This document is the accepted manuscript version of the following article: Rodríguez-Lado, L., Sun, G., Berg, M., Zhang, Q., Xue, H., Zheng, Q., & Johnson, C. A. (2013). Groundwater arsenic contamination throughout China. Science, 341(6148), 866-868. https://doi.org/10.1126/science.1237484

(revised manuscript for Science)

Groundwater arsenic contamination throughout China

Authors: Luis Rodríguez-Lado^{1†}, Guifan Sun², Michael Berg¹, Qiang Zhang², Hanbin Xue¹, Quanmei Zheng² and C. Annette Johnson¹*

Affiliations:

¹Eawag, Swiss Federal Institute of Aquatic Science and Technology, 8600 Dübendorf, Switzerland.

²Department of Occupational and Environmental Health, School of Public Health, China Medical University, No. 92 Bei Er Road, Heping District, 110001 Shenyang, China.

*Correspondence to: <u>annette.johnson@eawag.ch</u>

[†]Current Institution: Departamento de Edafoloxía e Química Agrícola, Universidade de Santiago de Compostela, Fac. de Bioloxía, Campus Vida, 15782 Santiago de Compostela (A Coruña), Spain.

Abstract

Arsenic-contaminated groundwater used for drinking in China is a health threat that was first recognized in the 1960s. However, due to the sheer size of the country, millions of groundwater wells remain to be tested in order to determine the magnitude of the problem. We developed a statistical risk model that classifies safe and unsafe areas with respect to geogenic arsenic contamination in China, using the threshold of 10 μ g L⁻¹. We estimate that 19.6 million people are at risk of being affected by the consumption of arsenic-contaminated groundwater. Although the results must be confirmed with additional field measurements, our risk model identifies numerous arsenic-affected areas and highlights the potential magnitude of this health threat in China.

One Sentence Summary

A predictive map of high levels of geogenic arsenic in groundwater systems in China reveals a potential health risk to 19.6 million people.

Main Text:

China faces groundwater quality problems of enormous proportions from both industrial and natural sources (<u>1</u>). Populations at risk of exposure to excessive levels of arsenic (As), a natural groundwater contaminant, have been continuously emerging since the 1960s (<u>2-7</u>). With the exception of Guizhou province, where endemic arsenicosis is mainly due to coal burning (<u>7</u>, <u>8</u>), most cases of water-related arsenic poisoning have

occurred in arid regions of the Northern provinces, probably because of the scarcity of clean water for cooking and drinking from sources other than groundwater.

Apart from geothermal and mining environments, two main environmental conditions are known to be linked to natural As enrichment in groundwater systems (9): i) aerobic alkaline environments in closed basins in arid and semi-arid regions, where high pH leads to alkaline desorption of As from mineral oxides, and ii) aquifers under strongly reducing conditions, where the release of As is related to reductive dissolution of As-bearing iron (hydr)oxides in sediments. Numerous studies in China have related arsenic-enriched groundwater to reducing conditions (10-16), particularly in arid regions, where such conditions in alluvial and lacustrine aquifer sediments are often concomitant with high alkalinity and/or salinity. In other areas, such as Taiwan, arsenic release has also been associated with anoxic aquifers (17). The few studies reporting arsenic measurements in oxic aquifers in China (18-20) indicate that As concentrations in oxic groundwaters are mostly below 10 μ g L⁻¹.

Between 2001 and 2005, the "Chinese National Survey Program", conducted by the Chinese Ministry of Health, tested some 445,000 wells in 20,517 villages of 292 counties (12% of all counties in China) for arsenic contamination. In almost 5% of wells, arsenic levels were higher than the Chinese standard guideline of 50 μ g L⁻¹, and about 10,000 individuals were found to be affected by arsenicosis in known and suspected endemic areas (*21*). Screening of wells has continued and it has been estimated that roughly 5.6 million people are exposed to high concentrations of arsenic in drinking water (>50 μ g L⁻¹) and that some 14.7 million are exposed to As concentrations >10 μ g L⁻¹(*22*). Due to the sheer size of China, it will take several decades to complete the screening of millions of wells to determine the spatial occurrence and magnitude of arsenic contamination throughout the country.

Here we present a predictive model that uses survey data and geological and hydrogeochemical parameters as proxies of processes that affect arsenic solubility in groundwater aquifers to pinpoint safe and unsafe areas with respect to natural or geogenic groundwater arsenic contamination. The model was validated with an independent dataset and an optimal cut-off probability value of 0.46 (SM, section 3.3), revealing fairly good agreement (Cohen's Kappa = 0.51), with 77% of the samples correctly classified in the high and low risk classes. The model sensitivity (ability to correctly classify samples with As >10 µg L⁻¹) and specificity (ability to correctly classify samples with As ≤ 10 µg L⁻¹) were 83% and 75% respectively, which indicates good performance for the prediction of both safe and unsafe areas with respect to As. The model uncertainty is predominantly low (SM, Fig. S5), but >0.5 in the sparsely inhabited sand deserts.

Previous studies in Southeast Asia have identified a relatively small number of geological and hydro-geochemical parameters as significant spatial predictors that can be used to characterize the regional distribution of high As (23-26). Holocene sediments, for example, have been found to be strongly associated with arsenic contamination. Such coincidence can be also observed in areas reported to suffer from As contamination in China (Fig.1). We initially considered 16 environmental proxies, in the form of 30 arcsecond ($\approx 1 \text{ km}^2$) digital raster maps, as potential factors for identifying areas affected by geogenic arsenic hazard. Only eight proxies were significantly linked to the regional distribution of high As occurrence (SM, section 2) and were used, together with 2,668 georeferenced As measurements, to develop an ensemble model with binomial logistic regression as the base classifier (SM, section 3). The model (Fig. 2-A) forecasts that the area at risk of groundwater arsenic contamination (>10 µg L⁻¹) may encompass more than 580,000 km², and it pinpoints elevated arsenic concentrations in regions where endemic arsenic poisoning has already been detected (27-30). Large areas such as Tarim basin (Xinjiang), Ejina basin (Inner Mongolia), Heihe basin (Gansu), Qaidam basin (Qinghai), the North-eastern Plain (Inner Mongolia, Jilin and Liaoning) and the North China Plain (Henan and Shandong) were identified for the first time as being potentially affected, although these results must be confirmed with additional field measurements.

The mean model coefficients (Table S1) indicate a positive correlation of Holocene sediments, soil salinity, fine subsoil texture, Topographic Wetness Index (TWI) and density of rivers with high As and a negative contribution of slope, distance to rivers and gravity. Holocene sediments, soil salinity, subsoil texture and TWI were significant in most of the ensemble members, highlighting their relative importance in predicting the occurrence of groundwater arsenic in relation to the remaining variables. This is consistent with the conditions found in most studies performed in high risk areas in China, such as Xinjiang Province (model sensitivity = 85%) and Hetao-Huhhot Basin (model sensitivity = 95%), where high As tends to be associated with the anoxic environments of poorly drained aquifers in young sediments with flat topography and high salinity (Fig. 2B) (16). The model also performed well for oxic aquifers, such as Minqin Basin, Chahaertan oasis and Liao-Ho Basin (Figs. 2-B3 & 2-B4). Surveys in these areas (21, 22) revealed that these aquifers bear low arsenic concentrations, as correctly predicted by our model with an accuracy ranging from 83-98%.

We categorized areas on the probability map as low-risk ($p\leq0.46$) and high-risk (p>0.46), and cross-referenced the map with a population distribution map for the year 2000. Correspondingly, an estimated 19,580,000 people in China live in high-risk areas, mainly in Xinjiang, Inner Mongolia, Henan, Shandong and Jiangsu provinces (Fig. 3). \pm This may be an overestimation of the actual population at risk because treated and piped water may be used in some places, but water use statistics are not available. However, most of the As affected areas in China correspond to arid/semiarid regions, where groundwater is the predominant source of drinking water. Furthermore, there was insufficient survey data to develop a risk model with a threshold of 50 µg L⁻¹, the Chinese standard, as only 7% of the elevated As concentrations in our calibration dataset were above this value.

Our approach complements traditional groundwater quality surveys, which are expensive and time-consuming. The model requires only a small number of geospatial parameters to provide a preliminary assessment of affected areas, reducing the area that needs to be screened. It may also be appropriate for use in other parts of the world, especially in arid regions such as the Southwestern United States, where high arsenic concentrations have been reported, or in countries such as Mongolia, Kazakhstan and Kyrgyzstan where risk assessments for groundwater As contamination have not yet been performed.

References

- 1. J. Qiu, China to spend billions cleaning up groundwater. *Science* **334**, 745-745 (2011).
- 2. G. Sun, Arsenic contamination and arsenicosis in China. *Toxicol. Appl. Pharm.* **198**, 268-271 (2004).
- 3. G. Sun, Y. Xu, Q. Zheng, S. Xi, Arsenicosis history and research progress in Mainland China. *Kaohsiung J. Med Sci.* **27**, 377-381 (2011).
- 4. J. X. Guo, L. Hu, P. Z. Yand, K. Tanabe, M. Miyatalre, Y. Chen, Chronic arsenic poisoning in drinking water in Inner Mongolia and its associated health effects. *J. Environ. Sci. Heal. Part A* **42**, 1853-1858 (2007).
- 5. S. Li, T. Xiao, B. Zheng, Medical geology of arsenic, selenium and thallium in China. *Sci. Total Environ.* **421-422**, 31-40 (2012).
- 6. W. P. Tseng, H. M. Chu, S. W. How, J. M. Fong, C. S. Lin, S. Yeh, Prevalence of Skin Cancer in an Endemic Area of Chronic Arsenicism in Taiwan. *J. Natl. Cancer Inst.* **40**, 453-463 (1968).

- 7. Y. Jin, C. Liang, G. He, J. Cao, Study on distribution of endemic arsenism in China. J. *Hyg. Res.* **32**, 519-540 (2003).
- 8. J. Liu, B. Zheng, H. V. Aposhian, Y. Zhou, M.-L. Chen, A. Zhang, M. P. Waalkes, Chronic arsenic poisoning from burning high-arsenic-containing coal in Guizhou, China. *Environ. Health Perspect.* **110**, 119-122 (2002).
- 9. P. L. Smedley, D. G. Kinniburgh, A review of the source, behaviour and distribution of arsenic in natural waters. *Appl. Geochem.* **17**, 517-568 (2002).
- Y. Deng, Y. Wang, T. Ma, Isotope and minor element geochemistry of high arsenic groundwater from Hangjinhouqi, the Hetao Plain, Inner Mongolia. *Appl. Geochem.* 24, 587-599 (2009).
- 11. H. Guo, Y. Wang, Geochemical characteristics of shallow groundwater in Datong basin, northwestern China. J. Geochem. Explor.87, 109-120(2005).
- 12. H. Guo, B. Zhang, G. Wang, Z. Shen, Geochemical controls on arsenic and rare earth elements approximately along a groundwater flow path in the shallow aquifer of the Hetao Basin, Inner Mongolia. *Chem. Geol.* **270**, 117-125 (2010).
- 13. Q. Guo, L. Zhang, Y. Wang, T. Bullen, Eds. (Taylor, Kunming, China, 2007), pp. 1299-1304.
- P. L. Smedley, M. Zhang, G. Zhang, Z. Luo, Mobilisation of arsenic and other trace elements in fluviolacustrine aquifers of the Huhhot Basin, Inner Mongolia. *Appl. Geochem.* 18, 1453-1477 (2003).
- 15. L. Wang, X. Sun, X. Xu, H. Ai, Arsenic in natural water system and endemic arsenicosis, Kuitun area, Xinjiang. *Endemic Disease Communication* **8**, 88-92 (1993).
- 16. X. Xie, T. M. Johnson, Y. Wang, C. C. Lundstrom, A. Ellis, X. Wang, M. Duan, Mobilization of arsenic in aquifers from the Datong Basin, China: Evidence from geochemical and iron isotopic data. *Chemosphere* **90**, 1878-1884 (2013).
- 17. J.-J. Lee, C.-S. Jang, C.-W. Liu, C.-P. Liang, S.-W. Wang, Determining the probability of arsenic in groundwater using a parsimonious model. *Environ. Sci. Technol.* **43**, 6662-6668 (2009).
- M. Currell, I. Cartwright, M. Raveggi, D. Han, Controls on elevated fluoride and arsenic concentrations in groundwater from the Yuncheng Basin, China. *Appl. Geochem.* 26, 540-552 (2011).
- W. M. Edmunds, J. Ma, W. Aeschbach-Hertig, R. Kipfer, D. P. F. Darbyshire, Groundwater recharge history and hydrogeochemical evolution in the Minqin Basin, North West China. *Appl. Geochem.* 21, 2148-2170 (2006).
- 20. B. E. O Dochartaigh, A. M. MacDonald, "Groundwater degradation in the Chahaertan Oasis, Alxa League, Inner Mongolia" (British Geological Survey, Nottingham, 2006).
- 21. G. Yu, D. Sun, Y. Zheng, Health Effects of Exposure to Natural Arsenic in Groundwater and Coal in China: An Overview of Occurrence. *Environ. Health Perspect.* **115**, 636-642 (2007).
- 22. G. F. Sun, J. B. Pi, B. Li, X. Y. Guo, H. Yamauchi, T. Yoshida, in Arsenic Exposure and Health Effects IV, W. R. Chappell, C. O. Abernathy, R. L. Calderon, Eds. (Elsevier, Amsterdam, 2001), pp. 79-85.
- 23. M. Amini, K. C. Abbaspour, M. Berg, L. Winkel, S. J. Hug, E. Hoehn, H. Yang, C. A. Johnson, Statistical modeling of global geogenic arsenic contamination in groundwater. *Environ. Sci. Technol.* **42**, 3669-3675 (2008).
- L. R. Lado, D. Polya, L. Winkel, M. Berg, A. Hegan, Modelling arsenic hazard in Cambodia: A geostatistical approach using ancillary data. *Appl. Geochem.* 23, 3010-3018 (2008).
- L. Winkel, M. Berg, M. Amini, S. J. Hug, C. A. Johnson, Predicting groundwater arsenic contamination in Southeast Asia from surface parameters. *Nature Geosci.* 1, 536-542 (2008).

- 26. L. H. E. Winkel, T. K. T. Pham, M. L. Vi, C. Stengel, M. Amini, T. H. Nguyen, H. V. Pham, M. Berg, Arsenic pollution of groundwater in Vietnam exacerbated by deep aquifer exploitation for more than a century. *Proc. Natl. Acad. Sci. USA* **108**, 1246-1251 (2011).
- 27. J. Bian, J. Tang, L. Zhang, H. Ma, J. Zhao, Arsenic distribution and geological factors in the western Jilin province, China. J. Geochem. Explor. **112**, 347-356 (2012).
- 28. S. Han, H. Zhang, M. Zhang, Hydrogeological and hydrochemical characterization of shallow high arsenic and deep low arsenic aquifers in Yinchuan Plain: A case study of deep aquifer development for domestic water supply. *Geochim. Cosmochim. Ac.* **74**, A378-A378 (2010).
- 29. M. N. Mead, Arsenic: in search of an antidote to a global poison. *Environ. Health Perspect.* **113**, A378--386 (2005).
- 30. Y. Xia, J. Liu, An overview on chronic arsenism via drinking water in PR China. *Toxicology* **198**, 25-29 (2004).

Acknowledgements:

This study was funded by the Sino-Swiss Science and Technology Cooperation Program of the Swiss State Secretariat for Education and Research, project No. *IZLCZ2 123971*, and by the External Cooperation Program of the Chinese Academy of Sciences. The authors wish to acknowledge Gui-Bin Jiang and Jianbo Shi of the Research Center for Eco-Environmental Sciences, Beijing for support and data collection in the initial phase of the project. We also wish to acknowledge the National Science and Technology Pillar Program of China during the 11th Five-Year Plan Period (Contract grant number: 2006BAI06B04). L.R.L. also acknowledges the Isidro Parga Pondal research program from the Xunta de Galicia (Spain) for current financial support. Data is available at www.eawag.ch/repository/remarc.

Supplementary Materials

Materials and Methods Figs. S1 to S5 Tables S1 to S7 References (*31–50*)

Additional data is available through www.eawag.ch/repository/remarc

Probability map for As > 10 μ g/L (raster data) Binary risk map for As > 10 μ g/L using the 0.46 probability threshold (raster data) Population at risk map (raster data) Map of model uncertainty (raster data) As database used to calibrate the model (ESRI shapefile) As database used to validate the model (ESRI shapefile)

Figures



Fig. 1: Location of known and potential arsenic-affected basins. Areas with high levels of arsenic are generally characterized by Holocene sediments (green) and large basins may be affected.



Fig. 2: Arsenic predictive model and validation maps. Modelled probability of As concentrations exceeding $10 \ \mu g \ L^{-1}$ in groundwaters (A). The boxes indicate the location of the basins used to describe the model performance (B).



Fig. 3: Estimated population at potential risk of exposure to As >10 μ g L⁻¹. Calculation is based on the model results and the population density estimates.