Screening for lead: Predictive modeling of indoor dust lead concentrations and possible effects of intervention

Matthew Dietrich*[1], Cynthia F. Isley[2], Jane A. Entwistle[3], Diana Meza-Figueroa[4], Chenyin Dong[5], Peggy Gunkel-Grillon[6], Khadija Jabeen[3], Lindsay Bramwell[3], John T. Shukle[1], Leah R. Wood[1], Ravi Naidu[7,8], Kara Fry[9], Mark Patrick Taylor,[9,10] and Gabriel M. Filippelli[1,11]

*Corresponding author: Matthew Dietrich; mjdietri@iu.edu; 723 W. Michigan Street, SL 118, Indianapolis, IN 46202

[1]Department of Earth Sciences, Indiana University–Purdue University Indianapolis, Indianapolis, IN, USA

[2]The Australian Centre for Housing Research, Faculty of Arts, Business, Law and Economics, University of Adelaide, SA, 5000, Australia; School of Natural Sciences, Faculty of Science and Engineering, Macquarie University, North Ryde, Sydney, NSW 2109, Australia

[3]Department of Geography and Environmental Sciences, Northumbria University, Newcastle-upon-Tyne NE1 8ST, U.K.

[4]Universidad de Sonora, 83067 Hermosillo, Mexico

[5]State Environmental Protection Key Laboratory of Environmental Pollution Health Risk Assessment, South China Institute of Environmental Sciences, Ministry of Ecology and Environment, Guangzhou 510655, China

[6]Institute of Exact and Applied Sciences (ISEA), University of New Caledonia, BPR4, 98851 Nouméa cedex, New Caledonia, France

[7]Global Centre for Environmental Remediation (GCER), College of Engineering Science and Environment, The University of Newcastle, Callaghan, NSW 2308, Australia

[8]Cooperative Research Centre for Contamination Assessment and Remediation of the Environment (CRC CARE), ATC Building, The University of Newcastle, Callaghan, NSW 2308, Australia

[9]Environment Protection Authority, EPA Science, Centre for Applied Sciences, Ernest Jones Drive, Macleod, Melbourne, VIC 3085, Australia

[10]School of Natural Sciences, Faculty of Science and Engineering, Macquarie University, North Rdy, Sydney, NSW 2109, Australia

[11]Environmental Resilience Institute, Indiana University Bloomington, Bloomington, IN, USA

Abstract:

Lead (Pb) pollution continues to contribute to world-wide morbidity in all countries, particularly low- and middle-income countries. Despite its continued widespread adverse effects on global populations, particularly children, accurate prediction of elevated household dust Pb and the potential implications of simple household interventions at national and global scales have been lacking. A global dataset (~40 countries, n = 1951) of community sourced household dust samples were used to predict whether indoor dust was elevated in Pb, expanding on recent work in the United States (U.S.). Binned housing age category alone was a significant ($p<0.01$) predictor of elevated dust Pb, but only generated effective predictive accuracy for the U.K. and Australia (sensitivity of ~80%), similar to previous results in the U.S. This likely reflects comparable Pb pollution legacies between these three countries, particularly with residential Pb paint. We also find that the heterogeneity associated with Pb pollution at a global scale can complicate the predictive accuracy of our model, which is lower for countries outside the U.K., U.S., and Australia. This is likely due to differing environmental Pb regulations, sources, and the paucity of dust samples available outside of these three countries. In the U.K., U.S., and Australia, straightforward household intervention could conservatively save $70 billion USD within a four-year period, and as much as $1.68 trillion USD globally with universal household remediation based on our predictive results.

Introduction

Lead (Pb) contamination affects millions of people adversely across the world, particularly children, because of their greater susceptibility to Pb poisoning due to their activities (i.e., hand-to-mouth behavior), developing bodies, and greater ability to absorb Pb relative to adults (e.g., Egendorf et al., 2020; Gundacker et al., 2021; Mielke et al., 1999). This has resulted in high global morbidity, evidenced through diminished IQ levels and other neurocognitive impairment (e.g., Meyer et al., 2008). While blood lead levels (BLLs) have rapidly declined in many countries following the phase-out of leaded gasoline, particularly in developed/high-income countries, BLLs continue to be elevated in many low- and middle-income countries (LMICs) and there is no known safe level of Pb exposure (e.g., Meyer et al., 2008, Ericson et al., 2021a).

Conservatively, nearly $1 trillion USD in potential life earnings is lost annually due to Pb-related IQ detriment in low- and middle-income countries (LMICs), with higher-income countries sharing less of the global Pb burden (Attina and Trasande, 2013). Lead sources also differ, with LMICs predominantly having BLLs influenced by Pb sources other than paint and leaded petrol, such as industrial sources of battery manufacturing or recycling (Ericson et al., 2021a). Recent estimates in the United States (U.S.) of income potential lost due to Pb exposure is around $46.2 billion USD/year for the years 1999-2010 and is disproportionately shouldered by Black (non-Hispanic) infants (Boyle et al., 2021). For example, Boyle et al. (2021) estimated 46–55% greater average IQ points lost due to blood Pb exposure in Black infants relative to Hispanic or White infants based on cross-sectional National Health and Nutrition Examination

Survey (NHANES) results in the U.S. Thus, in addition to uneven global Pb exposure, there can be disproportionate Pb exposure at the national scale as well.

To combat global Pb pollution, reduce exposure, and measure and educate the community about exposures and what they could do, an international collaboration of scientists came together to begin an initiative called "DustSafe" (also known as "360 Dust Analysis"). This initiative utilizes community scientists to collect household dust for trace metal(loid) screening (Isley et al., 2022). Results obtained through this program are used to better assess exposure sources and routes, and simplified versions of the results are communicated back to the community participants who supplied the samples. Participants are informed of any potential hazards and learn of steps they may take to reduce their trace metal exposure. In addition to informing community members, the collective results of this work have been used to inform researchers of similarities and dissimilarities in household dust pollution at national and global scales (Isley et al., 2022). Given that BLLs have been shown to relate strongly to household dust Pb (e.g., Lanphear et al., 1996; Gulson and Taylor, 2017; Rhoads et al., 1999), these dust data can assist with direct intervention to reduce potentially elevated BLLs. For example, a simple logistic regression model based on "DustSafe" Pb data in North America (predominantly the U.S.) was able to correctly classify elevated (≥80 mg/kg) or low (<80 mg/kg) dust Pb samples 75% of the time, with a sensitivity of 82% (Dietrich et al., 2022). This model was then incorporated into an interactive online app (Dietrich et al., 2022) so the general public can more easily participate in the "DustSafe" program and take intervention steps if necessary.

We sought to expand this model to the much larger global dust dataset to evaluate if and where it would be effective, and whether adjusting the model would be more effective in particular regions due to differing legacies/sources of Pb pollution worldwide (e.g., Ericson et al., 2021a). Predictive modeling of indoor dust Pb concentrations in general has been sparse (Dietrich et al., 2022). A growing number of predictive models for Pb have appeared in different environmental media, such as soil (e.g., Obeng-Gyasi et al., 2021; Schwarz et al., 2013), BLLs and water infrastructure (e.g., Gibson et al., 2020; Mulhern et al., 2022), and even predictive models for BLLs based on spatial and spatiotemporal data (e.g., Potash et al., 2020). However, many predictive models are complex and require extensive datasets with multiple variables for input. Several models also require complex machine-learning techniques for the best outcomes (e.g., Obeng-Gyasi et al., 2021; Potash et al., 2020). Our recent work has shown that a simple model with only a few important variables perform well at predicting elevated Pb in household dust (Dietrich et al., 2022), which may help to inform risk analysis and interventions.

To assess the usefulness of a global predictive indoor dust Pb model, we: (1) tested the U.S. based model (Dietrich et al., 2022) on global dust Pb data to determine its efficacy; (2) identified modifications required to improve predictive ability; (3) determined differences in model accuracy based on different country groupings; and (4) estimated the potential effects of household intervention based on modeling results. The purpose of this work was not to determine exact sources of Pb and make exposure estimates, but to use crowd-sourced environmental data to help better understand risk factors for indoor dust Pb in multiple countries.

Methods

Sampling Collection and Analysis

The DustSafe sampling and data protocols were subject to ethical review and approval at Macquarie University, Australia (project #2446); Indiana University, U.S. (project #1810831960); and Northumbria University, U.K. (project #2598). All dust samples were provided by community participants via post between 2018-2021 from 39 countries (Table 1; n = 1951), predominantly the U.K. and Australia (n = 1524), following the emptying of household vacuum cleaner contents into a clean, polyethylene bag. Participation was promoted through campaigns online, such as twitter and email, as well as via radio and open house days. Household dust samples are representative of composite household dust and uniform instructions for sampling were provided to all participants. Community participants also completed an online questionnaire (e.g., https://www.360dustanalysis.com/soil/get-started) that collected household data on potentially relevant matters (e.g., renovations, age of home, occupation, etc.). Household dust were sieved to <250 µm using either a pre-cleaned stainless-steel sieve or single-use polypropylene mesh. Pb concentrations were determined with X-ray fluorescence spectrometry (portable (pXRF) and energy-dispersive (ED-XRF)) for all samples except for a small subset of samples from China (inductively coupled plasma atomic emission spectrometry (ICP-AES)), outlined in Isley et al. (2022). Additionally, a small subset of samples from China were sieved to 150 µm instead of 250 µm, and the limit of detection (LOD) for Pb ranged from 0.1-2 mg/kg depending on the country conducting the analysis (Isley et al., 2022). Additional details on analytical procedures and quality control are provided in Isley et al. (2022). U.S. data were also collected following the same method as reported in Dietrich et al. (2022) and Isley et al. (2022), with 23 additional samples reported for this work (4 of the 365 samples are from Canada and are included in the "U.S. Model").

A detailed longitudinal study in one home within the U.K. was conducted to evaluate month to month (March 2020-October 2021) variability of reported indoor dust Pb concentrations and the effects of simple household washable door mat intervention on Pb concentrations. The same vacuum cleaner was used throughout the study, used across all rooms within the home each month, and no DIY or internal home improvements were undertaken during the longitudinal study. Prior to the washable door mat intervention, the vacuum dust samples included soil/dust from the non-washable door mat. Washable door mats were cleaned and changed every 1-3 weeks and not vacuumed.

Metadata Analysis

Metadata were provided via an online questionnaire (e.g., https://www.360dustanalysis.com/soil/get-started). Slight differences in questionnaires based on location are described in more detail in Isley et al. (2022). Data of hobbies involving metal exposure was omitted from this work because of the large number of hobby types (n = 8), and lack of data provided for most hobby types [Isley et al. (2022)—Fig. 9.7 (n < 40 for all but 2 hobby types in global data)].

All "Yes" responses were converted to "1," and all "No" responses were converted to "0" (Table S1). Housing age data was converted into binned housing age categories based on Dietrich et al. (2022), and ages were calculated assuming a sampling date of 2019, as this was when most samples were collected and date of actual sample collection was not directly available. They were reclassified as numeric variables of 0, 1, 2, 3 for the responses, "1980-Present," "1960-1979," "1940-1959," and "Pre-1940," respectively (Table S1). These groupings of housing age were selected based on the common phase-out history of Pb paint in countries such as the U.S., U.K. and Australia, and because the binned categories make it easier for community engagement when developing this variable into a predictive, interactive model/app. While these housing age categories do not necessarily follow Pb regulatory practices in many LMICs, we elected to base our model originally on these categories because it has been shown to be effective in the U.S. (Dietrich et al., 2022) and the bulk (>50%) of studies included in this work were collected in countries with similar Pb regulatory legacies to the U.S. (U.K. and Australia). Thus, if these housing groupings are found not to be effective in other country groupings, this would suggest closer examination of the nuances associated between housing age and Pb sources in other countries for future work, as the exploratory breakdown of best housing age categories by individual country is beyond the scope of this work.

Logistic Regression Modeling

Predictive logistic regression modeling was performed in RStudio (R Core Team, 2021) using the glm() function and the general equation:

$$\log\left[\frac{p}{1-p}\right] = b_0 + b_1 * x_1 + b_2 * x_2 \ldots + b_n * x_n \quad (1)$$

Where p is the probability of an event occurring, b_0 is the intercept, b_n is the regression beta coefficient, and x_n is a given predictor variable.

A stepwise algorithm to help identify best logistic regression models was run using the stepAIC() function in R, based on the MASS package (Venables and Ripley, 2002). Modeling was based on classifying samples as "Elevated" or "Low" Pb, with the cutoff as ≥80 mg/kg for "Elevated" Pb. This is based on California's (U.S.) human health screening level for soil Pb, which albeit more conservative, is more preventative than outdated Pb guidelines such as the U.S. EPA's 400 mg/kg residential soil standard (e.g., Gailey et al., 2020) and almost certainly represents an anthropogenic source of Pb in most areas, as average Pb in upper continental crust is only ~17 mg/kg (Rudnick and Gao, 2003). All data input into the modeling is freely available, including essential variables used for the best predictive modeling from the U.S. dataset (Table S1).

Given that Australia and the U.K. have similar Pb legacies and regulatory practices over the past century and comprised the majority of our DustSafe data, we elected to evaluate our predictive Pb logistic regression models both on the collective global dataset, as well as a subset of Australian and U.K. data to determine whether there were significant differences worth noting. We began with the U.S.-based predictive model (Dietrich et al., 2022) for evaluation, then, based on those results, refined our models based on the global dataset. Only samples with metadata responses were used in the modeling.

Online App Development

The online mobile app for Pb screening built upon the previous version in Dietrich et al. (2022) for the U.S. The goal was to provide an easily accessible, user-friendly way for people to evaluate likelihood for elevated dust Pb in their home, while also learning about Pb in the environment. The application was built using the shiny, shinydashboard, shinydashboardPlus, and shinyjs packages in R (Attali, 2020; Chang et al., 2021; Chang and Borges Ribeiro, 2018; Granjon, 2021).

Results/Discussion

Modeling Results

The Pb dust predictive model from the U.S. (Dietrich et al., 2022) resulted in a mean predictive accuracy of 73% Elevated/Low correct classification of Pb dust concentrations (probability threshold of 0.85) and a sensitivity of 80% on the global dataset (n = 1653; not including the U.S.). When omitting Australia and the U.K. the model performed at 64% accuracy with a sensitivity of 39% (n = 267, 0.8 probability threshold). U.K. alone (n = 132) had 75% predictive accuracy with the model and 92% sensitivity (0.85 probability threshold). Australia alone (n = 1254) had a 76% predictive accuracy and 82% sensitivity (0.85 probability threshold). U.K. and Australia combined (n = 1386) had a predictive accuracy of 76% and sensitivity of 83% (probability threshold of 0.85). Summary outputs from all scenarios are available in the Supplement (Supplementary Text S1).

When utilizing global, non-U.K./Australia, and U.K./Australia data for training and testing datasets, no additional significant ($p < 0.05$) predictor variables could be identified besides housing age category, which alone provided the best modeling outcomes (i.e., based on overall predictive accuracy, sensitivity, area under the ROC curve (AUC)). The U.K./Australian testing dataset (n = 421; based on 0.7 training/0.3 testing data ratio) produced a predictive accuracy of 76% and sensitivity of 80% with a probability threshold of 0.55 based on the housing age model and U.K./Australian training dataset (Table 2). For non-U.K. and Australian countries, the housing age predictive model based on the training dataset predicted accurately 74% of Elevated vs. Low Pb classification (probability threshold of 0.5), but with a sensitivity of only 38% (n = 84; Table 2).

Modifying the logistic model from Dietrich et al. (2022) (based predominantly on U.S. housing dust data with 23 samples added to the Dietrich et al. (2022) dataset) to include only the housing age category as a predictive variable improved the predictive accuracy slightly and maintained sensitivity of the model, even though interior peeling paint was a highly significant variable ($p < 0.01$) in the original model. Overall model predictive accuracy on the testing dataset (n = 109) slightly increased to 85%, while sensitivity remained at 82% (probability threshold of 0.8). This modified equation became:

$$log\left[\frac{p}{1-p}\right] = 2.5632 - 0.9551\ (Housing) \quad (2)$$

Where "Housing" is the housing age category (model output in Supplementary Text S2). Applying this model to all U.K./Australian data (n = 1386) resulted in a predictive accuracy of 75% and sensitivity of 81% (probability threshold of 0.85) (Table 2). Usage of the model on non-U.K./Australian data (n = 269) produced a predictive accuracy of 70%, with a sensitivity of 31% (probability threshold of 0.8) (Table 2).

The most effective logistic regression model contains only one variable. While we still contend this is a predictive model by convention (i.e., an equation that makes the prediction of an outcome based on sample data), it essentially boils down to a housing age threshold for determining whether house dust is likely to be elevated in Pb or not. Basically, any sample that falls in a housing age bin earlier than 1980-Present will result in a predictive outcome of elevated dust Pb. As discussed later, this corresponds with Pb regulatory history in the U.S., U.K., and Australia, where Pb paint was largely outlawed/reduced for home application in the 1970s.

Modelling Usefulness and Effectiveness

While the metadata questionnaire response to interior peeling paint was a significant predictive variable ($p<0.01$) in our North American dataset (Dietrich et al., 2022), inclusion of this variable was not significant at the global level ($p>0.05$), even with countries relatively analogous (economically and regarding Pb regulatory history) to the U.S., such as the U.K. and Australia. Furthermore, we discovered that although this interior peeling paint variable was highly significant ($p < 0.01$) in our North American model (Dietrich et al., 2022), omission of the variable and inclusion of only housing age category in a model slightly improved overall predictive accuracy (but not sensitivity) with predominantly the same testing dataset as used in Dietrich et al. (2022).

At the global level, housing age category was the most (and only) significant predictive factor, resulting in a predictive accuracy ≥75% and sensitivity ≥80% in grouped U.K. and Australia data (Table 2)—this is the case when using both the modified model developed from predominantly U.S. data [Equation (2)] and a model based on a training dataset of U.K. and Australian data (Supplementary Text S3). This is similar to the predictive accuracy of the housing category only model [Equation (2)] used on the predominantly Dietrich et al. (2022) testing dataset (n = 109), which resulted in a sensitivity of 82% and predictive accuracy of 85%. Graphing the distributions of Pb indoor dust data by housing age category demonstrates this, particularly in the U.K. and Australia (Fig. 1). This illustrates that housing age category alone can provide helpful information regarding which homes in the U.S., Australia, and the U.K. contain indoor dust Pb ≥ 80 mg/kg. The importance of housing age and Pb concentrations has been well-established in the literature for both soils (e.g., Taylor et al., 2021, Yesilonis et al., 2008) and house dusts (e.g., Isley et al., 2022; Rasmussen et al., 2011). Chance alone would result in a sensitivity and predictive accuracy of ~50% for the logistic regression model, but by just knowing relative housing age (not even the exact housing age), we can improve average predictive accuracy to ~75% and sensitivity to ~80% (Table 2).

The housing age category is less useful when grouping together results from countries outside of the U.S., Australia, and the U.K. Sensitivity drops to <40% when both types of housing age models (U.S.-based and non-U.K. and Australian-based) are used (Table 2), greatly reducing any real-world usefulness of the models. This is because this results in false-negative rates of >60%, where many homes with actual dust Pb ≥ 80 mg/kg will be incorrectly classified as "Low" Pb. In fact, this would be detrimental from an intervention standpoint, because the probability by pure chance of correctly classifying elevated versus low Pb homes would be greater, at 50%.

Because of small sampling size (i.e., n < 15) of paired Pb data and questionnaire responses in most countries outside of the U.S., Australia, and the U.K., we could not effectively examine the nuances between countries grouped together as non-U.K. and Australian data. Logistic regression requires large datasets, and we wanted to avoid making extrapolations of predictive accuracy on any sampling subsets where n < 100, as even our testing dataset in Dietrich et al. (2022) (n = 102) was subject to sampling size effects depending on the random subset of testing data chosen. Given the current data available, we suggest that housing age is not as important of a determinant of elevated household dust Pb in many samples in countries outside the U.S., U.K., and Australia, and that alternative sources typically not associated with housing age may be responsible for interior dust Pb concentrations.

A recent literature review compiled by Ericson et al. (2021a) supports this contention, as they found in LMICs that most studies of BLLs attributed predominant Pb sources to non-Pb paint sources, such as industrial emissions. This was further backed in a commentary reply by Ericson et al. (2021b), where they reemphasized the role of industrial Pb as a main source of elevated BLLs in LMICs, with only 1.5% of their study (Ericson et al., 2021a) subsamples reporting lead-paint as a likely exposure source. In high-income countries such as the U.S., Australia, and the U.K., Pb paint is likely still a large contributor of current household dust Pb because it still resides in many older homes and soils (e.g., Dietrich et al., 2022), which explains why housing age category alone remains a significant predictor variable. Additionally, it is important to note that Pb paint can end up in household dust from both inside or outside the home, as exterior peeling paint may also be tracked in from outdoor soils/dusts. These outdoor soils/dusts may also contain Pb from gasoline/industry sources, and it is noted that there is likely some covariance with housing age and sourcing of Pb from historic gasoline in soils that are trekked inside.

While not all our non-U.K. and Australian samples were from LMICs (i.e., Ireland, Greece, New Zealand), many were, such as China, Bangladesh, Iran, India, and Mexico, and 110 (41%) of our non-U.K. and Australian paired housing age and Pb concentration samples (used in modeling) were from countries also included in Ericson et al.'s (2021a) metanalysis of LMICs. Thus, it is reasonable to conclude that there are significant differences of controls on household dust Pb concentrations in homes based on country, particularly in LMICS where Pb pollution legacy often differs from countries such as the U.K., U.S., and Australia.

Online App Update for Pb Screening and Potential Application and Development

Our previous modeling based on indoor vacuum dust Pb concentrations in predominantly U.S. household dust samples led to the development of an interactive online app (for computers or mobile devices; https://iupui-earth-science.shinyapps.io/IUPUI-LeadRiskApp/) where users could input information about their home (housing age, interior peeling paint) and our model would then let the user know whether their home was likely to contain elevated (≥80 mg/kg) Pb in indoor dust (Dietrich et al., 2022). The app links to the "MapMyEnvironment" website (https://www.mapmyenvironment.com/), which contains a link to the "360 Dust Analysis" project page (where users could register for our free testing program) as well as links to other free testing programs for environmental media such as soil and water. Based on its success in predicting elevated Pb in U.K. and Australian house dust samples (Table 2), we have expanded the app to now include these countries. Additionally, because the response of interior peeling paint was deemed not sufficiently significant in predictive power, we have eliminated this question for users. While the previous model contained an option of "not sure" regarding housing age category, we have also elected to remove it from the app, as it was not a significant individual predictor in the U.S.-based model ($p = 0.12$) and none of the U.K. nor Australian samples contained this response. The logistic regression model currently used in the app is based on Equation (2). The results page of the app now links directly to the 360 Dust Analysis page as well as the MapMyEnvironment sampling map. While still in early roll-out, we envision that the binned housing age categories will make it easy enough for users to determine which category to select, even if they are unsure of their exact home age. This is particularly important for renters, who often have less knowledge of building information. Future work will evaluate whether the binned housing age categories are sufficient for the best user participation through examination of mobile app data and post-hoc survey responses from users who complete the community science process from start to finish.

While the predictive modeling for countries outside of the U.S., Australia, and the U.K. did not perform effectively enough to warrant inclusion into an interactive app for Pb screening (sensitivity <50%; Table 2), we envision that through continued sampling and assessment of results from these countries, we may eventually gather enough information to tailor specific predictive models that contain variables other than housing age. A key component of this may be different survey questions for specific countries, such as distance to metal smelters, distance to battery recycling plants, distance to mining sites, etc., as these industrial Pb sources are more common in LMICs (Ericson et al., 2021a). Continued global partnerships with LMIC communities are key to addressing these current knowledge gaps, particularly because those in LMICs are the ones mostly adversely affected by elevated BLLs (e.g., Attina and Trasande, 2013, Ericson et al., 2021a).

Potential Economic Impact of Simple Household Intervention Based on Modeling Results

One of the key objectives of our international DustSafe collaboration is to provide participants with information on how they can reduce their Pb exposure (Isley et al., 2022), which is particularly relevant where no government remediation services are provided. The online app provides an easy way to participate in DustSafe, and model results can provide users with key data they need for intervention without waiting for formal dust Pb analysis.

Using the geometric mean Pb dust concentration of all our global dust samples ≥80 mg/kg (225 mg/kg; Fig. 2), and assumptions of initial BLLs based on that mean, effects of household intervention on children's (<5 years) BLLs can be estimated (Table 3). Based on our estimations, which we deem conservative because of using U.S. baseline BLLs instead of global baseline BLLs, the effects of household intervention (e.g., wiping, high filter vacuuming) such as that done in Rhoads et al. (1999) in multiple homes could result in up to $70 billion USD saved within a four-year cohort within the U.K., Australia, and U.S. (Table 3). This estimate arises if every family with children <5 years old uses our current model [Equation (2)] at a sensitivity of 80% and acts on the results (Table 3). These cost savings are based on the prevention of IQ points lost due to Pb poisoning, which adversely affects lifetime earnings potential (e.g., Attina and Trasande, 2013; Boyle et al., 2021). If our model worked at the global scale with the same sensitivity of ~80%, the earnings potential saved could be up to $1.68 trillion USD within a four-year cohort following household intervention (Table 3).

Household interventions are a temporary solution to environmental Pb exposures, as cleaning and door mats do not necessarily remove the ultimate sources of Pb in the environment (internal and external), and Pb can persist in the home at elevated concentrations even following intervention (Fig. S1). Although this short-term solution may reduce the individual household Pb burden, effective remediation at the primary source of Pb (i.e., paint, outdoor soils, mining sites, etc.) is what will ultimately prevent childhood Pb exposure and poisoning. Nevertheless, simple household efforts such as utilizing washable door mats can effectively trap dust that is elevated in Pb before it enters the home, and regularly cleaning these door mats can therefore reduce overall household Pb dust concentrations (Fig. S1). Our case study example in the U.K. (~270-year-old home) demonstrates this, as the geometric mean monthly indoor dust Pb concentration was 437.5 mg/kg (n = 4) prior to the use of washable door mats. Using washable door mats resulted in household vacuum dust Pb concentrations dropping by an average of 55.1% to a geometric mean of 196.5 mg/kg (n = 12), albeit there was about a two-month lag before the reduced Pb concentrations stabilized (Fig. S1; Table S2).

Conclusions

Lead pollution persists globally, and adversely affects children. In analogous high-income countries such as the U.S., U.K., and Australia, similarities in Pb pollution legacy and sources enable simplistic predictive modeling to accurately assess which homes likely contain elevated dust Pb based on housing age. However, this does not necessarily work well in other countries, particularly LMICs because of differing Pb sources such as mining and industry. Thus, although household intervention based on usage of our predictive model could potentially save trillions of USD throughout the world, more refined data is needed in countries outside the U.S., U.K., and Australia to develop more effective predictive models of country specific household indoor dust Pb. Additionally, paired household indoor dust, outdoor soil, and house paint data in future community science projects along with important metadata such as housing age will further help elucidate ultimate sources of Pb in household environments throughout the world.

Acknowledgements

We are deeply grateful to the community scientists who provided dust samples and the lab techs who helped process samples. Support for this work to M.D. was from the U.S. National Science Foundation (NSF-EAR-PF Award #2052589); to G.M.F. from the Environmental Resilience Institute, funded by Indiana University’s Prepared for Environmental Change Grand Challenge Initiative, the U.S. National Science Foundation (NSF-ICER Award #1701132), and the U.S. Housing and Urban Development Agency; to J.E. from the Natural Environment Research Council (Research Grant NE/T004401/1, U.K.); and to M.P.T. from the Australian Government Citizen Science Grant, CSG55984.

Figures and Tables

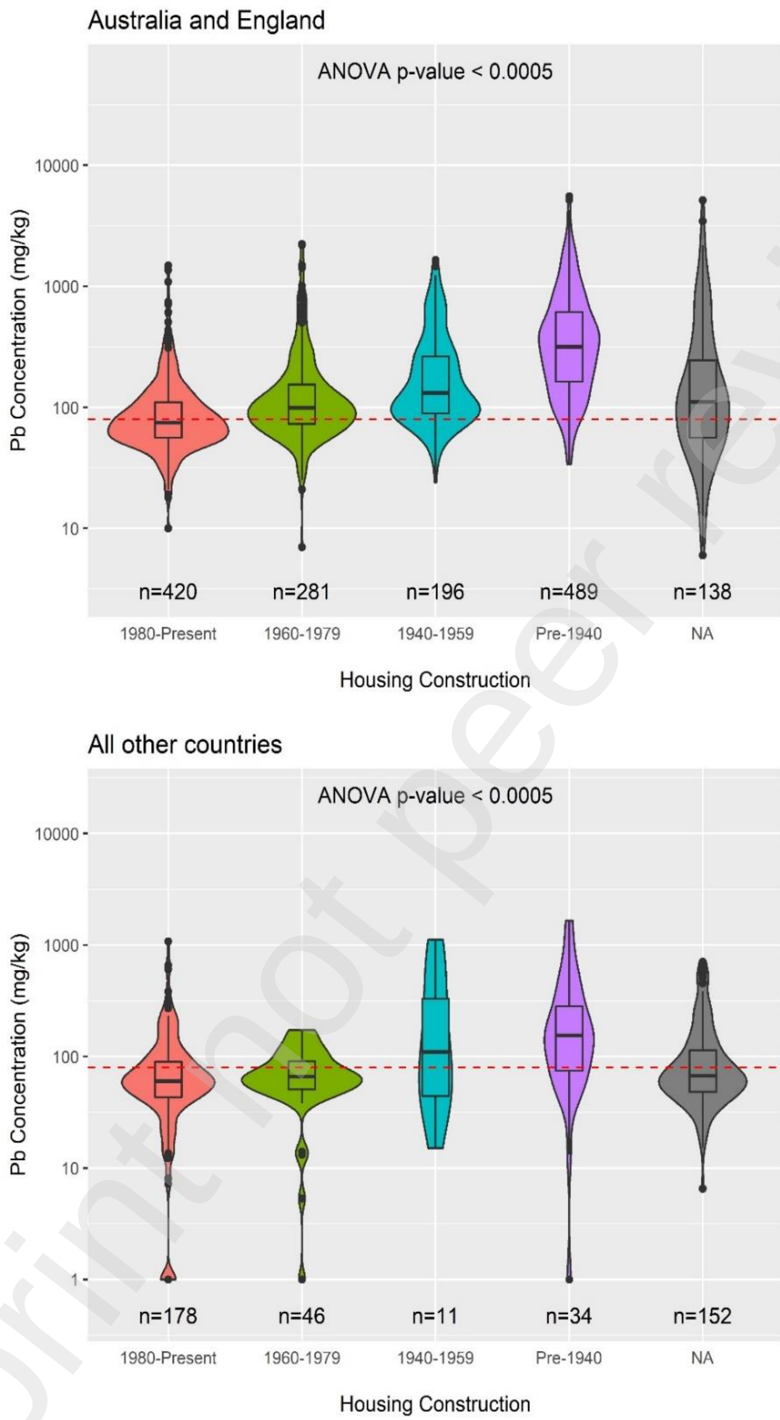

Figure 1 Embedded boxplots within violin plots for housing age categories used in the predictive models, as well as N/A housing age values (no survey responses). The boxes represent the interquartile range (IQR) of 25th-75th percentiles of data, the solid horizontal line is the median, and the whiskers represent 1.5 times the IQR. Analysis of variance (ANOVA) test associated p-values between all housing age categories are provided. The y-axis is transformed on a $\log_{10}$ scale, and the dashed red line represents California's human health screening level of 80 ppm for soil Pb.

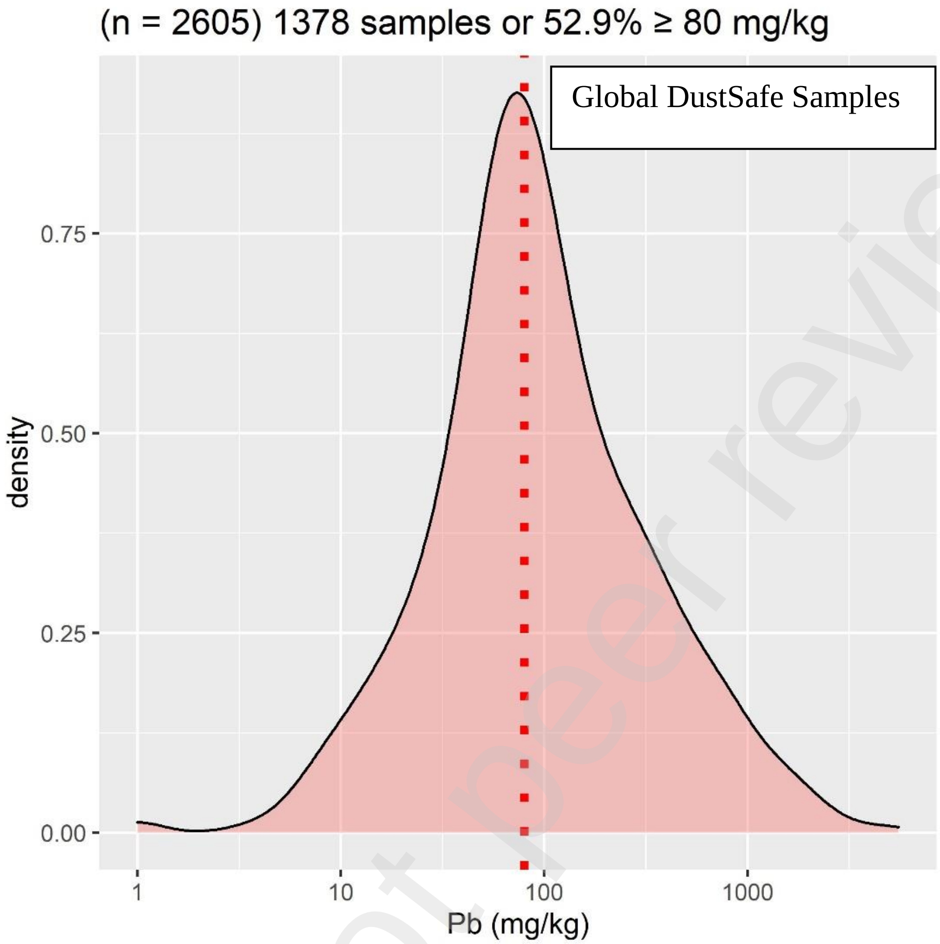

Figure 2 Proportion of global DustSafe samples ≥ 80 mg/kg Pb [*North America (Dietrich et al., 2022; 23 additional samples with survey responses in this study, and all samples analyzed without survey responses as well), and Nigeria (Isley et al., 2022)]*, with the corresponding smoothed density plot on a $\log_{10}$ scale x-axis. The dotted vertical line denotes 80 mg/kg.

Table 1: Summary data (sample size (n), median and standard deviation (S.D.) of Pb concentrations and housing age) of DustSafe samples with complete or nearly complete questionnaire responses to accompany Pb concentration measurements. United States samples and modeling results are presented in Dietrich et al. 2022, with an additional 19 U.S. samples presented in this work (n = 361 total with survey data and Pb concentrations) and 4 Canadian samples (n = 15 total).

Country	n	Median Pb (mg/kg)	S.D. Pb	Median House Age	S.D. House Age
Australia	1254	125	459	1966	37
U.S.	361	31	187	1985	34
England	132	113	523	1939	38
China	49	76	122	2004	10
New Zealand	42	79	313	1969	86
Greece	35	57	52	1993	19
Mexico	33	13	30	1989	18
Croatia	27	61	87	1979	22
Canada	15	54	46	1993	26
Ghana	14	62	96	2007	13
Scotland	5	83	528	1943	64
Wales	5	40	71	1929	15
France	4	102	65	1958	40
Bangladesh	3	159	50	1999	0
Belgium	3	178	108	1889	79
Cyprus	3	56	17	2004	15
Estonia	3	69	27	1979	54
Germany	3	65	63	1889	76
Iran	3	68	77	2001	14
Malaysia	3	51	9	2007	4
N. Ireland	3	83	51	1990	81
Nepal	3	101	23	1993	14
Netherlands	3	179	225	1904	51
South Korea	3	60	13	1992	11
Barbados	2	87	40	1992	18
Czech Republic	2	38	22	1997	11
Switzerland	2	742	525	1929	42
India	1	55		1998	
Italy	1	272		1994	
Northern Ireland	1	43		1934	
Slovakia	1	50		2017	
Thailand	1	109		2007	

Table 2: Confusion matrix output results for Pb household dust predictive models using the housing age category variable only. Grey highlighted outputs are based on models from training datasets of data from this study, while non-highlighted outputs are based on Equation (2).

Testing dataset of U.K. and Australia data (n = 421)	**Actual Elevated Pb**	**Actual Low Pb**	**Sensitivity**	**Mean Proportion Predicted Correctly**
Predicted Elevated Pb	243	42	0.80	0.76
Predicted Low Pb	61	75		
Testing dataset of non-U.K. and Australia data (n = 84)	**Actual Elevated Pb**	**Actual Low Pb**		
Predicted Elevated Pb	11	4	0.38	0.74
Predicted Low Pb	18	51		
Testing dataset of U.K. and Australia data (n = 1,386)	**Actual Elevated Pb**	**Actual Low Pb**		
Predicted Elevated Pb	813	153	0.81	0.75
Predicted Low Pb	188	232		
Testing dataset of non-U.K. and Australia data (n = 269)	**Actual Elevated Pb**	**Actual Low Pb**		
Predicted Elevated Pb	30	15	0.31	0.70
Predicted Low Pb	67	157		

Table 3: Estimate of potential life earnings lost from IQ detriment that would be saved within a four-year cohort of children due to correct household intervention based on predictive modeling results when Pb household dust concentrations are ≥ 80 mg/kg. Uncertainty is propagated based on the original range in starting BLLs and intervention reduction.

	Starting Pb concentration (mg/kg)[a]	**Starting BLL (μg/dL)[b]**	**Intervention reduction (%)[c]**	**BLL Decline (μg/dL)**	**Population <5 yrs old exposed to household Pb ≥ 80 mg/kg[d] (millions)**	**Model Sensitivity**	**IQ points saved (millions)[e]**	**Earnings potential saved (trillions USD)[f]**
Global	225	2.4 ± 1.2	15 ± 10	0.36 ± 0.12	358	0.7	48.7 ± 16.2	1.10 ± 0.37
						0.8*	55.7 ± 18.6	1.26 ± 0.42
						0.9	62.6 ± 20.9	1.42 ± 0.47
Australia, U.K., U.S.	225	2.4 ± 1.2	15 ± 10	0.36 ± 0.12	13	0.7	1.77 ± 0.59	0.04 ± 0.01
						0.8*	2.0 ± 0.67	0.05 ± 0.02
						0.9	2.3 ± 0.76	0.05 ± 0.02

*Our current models for U.K., Australia, U.S.

[a]Based on geometric mean of Global DustSafe Pb data ≥ 80 mg/kg from this study, all North American samples (even those without survey responses), and Nigeria (Isley et al., 2022)—n = 1378.

[b]Uses conservative baseline of 0.7 μg/dL based on U.S. median BLLs of children 1-5 years in 2015-2016 (U.S. EPA, 2019), which is likely much higher in low- and middle-income countries (e.g., Ericson et al., 2021a), then the relationship between soil Pb concentrations and increases of BLLs over background for 200 mg/kg Pb from Lanphear et al. (1998)

[c]Based on 17% average BLL reduction through household Pb intervention in Rhoads et al. (1999).

[d]**(United Nations – Population Division, 2019)**, based on assumption of 52.9% of global population <5 years old exposed to household dust Pb ≥ 80 mg/kg (Fig. 2)—from 2020 data (global data rounded down from 359 million to be conservative)

[e]Based on IQ points lost per μg/dL of BLL for the range of 2–10 μg/dL from Boyle et al. (2021): [$\mu = 0.54$] * BLL = IQ points lost

[f]Based on estimates of lifetime earnings for males ($1,413,313) and females ($1,156,157), and lifetime productivity decrease between 1.76% to 2.37% for each IQ point lost, used in Boyle et al. (2021) and Attina and Trasande (2013). Here, we used the minimum productivity decrease of 1.76% per IQ point lost to be conservative, which is $24,874 for males, and $20,348 for females

per IQ point. Because global population is roughly 1 male:1 female (~1.02 male:female), we took the arithmetic mean between both monetary values for $22,611.

References

Attali,, D. (2020). shinyjs: Easily Improve the User Experience of Your Shiny Apps in Seconds. R package version 2.0.0. https://CRAN.R-project.org/package=shinyjs

Attina, T. M., & Trasande, L. (2013). Economic costs of childhood lead exposure in low-and middle-income countries. *Environmental health perspectives*, *121*(9), 1097-1102.

Boyle, J., Yeter, D., Aschner, M., & Wheeler, D. C. (2021). Estimated IQ points and lifetime earnings lost to early childhood blood lead levels in the United States. *Science of the Total Environment*, *778*, 146307.

Chang, W., Cheng, J., Allaire, J.J., Sievert, C., Schloerke, B., Xie, Y., Allen, J., McPherson, J., Dipert, A., Borges, B. (2021). shiny: Web Application Framework for R. R package version 1.6.0. https://CRAN.R-project.org/package=shiny

Chang, W. & Borges Ribeiro, B. (2018). shinydashboard: Create Dashboards with 'Shiny'. R package version 0.7.1. https://CRAN.R-project.org/package=shinydashboard

Dietrich, M., Shukle, J. T., Krekeler, M. P., Wood, L. R., & Filippelli, G. M. (2022). Using Community Science to Better Understand Lead Exposure Risks. *GeoHealth*, 6(2), e2021GH000525.

Egendorf, S. P., Gailey, A. D., Schachter, A. E., & Mielke, H. W. (2020). Soil toxicants that potentially affect children's health. *Current problems in pediatric and adolescent health care*, *50*(1), 100741.

Ericson, B., Hu, H., Nash, E., Ferraro, G., Sinitsky, J., & Taylor, M. P. (2021a). Blood lead levels in low-income and middle-income countries: a systematic review. *The Lancet Planetary Health*, *5*(3), e145-e153.

Ericson, B., Hu, H., Nash, E., Ferraro, G., Sinitsky, J., & Taylor, M. P. (2021b). All lead exposures matter–Authors' reply. *The Lancet Planetary Health*, *5*(12), e860.

Gailey, A. D., Schachter, A. E., Egendorf, S. P., & Mielke, H. W. (2020). Quantifying soil contamination and identifying interventions to limit health risks. *Current Problems in Pediatric and Adolescent Health Care*, *50*(1), 100740.

Gibson, J. M., Fisher, M., Clonch, A., MacDonald, J. M., & Cook, P. J. (2020). Children drinking private well water have higher blood lead than those with city water. *Proceedings of the National Academy of Sciences*, *117*(29), 16898-16907.

Granjon, D. (2021). shinydashboardPlus: Add More 'AdminLTE2' Components to 'shinydashboard'. R package version 2.0.2. https://CRAN.R-project.org/package=shinydashboardPlus

Gulson, B., & Taylor, A. (2017). A simple lead dust fall method predicts children's blood lead level: New evidence from Australia. *Environmental Research*, *159*, 76-81.

Gundacker, C., Forsthuber, M., Szigeti, T., Kakucs, R., Mustieles, V., Fernandez, M. F., ... & Saber, A. T. (2021). Lead (Pb) and neurodevelopment: A review on exposure and biomarkers of effect (BDNF, HDL) and susceptibility. *International Journal of Hygiene and Environmental Health*, *238*, 113855.

Isley, C. F., Fry, K.L., Liu, X., Filippelli, G.M., Entwistle, J.A., Martin, A.P., Kah, M., Figueroa, D.M., Shuckle J., Jabeen, K., Famuyiwa, M.A., Wu, L., Sharifi Soltani, N., Doyi, I.N.Y., Argyraki, A., Ho, K.F., Dong, C., Gunkel-Grillon, P., Aelion, C. Marjorie, Taylor, M. P. (2022). Global analysis of sources and human health risk associated with trace metal contaminants in residential indoor dust. *Environmental Science & Technology*. https://doi.org/10.1021/acs.est.1c04494

Lanphear, B.P., Weitzman, M., Winter, N.L., Eberly, S., Yakir, B., Tanner, M., Emond, M. and Matte, T.D. (1996). Lead-contaminated house dust and urban children's blood lead levels. *American Journal of Public Health*, *86*(10), 1416-1421.

Lanphear, B. P., Burgoon, D. A., Rust, S. W., Eberly, S., & Galke, W. (1998). Environmental exposures to lead and urban children's blood lead levels. *Environmental Research*, *76*(2), 120-130.

Lioy, P. J., Yiin, L. M., Adgate, J., Weisel, C., & Rhoads, G. G. (1998). The effectiveness of a home cleaning intervention strategy in reducing potential dust and lead exposures. *Journal of Exposure Analysis and Environmental Epidemiology*, 8, pp. 17-35

Meyer, P. A., Brown, M. J., & Falk, H. (2008). Global approach to reducing lead exposure and poisoning. *Mutation research/reviews in mutation research*, *659*(1-2), 166-175.

Mielke, H. W., Gonzales, C. R., Smith, M. K., & Mielke, P. W. (1999). The urban environment and children's health: soils as an integrator of lead, zinc, and cadmium in New Orleans, Louisiana, USA. *Environmental research*, *81*(2), 117-129.

Mulhern, R., Roostaei, J., Schwetschenau, S., Pruthi, T., Campbell, C., & Gibson, J. M. (2022). A new approach to a legacy concern: Evaluating machine-learned Bayesian networks to predict childhood lead exposure risk from community water systems. *Environmental Research*, *204*, 112146.

Obeng-Gyasi, E., Roostaei, J., & Gibson, J. M. (2021). Lead Distribution in Urban Soil in a Medium-Sized City: Household-Scale Analysis. *Environmental Science & Technology*, *55*(6), 3696-3705.

Potash, E., Ghani, R., Walsh, J., Jorgensen, E., Lohff, C., Prachand, N., & Mansour, R. (2020). Validation of a Machine Learning Model to Predict Childhood Lead Poisoning. *JAMA network open*, *3*(9), e2012734-e2012734.

R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/

Rasmussen, P. E., Beauchemin, S., Chénier, M., Levesque, C., MacLean, L. C., Marro, L., ... & Gardner, H. D. (2011). Canadian house dust study: lead bioaccessibility and speciation. *Environmental science & technology*, *45*(11), 4959-4965.

Rhoads, G. G., Ettinger, A. S., Weisel, C. P., Buckley, T. J., Goldman, K. D., Adgate, J., & Lioy, P. J. (1999). The effect of dust lead control on blood lead in toddlers: a randomized trial. *Pediatrics*, *103*(3), 551-555.

Rudnick, R. L., & Gao, S. (2003). Composition of the continental crust. Treatise on geochemistry, 3, 659.

Schwarz, K., Weathers, K. C., Pickett, S. T., Lathrop, R. G., Pouyat, R. V., & Cadenasso, M. L. (2013). A comparison of three empirically based, spatially explicit predictive models of residential soil Pb concentrations in Baltimore, Maryland, USA: understanding the variability within cities. *Environmental geochemistry and health*, *35*(4), 495-510.

Taylor, M. P., Isley, C. F., Fry, K. L., Liu, X., Gillings, M. M., Rouillon, M., ... & Filippelli, G. M. (2021). A citizen science approach to identifying trace metal contamination risks in urban gardens. *Environment International*, *155*, 106582.

U.S. EPA [U.S. Environmental Protection Agency]. 2019. America's Children and the Environment | Third Edition. https://www.epa.gov/sites/default/files/2019-07/documents/ace3-lead-updates.pdf

United Nations – Population Division. (2019). Our World in Data: Number of children under 5 years old, World. https://ourworldindata.org/grapher/under-5-population?time=2020..latest&country=~OWID_WRL

Venables W.N., Ripley B.D. (2002). *Modern Applied Statistics with S*, Fourth edition. Springer, New York. ISBN 0-387-95457-0, https://www.stats.ox.ac.uk/pub/MASS4/.

Yesilonis, I. D., Pouyat, R. V., & Neerchal, N. K. (2008). Spatial distribution of metals in soils in Baltimore, Maryland: role of native parent material, proximity to major roads, housing age and screening guidelines. *Environmental Pollution*, *156*(3), 723-731.