

Article

Groundwater Arsenic-Attributable Cardiovascular Disease (CVD) Mortality Risks in India

Ruohan Wu *, Lingqian Xu and David A. Polya * 

Department of Earth and Environmental Sciences, School of Natural Sciences and Williamson Research Centre for Molecular Environmental Sciences, University of Manchester, Manchester M13 9PL, UK; lingqian.xu@postgrad.manchester.ac.uk

* Correspondence: ruohan.wu@postgrad.manchester.ac.uk (R.W.); david.polya@manchester.ac.uk (D.A.P.)

Abstract: Cardiovascular diseases (CVDs) have been recognized as the most serious non-carcinogenic detrimental health outcome arising from chronic exposure to arsenic. Drinking arsenic contaminated groundwaters is a critical and common exposure pathway for arsenic, notably in India and other countries in the circum-Himalayan region. Notwithstanding this, there has hitherto been a dearth of data on the likely impacts of this exposure on CVD in India. In this study, CVD mortality risks arising from drinking groundwater with high arsenic ($>10 \mu\text{g/L}$) in India and its constituent states, territories, and districts were quantified using the population-attributable fraction (PAF) approach. Using a novel pseudo-contouring approach, we estimate that between 58 and 64 million people are exposed to arsenic exceeding $10 \mu\text{g/L}$ in groundwater-derived drinking water in India. On an all-India basis, we estimate that 0.3–0.6% of CVD mortality is attributable to exposure to high arsenic groundwaters, corresponding to annual avoidable premature CVD-related deaths attributable to chronic exposure to groundwater arsenic in India of between around 6500 and 13,000. Based on the reported reduction in life of 12 to 28 years per death due to heart disease, we calculate value of statistical life (VSL) based annual costs to India of arsenic-attributable CVD mortality of between USD 750 million and USD 3400 million.

Keywords: arsenic; groundwater; cardiovascular disease (CVD); population-attributable fraction (PAF); geospatial statistical



Citation: Wu, R.; Xu, L.; Polya, D.A. Groundwater Arsenic-Attributable Cardiovascular Disease (CVD) Mortality Risks in India. *Water* **2021**, *13*, 2232. <https://doi.org/10.3390/w13162232>

Academic Editor: Aldo Fiori

Received: 21 July 2021

Accepted: 12 August 2021

Published: 17 August 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Geogenic or anthropogenic arsenic has long been recognized as an important human carcinogen [1,2]. Globally, and notably, particularly in India, the utilization of high arsenic hazard groundwaters as drinking and irrigation water has posed, and still does pose, a serious threat to public health [3–5]. The major human exposure pathways to groundwater arsenic are drinking arsenic-contaminated groundwaters and consuming inorganic arsenic-bearing crops, notably including rice [6,7], due, albeit only in part, to irrigation with high arsenic groundwater [8]. Chronic exposure to inorganic arsenic may give rise to internal cancers (bladder, lung, kidney, and liver), skin cancers, skin lesions, and, importantly, cardiovascular diseases (CVDs), which are considered as the most serious non-carcinogenic detrimental health outcome [9–12].

In India, arsenic contamination of groundwaters takes place mainly in the alluvial deposits of major rivers (e.g., Ganges and Brahmaputra) flowing from the Himalayas and Tibetan plateau [13–15]. The dominant mechanism of arsenic enrichment and mobilization in these groundwaters in alluvial aquifers is the microbially mediated reductive dissolution of arsenic-bearing Fe(III) oxide (oxyhydroxide, hydroxide, and oxide) minerals [16,17]. Other processes, including complexation of arsenic by dissolved humic substances, competitive desorption, oxidation of sulfide minerals, geothermal activities, and mining-related activities, may also influence arsenic concentrations in groundwaters [18,19].

Spatial geostatistical models of the distribution of groundwater arsenic have been produced at both state (Gujarat [20], Uttar Pradesh [21]) and national scales [22,23] to help better fully understand the distribution of groundwater arsenic hazard in India, particularly given the lack of all-India systematic groundwater arsenic testing. Podgorski et al. [22] generated a random forest model with over a hundred thousand arsenic concentration observations and over two dozen environmental parameter predictors to identify high groundwater arsenic ($>10 \mu\text{g/L}$) in well-known arsenic-contaminated areas in the alluvial sediments along the Ganges and Brahmaputra plains and in Gujarat and Punjab states and to capture less-known or previously unrecorded areas. Estimates of the population numbers in India exposed to high arsenic through drinking groundwaters range from 18 to 30 million [22] to 90 million [23].

Notwithstanding recent spatial distribution models of groundwater arsenic for India, a country-wise and country-specific public health risk model, notably for cardiovascular diseases has not yet been created. The association of CVDs with chronic exposure to arsenic in drinking water have been widely discussed [24–28]. Meta-analysis is a recommended method to quantify the dose–response relationship between CVDs and arsenic exposure, suitably weighting the results of a range of studies [24,29]. Meanwhile, the population-attributable fraction (PAF) approach is a common method used to quantify potential disease burden reduction in population by eliminating a certain risk factor [30,31].

The aim of this study was to estimate the magnitude of groundwater arsenic-attributable cardiovascular disease (CVD) mortality in India. These aims were achieved by (i) generating machine learning models of the probability distribution of groundwater arsenic concentrations exceeding four different thresholds (10, 20, 50, 80 $\mu\text{g/L}$); (ii) thereby creating a pseudo-contour map of groundwater arsenic concentrations; (iii) estimating the state/district-level population exposed to these four arsenic concentration ranges (10–20, 20–50, 50–80, $>80 \mu\text{g/L}$) based on simple exposure models; (iv) determining the state/district-level population-attributable fraction (PAF) of CVD mortality attributable to exposure to groundwater arsenic; and hence (v) determining the state-level CVD mortality attributable to groundwater arsenic. Lastly, (vi) the costs of arsenic-attributable CVD mortality were estimated for the whole of country.

2. Materials and Methods

2.1. Hazard Models of Groundwater Arsenic

A pseudo-contour map of groundwater arsenic concentrations for the whole of India was produced, based on the machine learning approach of Podgorski et al. [22] incorporating the novel pseudo-contour approach of Wu et al. [20]. The reader is referred to those studies [20,22] and the references therein for a brief account of the advantages, disadvantages, and limitations of machine learning models. In this study, four random forest models, encompassing 1001 trees, were conducted to comprehensively characterize the distribution of groundwater arsenic by setting four different thresholds (10, 20, 50, and 80 $\mu\text{g/L}$) and using the package ‘Random Forest’ in R statistical software. The prediction results of the four machine learning models were applied to map the probability of groundwater arsenic concentrations exceeding 10, 20, 50, and 80 $\mu\text{g/L}$ in India. The cross-validations of the random forest models were executed by area (AUC) under the receiver operating characteristic (ROC) curve (0.5 for random model, 1 for perfect model) [32–34] calculated on the testing dataset, while, the importance of the predictors in random forest models was assessed by considering the decrease in accuracy as well as in Gini node impurity.

Compared with the random forest model of groundwater arsenic exceeding 10 $\mu\text{g/L}$ for India conducted by Podgorski et al. [22], 716 more geolocated (longitude/latitude) groundwater arsenic concentration measurements [35–47] (Figure 1a and Table S1) and 5 more potential environmental predictors (Table S2) were included in the dataset; then, the averaged 24,297 arsenic data were classified into 0–10 (59%, $n = 14,192$), 10–20 (9%, $n = 2233$), 20–50 (14%, $n = 3456$), 50–80 (6%, $n = 1536$), and >80 (12%, $n = 2880$) $\mu\text{g/L}$ concentration categories (Figure S1). Considering the five arsenic concentration categories,

which are different from the arsenic classification of Podgorski et al. [22], the entire dataset was randomly split into training (80%) and testing (20%) datasets by stratified sampling to ensure that the same ratios of the arsenic concentration categories were present in the entire, training, and testing datasets.

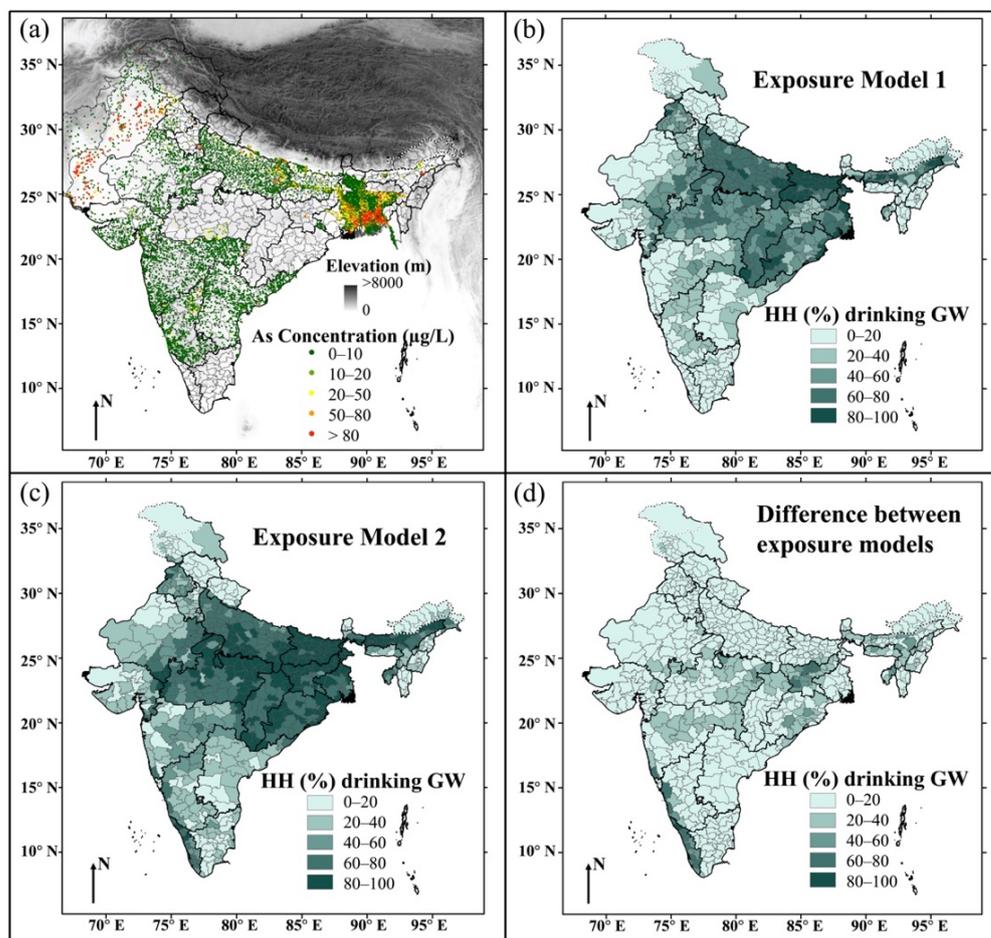


Figure 1. District-wise map of spatial distribution of India groundwater arsenic and exposure data utilized in this study. (a) District averaged arsenic concentrations and topographic elevation. (b) Exposure model 1: district-wise percentage of households (HH) drinking groundwater via hand pumps and tube wells/boreholes. (c) Exposure model 2: district-wise percentage of households (HH) drinking groundwater via covered/uncovered wells, hand pumps, and tube wells/boreholes. (d) Difference (expressed as percentage of households) between exposure models 1 and 2.

Then, these probability maps were converted into a pseudo-contour map of groundwater arsenic concentrations by using a default probability cutoff value of 0.5. Such a machine learning approach may give rise to maps in which the areas ascribed to higher concentration categories may not entirely map within those ascribed to lower concentration categories. In this study, this was minimized by appropriate stratified sampling on splitting the overall dataset into training and testing datasets, but in the limited number of areas for which this still arose, higher concentration areas were still recognized to exist for the purposes of modelling arsenic-attributable health risks.

The district-wise proportion of areas with groundwater arsenic concentration ranges of 10–20, 20–50, 50–80, and >80 $\mu\text{g/L}$ were calculated based on the pseudo-contour map of groundwater arsenic concentrations, and the error boundaries of the area proportion were estimated by Monte Carlo simulations (1000 times) [48].

2.2. Hazard and Population Exposure Maps

The district-wise proportion of areas with various groundwater arsenic concentration ranges was combined with total/rural/urban population [49] and household proportion of using groundwater as drinking water [50] to estimate district-wise total/rural/urban population exposed to arsenic concentration ranges of 10–20, 20–50, 50–80, and >80 µg/L by drinking groundwaters. The proportion of using groundwater as drinking water were calculated by two model approaches: (exposure model 1) regarding water from hand pumps and tube wells/boreholes as groundwater and (exposure model 2) considering water from wells (covered + uncovered), hand pumps, and tube wells/boreholes as groundwater [51]. The spatial distribution of these values, together with that of the difference between two values, is shown in Figure 1b–d.

Assuming that the fraction of district rural ($\frac{A_{d,r,\ell}}{A_d}$) and urban ($\frac{A_{d,u,\ell}}{A_d}$) areas with groundwater arsenic level ℓ ($\ell = 1$ (0–10 µg/L), 2 (10–20 µg/L), 3 (20–50 µg/L), 4 (50–80 µg/L), and 5 (> 80 µg/L)) are equal, the fraction of the district rural ($F_{d,r,gw,n,\ell}$) and urban ($F_{d,u,gw,n,\ell}$) populations exposed to groundwater arsenic exposure levels were calculated using Equations (1) and (2):

$$F_{d,r,gw,n,\ell} = \frac{A_{d,r,\ell}}{A_d} \times X_{d,r,gw,n} \quad (1)$$

$$F_{d,u,gw,n,\ell} = \frac{A_{d,u,\ell}}{A_d} \times X_{d,u,gw,n} \quad (2)$$

where $X_{d,r,gw,n}$ and $X_{d,u,gw,n}$ are the fraction of the households drinking groundwater in rural and urban areas of a district, respectively, and n is the exposure model approach number to calculate proportion of using groundwater as drinking water, $n = 1$ (groundwater = hand pump + tube well/borehole) or 2 (groundwater = well (covered + uncovered) + hand pump + tube well/borehole).

The fraction of the total district population exposed to groundwater arsenic exposure level ℓ ($F_{d,t,gw,n,\ell}$) was then calculated by:

$$F_{d,t,gw,n,\ell} = \frac{(F_{d,r,gw,n,\ell} \times P_{d,r}) + (F_{d,u,gw,n,\ell} \times P_{d,u})}{P_{d,t}} \quad (3)$$

where $P_{d,t}$, $P_{d,r}$, $P_{d,u}$ are the total, rural, and urban population respectively of a district.

Thus, district-wise rural ($P_{d,r,gw,n,\ell}$), urban ($P_{d,u,gw,n,\ell}$), and total ($P_{d,t,gw,n,\ell}$) population exposed to groundwater arsenic exposure level ℓ were estimated by Equations (4)–(6), respectively:

$$P_{d,r,gw,n,\ell} = F_{d,r,gw,n,\ell} \times P_{d,r} \quad (4)$$

$$P_{d,u,gw,n,\ell} = F_{d,u,gw,n,\ell} \times P_{d,u} \quad (5)$$

$$P_{d,t,gw,n,\ell} = F_{d,r,gw,n,\ell} \times P_{d,r} + F_{d,u,gw,n,\ell} \times P_{d,u} \quad (6)$$

2.3. Pooled Relative Risks of Cardiovascular Disease (CVD) Related to Arsenic Concentrations in Drinking Water

Incorporating the dose–response relationships between CVD mortality risk and arsenic in drinking water from a recent meta-analysis by Xu et al. [24], the pooled log-linear and non-linear relative risks for each CVD endpoint were calculated for arsenic concentrations of 15 µg/L (mid-point of the 10–20 µg/L category), 35 µg/L (mid-point; 20–50 µg/L), 65 µg/L (mid-point; 50–80 µg/L), and 131 µg/L (a weighted mid-point for the >80 µg/L category), using a drinking water arsenic concentration of 5 µg/L (mid-point for the 0–10 µg/L category) as the reference exposure. The goodness-of-fit of the model was evaluated using deviance tests, coefficient of determination (R^2 and adjusted R^2) and Akaike's information criterion (AIC) values. The heterogeneity was illustrated by P-

heterogeneity ($p < 0.05$ for significant), I^2 -statistics (<25% for low heterogeneity, 25–50% for moderate, and >50% for high) and Cochran's Q-statistic [52].

2.4. Cardiovascular Disease (CVD) Health Risk Estimation

The contribution of drinking groundwater with high arsenic (>10 $\mu\text{g/L}$) to CVD mortality was quantified using the population-attributable fraction (PAF) approach. The state-wise and district-wise PAFs of high groundwater arsenic-attributable CVD mortality were calculated by combining the population exposure distribution and dose–response relationship using Equation (7) based on the formula given by Rockhill et al. [30] and Mondal et al. [31]:

$$PAF_n = \frac{\sum_{\ell=1}^{\ell=5} F_{d,t,gw,n,\ell} RR_{\ell} - \sum_{\ell=1}^{\ell=5} F'_{d,t,gw,n,\ell} RR_{\ell}}{\sum_{\ell=1}^{\ell=5} F_{d,t,gw,n,\ell} RR_{\ell}} \quad (7)$$

where RR_{ℓ} is the relative (to reference exposure level, $\ell = 1$, (0–10 $\mu\text{g/L}$)) risk of CVD mortality at groundwater arsenic exposure level ℓ ; $\ell = 1$ (0–10 $\mu\text{g/L}$), 2 (10–20 $\mu\text{g/L}$), 3 (20–50 $\mu\text{g/L}$), 4 (50–80 $\mu\text{g/L}$), and 5 (> 80 $\mu\text{g/L}$); $F_{d,t,gw,n,\ell}$ and $F'_{d,t,gw,n,\ell}$ are the fraction of the total district population exposed to groundwater arsenic exposure level ℓ for actually current situation and ideal situation, respectively; and n is the number of the exposure model for calculating fraction of the population drinking groundwater. The error boundaries of the PAF values were estimated by 1000 Monte Carlo simulations.

The PAF of groundwater arsenic-attributable CVD mortality was combined with published data [53] on CVD mortality in India on a state-wise basis to quantify the state-wise annual CVD mortality due to consuming high groundwater arsenic. Finally, the total annual hospitalization costs due to high groundwater arsenic-attributable CVD mortality were estimated for the whole of country based on the estimated PAFs, cost of hospitalization [54], prevalent cases [55], and deaths [55] for CVDs. Value of statistical life (VSL)-based annual costs of arsenic-attributable CVD mortality were estimated for India based on our estimated CVD mortality, reported Indian individual VSL [56], and estimates of years of life lost due to heart diseases [57,58].

3. Results and Discussion

3.1. Random Forest Models for Hazard Maps

The assembled dataset of 24,297 averaged arsenic concentration with potential environmental predictors were utilized to create four random forest models of probability of groundwater arsenic exceeding 10, 20, 50, and 80 $\mu\text{g/L}$ for the whole of India. The optimal numbers with the smallest out-of-bag (OOB) error of predictors at individual tree branch in random forest models were 16, 18, 22, and 21, separately (Table S3). There was no predictor with negative mean decreases in both accuracy and Gini node impurity, so all predictors benefited the models and so were retained in the models. The cross-validation of the models on their testing dataset showed that the area (AUC) under the ROC curves were 0.84, 0.84, 0.83, and 0.83, indicating good model prediction performance (Figure S2). The importance of the predictors in the four random forest models is shown in Figure S3, showing that some climate (e.g., temperature, actual evapotranspiration, aridity, precipitation) and soil (e.g., fluvisols) predictors tend to have relatively higher importance and rank than others.

The probability maps (Figure 2a–d) of arsenic concentration exceeding 10, 20, 50, and 80 $\mu\text{g/L}$ in groundwaters were transformed into a pseudo-contour map (Figure 2e) of groundwater arsenic concentrations using a probability cutoff value of 0.5. The concentration map captured a significantly high hazard of groundwater arsenic in the north and northeast India, particularly in Assam and West Bengal, mainly consistent with previously well-known documented high hazard areas [22]. The district-wise proportions of areas with various groundwater arsenic concentration ranges were calculated (Figure 3, and Table S4), showing over 50% of areas of several districts in the Ganges and Brahmaputra

alluvial plains, in Punjab, and in less known areas of Haryana and NCT of Delhi with groundwater arsenic greater than $10 \mu\text{g/L}$, the provisional guideline value of arsenic in drinking water recommended by the World Health Organization (WHO) [59].

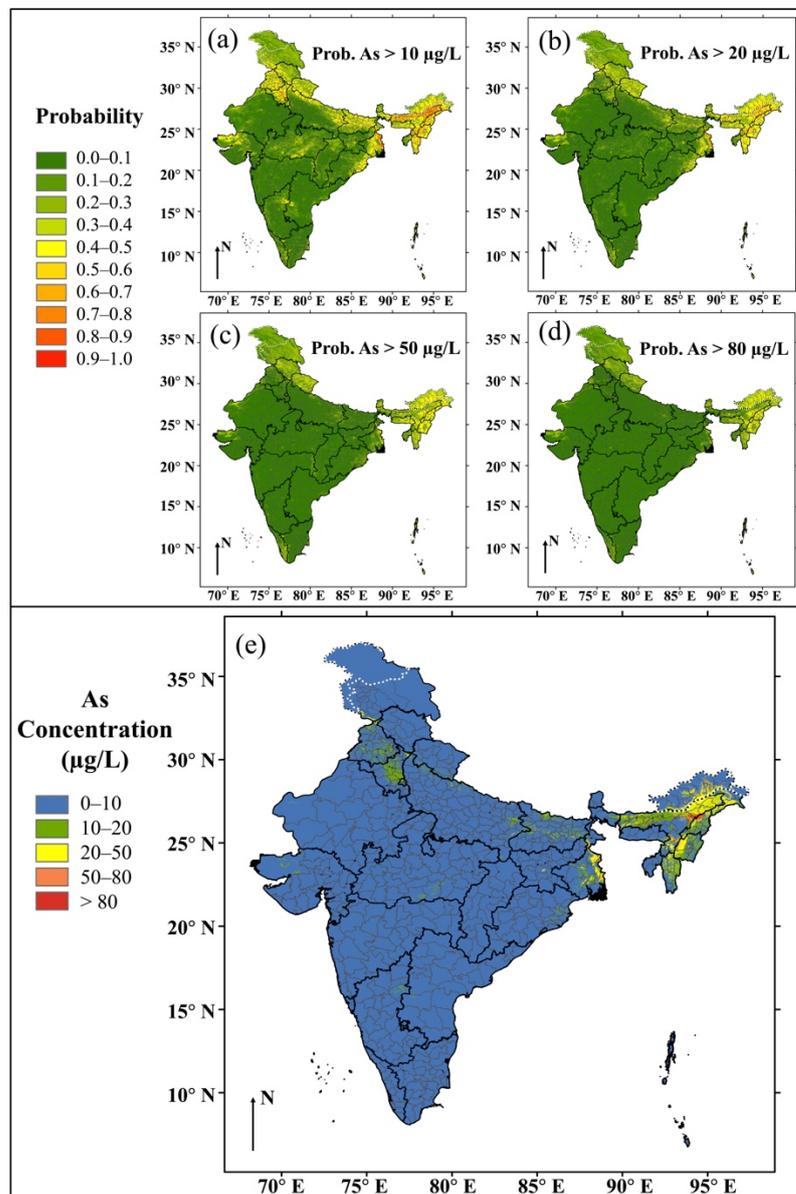


Figure 2. Modelled hazard maps of groundwater arsenic in India. Probability of groundwater arsenic exceeding (a) $10 \mu\text{g/L}$, (b) $20 \mu\text{g/L}$, (c) $50 \mu\text{g/L}$, and (d) $80 \mu\text{g/L}$. (e) Pseudo-contour map of groundwater arsenic concentrations defined here using a probability cutoff value of 0.5.

The district-wise statistics of modelled high groundwater arsenic areas were associated with greater errors than those at the state and country levels; however, most of those districts having high groundwater arsenic listed by India Central Ground Water Board (CGWB) [60] were captured by our models. The produced random forest models are only 2D models, not 3D models, and common limitations of such 2D models are illustrated by Podgorski et al. [22]), and Wu et al. [20,61]. Clearly, in areas where there are multiple aquifers, substantial differences between the arsenic concentrations between aquifers may exist, indicating the requirement for detailed local studies and tests. However, the models here may broadly reflect the relative sampling densities in these different aquifers, so they could arguably be used to estimate current exposures. However, as the balance

of sampling/supply between different aquifers changes, then these 2D models would necessarily become dated and unrepresentative.

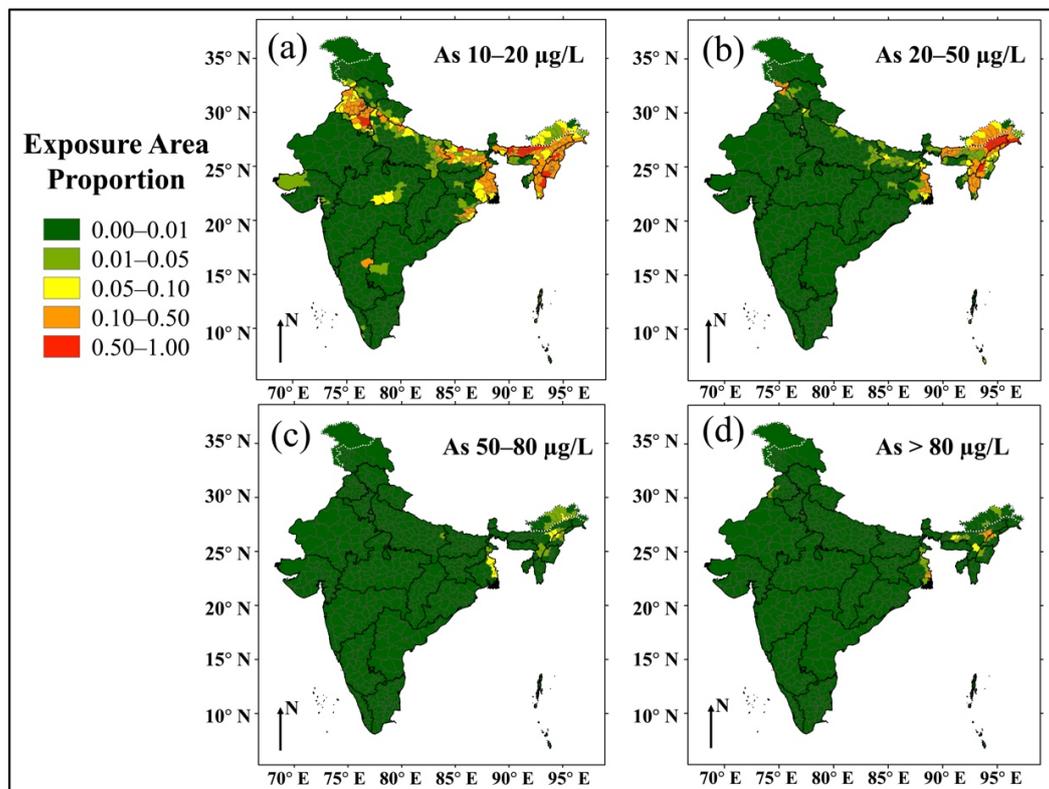


Figure 3. District-wise distribution of modelled percentage of district areas with groundwater arsenic in the concentration ranges of (a) 10 to 20 $\mu\text{g/L}$, (b) 20 to 50 $\mu\text{g/L}$, (c) 50 to 80 $\mu\text{g/L}$, and (d) >80 $\mu\text{g/L}$.

3.2. Population Exposure Maps

The district-wise total/rural/urban population exposed to arsenic concentration ranges of 10–20, 20–50, 50–80, and >80 $\mu\text{g/L}$ by drinking groundwater are shown in Figure 4 (exposure model 1), Figure S4 (exposure model 2), and Table S4 (exposure model 1). In India, totally, we estimated that between 58 (exposure model 1) and 64 (exposure model 2) million inhabitants are exposed to arsenic exceeding 10 $\mu\text{g/L}$ in drinking water. This estimation is higher than the estimations of 31 million people by Ravenscroft et al. [62] and 18–30 million people by Podgorski et al. [22], whilst it is lower than the estimation of 90 million people exposed to high arsenic in groundwater by Mukherjee et al. [23]. We prefer our estimate of 58 million to 64 million as it is based on a more granular and detailed assessment of district-level exposures at a range of arsenic concentration levels than the previous estimates—we also note that this estimated range is consistent with the arithmetic mean of those of Podgorski et al. [22] and Mukherjee et al. [23]. Meanwhile, about between 39.1 (exposure model 1) and 43.4 (exposure model 2) million (i.e., 3.2–3.6% of the population of India), 15.1–17.0 million (1.3–1.4%), 1.5–1.6 million (0.1%), and 2.3–2.4 million (0.2%) people are exposed to high groundwater arsenic of 10–20 $\mu\text{g/L}$, 20–50 $\mu\text{g/L}$, 50–80 $\mu\text{g/L}$, and >80 $\mu\text{g/L}$, respectively, with more rural populations (32.9–36.6 million for 10–20 $\mu\text{g/L}$; 12.3–13.8 million for 20–50 $\mu\text{g/L}$; 1.2 million for 50–80 $\mu\text{g/L}$; 1.6–1.7 million for >80 $\mu\text{g/L}$) currently consuming high arsenic in groundwaters than urban populations (6.1–6.8 million for 10–20 $\mu\text{g/L}$; 2.9–3.1 million for 20–50 $\mu\text{g/L}$; 0.3 million for 50–80 $\mu\text{g/L}$; 0.6 million for >80 $\mu\text{g/L}$). A large number of districts in Assam, Bihar, and West Bengal states had a large population (>1 million people) drinking groundwater with high arsenic concentration (>10 $\mu\text{g/L}$). The population exposures were based on two contrasting end-member assumptions as to the arsenic status of shallow wells: (exposure model 1) shallow dug wells

are largely groundwater arsenic-free and (exposure model 2) shallow dug wells contain groundwater arsenic, posing a threat to human health. Thus, the range of values presented here reflect in part uncertainties in the relative validities of exposure Models 1 and 2. It is recognized both that (i) the validity of the models may be a function of geology and consumer behavior and (ii) the validity of the models may change with time, indeed may have changed materially since 2011, the date of the basis of the district-wise rural/urban population data used in this study [49,50].

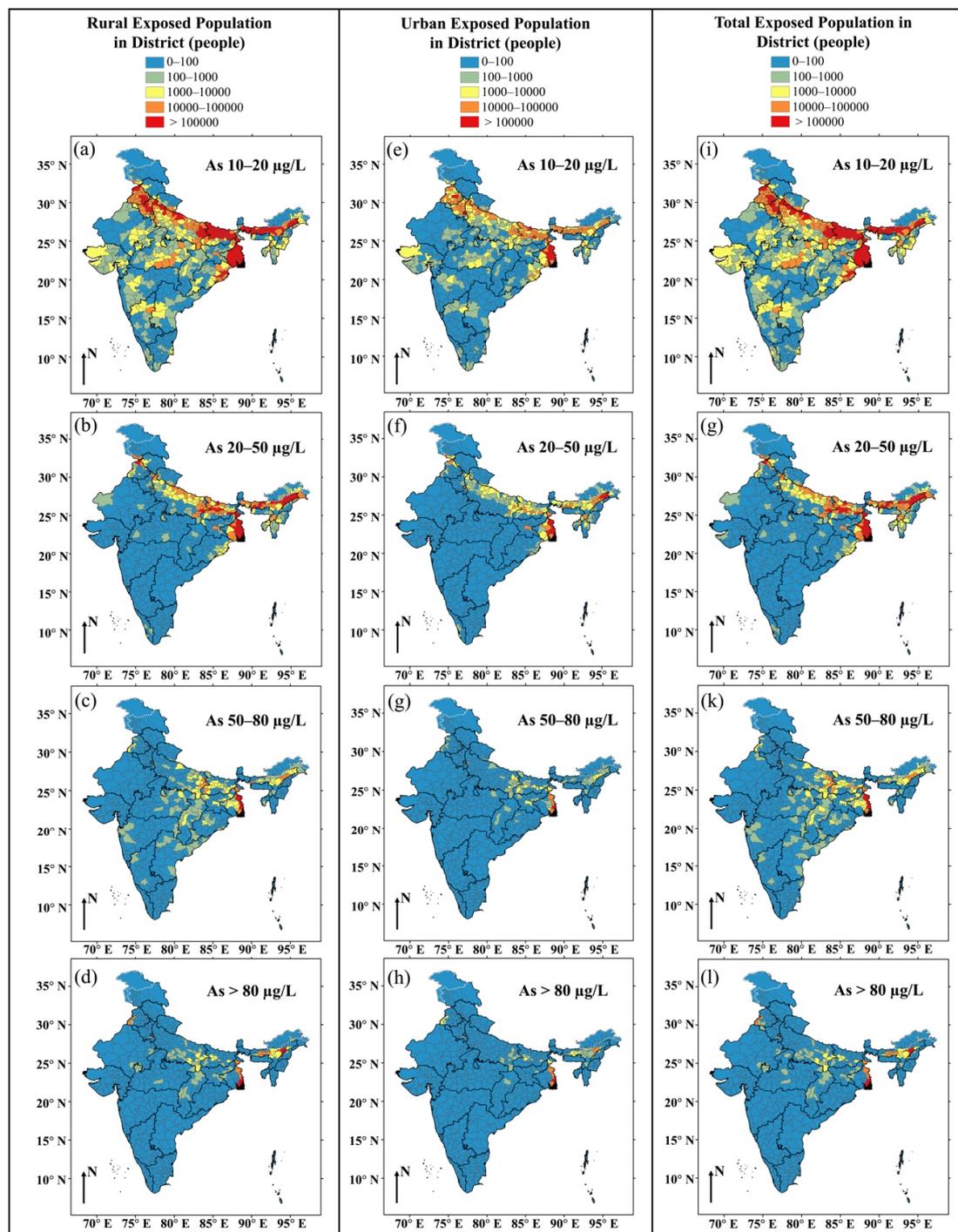


Figure 4. Maps of modelled district-wise population exposure to groundwater arsenic in specific concentration ranges and based on exposure model 1 (see text). District-wise rural (a–d), urban (e–h), and total (i–l) population exposed to arsenic concentration ranges of 10 to 20 µg/L, 20 to 50 µg/L, 50 to 80 µg/L, and >80 µg/L by drinking groundwater.

3.3. Pooled Relative Risks of Cardiovascular Disease (CVD) Related to Arsenic Concentrations in Drinking Water

The utilized linear and non-linear positive associations between mortality risk of cardiovascular diseases (CVDs) and arsenic concentrations in drinking water with significant overall trends ($p < 0.05$) were justified based on the recent meta-analysis by Xu et al. [24] (Table 1). The relative mortality risks of CVDs of arsenic concentration ranges of 10–20, 20–50, 50–80, and >80 $\mu\text{g/L}$ compared with the reference level of 0–10 $\mu\text{g/L}$ are listed in Table 1. The association models have great heterogeneity due to considerable variations of the selected studies in terms of study area, study design, exposure assessment, ascertainment of outcome, and adjustment of confounders.

Table 1. Dose–response relationships between cardiovascular disease mortality and arsenic concentration in drinking water. Pooled relative risks (95% CIs) for cardiovascular disease (CVDs) mortality for the median concentration in each arsenic concentration category relative to the median concentration in the reference category of 0–10 $\mu\text{g/L}$. Calculations after Xu et al. [24]. Linear model: Coefficient for trend: 0.070; p -value for trend: 0.005; deviance: 22.58; AIC: -6.77 ; R^2 : 0.258; adjusted R^2 : 0.225; I^2 : 77.9%; Cochran’s Q-statistic: 31.70; P-heterogeneity: <0.001 . Non-linear model: p -value for trend: <0.001 ; deviance: 22.81; AIC: 5.89; R^2 : 0.620; adjusted R^2 : 0.584; I^2 : 35.3%; Cochran’s Q-statistic: 21.65; P-heterogeneity: 0.086.

Arsenic Concentration Category ($\mu\text{g/L}$)	Median Concentration ($\mu\text{g/L}$)	CVD Mortality Risk (95% CI)	
		Linear Model	Non-Linear Model
[0, 10]	5	1.000	1.000
(10, 20]	15	1.079 (1.023–1.139)	1.028 (1.016–1.040)
(20, 50]	35	1.145 (1.041–1.259)	1.086 (1.057–1.116)
(50, 80]	65	1.195 (1.055–1.355)	1.154 (1.102–1.208)
(80, $+\infty$]	131	1.255 (1.071–1.472)	1.252 (1.163–1.348)

Although the dose–response relationship utilized does not explicitly take into account any of the factors known to be strongly associated with CVD (for example, gender, age, obesity, smoking habits, and income), the PAF approach used in this study does mean that these factors are largely implicitly accounted for in the baseline CVD mortality risks. Despite this, in due course, we still expect more refined dose–response models to be published, and the estimates should therefore be accordingly updated to take these into account.

3.4. Cardiovascular Disease (CVD) Health Risk Estimation

CVD mortality risks arising from drinking groundwater with high arsenic (>10 $\mu\text{g/L}$) in India was quantified using the population-attributable fraction (PAF) approach. For the whole of country, the PAF of groundwater arsenic-attributable CVD mortality was estimated to be between 0.26% (exposure model 1, non-linear dose–response relationship) and 0.51% (exposure model 1, linear dose–response relationship) or between 0.29 (exposure model 2, non-linear dose–response relationship) and 0.56% (exposure model 2, linear dose–response relationship), of which, PAFs of 0.09–0.26% for exposure model 1 (or 0.10–0.28% for exposure model 2), 0.11–0.18% (or 0.12–0.20%), 0.02% (or 0.02–0.03%), and 0.05% (or 0.05%) were contributed by arsenic concentration categories of 10–20, 20–50, 50–80, and >80 $\mu\text{g/L}$, respectively. Thus about 0.3–0.6% of CVD mortality in India could be eliminated if current exposure to groundwater with arsenic > 10 $\mu\text{g/L}$, the WHO provisional guideline value for drinking water arsenic, were eliminated.

Based upon the PAF estimates by two exposure models and linear and non-linear dose–response relationships, recent annual hospitalization cost of USD 14.4 billion [54], prevalent cases of 54.5 million [55], and deaths of 2.8 million [55] for CVDs, annual total hospitalization costs due to high groundwater arsenic-attributable CVD mortality were estimated to be USD 1.95–3.74 million (exposure model 1) or USD 2.14–4.13 million (exposure model 2) in India. Meanwhile, the value of statistical life (VSL)-based annual costs of arsenic-attributable CVD mortality were estimated to be USD 749–3127 million (exposure model 1) or USD 813–3408 million (exposure model 2) in India, according to our estimated

mortality, reported individual VSL of USD 0.64 million [56], and year of life lost due to heart diseases per death of 12 years [58] or 28 years [57]. The produced VSL is much higher than our estimated hospitalization costs—this may be because many people, notably in poorer rural areas, do not get hospital treatments or only get limited and incomplete treatments for their CVD. Although CVD cases that are untreated by hospitals do not give rise to hospitalization costs, they still impact the country's economy, this being more accurately reflected in VSL-based cost estimates. Lastly, on this point, we note that the estimated costs here are only in relation to CVD mortality—costs related otherwise to morbidity, though not quantified here, are likely also to be very substantial.

At the state level, the PAFs (Tables S5 and S6) of groundwater arsenic-attributable CVD mortality were calculated to be highest in Assam (2.5% to 5.7%) and West Bengal (1.6% to 2.7%). The PAFs of most entire states were estimated to be below 1%. Combined with recently reported disease burden of CVDs for India [53], the annual state-wise CVD mortalities due to consuming high groundwater arsenic were estimated and are listed in Table S7 (exposure model 1) and Table S8 (exposure model 2). For the whole of India, in total, the modelled annual CVD mortality attributable to chronic consumption of high groundwater arsenic was estimated to be about 6510–12,132 (exposure model 1) or 7067–13,224 (exposure model 2). The country-wise cost and state-wise mortality estimates were completed based on reported data of from somewhat different years (PAF: 2011; hospitalization cost, death, and cases: 2016, disease burden: 2019), which may cause some estimation bias; however, the magnitude of this bias is considerably less than the combined uncertainties arising from hazard distribution, exposure and dose–response models, and documented all-cause CVD disease burden in India.

At a district level (Figures 5 and S5 and Table S4), the estimated PAF of groundwater arsenic-attributable CVD mortality was as high as between 6.0 (exposure model 1, non-linear dose–response relationship) and 9.8% (exposure model 1, linear dose–response relationship) or between 6.5 (exposure model 2, non-linear dose–response relationship) and 10.7% (exposure model 2, linear dose–response relationship) in Dhemaji district (Assam), indicating that drinking high arsenic groundwater is likely to be one of most important factors causing CVD mortality in some Indian districts. Relatively higher PAF (>3%, about 5 or 10 times the country-wise overall PAF) was identified in some further districts in Arunachal Pradesh, Assam, Bihar, Punjab, Nagaland, and West Bengal states. This is consistent with the recorded distribution of well-known relatively higher groundwater arsenic [22]. District-wise PAF estimates exhibit a wider range of modelled CVD mortality attributable to drinking high arsenic groundwater than state-wise PAF estimations, however, in part because of the greater number of districts than state, but further it is reiterated that the reliability of the model decreases as it becomes more granular. Overall modelled mean hazard and ultimately CVD risk for the whole of India are associated with smaller relative errors than for the same estimates at the state level, more so at the district level, whilst very substantial errors might be expected if the model were used to predict arsenic concentrations at one well/household level, an application of the model that we do not recommend unless supplemented by well-specific testing.

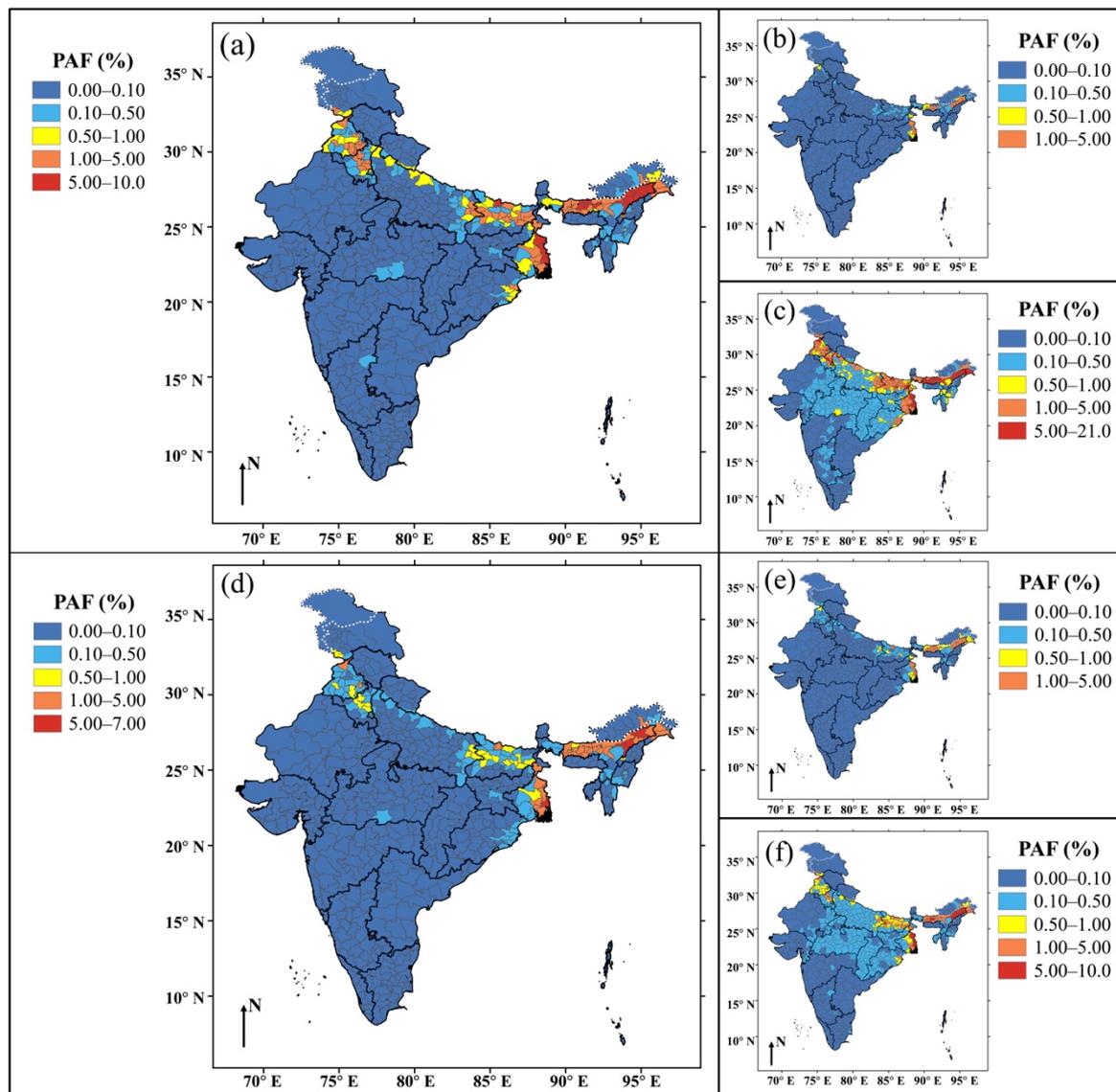


Figure 5. District-level maps of India showing (a,d) population-attributable fraction (PAF) of high groundwater arsenic-attributable CVD mortality with error boundaries (lower (b,e) and upper (c,f) limits) based on exposure model 1 (see text). PAF estimations were based on both linear (a–c) and non-linear (d–f) dose–response relationships between cardiovascular disease mortality and arsenic concentration in drinking water.

4. Conclusions

By combining machine learning models of groundwater arsenic distribution with basic exposure route models and CVD-specific dose–response models, we provide for the first time an estimate of both the magnitude and distribution of groundwater arsenic-attributable CVD mortality risks in India. The data were mapped at country, state, and district scales, although the higher resolution (e.g., district-level) mapping and modelling is likely to be associated with higher relative model errors.

As the basis of this, we generated for the first time a pseudo-contour map of model groundwater arsenic concentrations for the whole of India, based upon the machine learning approach of Podgorski et al. [22] and the pseudo-contour approach of Wu et al. [20]. Incorporation of CVD dose–response relationships after a recent meta-analysis by Xu et al. [24], using the PAF approach rather than an absolute risk vs. dose–response (cf. Mondal et al. [31]) has enabled major spatially variable drivers of CVD in India, to a large extent, to be implicitly taken into account in the health-risk modelling.

On a country-wise basis, groundwater arsenic is calculated to contribute to around 0.3–0.6% of all CVD mortality, but this figure is much higher (>3%) for many individual districts in Arunanchal Pradesh, Assam, Bihar, Punjab, Nagaland, and West Bengal and indicates that the continued focusing of government and NGO resources on mitigation of groundwater-derived drinking water supplies in these areas is indicated to protect public health.

On a country-wise basis, we estimate around 6500 to 13,000 premature avoidable CVD deaths may be ascribed to chronic exposure from groundwater-derived drinking water supplies with high arsenic concentrations. The cost of this on a VSL (value of a statistical life) basis is conservatively estimated to be at least on the order of USD 750 million to USD 3400 million per annum. These costs could be used to inform what magnitude of investment might reasonably be warranted to mitigate high arsenic groundwater-derived drinking water supplies in India.

Lastly, we note that the models presented are also limited both in terms of data availability and the nature of the models themselves. In particular, (i) the hazard models are dependent on the quality of the published data upon which they are based [20,22,61]; (ii) the geospatial hazard models produced here are 2D models, not 3D models, that do not consider comprehensively the often substantial differences between the arsenic concentrations in shallow aquifers often exploited via dug wells and deeper aquifers often exploited by tube wells or boreholes [22,61]; (iii) in the absence of more detailed readily available public domain data, the exposure models make plausible but untested assumptions about the distribution of groundwater arsenic concentrations within districts and in particular between areas dominated by rural and urban populations; (iv) the data on district-wise drinking water supply resources and total/rural/urban populations were produced in 2011 [49,50] and so the exposure and health risk estimates may be dated as a result of changes in population distribution and water supply type since that time; (v) model imprecision is likely to be greater at higher spatial resolutions—thus country-wise estimates are likely to be more precise than, in turn, state-wise and district-wise estimates, and extending the model to individual wells and households is not recommended—indeed, for the latter, any material decision-making would be best advised to be based on due consideration by appropriate professionals of individual/household level circumstances and data.

Supplementary Materials: The following are available online at <https://www.mdpi.com/article/10.3390/w13162232/s1>, Figure S1: Frequency distribution of averaged groundwater arsenic concentrations used in this study, Figure S2: Cross-validation results. ROC curves and AUC values of random forest models of groundwater arsenic exceeding (a) 10, (b) 20, (c) 50, and (d) 80 µg/L, Figure S3: Normalized importance of predictors in terms of mean decrease values in accuracy and in Gini node impurity in the random forest models of groundwater arsenic exceeding (a) 10, (b) 20, (c) 50, and (d) 80 µg/L. Both mean decrease in accuracy and mean decrease in Gini node impurity were normalized by their largest values, respectively, Figure S4: Maps of modelled district-wise population exposure to groundwater arsenic in specific concentration ranges and based on exposure model 2 (see text). District-wise rural (a–d), urban (e–h), and total (i–l) population exposed to arsenic concentration ranges of 10 to 20 µg/L, 20 to 50 µg/L, 50 to 80 µg/L, and >80 µg/L by drinking groundwater, Figure S5: District-level maps showing (a,d) population-attributable fraction (PAF) of high groundwater arsenic-attributable CVD mortality with error boundaries (lower (b,e) and upper (c,f) limits), based on exposure model 2 (see text). PAF estimations were based on both linear (a–c) and non-linear (d–f) dose–response relationships between cardiovascular disease mortality and arsenic concentration in drinking water, Table S1: Groundwater arsenic concentration measurements in random forest models. Existing arsenic measurements taken from over 50 sources, mainly from India but also some neighboring South Asian countries. Summaries are given for before and after spatial averaging (geometric mean of arsenic concentrations within a 1 km² pixel), Table S2: Potential predictors used in the machine learning models. Brief descriptions and data sources are listed, Table S3: Comparison of out-of-bag (OOB) errors as a function of number of available predictors at each branch in random forest models (note 31 is the total number of potential predictors). The number

shown in grey shadow is the optimum number of predictors at each branch in the model, Table S4: Modelled district-wise exposures and cardiovascular disease (CVD) health risk. District-wise rural, urban, and total population exposed to arsenic concentration ranges of 10 to 20, 20 to 50, 50 to 80, and >80 µg/L by drinking groundwater. Population-attributable fraction (PAF) of high groundwater arsenic-attributable CVD mortality. The estimations were based on exposure model 1 regarding water from hand pumps and tube wells/boreholes as groundwater, and both linear and non-linear dose–response relationships between cardiovascular disease mortality and arsenic concentration in drinking water, Table S5: Modelled state-wise cardiovascular disease (CVD) health risk. State-wise rural, urban, and total population exposed to arsenic concentration ranges of 10 to 20, 20 to 50, 50 to 80, and >80 µg/L by drinking groundwater. Population-attributable fraction (PAF) of high groundwater arsenic-attributable CVD mortality. The estimations were based on exposure model 1 regarding water from hand pumps and tube wells/boreholes as groundwater, and both linear and non-linear dose–response relationships between cardiovascular disease mortality and arsenic concentration in drinking water, Table S6: Modelled state-wise cardiovascular disease (CVD) health risk. State-wise rural, urban, and total population exposed to arsenic concentration ranges of 10 to 20, 20 to 50, 50 to 80, and >80 µg/L by drinking groundwater. Population-attributable fraction (PAF) of high groundwater arsenic-attributable CVD mortality. The estimations were based on exposure model 2 considering water from wells (covered + uncovered), hand pumps, and tube wells/boreholes as groundwater, and both linear and non-linear dose–response relationships between cardiovascular disease mortality and arsenic concentration in drinking water, Table S7: India state-wise modelled annual CVD mortality attributable to chronic consumption of high groundwater arsenic in India based on exposure model 1, including total CVD mortality for all high arsenic concentrations and CVD mortality for four different arsenic concentration categories (10–20, 20–50, 50–80, and >80 µg/L). These mortality estimations were based on both linear and non-linear dose–response relationships between cardiovascular disease mortality and arsenic concentration in drinking water. By definition only there is no modelled arsenic-attributable CVD mortality in the reference arsenic concentration category. Exposure model 1 assumes that modelled groundwater arsenic hazard data apply to and exposure arises from only hand pumps, tube wells, and boreholes. Although modelled mortalities for model 1 are somewhat lower than for model 2, these differences are of the same order as the uncertainties in baseline all-cause CVD mortality data for the states, Table S8: India state-wise modelled annual CVD mortality attributable to chronic consumption of high groundwater arsenic in India based on exposure model 2, including total CVD mortality for all high arsenic concentrations and CVD mortality for four different arsenic concentration categories (10–20, 20–50, 50–80, and >80 µg/L). These mortality estimations were also based on both linear and non-linear dose–response relationships between cardiovascular disease mortality and arsenic concentration in drinking water. By definition only there is no modelled arsenic-attributable CVD mortality in the reference arsenic concentration category. Exposure model 2 assumes that modelled groundwater arsenic hazard data apply to and exposure arises from covered/uncovered hand pumps, tube wells, and boreholes. Although modelled mortalities for model 1 are somewhat lower than for model 2, these differences are of the same order as the uncertainties in baseline all-cause CVD mortality data for the states.

Author Contributions: Conceptualization, D.A.P., R.W. and L.X.; methodology, software, validation, formal analysis in relation to As–CVD dose–response relationships: L.X.; all other methodology, software, validation, formal analysis, R.W.; data curation, R.W.; writing—original draft preparation, R.W.; writing—review and editing R.W., D.A.P. and L.X.; supervision and project administration, D.A.P.; funding acquisition, D.A.P. and L.X. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded in part by the Newton Fund, Natural Environmental Research Council (UK) (NE/R003386/1), and Department of Science and Technology (India) (DST/TM/INDO-UK/2K17/55(C) & 55(G)) and a University of Manchester President’s Doctoral Scholarship to L.X.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data presented in this study not otherwise available from the references and organizations indicated in the text may be available on request from the corresponding authors.

Acknowledgments: We thank and acknowledge the Newton Fund, NERC (UK), and DST (India) for their joint funding of the Indo-UK Water Quality project, FAR-GANGA (www.farganga.org, (accessed on 20 February 2021)). L.X. acknowledges with thanks the receipt of a University of Manchester President’s Doctoral Scholarship. We thank Biswajit Chakravorty, Himanshu Joshi, Debapriya Mondal, Abhijit Mukherjee, Joel Podgorski, Peter Ravenscroft, and Dipankar Saha for discussions. The opinions expressed in this paper however do not necessarily reflect those of any of the organizations or individuals whom we acknowledge here. This paper is dedicated to the memory of Sanjay Bajpai, former Head of Technology Mission Division at the DST and to whom we are indebted for his support and advice for the Indo-UK FAR-GANGA project and more widely for the development of applied research in water and energy in India (<https://www.psa.gov.in/article/memory-dr-sanjay-bajpai/2868>, (accessed on 20 February 2021)).

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results. We note that no jurisdictional claims are made or implied in the boundaries presented in figures in this article.

References

- Pinto, S.S.; Bennett, B. Effect of arsenic trioxide exposure on mortality. *Arch. Environ. Health Int. J.* **1963**, *7*, 583–591. [[CrossRef](#)]
- Currie, A. The role of arsenic in carcinogenesis. *Br. Med. Bull.* **1947**, *4*, 402–405. [[CrossRef](#)]
- Rodell, M.; Velicogna, I.; Famiglietti, J.S. Satellite-based estimates of groundwater depletion in India. *Nature* **2009**, *460*, 999. [[CrossRef](#)]
- Famiglietti, J.S. The global groundwater crisis. *Nat. Clim. Chang.* **2014**, *4*, 945–948. [[CrossRef](#)]
- Polya, D.A.; Sparrenbom, C.; Datta, S.; Guo, H. Groundwater arsenic biogeochemistry—Key questions and use of tracers to understand arsenic-prone groundwater systems. *Geosci. Front.* **2019**, *10*, 1635–1641. [[CrossRef](#)]
- Mondal, D.; Polya, D.A. Rice is a major exposure route for arsenic in Chakdaha block, Nadia district, West Bengal, India: A probabilistic risk assessment. *Appl. Geochem.* **2008**, *23*, 2987–2998. [[CrossRef](#)]
- Williams, P.N.; Islam, M.R.; Adomako, E.E.; Raab, A.; Hossain, S.A.; Zhu, Y.G.; Feldmann, J.; Meharg, A.A. Increase in rice grain arsenic for regions of Bangladesh irrigating paddies with elevated arsenic in groundwaters. *Environ. Sci. Technol.* **2006**, *40*, 4903–4908. [[CrossRef](#)] [[PubMed](#)]
- Dittmar, J.; Voegelin, A.; Roberts, L.C.; Hug, S.J.; Saha, G.C.; Ali, M.A.; Badruzzaman, A.B.M.; Kretzschmar, R. Spatial distribution and temporal variability of arsenic in irrigated rice fields in Bangladesh. 2. Paddy soil. *Environ. Sci. Technol.* **2007**, *41*, 5967–5972. [[CrossRef](#)]
- Polya, D.A.; Middleton, D.R. Arsenic in drinking water: Sources & human exposure. In *Best Practice Guide on the Control of Arsenic in Drinking Water*, 1st ed.; Bhattacharya, P., Polya, D.A., Draganovic, D., Eds.; International Water Association Publishing: London, UK, 2017; Chapter 1; ISBN 9781843393856.
- Ferreccio, C.; Smith, A.H.; Duran, V.; Barlaro, T.; Benitez, H.; Valdes, R.; Aguirre, J.J.; Moore, L.E.; Acevedo, J.; Vasquez, M.I.; et al. Case-Control Study of Arsenic in Drinking Water and Kidney Cancer in Uniquely Exposed Northern Chile. *Am. J. Epidemiol.* **2013**, *178*, 813–818. [[CrossRef](#)] [[PubMed](#)]
- Chen, Y.; Graziano, J.H.; Parvez, F.; Liu, M.; Slavkovich, V.; Kalra, T.; Argos, M.; Islam, T.; Ahmed, A.; Rakibuzzaman, M. Arsenic exposure from drinking water and mortality from cardiovascular disease in Bangladesh: Prospective cohort study. *Brit. Med. J.* **2011**, *342*, d2431. [[CrossRef](#)]
- Farzan, S.F.; Chen, Y.; Rees, J.R.; Zens, M.S.; Karagas, M.R. Risk of death from cardiovascular disease associated with low-level arsenic exposure among long-term smokers in a US population-based study. *Toxicol. Appl. Pharmacol.* **2015**, *287*, 93–97. [[CrossRef](#)] [[PubMed](#)]
- Ahmed, K.M.; Bhattacharya, P.; Hasan, M.A.; Akhter, S.H.; Alam, S.M.; Bhuyian, M.H.; Iman, M.B.; Khan, A.A.; Sracek, O. Arsenic enrichment in groundwater of the alluvial aquifers in Bangladesh: An overview. *Appl. Geochem.* **2004**, *19*, 181–200. [[CrossRef](#)]
- Chakraborti, D.; Rahman, M.M.; Das, B.; Nayak, B.; Pal, A.; Sengupta, M.K.; Hossain, M.A.; Ahamed, S.; Sahu, M.; Saha, K.C.; et al. Groundwater arsenic contamination in Ganga–Meghna–Brahmaputra plain, its health effects and an approach for mitigation. *Environ. Earth Sci.* **2013**, *70*, 1993–2008. [[CrossRef](#)]
- McArthur, J.M.; Ravenscroft, P.; Safiulla, S.; Thirlwall, M.F. Arsenic in groundwater: Testing pollution mechanisms for sedimentary aquifers in Bangladesh. *Water Resour. Res.* **2001**, *37*, 109–117. [[CrossRef](#)]
- Islam, F.S.; Gault, A.G.; Boothman, C.; Polya, D.A.; Charnock, J.M.; Chatterjee, D.; Lloyd, J.R. Role of metal-reducing bacteria in arsenic release from Bengal delta sediments. *Nature* **2004**, *430*, 68–71. [[CrossRef](#)] [[PubMed](#)]
- Berg, M.; Stengel, C.; Trang, P.T.K.; Viet, P.H.; Sampson, M.L.; Leng, M.; Samreth, S.; Fredericks, D. Magnitude of arsenic pollution in the Mekong and Red River Deltas—Cambodia and Vietnam. *Sci. Total Environ.* **2007**, *372*, 413–425. [[CrossRef](#)] [[PubMed](#)]
- Guo, H.; Zhang, B.; Li, Y.; Berner, Z.; Tang, X.; Norra, S.; Stüben, D. Hydrogeological and biogeochemical constrains of arsenic mobilization in shallow aquifers from the Hetao basin, Inner Mongolia. *Environ. Pollut.* **2011**, *159*, 876–883. [[CrossRef](#)]

19. Mladenov, N.; Zheng, Y.; Simone, B.; Bilinski, T.M.; McKnight, D.M.; Nemergut, D.; Radloff, K.A.; Rahman, M.M.; Ahmed, K.M. Dissolved organic matter quality in a shallow aquifer of Bangladesh: Implications for arsenic mobility. *Environ. Sci. Technol.* **2015**, *49*, 10815–10824. [[CrossRef](#)]
20. Wu, R.; Podgorski, J.; Berg, M.; Polya, D.A. Geostatistical model of the spatial distribution of arsenic in groundwaters in Gujarat State, India. *Environ. Geochem. Health* **2020**, *43*, 2649–2664. [[CrossRef](#)]
21. Bindal, S.; Singh, C.K. Predicting groundwater arsenic contamination: Regions at risk in highest populated state of India. *Water Res.* **2019**, *159*, 65–76. [[CrossRef](#)]
22. Podgorski, J.E.; Wu, R.; Chakravorty, B.; Polya, D.A. Groundwater Arsenic Distribution in India by Machine Learning Geospatial Modeling. *Int. J. Environ. Res. Public Health* **2020**, *17*, 7119. [[CrossRef](#)]
23. Mukherjee, A.; Sarkar, S.; Chakraborty, M.; Duttagupta, S.; Bhattacharya, A.; Saha, D.; Bhattacharya, P.; Mitra, A.; Gupta, S. Occurrence, predictors and hazards of elevated groundwater arsenic across India through field observations and regional-scale AI-based modeling. *Sci. Total Environ.* **2021**, *759*, 143511. [[CrossRef](#)]
24. Xu, L.; Mondal, D.; Polya, D.A. Positive association of Cardiovascular Disease (CVD) with chronic exposure to drinking water Arsenic (As) at concentrations below the WHO provisional guideline value: A systematic review and meta-analysis. *Int. J. Environ. Res. Public Health* **2020**, *17*, 2536. [[CrossRef](#)]
25. Rahman, M.; Sohel, N.; Yunus, M.; Chowdhury, M.E.; Hore, S.K.; Zaman, K.; Bhuiya, A.; Streatfield, P.K. A prospective cohort study of stroke mortality and arsenic in drinking water in Bangladeshi adults. *BMC Public Health* **2014**, *14*, 174. [[CrossRef](#)] [[PubMed](#)]
26. Wu, F.; Jasmine, F.; Kibriya, M.G.; Liu, M.; Wójcik, O.; Parvez, F.; Rahaman, R.; Roy, S.; Paulbrutus, R.; Segers, S. Association Between Arsenic Exposure from Drinking Water and Plasma Levels of Cardiovascular Markers. *Am. J. Epidemiol.* **2012**, *175*, 1252–1261. [[CrossRef](#)] [[PubMed](#)]
27. Chen, Y.; Wu, F.; Liu, M.L.; Parvez, F.; Slavkovich, V.; Eunos, M.; Ahmed, A.; Argos, M.; Islam, T.; Rakibuz-Zaman, M.; et al. A Prospective Study of Arsenic Exposure, Arsenic Methylation Capacity, and Risk of Cardiovascular Disease in Bangladesh. *Environ. Health Perspect.* **2013**, *121*, 832–838. [[CrossRef](#)] [[PubMed](#)]
28. Islam, M.S.; Mohanto, N.C.; Karim, M.R.; Aktar, S.; Hoque, M.M.; Rahman, A.; Jahan, M.; Khatun, R.; Aziz, A.; Salam, K.A.; et al. Elevated concentrations of serum matrix metalloproteinase-2 and-9 and their associations with circulating markers of cardiovascular diseases in chronic arsenic-exposed individuals. *Environ. Health* **2015**, *14*, 92. [[CrossRef](#)] [[PubMed](#)]
29. National Research Council. *Critical Aspects of EPA's IRIS Assessment of Inorganic Arsenic: Interim Report*; The National Academies Press: Washington, DC, USA, 2013; p. 127, ISBN 978-0-309-29706-6.
30. Rockhill, B.; Newman, B.; Weinberg, C. Use and misuse of population attributable fractions. *Am. J. Public Health* **1998**, *88*, 15–19. [[CrossRef](#)]
31. Mondal, D.; Adamson, G.C.; Nickson, R.; Polya, D.A. A comparison of two techniques for calculating groundwater arsenic-related lung, bladder and liver cancer disease burden using data from Chakdha block, West Bengal. *Appl. Geochem.* **2008**, *23*, 2999–3009. [[CrossRef](#)]
32. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [[CrossRef](#)]
33. Ho, T.K. Random decision forests. In Proceedings of the 3rd International Conference on Document Analysis and Recognition, Montreal, QC, Canada, 14–16 August 1995; Volume 1, pp. 278–282. [[CrossRef](#)]
34. Fawcett, T. An introduction to ROC analysis. *Pattern Recognit. Lett.* **2006**, *27*, 861–874. [[CrossRef](#)]
35. Sharma, P. A Study on the Groundwater Quality of a Part of Brahmaputra Floodplain in Nalbari District Assam with Special Reference to Arsenic and Fluoride. 2011. Available online: <http://hdl.handle.net/10603/114375> (accessed on 2 May 2019).
36. Satyapal, G.K.; Mishra, S.K.; Srivastava, A.; Ranjan, R.K.; Prakash, K.; Haque, R.; Kumar, N. Possible bioremediation of arsenic toxicity by isolating indigenous bacteria from the middle Gangetic plain of Bihar, India. *Biotechnol. Rep.* **2018**, *17*, 117–125. [[CrossRef](#)] [[PubMed](#)]
37. Dubey, C.S.; Mishra, B.K.; Shukla, D.P.; Singh, R.P.; Tajbakhsh, M.; Sakhare, P. Anthropogenic arsenic menace in Delhi Yamuna flood plains. *Environ. Earth Sci.* **2012**, *65*, 131–139. [[CrossRef](#)]
38. Mishra, B.K.; Dubey, C.S.; Shukla, D.P.; Bhattacharya, P.; Usham, A.L. Concentration of arsenic by selected vegetables cultivated in the Yamuna flood plains (YFP) of Delhi, India. *Environ. Earth Sci.* **2014**, *72*, 3281–3291. [[CrossRef](#)]
39. Tirkey, P.; Bhattacharya, T.; Chakraborty, S. Arsenic and other metals in the groundwater samples of Ranchi city, Jharkhand, India. *Curr. Sci.* **2016**, *110*, 76. [[CrossRef](#)]
40. Kaur, T.; Bhardwaj, R.; Arora, S. Assessment of groundwater quality for drinking and irrigation purposes using hydrochemical studies in Malwa region, southwestern part of Punjab, India. *Appl. Water Sci.* **2017**, *7*, 3301–3316. [[CrossRef](#)]
41. Coyte, R.; Singh, A.; Furst, K.; Mitch, W.; Vengosh, A. Co-occurrence of geogenic and anthropogenic contaminants in groundwater from Rajasthan, India. *Sci. Total Environ.* **2019**, *688*, 1216–1227. [[CrossRef](#)] [[PubMed](#)]
42. Singh, C.K.; Kumar, A.; Bindal, S. Arsenic contamination in Rapti River basin, Terai region of India. *J. Geochem. Explor.* **2018**, *192*, 120–131. [[CrossRef](#)]
43. Mukherjee, A.; Fryar, A.E. Deeper groundwater chemistry and geochemical modeling of the arsenic affected western Bengal basin, West Bengal, India. *Appl. Geochem.* **2008**, *23*, 863–894. [[CrossRef](#)]

44. Nath, B.; Sahu, S.J.; Jana, J.; Mukherjee-Goswami, A.; Roy, S.; Sarkar, M.J.; Chatterjee, D. Hydrochemistry of arsenic-enriched aquifer from rural West Bengal, India: A study of the arsenic exposure and mitigation option. *Water Air Soil Pollut.* **2008**, *190*, 95–113. [[CrossRef](#)]
45. Singh, U.K.; Ramanathan, A.L.; Subramanian, V. Groundwater chemistry and human health risk assessment in the mining region of East Singhbhum, Jharkhand, India. *Chemosphere* **2018**, *204*, 501–513. [[CrossRef](#)]
46. Duggal, V.; Rani, A.; Mehra, R.; Balaram, V. Risk assessment of metals from groundwater in northeast Rajasthan. *J. Geol. Soc. India* **2017**, *90*, 77–84. [[CrossRef](#)]
47. Ghosh, M.; Pal, D.K.; Santra, S.C. Spatial mapping and modeling of arsenic contamination of groundwater and risk assessment through geospatial interpolation technique. *Environ. Dev. Sustain.* **2019**, *22*, 2861–2880. [[CrossRef](#)]
48. Raychaudhuri, S. Introduction to monte carlo simulation. In Proceedings of the 2008 Winter Simulation Conference, Miami, FL, USA, 7–10 December 2008; IEEE: Manhattan, NY, USA, 2008; pp. 91–100.
49. Government of India, Census of India: Population Enumeration Data. Available online: https://censusindia.gov.in/2011census/population_enumeration.html (accessed on 10 February 2020).
50. Government of India, Census of India: HH-6 Households by Main Source of Drinking Water and LOCATION. 2001. Available online: <https://www.censusindia.gov.in/2011census/Hlo-series/HH06.html> (accessed on 10 February 2020).
51. Chaudhuri, S.; Roy, M. Drinking water sources in India: How safe is safe? *Curr. Sci.* **2017**, *113*, 393–402. [[CrossRef](#)]
52. Quansah, R.; Armah, F.A.; Essumang, D.K.; Luginaah, I.; Clarke, E.; Marfoh, K.; Cobbina, S.J.; Nketiah-Amponsah, E.; Namuju, P.B.; Obiri, S.; et al. Association of Arsenic with Adverse Pregnancy Outcomes/Infant Mortality: A Systematic Review and Meta-Analysis. *Environ. Health Perspect.* **2015**, *123*, 412–421. [[CrossRef](#)]
53. Institute for Health Metrics and Evaluation. Global Burden of Disease Study 2019. Available online: <http://vizhub.healthdata.org/gbd-compare/> (accessed on 7 April 2021).
54. Walker, I.F.; Garbe, F.; Wright, J.; Newell, I.; Athiraman, N.; Khan, N.; Elsey, H. The economic costs of cardiovascular disease, diabetes mellitus, and associated complications in South Asia: A systematic review. *Value Health Reg. Issues* **2018**, *15*, 12–26. [[CrossRef](#)] [[PubMed](#)]
55. India State-Level Disease Burden Initiative CVD Collaborators. The changing patterns of cardiovascular diseases and their risk factors in the states of India: The Global Burden of Disease Study 1990–2016. *Lancet Glob. Health* **2018**, *6*, e1339–e1351. [[CrossRef](#)]
56. Majumder, A.; Madheswaran, S. Value of Statistical Life in India: A Hedonic Wage Approach. 2018. Available online: <http://isec.ac.in/WP%20407%20-%20Agamoni%20Majumder%20and%20Madheswaran%20-%20Final.pdf> (accessed on 10 April 2021).
57. Menon, G.R.; Singh, L.; Sharma, P.; Yadav, P.; Sharma, S.; Kalaskar, S.; Singh, H.; Adinarayanan, S.; Joshua, V.; Kulothungan, V.; et al. National burden estimates of healthy life lost in India, 2017: An analysis using direct mortality data and indirect disability data. *Lancet Glob. Health* **2017**, *7*, e1675–e1784. [[CrossRef](#)]
58. Maniecka-Bryła, I.; Pikala, M.; Bryła, M. Life years lost due to cardiovascular diseases. *Kardiol. Pol.* **2013**, *71*, 1065–1072. [[CrossRef](#)]
59. WHO/UNICEF. Arsenic Primer—Guidance on the Investigation & Mitigation of Arsenic Contamination. 2018. Available online: https://www.unicef.org/wash/files/UNICEF_WHO_Arsenic_Primer.pdf (accessed on 10 September 2020).
60. Central Ground Water Board. Ground Water Quality in Shallow Aquifers in India. 2018 CGWB, Faridabad. Available online: http://cgwb.gov.in/documents/Waterquality/GW_Quality_in_shallow_aquifers.pdf/ (accessed on 2 May 2019).
61. Wu, R.; Alvareda, E.M.; Polyá, D.A.; Blanco, G.; Gamazo, P. Distribution of Groundwater Arsenic in Uruguay Using Hybrid Machine Learning and Expert System Approaches. *Water* **2021**, *13*, 527. [[CrossRef](#)]
62. Ravenscroft, P.; Brammer, H.; Richards, K. *Arsenic Pollution: A Global Synthesis*; Wiley-Blackwell: Chichester, UK, 2009; p. 588.