patras, Greece

[2] Lab of Public Health, Epidemiology and Quality of Life, Medical School, University of Patras, 26504 Patras, Greece

Published online: 21 April 2026

human exposure to As is underscored (Frisbie and Mitchell 2022). Yet, little is known about what constitutes "biologically safe" As levels in topsoil, or how soil As that enters the food chain may influence cancer risk. Spatial epidemiologic analyses in Spain have demonstrated geographical correlations between As topsoil concentration and elevated cancer mortality (Núñez et al. 2016). In our recent work (Katsifara et al. 2025), we found significant association between six toxic elements in topsoil (i.e., Mn, Ni, Pb, Be, As, and Cd) and elevated observed cancer mortality in Greece. Specifically, we isolated the extremes of cancer mortality distribution across all age categories, i.e. the lower and the upper quartile, and by employing ensemble tree-based machine learning models, we demonstrated that the two groups were statistically distinguishable based solely on their topsoil elemental composition. Motivated by these findings, the current study attempts to quantify the risk of cancer mortality attributed to topsoil elemental concentrations, by adopting a regression modeling approach, where cancer mortality is expressed as event counting Poisson-based and nonlinear growth models.

Materials and Methods

Study Area and Population

This study targets small municipalities and communities with populations under 20,000 that collectively account for nearly half of Greece's population. We deliberately excluded larger municipalities to eliminate the confounding urban-life risk factors, like air and noise pollution or lifestyle variability. Our focus is on regions where people live and work in close connection with their land, consuming local food, engaging in agriculture and facing direct exposure to the elements present in the topsoil. Another potential source of human exposure to toxic elements is the consumption of fish and seafood - key components of the Greek diet. However, there is no evidence indicating that individuals living in coastal areas of Greece consume greater quantities of these foods than those residing in inland regions (Kotopoulou et al. 2025). Therefore, it is unlikely that geographical differences in place of residence would result in differences in exposure to these elements through fish consumption in our study population, especially given that urban areas were excluded.

Data of topsoil element concentrations were obtained from the GEMAS (Geochemical Mapping of Agricultural and Grazing land Soil) database (Reimann et al. 2014; GEMAS 2021). The GEMAS project is a collaborative initiative between the Geochemistry Expert Group of EuroGeoSurveys and Eurometaux that provides fully harmonized, high-quality geochemical data on element concentrations and bioavailability across the European continent. The sampling conducted in 2008 and employed a regular grid covering the area of whole Europe, resulting in the analysis of 53 chemical elements in 2,211 agricultural soil samples (0–20 cm) and 2,118 grazing land soil samples (0–10 cm), with an average sampling density of one site per 2,500 km^2. These samples were collected under a strict and uniform field protocol explicitly designed to minimize anthropogenic influence and contamination. Specifically, sample sites were located at least: (a) 2,000 m from known contaminated areas, industrial facilities, or power plants, (b) 200 m from major transport infrastructure such as highways and railways, and (c) 100 m from high-voltage power lines or pylons. Regarding the laboratory/chemical analysis of the soil samples is concisely outlined as follows (Reimann et al. 2014): samples underwent air-drying, followed by sieving to a particle size of <2 mm using a nylon mesh. The resulting material was homogenized and subsequently subjected to aqua regia digestion. Quantitative determination of 53 chemical elements was then performed using inductively coupled plasma mass spectrometry (ICP-MS) (Reimann et al. 2014; GEMAS 2021).

In this study, 60 rural municipalities and communities, with populations under 20,000 and containing the corresponding GEMAS sampling points were selected, representing 9.5% of the total rural population of Greece. Across the study period, the total number of person-years is 7,260,757. Detailed municipal population data, stratified by ten-year age groups, and structured according to the "Kapodistrias" administrative reform, were obtained from the Hellenic Statistical Authority (ELSTAT) for each municipality and for every year of the study period. Observed municipal level mortality data, disaggregated by age, were also retrieved directly from ELSTAT for the years 2000–2015. These data correspond specifically to deaths attributed to neoplasms (i.e. 18,360), excluding non-melanoma skin cancer, and were obtained through a custom request tailored to the requirements of this investigation. To improve statistical robustness due to the low frequency of cancer mortality especially in younger ages, the first three age groups (i.e. 0–9, 10–19, and 20–29 years) were aggregated into a single stratum (i.e. 0–29 years age group).

Figure 1 presents the locations of the GEMAS soil samples. Using a Geographic Information System (GIS) platform, each GEMAS sampling point was geospatially linked to its corresponding municipality or community. Demographic and cancer mortality data were then retrieved for those specific regions, and cancer mortality rates were subsequently calculated.

Regarding the population at risk and the associated mortality for the first two age groups (i.e., 0–29 and 30–39

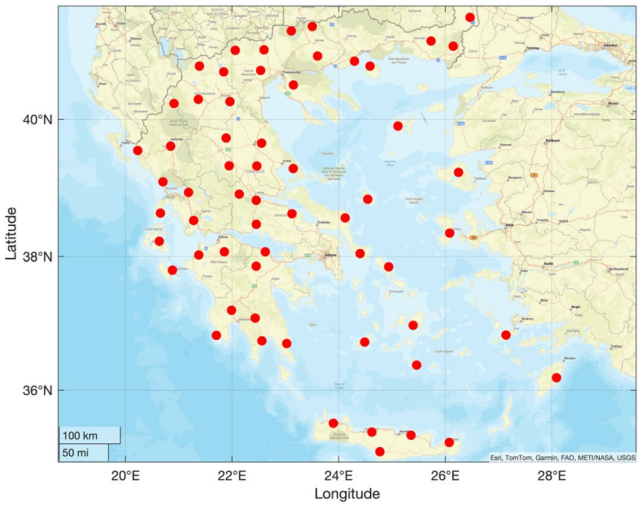

Fig. 1 Spatial distribution of soil sampling points in Greece

years), which are the focus of this study, the results are shown in Fig. 2. These visualizations indicate that small population sizes constitute the majority of observations, and a similar pattern holds for the number of deaths (grey zones in Fig. 2). At higher values of the population at risk, a nearly linear relationship is observed (Fig. 2), accompanied by relatively high R^2 values. The probability density function of the dominant chemical element, namely As, is shown in the Appendix (Fig. 9).

Modeling

Our objective is to combine topsoil element concentrations with cancer mortality rate by age group, using fully explainable statistical predictive models. That is, models whose mathematical formulation is a single formula, with parameters that are interpretable in the context of the epidemiological problem under study.

A wide variety of mathematical models have been proposed and employed for scientific purposes in general (e.g. Eck et al. 2017) and in epidemiology in particular (e.g. Brauer et al. 2019). In this study, we opt for models that are: (i) interpretable by epidemiologists, and (ii) frugal in terms of parameterization, making them well-suited to the limited available data and inherently resistant to overfitting, i.e. they lack the capacity to memorize data by design. In data driven modeling, i.e. where the data define the model parameters (i.e., model fit by using statistical methods), the structure of the chosen model family is a critical factor. In this study we propose models that are consistent with the following assumptions: (a) health effects become increasingly harmful as exposure to As increases (rational: higher concentrations of As elevate biological exposure and increase the likelihood of cellular damage and systematic toxicity), and (b) there is an upper limit to the observed cancer mortality rate. Otherwise, the area would likely be uninhabitable, i.e., historically people would not continue to reside in this rural region.

A hierarchy of family models will be used for this study. We begin with simpler Poisson mortality models, in which

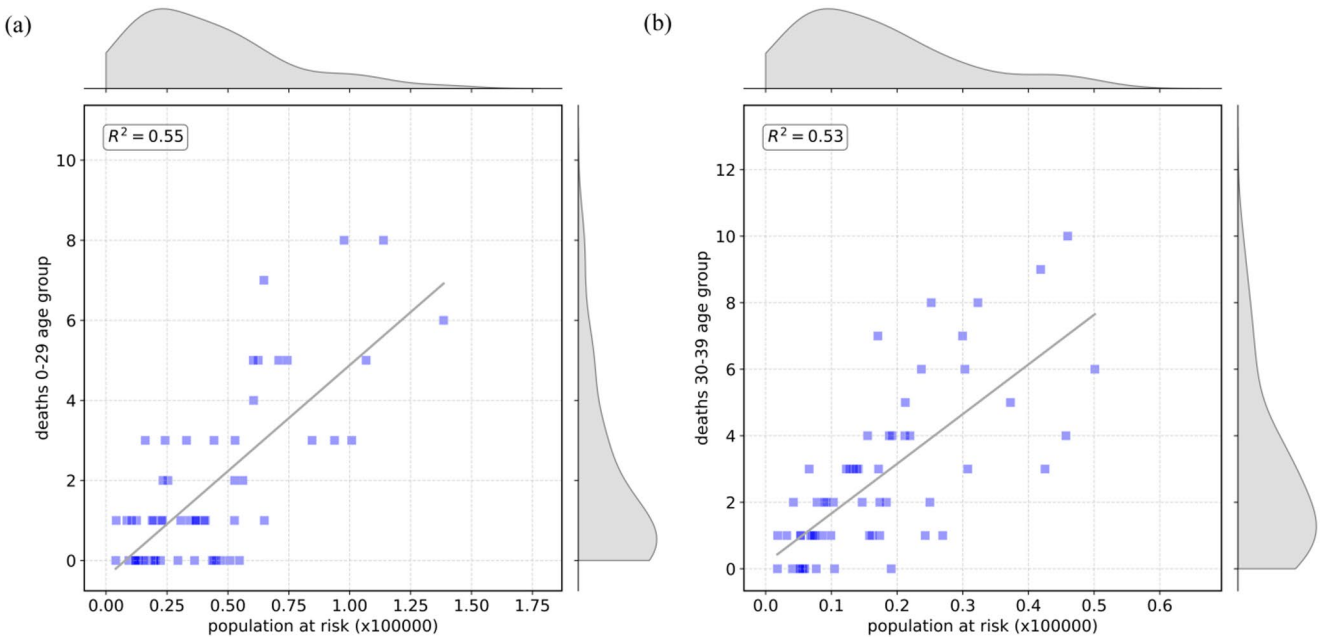

Fig. 2 Association of population at risk and observed cancer mortality in 0–29 (**a**) and 30–39 (**b**) age groups. The grey zones depict the corresponding probability density functions of each axis

the independent variables correspond to the concentrations of individual chemical elements. Model selection across elements will be conducted using information-based criteria (i.e., the Akaike Information Criterion, AIC), and model adequacy will be assessed using the index of dispersion (Cox and Lewis 1966). This approach will provide insight into the relative importance of the most influential chemical elements.

Based on the chemical elements selected by the simpler models, we extend the complexity of the Poison model to incorporate the two assumptions described above. Specifically, we employ Gompertz growth models (Tjørve and Tjørve 2010) to describe the mean response of the Poisson model. This more complex formulation provides realistic growth curves that satisfy the stated growth assumptions, while simultaneously adhering to the epidemiological framework by explicitly accounting for the number of deaths and the population at risk.

Upon constructing the quasi-Poisson-Gompertz models for the different age groups, we attempt to reduce/simplify them to an ordinary least squares (OLS) framework. The primary objective of this simplification is to evaluate their predictive performance in terms of explained variance and to quantify model uncertainty. To this end, we propose Gompertz growth models fitted to As concentrations and estimated mortality rates, rather than explicitly modelling the number of deaths and the population at risk.

Proposed Model Families

Poisson Model

We assume that the number of deaths (y_i) of a rural region follows the Poisson distribution, $y_i \sim Poisson\,(\mu_i)$, where μ_i is the mean value of deaths (Andersen et al. 2012; McCullagh and Nelder 1989). This mean value is given by the multiplication of the population at risk (P_i) with the baseline mortality and the relative risk, i.e. $\mu_i = P_i \exp(\beta_o + \beta_1 x_i)$, where x_i is the concentration of the element in region i.

Poisson-Gompertz Model

An extended form of the above model, which incorporates by design the assumptions of Sect. "Modeling", can be obtained by modifying the mean value function. The Richards family of models constitute growth models of sigmoid shape, that fulfil the above two assumptions, and include models such as the negative exponential, the logistic, the Bertalanffy and the Gompertz among others. In the meticulously crafted study of Tjørve and Tjørve (2010), a unified approach to the members of the Richards family was presented, identifying two distinct forms of interest among the multitude of Richards model-versions proposed in previous studies. One of the advantages of the Unified-Richards models use, is that they have a set of parameters which are comparable across models in the family, without conversion equations. Moreover, the proposed taxonomy of Tjørve and Tjørve (2010), brings clarity to the alternative formulations and their suitability for the purposes of a given study. Here we focus on the X_i form, i.e. chemical element concentration at maximum growth, rather the alternative W_0 form, which expresses explicitly the starting value. The reference Unified-Richards model of four parameters has the following form:

$$f(x) = A\left(1 + (d-1)\ exp\left(\frac{-K_u\,(x - X_i)}{d^{d/(1-d)}}\right)\right)^{1/(1-d)} \tag{1}$$

where, x the chemical element concentration, A is the upper asymptote (i.e. maximum cancer mortality), K_u is the slope of inflection (maximum relative growth rate), X_i is chemical element concentration that the inflection occurs, and $d^{1/(1-d)}$ is the relative value of maximum growth.

Due to the limited number of rural regions in this study (i.e. 60, primarily determined by the GEMAS geospatial sampling grid) and our requirement/intention for low model complexity to ensure optimal interpretability and minimize overfitting, we propose the use of the following simplification of model (1), the Unified-Gompertz model introduced by Tjørve and Tjørve (2017):

$$f(x) = A \exp\left(-\exp\left(-e\ K_u\ (x - X_i)\right)\right) \tag{2}$$

where the parameters and variables are the same as in model (1).

Model (2) is the minimal model that satisfies the two assumptions above, while retaining most of the properties of the reference model (1).

Thus, the proposed Poisson-Gompertz takes the following form:

$$y_i \sim Poisson\,(\mu_i)$$

$$\mu_i = P_i\, A \exp\left(-\exp\left(-e\ K_u\ (x_i - X_i)\right)\right) \tag{3}$$

Simplified Gompertz Growth Model

A reduced/simplified version of model (3) can be obtained by modeling directly the mortality rates. Given the similar population-at-risk sizes and numbers of deaths across the rural regions considered in this study as visualized in Fig. 2, a Gompertz growth model is obtained directly from Eq. (2):

$$mr_i = A \exp\left(-\exp\left(-e\ K_u\,(x_i - X_i)\right)\right) \tag{4}$$

where $mr_i = (number\ of\ deaths)\ /\ (population\ at\ risk)$ is the mortality rate of region i.

Model Selection and Comparison

For each combination of age group and chemical element concentration, we estimate the parameters of the Poisson models. For each age group, we retain only those models that do not significantly deviate from the Poisson assumption, as assessed by their index of dispersion (i.e. a variance-to-mean ratio within 1 ± 0.2). The retained models are then ranked according to their AIC (Aho et al. 2014). In this way, the best-performing Poisson model is subsequently examined with respect to its statistical properties. The same approach is used for the Poisson-Gompertz model.

Although these models are optimal in terms of AIC and constitute valid Poisson models with nearly equal mean and variance within the corresponding families, their overall adequacy cannot be fully assessed on this basis alone. To quantify the performance of the Poisson-Gompertz models in terms of explained variance, we will compare them both visually and statistically with the corresponding Gompertz growth models. For this reason, we evaluate the performance of the fitted Gompertz growth model (as an attempt to implicitly evaluate the performance of the Poisson-Gompertz model), using the nonlinear least squares method (Strutz 2016), based on the adjusted R^2 metric. The main advantage over the coefficient of determination is that it penalizes model complexity while accounting for the number of data points. This quality is important not only for avoiding overfitting but also for enabling comparisons with different (future) studies and sample sizes. For the purposes of this study, we will focus on models that demonstrate significant performance (i.e. $R^2_{adj} > 0.2$).

Model Comparison

After fitting both the Poisson-Gompertz and Gompertz growth models and taking into account that the former is directly related to the epidemiological problem while the latter represents a simplification that facilitates performance assessment and interpretation, the two models should be compared to assess their relevance. This task is challenging because it involves comparing a Poisson-type model with an ordinary least squares model. Insight can be gained by examining and comparing the model residuals, in particular the deviance residuals (McCullagh and Nelder 1989) of the Poisson-Gompertz model normalized by the population at risk of each region, and the raw residuals of the Gompertz growth model. A visual comparison of the distributions of the residuals provides an overall qualitative assessment.

Model Sensitivity

The purpose of the sensitivity analysis of the Gompertz growth models is to assess their stability with respect to alternative sample points (i.e., pairs of mortality rate and chemical element concentration). Some variability is expected in the fitted models, as they depend on the specific sample drawn from the population. To explore this, we attempt to visualize the "alternative" models that could have been obtained under different sampling scenarios by simulating them using the bootstrap method (Efron and Tibshirani 1994). Bootstrap resampling generates multiple versions of the training dataset by sampling with replacement (in our case, 500 iterations). Fitting the model to each resample reveals the extent to which the fitted models vary due to data fluctuations. We visualize all resulting models in a single graph using semi-transparent lines, where darker areas indicate stronger model agreement. This approach can also highlight the most dominant model structures.

Results and Discussion

Selecting Models: The Dominance of Arsenic

Poisson Models

For all age groups, we fitted separate models for the full list of chemical elements reported in GEMAS (2021), rather than restricting the analysis to only known toxic elements. This comprehensive approach was intended to minimize selection bias and ensure that the results were not chosen with confirmation bias (a.k.a. cherry picked). Valid Poisson models with approximately equal mean and variance were obtained only for the 0–29 and 30–39 age groups. For both age groups, As ranked first according to AIC.

Notably, the relative risk parameter β_1 for As model is statistically significant (at the 0.05 significance level), while the corresponding parameters for the second-ranked models (Pb and U for 20–29 and 30–39 age group, respectively) are not statistically significant for these age groups (numerical details are shown in the Appendix). Graphical representations of these models are presented in Fig. 3. A clear positive association is observed between cancer mortality and As concentration. The estimated relative risk increase associated with As is 2.7% for the 0–29 age group and 2.2% for the 30–39 age group.

Poisson-Gompertz Models

Using a similar AIC based comparison, in combination with an adequate index of dispersion for the fitted models, we

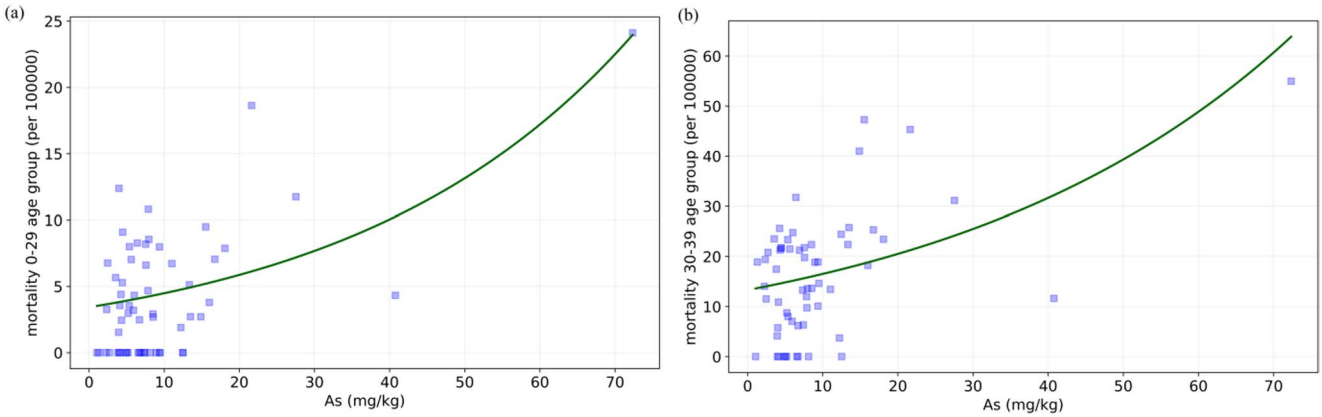

Fig. 3 Cancer mortality estimation using the Poisson model based on As topsoil concentration, for 0–29 (**a**) and 30–39 (**b**) age groups. The green line indicates the estimated cancer mortality, while the blue squares denote the data points

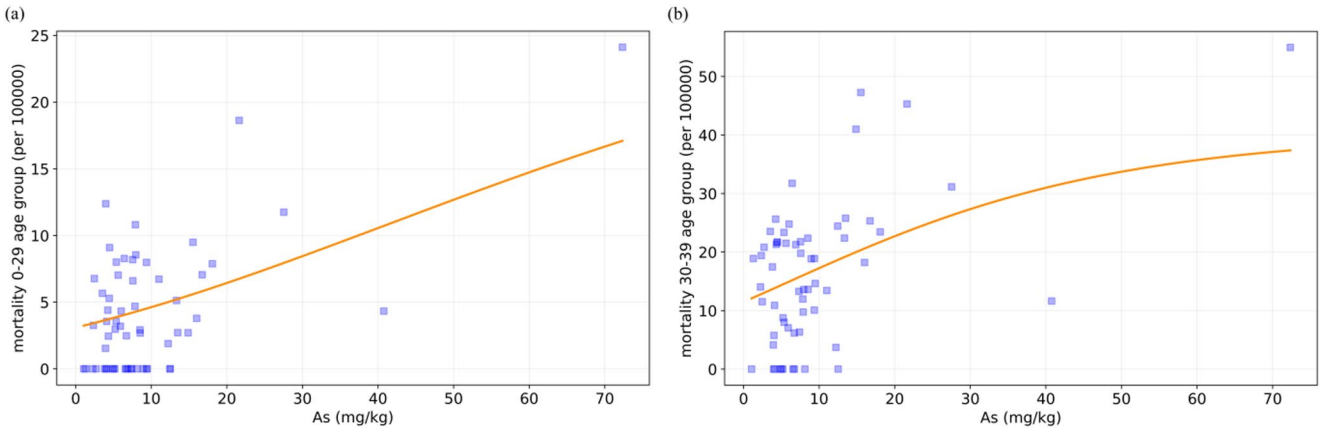

Fig. 4 Cancer mortality estimation using the Poisson-Gompertz model based on As topsoil concentration, for 0–29 (**a**) and 30–39 (**b**) age groups. The orange line indicates the estimated cancer mortality, while the blue squares denote the data points

ranked the Poisson-Gompertz models of all elements, with As again appearing at the top of the list. Figure 4 illustrates how the resulting curves conform to the epidemiological assumptions outlined in Sect. "Modeling", representing an improvement over the unconstrained growth exhibited by the initial Poisson models.

For the 0–29 age group, we observe a subtle sigmoidal pattern, in which a relatively short low-mortality region is followed by a slow transitional phase, while a stabilization pattern emerges at very high As concentrations. The 30–39 age group exhibits a similar pattern, with a shorter transitional phase and stabilization occurring at approximately 40 deaths per 100,000. The fitted model parameters are provided in the Appendix.

For both age groups, we observe a visual resemblance to the simpler Poisson models. This provides an indication that higher As concentrations in topsoil are associated with increased mortality rates. The Poisson-Gompertz models possess the internal structure and flexibility to further differentiate the rate of change across different As concentration levels. This differentiation is depicted on the three phases of these models.

Simplified Gompertz Growth Models

Model selection among chemical elements for a given age group in the simplified Gompertz growth family, differs substantially from that used for the Poisson-based models. Because the current models are formulated explicitly in terms of mortality rates, ordinary least squares estimation can be employed. Model selection is based on the adjusted R^2 metric, which facilitates model ranking while simultaneously quantifying how well each model describes the observed data.

Up to this point, using the Poisson and Poisson-Gompertz models, which explicitly account for the number of deaths and the population at risk, we identified As as the element yielding the best performance. However, these models did not provide a direct measure of how well these As-based models explain variability in mortality. By reducing the

Poisson-Gompertz models to simplified Gompertz models, we sacrifice their explicit count-based formulation but gain access to explained-variance metrics. In Sect. "Comparing Poisson-Gompertz Model with Gompertz Simplified Model", we compare these models and quantify their level of agreement in order to assess whether in practice they address the same modelling objective.

In Fig. 5 we show the growth curves for the first two age groups. The model selection was based on the R^2_{adj} on all GEMAS (2021) elements as shown in Fig. 10 of the Appendix, while the model parameters, statistical properties and goodness of fit metrics are presented in the Appendix, as well. We observe that As and antimony (Sb) yield the higher performance, with relatively high proportion of the variation in the mortality rate (i.e. > 0.25 and as high as ≈ 0.34) that is predictable from the independent variable (i.e. concentration of As and Sb). This is a clear indication that As and Sb estimate relatively accurately the cancer mortality in these two age groups.

It should be noted that even high-performing models cannot demonstrate a causal effect solely based on their adjusted R^2, and thus should be examined through their known carcinogenic effect. A model could achieve high R^2_{adj} due to spurious correlation. For example, some elements may appear proportional to other carcinogenic ones due to geological factors, acting as confounding or proxy variables. It is therefore important to examine the role of the selected elements (i.e., As and Sb) in the higher-performing models, particularly in relation to their known association with cancer mortality. Under this prism, both elements are known to be carcinogenic to humans. It is documented the implication of As exposure to cancers of the skin, lung, liver and bladder, among other tissues (Cogliano et al. 2011; Naujokas et al. 2013), while Sb is recognized as carcinogenic by U.S. Department of Health and Human Services (ATSDR 2019).

Regarding the close performance of the models for As and Sb, which are statistically indistinguishable from each other (comparing their residuals using Wilcoxon signed rank test, McDonald 2014), we report p-values 0.171 and 0.624, for 0–29 and 30–39 age groups, respectively), which is not surprising. The concentrations of these elements often appear proportional in topsoil due to their similar geochemical behavior and shared sources in mineral deposits and geological formations (Kabata-Pendias 2000). This is also the case in Greece, where a strong Pearson correlation coefficient of approximately 0.85 is observed between As and Sb. Moreover, in Greece the concentrations of Sb are generally an order of magnitude lower than those of As. This strongly suggests that Sb acts as a proxy for As, rather than being the primary contributor in Greek rural areas.

The adjusted R^2 performance for the 40–49 years and older age groups drops to non-significant levels i.e. < 0.1, considerably smaller than the 0.2 threshold we adopted in this study. A possible explanation is a masking effect caused by lifestyle-related risk factors and other aging related factors which may become the primary causes of mortality. This hypothesis requires further investigation and falls outside the scope of this study.

Comparing Poisson-Gompertz Model with Gompertz Simplified Model

The model residuals of the As-based models for the 0–29 and 30–39 year age groups are shown in Fig. 6. We observe an intensive overlap between the distributions of the Poisson-Gompertz and Gompertz growth models. This overlap provides strong evidence that the simplified model (i.e., the Gompertz growth model) is highly consistent with the epidemiologically oriented model (i.e., the Poisson-Gompertz model, which accounts explicitly for both the number of deaths and the corresponding population at risk). On this

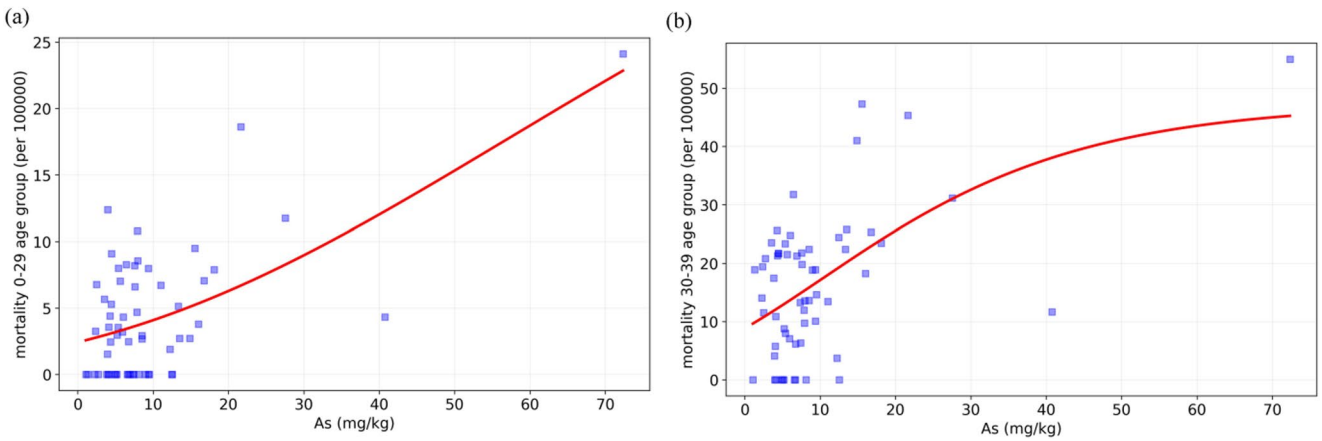

Fig. 5 Cancer mortality estimation using the Gompertz growth model based on As topsoil concentration, for 0–29 (**a**) and 30–39 (**b**) age groups. The red line indicates the estimated cancer mortality, while the blue squares denote the data points

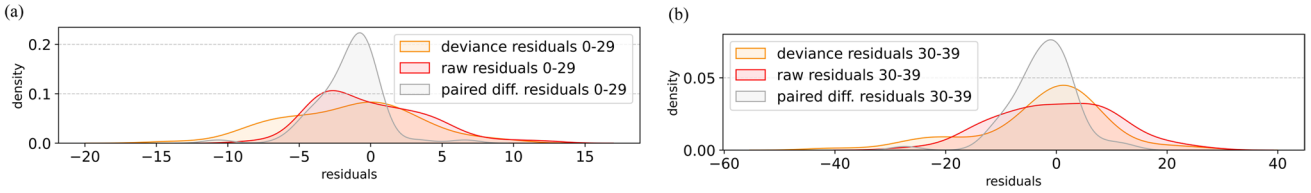

Fig. 6 Model residuals of Poisson-Gompertz (deviance residuals, orange curve) and Gompertz growth (raw residuals, red curve) model, for 0–29 (**a**) and 30–39 (**b**) age groups. The grey curve indicates the paired difference residuals

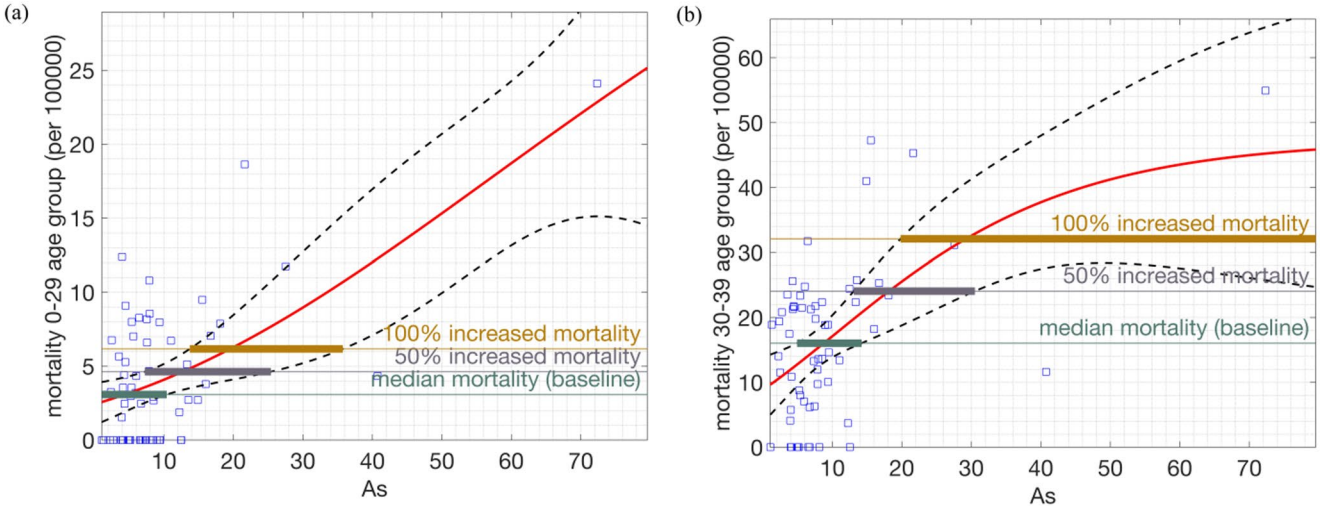

Fig. 7 Cancer mortality estimation based on As topsoil concentration (in mg kg^{-1}). The red line indicates the estimated cancer mortality, while the dotted black lines represent the 95% confidence intervals, and blue squares denote the data points. The thin horizontal green line corresponds to the median mortality of the age group (i.e. baseline), while the thick green line represents the confidence interval of the As concentrations associated with it. A 50% and a 100% increase in mortality relative to the baseline are shown with dark grey and gold lines, respectively, along with their confidence intervals. The left subplot (**a**) corresponds to the 0–29 years age group, while the right one (**b**) represents the 30–39 years age group

basis, we conclude that the two models are reasonably close to each other. Consequently, (i) the Poisson-Gompertz model exhibits sufficient descriptive capability, and (ii) the Gompertz model can be used as a suitable surrogate for further investigation.

Estimating Cancer Mortality Risk Based on Gompertz Growth Models

The model of Eq. (4) predicts cancer mortality for a given age group based on the concentration of a topsoil element, $f(x) \rightarrow cancer\ mortality$, where x denotes the concentration of a topsoil element. To estimate cancer mortality risk as a function of As concentration, we can invert the model $f(x)$, thereby identifying the concentrations of As that correspond to a given mortality rate: $f^{-1}(cancer\ mortality) \rightarrow x$, where x is the As concentration in topsoil. A simple and intuitive method for performing this inversion, while simultaneously incorporating model uncertainty (and thus deriving a confidence interval for the estimated As concentration), is through a graphical approach. Specifically, we plot As concentration against cancer mortality using the fitted model parameters, including 95% uncertainty bounds. We take the median cancer mortality rate for a particular age group as a robust baseline metric (i.e. not sensitive to extreme values). When we draw a horizontal line corresponding to this baseline mortality on the graph, it intersects the model prediction curve at a single point (the same applies to the 95% uncertainty bounds). This intersection point yields the estimated As concentration corresponding to the baseline mortality rate, along with its associated confidence interval. By raising the horizontal line to represent increased mortality, e.g. a 50% or 100% increase relative to baseline, we can similarly determine the corresponding As concentrations and their corresponding confidence intervals. Using this approach, we can estimate the ranges of As concentration associated with significant increases in cancer mortality, for example a 50% increase or doubling the risk.

Figure 7 shows the cancer mortality models for the 0–29 and 30–39 years age groups, based on As concentrations in topsoil. The red line represents the estimated cancer mortality, with dotted black lines indicating the 95% confidence intervals. Blue squares mark the observed data points. A thin

horizontal green line defines the baseline, i.e. median mortality for the age group, while the thick green line represents the range of As concentrations linked to it. Dark grey and gold lines mark 50% and 100% increases in mortality relative to the baseline (i.e. median mortality of the age group), each with their respective confidence intervals

Aiming to describe As concentration values (expressed in mg kg^{-1}) in topsoil (along with the associated confidence intervals), for which we observe an increase of the baseline cancer mortality, we summarize the estimated risk as follows:

For the 0–29 years age group:

- Baseline cancer mortality corresponds to As concentration of 4.4 mg kg^{-1}, with a confidence interval of [1.0, 10.4] mg kg^{-1}.
- A 50% increase in mortality is associated with As concentration of 12.7 mg kg^{-1} with confidence interval [7.3, 25.4] mg kg^{-1}.
- A 100% increase in mortality is associated with As concentration of 19.5 mg kg^{-1} with confidence interval [13.7, 35.8] mg kg^{-1}.

For the 30–39 years age group:

- Baseline cancer mortality corresponds to As concentration of 8.8 mg kg^{-1}, with a confidence interval of [4.9, 14.2] mg kg^{-1}.
- A 50% increase in mortality is associated with As concentration of 18.2 mg kg^{-1} with confidence interval [13.1, 30.5] mg kg^{-1}.
- A 100% increase in mortality is associated with As concentration of 29.2 mg kg^{-1} with confidence interval [19.8, 79.5] mg kg^{-1}.

The models presented in Fig. 7 can also be interpreted from the perspective of As concentration. For instance, if a specific region of interest reports an As concentration of approximately 20 mg kg^{-1}, a vertical move on the model curve yields an estimated cancer mortality rate of 6.0 per 100,000, for the 0–29 years age group. The corresponding confidence interval is [4.0, 8.5] per 100,000.

Applicable numerical standards for soil contaminants protective of human health have been established by the New Zealand Government (New Zealand Government 2012). Notably, the soil contaminant standard (SCS) for As in rural residential land is set at 17 mg kg^{-1}. In the present study, the modelled baseline concentrations of As for the two evaluated age groups were 4.4 mg kg^{-1} and 8.8 mg kg^{-1}, respectively, substantially below the established SCS value. This observation highlights the non-threshold mode of action associated with As toxicity, wherein adverse health effects may occur even at concentrations below established guideline levels.

Sensitivity Analysis of the Proposed Gompertz Growth Models

To assess model stability, bootstrap resampling with 500 iterations is used to simulate potential alternative datasets. Models are refitted to each resample with substitution to observe variability as discussed in Sect. "Model Sensitivity". The results are visualized with semi-transparent lines, highlighting model agreement and dominant structures through darker regions in the graph.

Figure 8a clearly shows that the proposed model for the 0–29 years age group closely follows the single dark/common region of the bootstrap results, indicating that it effectively captures the major trend. For the 30–39 years

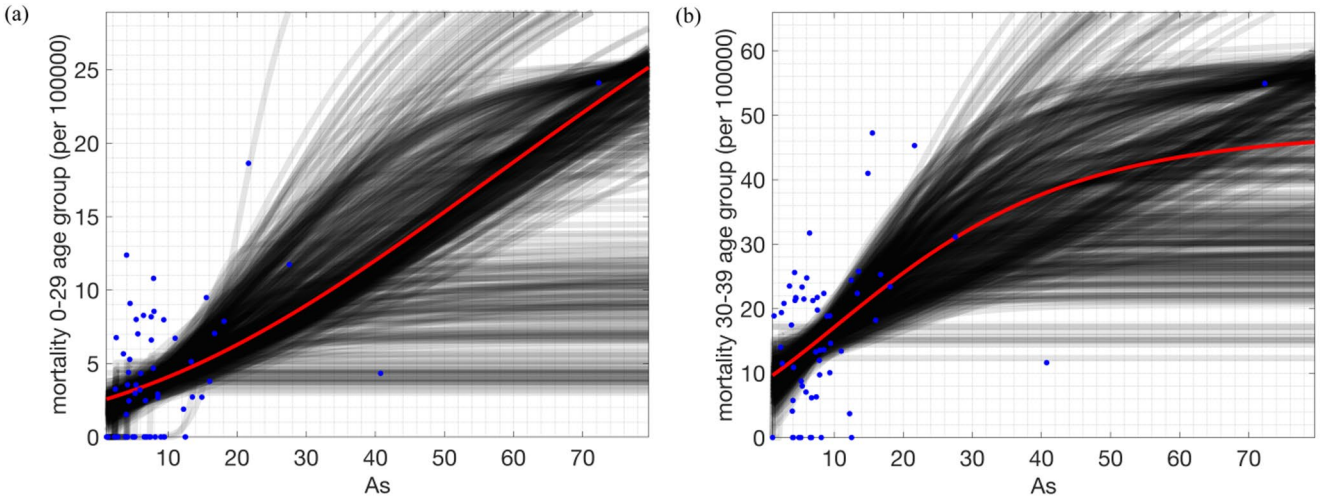

Fig. 8 Sensitivity analysis of the proposed models using the bootstrap method. The red line represents the proposed model, blue squares indicate observed data points, and black semi-transparent lines depict simulated alternative models. **(a)** 0-29 years age group, **(b)** 30-39 years age group

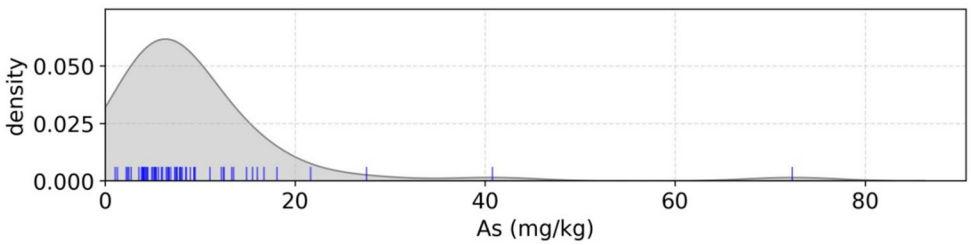

Fig. 9 Probability density function of As concentration in Greek topsoils

age group, two major paths appear as the most dominant (Fig. 8b). The proposed model generally follows the middle region between them but deviates at higher As concentrations, potentially indicating an underestimation in that range.

Conclusions

A statistically significant link was identified between As concentrations in rural Greek topsoils and cancer mortality risk for younger age groups. This link has been identified through different families of models. The proposed models reveal a rather strong association for 0–29 and 30–39 years age groups, where elevated As levels correspond to substantial increases in mortality. These results suggest a simplified rule: a three to five fold increase in As concentration relative to the baseline level is associated with a doubling of mortality. More precisely, the numerical thresholds derived from the proposed risk models indicate that 12.7 mg kg^{-1} and 19.5 mg kg^{-1} are associated with 50% and 100% increases in mortality for the 0–29 age group, respectively, while 18.2 mg kg^{-1} and 29.2 mg kg^{-1} correspond to comparable increases in the 30–39 age group. These findings should be considered for deriving As topsoil maximum allowable limit (MCL). Moreover, the proposed nonlinear models provide an actionable framework for cross-regional comparisons and epidemiological risk assessments in other populations.

Overall, the findings underscore the public health significance of As concentrations in topsoil and highlight the need for targeted environmental monitoring and preventive interventions, such as the remediation of highly contaminated areas. However, while the analysis was purposefully confined to small rural municipalities to limit regional heterogeneity, the ecological nature of the study inherently allows for residual differences in lifestyle and healthcare that may have influenced cancer mortality. To address these limitations, future studies could extend the analysis to broader geographical regions - ideally at the European or U.S. level -to better estimate cancer mortality associated with As concentrations in topsoil. Such expanded populations are expected to further refine As guideline thresholds and narrow the associated confidence intervals.

Appendix

See Fig. 9.

Poisson Model of As, 0-29 age group

Dep. Variable:	y	No. Observations:	60		
Model:	GLM	Df Residuals:	58		
Model Family:	Poisson	Df Model:	1		
Link Function:	Log	Scale:	1.0000		
Method:	IRLS	Log-Likelihood:	-89.895		
		Deviance:	75.390		
		Pearson chi2:	61.9		
No. Iterations:	5	Pseudo R-squ. (CS):	0.07366		
Covariance Type:	nonrobust				
	coef	std err	z	P>|z|	[0.025 0.975]
const	-10.2805	0.138	-74.380	0.000	-10.551 -10.010
x	0.0269	0.011	2.484	0.013	0.006 0.048

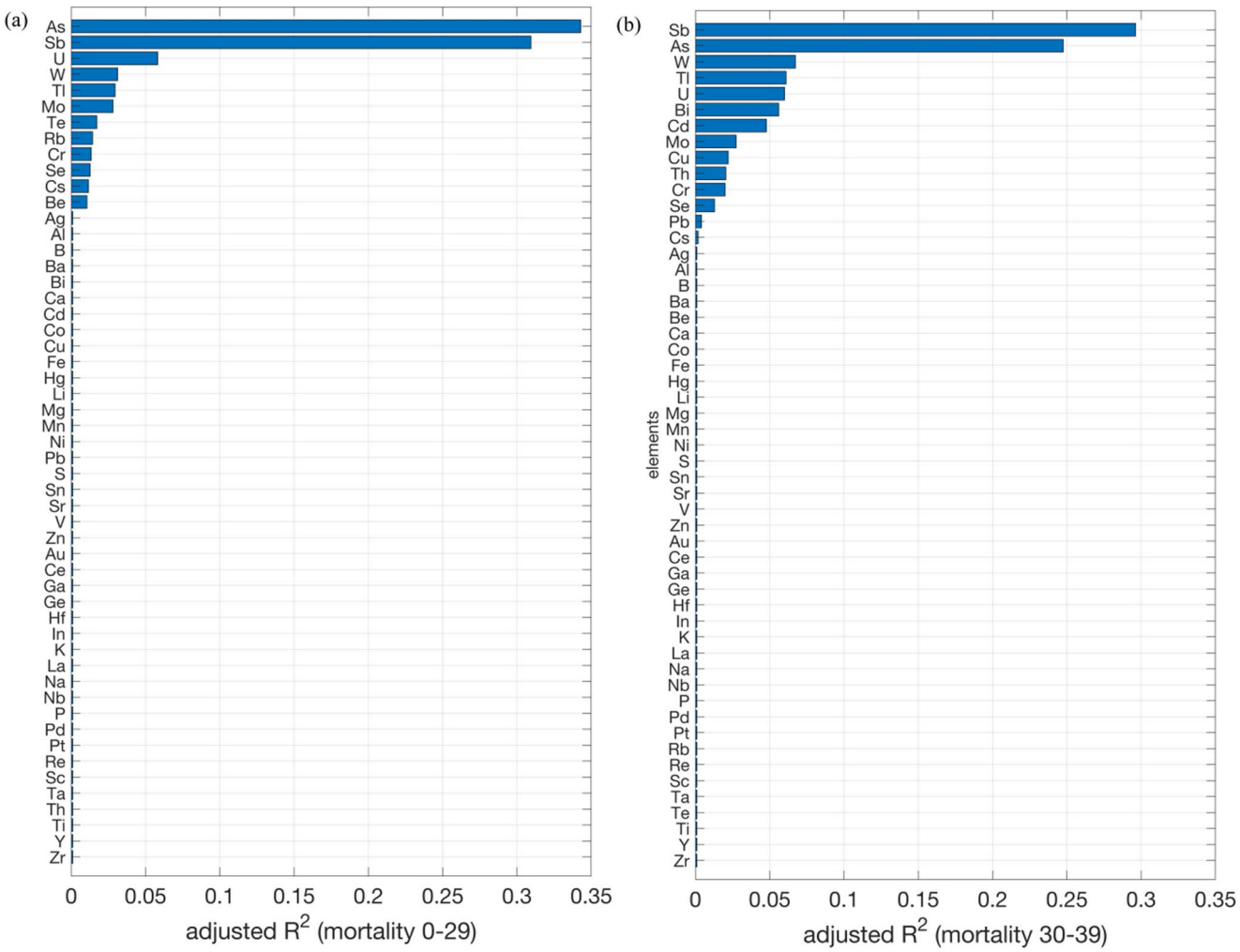

Fig. 10 Adjusted R^2 of the cancer mortality Gompertz growth models (**a**) for the 0–29 years age group and (**b**) for the 30–39 years age group

Poisson Model of As, 30-39 age group

Dep. Variable:	y	No. Observations:	60
Model:	GLM	Df Residuals:	58
Model Family:	Poisson	Df Model:	1
Link Function:	Log	Scale:	1.0000
Method:	IRLS	Log-Likelihood:	-101.23
		Deviance:	60.790
		Pearson chi2:	50.8
No. Iterations:	5	Pseudo R-squ. (CS):	0.06370
Covariance Type:	nonrobust		

	coef	std err	z	P>|z|	[0.025	0.975]
const	-8.9255	0.117	-76.475	0.000	-9.154	-8.697
x	0.0217	0.010	2.235	0.025	0.003	0.041

where const: βo, x: β1 of the Poisson distribution formula

Poisson-Gompertz Models of As

0–29 age group: A = 2.934e-04, K_u = 7.254e-03, X_i = 4.112e + 01.

30–39 age group: A = 4.020e-04, K_u = 1.444e-02, X_i = 5.762e + 00

(See Fig. 10).

Gompertz Growth Models of As

0 – 29 years age group	30 – 39 years age group
Coefficients (with 95% confidence bounds): A = 48.23 (-70.57, 167.00) K_u = 0.007051 (-0.005248, 0.01935) X_i = 57.13 (-77.26, 191.50) Goodness of fit metrics: sse: 862.7417 rsquare: 0.3652 dfe: 57 adjrsquare: 0.3429 rmse: 3.8905	Coefficients (with 95% confidence bounds): A = 47.30 (20.54, 74.07) K_u = 0.01842 (-0.0009369, 0.03777) X_i = 10.33 (-2.66, 23.33) Goodness of fit metrics: sse: 6.4812e+03 rsquare: 0.2731 dfe: 57 adjrsquare: 0.2476 rmse: 10.6633

Numerical Software

All numerical computations were carried out in Python 3.9.19 using the following libraries: statsmodels 0.14.2, scipy 1.13.1, seaborn 0.13.2, numpy 2.0.0. The numerical computations of the Gompertz growth models were performed using MATLAB 2024a and the Curve Fitting Toolbox 24.1 (MathWorks 2024).

Acknowledgements The authors thank two anonymous reviewers for their helpful comments and feedback, which have improved this manuscript.

Author Contributions George Tsirogiannis: Writing – original draft, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. Michalis Leotsinidis: Writing – review & editing, Data curation, Conceptualization. Eleni Sazakli: Writing – review & editing, Methodology, Data curation, Conceptualization.

Funding Open access funding provided by HEAL-Link Greece. This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Data Availability All data supporting our findings are available upon request.

Declarations

Competing interests The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Aho K, Derryberry D, Peterson T (2014) Model selection for ecologists: the worldviews of AIC and BIC. Ecology 95(3):631–636. https://doi.org/10.1890/13-1452.1

Andersen PK, Borgan O, Gill RD, Keiding N (2012) Statistical models based on counting processes. Springer Science & Business Media

ATSDR (Agency for Toxic Substances and Disease Registry) (2019) Toxicological profile for antimony and compounds. U.S. Department of Health and Human Services, Public Health

Bradl H (2005) Heavy Metals in the Environment: Origin. Interaction and Remediation, vol 6. Elsevier/Academic, London

Brauer F, Castillo-Chavez C, Feng Z (2019) Mathematical models in epidemiology, vol 32. Springer, New York

Cogliano VJ, Baan R, Straif K et al (2011) Preventable exposures associated with human cancers. J Natl Cancer Inst 103(24):1827–1839. https://doi.org/10.1093/jnci/djr483

Cox DR, Lewis PA (1966) The statistical analysis of series of events. Methuen, London and John Wiley & Sons, New York

Draper NR, Smith H (1998) Applied regression analysis, vol Vol. 326. John Wiley & Sons. https://doi.org/10.1002/9781118625590

Eck C, Garcke H, Knabner P (2017) Mathematical modeling. Springer International Publishing, Cham. https://doi.org/10.1007/978-3-319-55161-6

Efron B, Tibshirani RJ An introduction to the bootstrap. Chapman and Hall/CRC. EFSA CONTAM Panel (EFSA Panel on Contaminants in the Food Chain), Schrenk D, Bignami M, Bodin L et al (1994) 2024 Update of the risk assessment of inorganic arsenic in food. EFSA J 22(1):e8488. https://doi.org/10.2903/j.efsa.2024.8488 Accessed 13 October 2025

EFSA CONTAM Panel (EFSA Panel on Contaminants in the Food Chain), Schrenk D, Bignami M, Bodin L et al (1994) 2024 Update of the risk assessment of inorganic arsenic in food. EFSA J 22(1):e8488. https://doi.org/10.2903/j.efsa.2024.8488 Accessed 13 October 2025

Frisbie SH, Mitchell EJ (2022) Arsenic in drinking water: an analysis of global drinking water regulations and recommendations for updates to protect public health. PLoS One 17(4):e0263505. https://doi.org/10.1371/journal.pone.0263505

GEMAS (Geochemical Mapping of Agricultural and Grazing land Soil) (2021) Chemistry of Europe's Agricultural soils, Original dataset and DVD content. http://data.europa.eu/88u/dataset/20607a6b-f412-4777-9d43-67d734677d57 Accessed 13 October 2025

Hou D, Jia X, Wang L et al (2025) Global soil pollution by toxic metals threatens agriculture and human health. Science 388(6744):316–321. https://doi.org/10.1126/science.adr5214

Huang Y, Zhang N, Ge Z et al (2024) Determining soil conservation strategies: ecological risk thresholds of arsenic and the influence

of soil properties. Eco-Environ Health 3(2):238–246. https://doi.org/10.1016/j.eehl.2024.02.007

Hughes MF (2002) Arsenic toxicity and potential mechanisms of action. Toxicol Lett 133(1):1–16. https://doi.org/10.1016/s0378-4274(02)00084-x

IARC Working Group on the Evaluation of Carcinogenic Risks to Humans (2012) Arsenic, Metals, Fibres and Dusts. Lyon (FR): International Agency for Research on Cancer. (IARC Monographs on the Evaluation of Carcinogenic Risks to Humans, No. 100 C.) https://www.ncbi.nlm.nih.gov/books/NBK304375/ Accessed 13 October 2025

Kabata-Pendias A (2010) Trace elements in soils and plants, 4th ed. CRC Press. https://doi.org/10.1201/b10158

Katsifara A, Tsirogiannis G, Sazakli E, Jelastopulu E, Leotsinidis M (2025) Overall effects of topsoil elements on cancer mortality in rural Greece: a modeling approach. Chemosphere 370:144018. https://doi.org/10.1016/j.chemosphere.2024.144018

Kotopoulou S, Zampelas A, Magriplis E (2025) Impact of fish intake on dyslipidemia likelihood by cooking method: results from the Hellenic National Nutrition and Health Survey. Nutr Res 133:22–34. https://doi.org/10.1016/j.nutres.2024.11.004

Mandal BK, Suzuki KT (2002) Arsenic round the world: a review. Talanta 58(1):201–235. https://doi.org/10.1016/S0039-9140(02)00268-0

Masuda H (2018) Arsenic cycling in the Earth's crust and hydrosphere: interaction between naturally occurring arsenic and human activities. Prog Earth Planet Sci 5:68. https://doi.org/10.1186/s40645-018-0224-3

MathWorks I (2024) MATLAB (Version R2024a) [Computer software]. https://www.mathworks.com/products/matlab.html

McCullagh P, Nelder JA (1989) Generalized linear models, 2nd ed. Chapman and Hall, London

McDonald JH (2014) Handbook of biological statistics, 3rd ed. Sparky House Publishing, Baltimore, Maryland

Naujokas MF, Anderson B, Ahsan H et al (2013) The broad scope of health effects from chronic arsenic exposure: update on a worldwide public health problem. Environ Health Perspect 121(3):295–302. https://doi.org/10.1289/ehp.1205875

New Zealand Government (2012) Users' Guide: National Environmental Standard for Assessing and Managing Contaminants in Soil to Protect Human Health. Wellington: Ministry for the Environment. https://environment.govt.nz/publications/users-guide-national-environmental-standard-for-assessing-and-managing-contaminants-in-soil-to-protect-human-health/ Accessed 13 October 2025

Núñez O, Fernández-Navarro P, Martín-Méndez I et al (2016) Arsenic and chromium topsoil levels and cancer mortality in Spain. Environ Sci Pollut Res Int 23(17):17664–17675. https://doi.org/10.1007/s11356-016-6806-y

Poghosyan Z, Bigalke M, Gevorgyan A, Sahakyan L, Tepanosyan G (2025) Geochemical characterization and human health risk assessment of arsenic in soils of a high background area (Lori region, Armenia). Environ Res 285(2):122356. https://doi.org/10.1016/j.envres.2025.122356

Reimann C, Birke M, Demetriades A, Filzmoser P, O'Connor P (eds) (2014) Chemistry of Europe's Agricultural Soils—Part A: methodology and Interpretation of the GEMAS Data Set. Geologisches Jahrbuch (Reihe B 102). Schweizerbarth, Hannover

Rossman TG (2003) Mechanism of arsenic carcinogenesis: an integrated approach. Mutat Res 533(1–2):37–65. https://doi.org/10.1016/j.mrfmmm.2003.07.009

Service (2025) https://www.atsdr.cdc.gov/toxprofiles/tp23.pdf Accessed 13 October

Spaur M, Glabonjat RA, Schilling K et al (2024) Contribution of arsenic and uranium in private wells and community water systems to urinary biomarkers in US adults: the Strong Heart Study and the Multi-Ethnic Study of Atherosclerosis. J Expo Sci Environ Epidemiol 34(1):77–89. https://doi.org/10.1038/s41370-023-00586-2

Strutz T (2016) Data fitting and uncertainty: a practical introduction to weighted least squares and beyond, 2nd edn. Springer, Berlin

Tjørve E, Tjørve KM (2010) A unified approach to the Richards-model family for use in growth analyses: why we need only two model forms. J Theor Biol 267(3):417–425. https://doi.org/10.1016/j.jtbi.2010.09.008

Tjørve KM, Tjørve E (2017) The use of Gompertz models in growth analyses, and new Gompertz-model approach: an addition to the Unified-Richards family. PLoS One 12(6):e0178691. https://doi.org/10.1371/journal.pone.0178691

WHO (2022) Arsenic. https://www.who.int/news-room/fact-sheets/detail/arsenic Accessed 13 Oct 2025