nature sustainability

ARTICLES

https://doi.org/10.1038/s41893-022-00944-2

Check for updates

Reducing the burden of anaemia in Indian women of reproductive age with clean-air targets

Ekta Chaudhary[1], Sagnik Dey[1,2,3]✉, Santu Ghosh[4]✉, Sumit Sharma[5], Nimish Singh[5], Shivang Agarwal[5], Kushal Tibrewal[6], Chandra Venkataraman[6], Anura V. Kurpad[4], Aaron J. Cohen[7,8,9], Shuxiao Wang[10] and Srishti Jain[1,11]

India has one of the highest (53%) global prevalences of anaemia among women of reproductive age (WRA, 15–49 years). Long-term exposure to ambient fine particulate matter ($PM_{2.5}$), a type of air pollution, may increase the prevalence of anaemia through systemic inflammation. Using a linear mixed model adjusted for potential confounding factors, we show that for every 10 μg m^{-3} increase in ambient $PM_{2.5}$ exposure, the average anaemia prevalence among Indian WRA increases by 7.23% (95% uncertainty interval, 6.82–7.63). Among $PM_{2.5}$ species, sulfate and black carbon are more associated with anaemia than organics and dust. Among sectoral contributors, industry was the greatest, followed by the unorganized, domestic, power, road dust, agricultural waste burning and transport sectors. If India meets its recent clean-air targets, such anaemia prevalence among WRA will fall from 53% to 39.5%, taking 186 districts below the national target of 35%. Our results suggest that the transition to clean energy would accelerate India's progress towards the 'anaemia-free' mission target.

Anaemia is a leading cause of global disease burden that is characterized by a diminished blood haemoglobin (Hb) concentration, often accompanied by a decrease in red blood cells, resulting in a decrease in the oxygen-carrying capacity of the blood[1]. Women of reproductive age (WRA), aged 15–49 years, may suffer from regular iron deficiency due to menstruation and therefore are particularly prone to develop anaemia (from mild to severe)[2]. Dietary iron deficiency is another chief cause of anaemia[3], while other contributing factors include genetic disorders, parasitic infections and inflammation from infections[4] and chronic diseases[1]. The World Health Organization (WHO) has set a global nutrition target to reduce anaemia in WRA by 50% by 2025[3]. Anaemia prevalence in WRA is also an indicator to track progress in eradicating all forms of malnutrition[5].

Anaemia is highly prevalent in India. The National Family and Health Survey 2015–2016 (NFHS-4) reported that 53.1% of WRA and 58.5% of children under five were anaemic[6]. India launched a national programme under the POSHAN Abhiyaan ('nutrition mission') aiming to make the country 'anaemia-free' and set a target of reducing anaemia in WRA to below 35% by 2022. Because an iron-deficient diet is the primary cause of the large burden of anaemia, the Ministry of Health and Family Welfare is engaged in increasing the iron intake of the population through supplementation and fortification[7]. However, recent studies have shown that dietary supplementation alone may not substantially reduce anaemia among WRA in India and that the benefits and risks may vary across states[5,8].

One reason why increased iron intake may not reduce iron deficiency anaemia is if the ingested iron is not absorbed. This can happen in an inflammatory state when the liver produces an increased level of hepcidin[9]. Erythrocyte production and Hb level can be reduced as a result of its direct negative impact on erythropoiesis and reduced dietary iron absorption[10]. Long-term particulate matter ($PM_{2.5}$) exposure can also cause systemic inflammation and may increase the risk of developing anaemia[10,11]. However, the evidence for this link is confined to high-income countries and is lacking in low- and middle-income countries[9,12]. India has one of the world's highest levels of ambient $PM_{2.5}$ exposure, exceeding the WHO annual air quality guideline of 5 μg m^{-3} many times[13]. The Global Burden of Disease India study identified child and maternal malnutrition as the largest risk factor for ill health, followed by the risk of $PM_{2.5}$ exposure[14]. Epidemiological studies in India based on the impact of ambient $PM_{2.5}$ exposure on health outcomes are ever-widening[15,16]. Ambient $PM_{2.5}$ exposure is a leading risk factor for child anaemia in India[17], although whether the association holds true for WRA is not known. Moreover, the differential impacts, if any, of exposure to the composition of $PM_{2.5}$ on various health outcomes, including anaemia, are not known. Some studies of long- and short-term exposure have reported that particulates emitted from combustion processes have more adverse health effects compared with particulates from non-combustion sources, such as sea salt and crustal emissions[18,19], but other studies[20] and reviews have failed to find strong evidence of systematic differences in toxicity.

In this study, we estimated the effect of long-term exposure to ambient $PM_{2.5}$ on anaemia prevalence among WRA in India. We statistically matched two different Indian national surveys on anaemia and diet using 'case identification (CASEID)' of the participants (NFHS-4 for anaemia and the 68th round of the National Sample Survey Office (NSSO) for diet), conducted within a four-year interval (2013–2016), and evaluated their association over varying levels of dietary intake of iron and other related nutrients[21]. We estimated long-term exposure to annual ambient $PM_{2.5}$ using satellite-based

[1]Centre for Atmospheric Sciences, Indian Institute of Technology (IIT) Delhi, New Delhi, India. [2]Centre of Excellence for Research on Clean Air, IIT Delhi, New Delhi, India. [3]School of Public Policy, IIT Delhi, New Delhi, India. [4]St. John's Medical College, Bengaluru, India. [5]TERI, New Delhi, India. [6]Department of Chemical Engineering, IIT Bombay, Bombay, India. [7]Health Effects Institute, Boston, MA, USA. [8]Institute for Health Metrics and Evaluation, Seattle, WA, USA. [9]Boston University School of Public Health, Boston, MA, USA. [10]State Key Joint Laboratory of Environmental Simulation and Pollution Control, School of Environment, Tsinghua University, Beijing, China. [11]Department of Chemistry, University College Cork, Cork, Ireland. ✉e-mail: sagnik@cas.iitd.ac.in; santu.g@stjohns.in

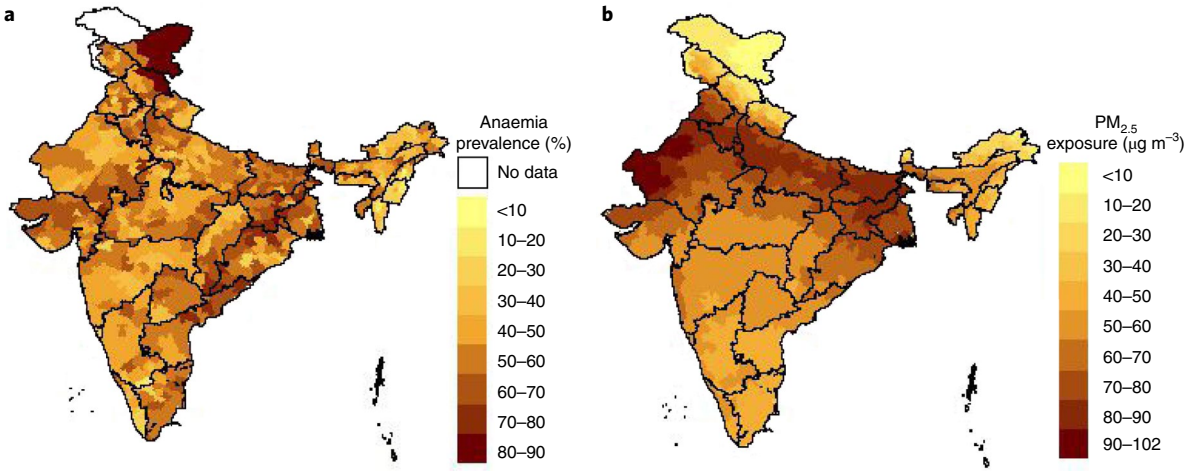

Fig. 1 | Spatial patterns of anaemia prevalence (%) and $PM_{2.5}$ exposure in India. a,**b**, Prevalence of anaemia amongst women of reproductive age by district within each state in India (**a**) and ambient $PM_{2.5}$ exposure (averaged from 2007–2016) by district in each state (**b**), measured in μg m^{-3}.

data[22] for the ten years (2007–2016) preceding the NFHS-4 survey that captured the anaemia prevalence data (Fig. 1) at the aggregate as well as the individual level. For $PM_{2.5}$ composition, we used MERRA-2 reanalysis data to estimate long-term exposure to the principal $PM_{2.5}$ components: black carbon (BC), organic carbon (OC), sulfate (SO_4^{2-}), soil dust and sea salt (Methods). Further, we examined the impacts of exposure to sector-specific $PM_{2.5}$ using data from two different chemical transport models (Methods). Finally, using the exposure–response function derived from these data, we estimated the health benefits (represented in terms of the expected reduction in anaemia prevalence) at the district level due to a decrease in long-term $PM_{2.5}$ exposure following the various clean-air targets set by the Government of India (GoI).

Results

Impact of long-term ambient $PM_{2.5}$ exposure on the burden of anaemia. We estimated the anaemia prevalence from 640,557 observations for WRA recorded in NFHS-4 data across 636 districts after excluding 59,129 observations that had missing Hb concentration, $PM_{2.5}$ exposure and covariate data (Supplementary Fig. 1). The anaemia prevalence across Indian districts in 2015–16 was 53.1% (range 22.4–80.0%) with a slightly lower prevalence in urban India (50.9%, range 11.5–75.4%) than in rural India (54.3%, range 9.2–83.9%)[6,21]. Anaemia prevalence varied substantially among states (Fig. 1), ranging from 22.6% in Nagaland to 64.4% in Jharkhand. Ambient $PM_{2.5}$ exposure ranged from 35 to 90 μg m^{-3} (Fig. 1).

We used a multivariate logistic regression model to estimate the effect of long-term ambient $PM_{2.5}$ exposure on anaemia prevalence in WRA. The model was adjusted for several risk factors: daily iron intake (from the National Sample Survey, NSSO-68), body mass index (BMI), smoking, second-hand smoke, education level, cooking fuel, wealth index and residence (urban/rural). At the ecological level, we estimated a 7.23% (95% uncertainty interval (UI), 6.82–7.63) increase in anaemia prevalence for every 10 μg m^{-3} increase in long-term ambient $PM_{2.5}$ exposure after adjusting for the leading confounders in the multivariate model (Methods). The corresponding decrease in average Hb level was 0.57 g L^{-1} (95% UI, 0.68–0.54). We assessed potential nonlinearities in the associations of the probability of anaemia with ambient $PM_{2.5}$ exposure using a generalized additive model (Fig. 2). Anaemia prevalence increased monotonically with $PM_{2.5}$ exposure.

We also carried out an aggregate-level analysis for 636 districts, excluding four districts (Lakshadweep, Nicobar, North and Middle Andaman, South Andaman) outside the continental Indian landmass (Supplementary Table 3). For every 10 μg m^{-3} increase in ambient $PM_{2.5}$ exposure, the adjusted average Hb level was found to decrease by 0.44 g L^{-1} (95% UI: 0.47–0.41). The adjusted odds ratio (OR) of anaemia was 1.07 (95% UI, 1.06–1.08) for every 10 μg m^{-3} increase in ambient $PM_{2.5}$ exposure when anaemia was modelled as a binary outcome across the multivariate logistic regression model.

Effect modification. Fig. 3 compares the adjusted OR with 95% UI of anaemia for every 10 μg m^{-3} increase in ambient $PM_{2.5}$ exposure stratified by different socio-demographic factors that potentially modify the effects of long-term $PM_{2.5}$ exposure on anaemia among WRA. The first to the fifth national wealth quintiles are categorized as the 'poorest' to the 'richest' wealth categories, respectively. The weight categories are based on BMI (kg m^{-2}): underweight <18.5; normal 18.5–24.9; overweight 25–29.9; obese ≥30. The impact of $PM_{2.5}$ on anaemia prevalence increased as socio-economic status improved; the OR ranged from 1.060 (95% UI, 1.052–1.069) for the poorest to 1.097 (95% UI, 1.088–1.106) for the richest for every 10 μg m^{-3} increase in ambient $PM_{2.5}$ exposure. Also, when we compared the adjusted changes in Hb level for every 10 μg m^{-3} increase in ambient $PM_{2.5}$ exposure, Hb level decreased with improved socio-economic status (Fig. 3). For example, the slope ranges from −0.24 (95% UI, −0.32 to −0.16) for the poorest to −0.82 (95% UI, −0.88 to −0.76) for the richest for every 10 μg m^{-3} increase in ambient $PM_{2.5}$ exposure. The relative effects are larger for the individuals not exposed to second-hand smoke (OR, 1.083; 95% UI, 1.077–1.089) relative to those who were exposed to second-hand smoke (OR, 1.063; 95% UI, 1.057–1.068). The urban population (OR, 1.078; 95% UI, 1.071–1.084) is more susceptible to the impact of ambient $PM_{2.5}$ exposure on anaemia prevalence than the rural population (OR, 1.069; 95% UI, 1.063–1.074). The households using kerosene are the most affected compared with households using other unclean fuel and clean fuel. The impact is also inversely related to socio-demographic factors such as 'education level' and 'BMI'. Additional covariates of interest such as age and vitamin A intake were adjusted into the model and assessed for effect size. However, these variables showed negligible change in effect size and weak significance; therefore, they were excluded from the model.

We also calculated the relative excess risk due to interaction (RERI) and its 95% UI using unit additive interaction estimated through a logistic regression model, to analyse whether there were cases of anaemia that were due to joint exposure to air pollution and other risk factors and, therefore, that acting on either of the two factors would reduce the effect of air pollution exposure. The

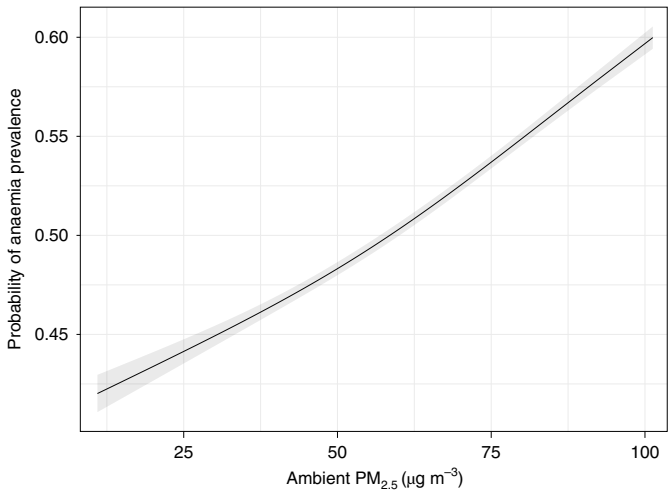

Fig. 2 | Exposure–response curve of ambient $PM_{2.5}$ long-term (2007–2016) exposure against anaemia prevalence. The curve indicates the probability of anaemia for every 10 μg m^{-3} increase in $PM_{2.5}$ at an ecological level. Shaded regions denote 95% UI.

higher-order polynomial regression coefficients were very small in our study (Supplementary Table 2). RERI was found to be positive, with a certain significance level, at less than 0.05 for all the risk factors considered in adjusting the model.

Differential impact of exposure to $PM_{2.5}$ species and sources. To assess the impact of long-term exposure to $PM_{2.5}$ composition on the burden of anaemia in WRA, we carried out an aggregate-level analysis. We found a 6.13% (95% UI, 5.20–7.12%) increase in anaemia prevalence for every 10 μg m^{-3} increase in long-term ambient $PM_{2.5}$ exposure (with MERRA-2 $PM_{2.5}$) using a multivariable logistic regression model adjusted for several known risk factors: daily iron intake (from NSSO-68), BMI, smoking, second-hand smoke, education level, cooking fuel, wealth index and residence (urban/rural). This was slightly lower than the effect observed with satellite-based exposure data, as $PM_{2.5}$ from MERRA-2 is underestimated in India compared with satellite-derived estimates[23]. The spatial patterns of long-term exposure (2007–2016) to individual $PM_{2.5}$ species from MERRA-2 reanalysis data are shown in Supplementary Fig. 3. The largest OR is found for the exposure to SO_4^{2-} (OR, 1.26; 95% UI, 1.25–1.27) and BC (OR, 1.25; 95% UI, 1.23–1.26), followed by dust (OR, 1.21; 95% UI, 1.20–1.22), OC (OR, 1.10; 95% UI, 1.09–1.11) and sea salt (OR, 1.04; 95% UI, 1.03–1.045).

To understand the differential exposure in view of the relative source contributions to ambient $PM_{2.5}$ species, we used two different chemical transport models: WRF-CMAQ and GEOS-chem (Methods). The uncertainty in sectoral $PM_{2.5}$ estimated from a chemical transport model depends on the representativeness of the emission inventory and the model calculation of secondary species. Sector-specific $PM_{2.5}$ cannot be directly validated; therefore, we used results from the two different models to check the consistency. The sectoral contributions to ambient $PM_{2.5}$ from the WRF-CMAQ model (2016) and GEOS-chem model (2015) are shown in Supplementary Figs. 4 and 9. We found that the largest share of $PM_{2.5}$ concentration was contributed by the domestic sector in India (Supplementary Fig. 9). To compare sectoral contributions to $PM_{2.5}$ from both the models and analyse them against anaemia prevalence, we adapted the GEOS-chem model concentrations from 50 km × 67.5 km to 36 km × 36 km^2 to match the WRF-CMAQ model resolution. The contributions of sectors such as transport and road dust concentration are high in urban regions[24], and a higher exposure is expected at local scales. The impact of transboundary pollutants is higher in the areas near the north-western political boundary of the country. The annual population-weighted contribution of sectors to $PM_{2.5}$ concentration are similar to the results reported earlier in the GBD-MAPS study[25]. Furthermore, in the case of WRF-CMAQ outputs, for every interquartile range increase in $PM_{2.5}$ exposure attributable to the sectoral emissions, the OR is the highest (OR, 1.22; 95% UI, 1.21–1.23) for the industrial sector, followed by the unorganized (for example, municipal waste burning and crematorium) sectors (OR, 1.19; 95% UI, 1,18–1.20) and an equal impact of the domestic (OR, 1.12; 95% UI, 1.10–1.13) and power (OR, 1.12; 95% UI, 1.11–1.12) sectors. The impacts of exposure to the road dust (OR, 1.09; 95% UI, 1.08–1.10) and agriculture waste burning sectors (OR, 1.08; 95% UI, 1.07–1.09) on anaemia prevalence are comparable in India, while the impacts of the transboundary (OR, 1.05; 95% UI, 1.04–1.06) and transport sector (OR, 1.07; 95% UI, 1.06–1.08) are slightly lower. The analysis of the outputs from the GEOS-chem model revealed similar patterns. For every interquartile range increase in $PM_{2.5}$ exposure attributable to the sectoral emissions from the GEOS-chem model, the OR is the highest (OR, 1.28; 95% UI, 1.26–1.29) for the industrial sector, followed by the power (OR, 1.27; 95% UI, 1.26–1.28) and other (comprising anthropogenic mineral matter, natural dust, residential lighting using kerosene lamps, informal industry, waste burning and fugitive dust) sectors (OR, 1.24; 95% UI, 1.23–1.26). The OR for the domestic (OR, 1.11; 95% UI, 1.10–1.12) and transport (OR, 1.11; 95% UI, 1.10–1.13) sectors are quite close to the results from the WRF-CMAQ model. The impact of the open burning (agricultural waste burning) sector (OR, 1.03; 95% UI, 1.02–1.04) is slightly lower in this model.

The benefits of meeting clean-air targets. The GoI announced three important social programmes in recent years which contribute to reductions in air pollution, both outdoor and indoor. The National Clean Air Programme (NCAP), launched in 2019, aimed at reducing ambient $PM_{2.5}$ levels by 30% by 2024 relative to the levels in 2017 (ref. [26]). The mitigation measures for the NCAP focused only on outdoor sources. The Pradhan Mantri Ujjwala Yojana (PMUY) was launched in 2016 to provide clean fuel to the 80 million households that depend on solid fuel for cooking and heating[27]. The Deen Dayal Upadhyay Grameen Jyoti Yojana (DDUGJY) was mandated to electrify rural India, thereby eradicating the use of kerosene for lighting. Household emissions were found to be the largest (about 30%) contributor to ambient $PM_{2.5}$ in India[28]. In fact, complete mitigation of household emissions in 2015 could reduce the ambient $PM_{2.5}$ exposure below the national ambient air quality standard of 40 μg m^{-3} (refs. [27,28]). Since these recent policies are expected to reduce ambient $PM_{2.5}$ exposure substantially, we analysed the potential co-benefits of successful implementation of these programmes on reducing the burden of anaemia in India using the exposure–response function discussed (Methods).

As per the NFHS-4, only 61 districts have anaemia prevalence below 35%, and 575 districts (90.4% of the total 636 districts considered here) do not meet the target of <35% anaemia prevalence in WRA set by the GoI for its 'anaemia-free' national mission (Fig. 1). If India could implement the NCAP successfully and reduce ambient $PM_{2.5}$ exposure by 30% across the districts, 110 districts (17.3%) would meet the target of <35% anaemia prevalence (Fig. 4a). By completely mitigating emissions from household sources, an additional 111 districts (17.4%) would meet the target (Fig. 4b). As a combined impact of exposure reduction through the NCAP, PMUY and DDUGJY programmes, 186 districts (29.2%) would be able to reduce anaemia prevalence to below 35% (Fig. 4c). If India meets the WHO air quality guideline of 5 μg m^{-3}, the national anaemia prevalence would come down to 34.6% (Fig. 4d), meeting the 'anaemia-free' mission target. In this scenario, anaemia prevalence would be below 35% in 340 districts (53.5%). District-level ambient

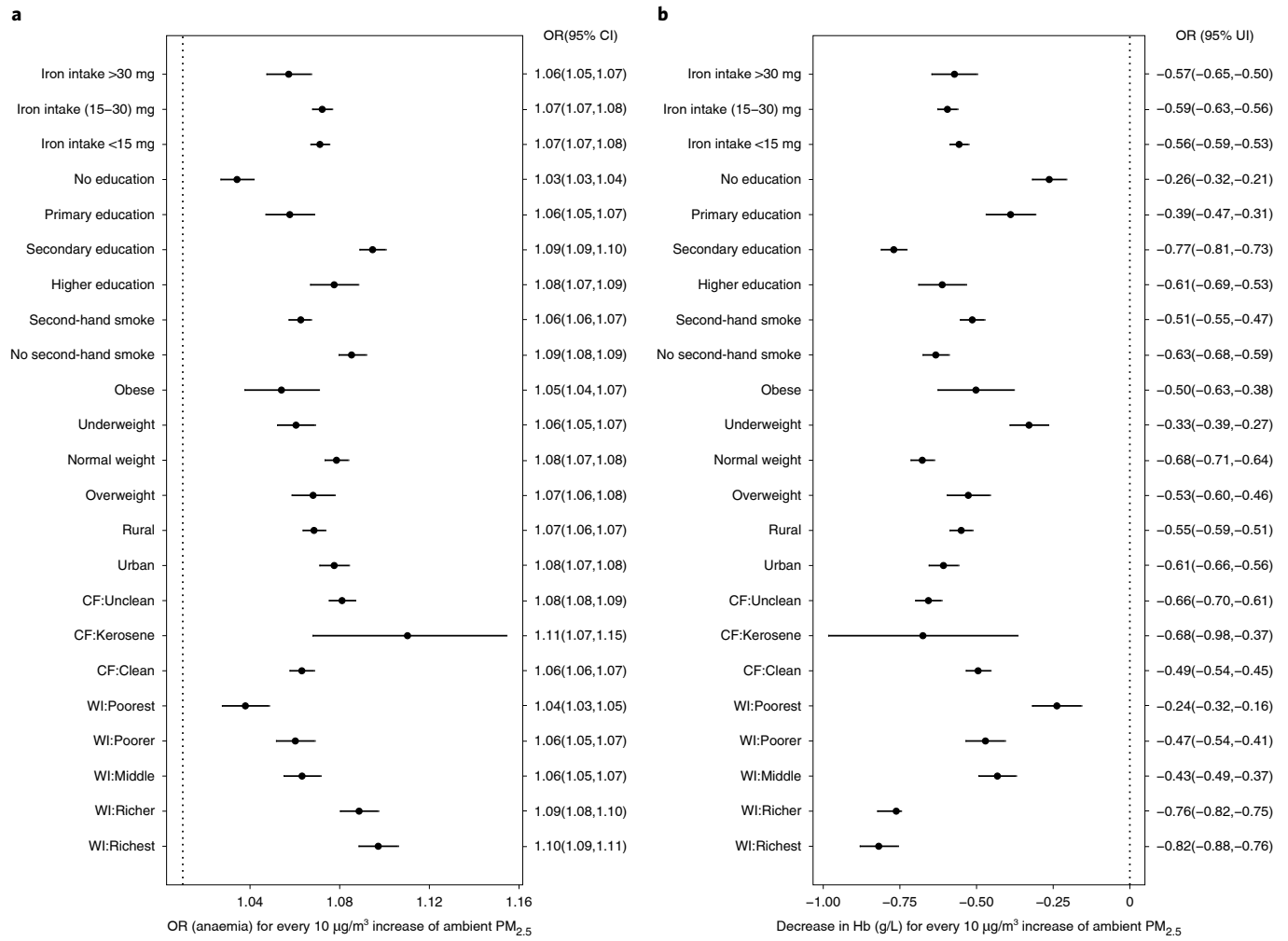

Fig. 3 | Effect of $PM_{2.5}$ exposure on odds ratio of anaemia and Hb level among WRA in India. a,**b**, Adjusted OR of anaemia in WRA (**a**) and decrease in Hb level (**b**) for every 10 µg m^{-3} increase of ambient $PM_{2.5}$ exposure stratified by different socio-demographic and individual-level factors.

$PM_{2.5}$ exposure for the period 2007–2016, anaemia prevalence (in per cent) captured by the NHFS-4, anaemia prevalence after meeting clean-air targets and the WHO air quality guideline are summarized in Supplementary Table 3.

Discussion

In India, there are no epidemiological studies that investigate associations between $PM_{2.5}$ components and sector-specific $PM_{2.5}$ and health effects, especially in the context of anaemia. Here we focus on the association between $PM_{2.5}$ and its components with anaemia prevalence among WRA. The ambient $PM_{2.5}$, adjusted for average district-level risk factors, estimated a 7.23% increase in average anaemia prevalence in WRA for every 10 µg m^{-3} increase in ambient $PM_{2.5}$. The impact was almost four times higher than that on children (under five years), as found recently by Mehta et al., which reported a 1.9% increase in child anaemia prevalence for every 10 µg m^{-3} increase in ambient $PM_{2.5}$ (ref.[17]). Similar associations for children under five were reported by Morales-Ancajima et al.[9] in Lima, Peru, where they observed a 39.6% (95% UI, 39.3–39.9) increase in anaemia prevalence and a slight decrease in Hb (OR, −0.03, 95% UI, −0.05 to −0.02) with an increase in $PM_{2.5}$ exposure. The smaller impact in children may imply a more adverse effect with longer exposure, or that adults are simply more likely to have greater cumulative exposure to outdoor air pollution.

Socio-economic status (SES) can be an important modifier of the effects of air pollution on health[29]. Many studies have found that poorer communities have greater susceptibility (stronger pollution–disease associations) to health impacts due to air pollution. However, this finding has not been consistent in the literature[29,30]. Inconsistencies may be due to nonlinearities in susceptibility, including potential threshold, saturation effects or simply a relatively lower exposure to other aetiologies of anaemia as wealth increases; alternatively, they may be due to differences in the SES indicators used, or in the relative distributions of SES among the individuals represented in any given population (especially when comparing populations across very different countries or communities). Many researchers have also explored the role of psychosocial stress, an important product of life in many lower SES settings, as an effect modifier of the air pollution–health association[31,32]. The effects have been inconsistent in other settings[33]. The other factor that may play a crucial role in effect modification is variation in $PM_{2.5}$ composition related to the source characteristics.

The reported RERI (Supplementary Table 2) and effect modifications (Fig. 3) suggest that there are excess cases of anaemia due to joint exposure to air pollution and other risk factors considered in our study. Hence, reducing these other risk factors will also reduce the number of cases caused by air pollution exposure. The additive interaction between smoking and $PM_{2.5}$ was positive with RERI 0.55 (95% UI, 0.54–0.55), indicating a sufficient-cause interaction.

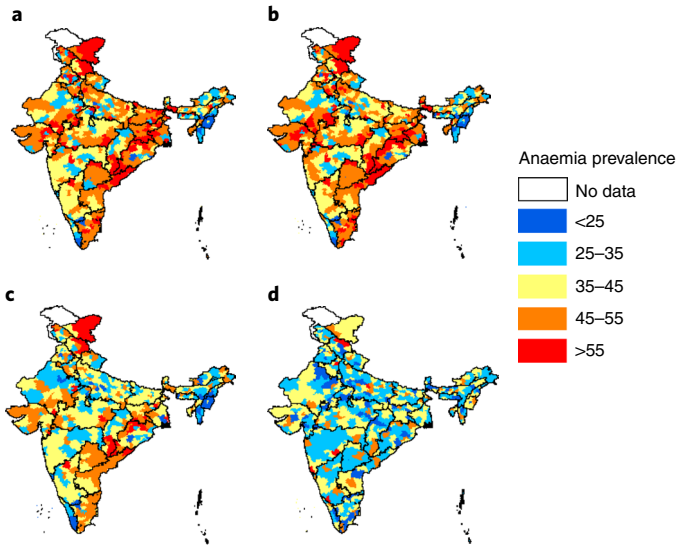

Fig. 4 | Anaemia prevalence and clean-air targets. a–d, Anaemia prevalence (%) (**a**) after the NCAP implementation where ambient $PM_{2.5}$ exposure is reduced by 30%, (**b**) after mitigating domestic emissions by implementing PMUY and DDUGJY, (**c**) after reduction in ambient $PM_{2.5}$ exposure by the combined impact of the NCAP, PMUY and DDUGJY programmes and (**d**) if India meets the WHO air quality guideline.

Similarly, for second-hand smoke (1.18; 95% UI, 1.17–1.18), education level (1.64; 95% UI, 1.64–1.65), residence (0.99; 95% UI, 0.99–1.00), wealth index (1.03; 95% UI, 1.02–1.03), BMI (0.95; 95% UI, 0.94–0.95), iron intake (1.00; 95% UI, 0.99–1.00) and cooking fuel (0.79; 95% UI, 0.78–0.80), the interaction was in a positive direction. Our results imply large health benefits of meeting the NCAP target of reducing $PM_{2.5}$ exposure by at least 30%. Further research is recommended to understand the underlying biological mechanism and its response to toxicity from differential exposures. The largest OR for anaemia prevalence was observed for exposure to SO_4^{2-} and BC, followed by dust, OC and sea salt.

An additional aspect of our study is that it is the first of its kind conducted among WRA across India that did not focus only on $PM_{2.5}$ but also on its components and sectoral contributions. Elbarbary et al.[12] assessed the association between ambient air pollution exposure and anaemia prevalence/Hb levels in older Chinese adults, and detected nonlinearities in dose-response curves for PM and NO_2 exposure with respect to Hb levels and anaemia prevalence in the study population. Sun et al.[34] suggested that some specific $PM_{2.5}$ constituents, such as organic carbon, nitrate, sulfate and elemental carbon, may have greater toxic effects than $PM_{2.5}$ itself. We found the largest odds ratios for exposure to SO_4^{2-} (OR, 1.26; 95% UI, 1.25–1.27) and BC (OR, 1.25; 95% UI, 1.23–1.26), followed by dust (OR, 1.21; 95% UI, 1.20–1.22), OC (OR, 1.10; 95% UI, 1.09–1.11) and sea salt (OR, 1.04; 95% UI, 1.03–1.045). Therefore, to trace the contributing sources of SO_4^{2-} and BC, we explored sectoral information. The results show that the industrial, domestic and power sectors have major contributions to sulfate and BC loading. This may explain the relatively higher impacts of these sectors across the models compared with others. The epidemiological and toxicological evidence provide little support for the differential toxicity of PM components. For the higher impact of SO_4^{2-}, Reiss et al.[35] suggested three possible indirect effects of particulate sulfate within the PM mixture in the atmosphere, including (1) catalysis of organic aerosol formation, (2) potential increase in bioavailability of metallic species and (3) enhancement of lung deposition of more toxic compounds. At present, there is insufficient evidence to draw general conclusions about these mechanisms. Our estimates will provide support for policymakers in targeting specific sectors that are contributing to exposure to $PM_{2.5}$ and other pollutants and, in turn, causing lethal health impacts.

We hypothesize that exposure to high levels of ambient $PM_{2.5}$ leads to inflammation through oxidative stress, which is believed to be one of the chief mechanisms responsible for adverse health outcomes caused by particulate matter exposure[36,37]. Chronic inflammation can impact erythrocytes and Hb concentration[38] through immune activation, which can further change the iron trafficking in the body by secreting cytokines (for example, interleukin-6). As a result, the liver is stimulated to produce hepcidin, which in turn binds the cellular transmembrane iron-exporter, ferroprotein. This binding inhibits the iron recycling from senescent erythrocytes through macrophages, and absorption of dietary iron in the duodenum. Subsequently, the inflammatory cytokines may independently suppress erythropoiesis owing to less iron availability for Hb formation. We note that inflammation can occur in a variety of clinical situations (for example, cancer and different organ dysfunctions), while the innate immune system can be activated by obesity[9].

One puzzling aspect was that the impact of ambient $PM_{2.5}$ on anaemia prevalence increased as SES improved. This is counterintuitive, as one would expect that the impact of pollution would be greater in poverty. One possible reason for this is that pollution-induced inflammation in wealthier individuals might have had a greater influence on their iron absorption, as this may have been already downregulated due to better body iron stores, or a higher degree of adiposity. It is known that adipose tissue is inflammatory[39], with hepcidin-related inhibition of iron absorption[40]. Pollution may have simply potentiated the existing inflammation with greater adiposity in the wealthier group. In contrast, the poorer individuals, with poorer quality diets and lower body fat, might have had an upregulated iron absorption, which offset the influence of pollution. This has been shown to be the situation in poorer, rural populations, where higher serum ferritin concentrations and body iron stores were observed in comparison to richer, urban populations[41]. In that study, groups with a high prevalence of inflammation (as measured by C-reactive protein), belonging to the poorer households with poor sanitation, had lower iron deficiency prevalence. This, too, is counterintuitive, but a systematic review has also reported a relatively lower prevalence of iron deficiency in rural than in urban populations[42].

The analysis presented here can be repeated using NFHS-V (2019–2021) data to observe whether anaemia prevalence has changed in recent years. Further analysis using the NFHS-V dataset would provide stronger evidence for the transition to a cleaner household energy ambient environment. We recommend further epidemiological and toxicological studies to understand this relationship and the biological mechanisms that drive the link between air pollution and its causal effects.

Methods

Study population. Our study analyses two Indian national-level survey datasets. Data on anaemia in WRA and their socio-demographic factors were obtained from the NFHS-4 2015–2016[6] and data on daily iron intake were obtained from the 68th round survey (2011–2012) of the NSSO by converting household-level food purchases to per capita micronutrient intake with the help of the Indian Food Composition table[43]. The NFHS-4 survey data provide information on household and individual socio-demographic characteristics and blood biomarker measurements, including Hb from capillary blood samples of 699,686 WRA residing in 601,509 households across 640 districts, covering 29 states and six union territories of India. Primary sampling units (PSUs) were selected based on the 2011 Indian census. PSUs were villages in rural areas and census enumeration blocks (CEBs) in urban areas. Selected PSUs with at least 300 households were subdivided into segments of approximately 100–150 households. For the survey, two of the segments were randomly selected using systematic sampling with probability proportional to segment size. Therefore, an NFHS-4 cluster will either be a PSU or a segment of a PSU. Later, 22 households were randomly selected with systematic sampling in every selected rural and urban cluster (Supplementary Fig. 2). More information about the NFHS-4 survey, sampling frame and survey design is given in Supplementary Note 1.

There were four main survey questionnaires administered in NHFS-4: a household questionnaire, a men's questionnaire, a women's questionnaire and a biomarker questionnaire. The women's questionnaire was given to women between the ages of 15 and 49 years, and yielded a 97% response rate. For more details on sampling design, see the *National Family and Health Survey IV Report* by the International Institute for Population Sciences[6]. The 9th quinquennial Household Consumer Expenditure survey of the 68th round of the NSSO covered all regions of India (29 states and six union territories). More details on the NSSO dataset are given in Supplementary Note 2.

Hb concentrations and anaemia prevalence estimates. Blood specimens were collected from all the women aged between 15 and 49 years who resided in the households selected for the survey. Blood samples were drawn to measure Hb concentration using the HemoCue Hb 201+ analyser in each district[44]. The systematic protocol was used for blood sample processing. The samples were then sent to the designated laboratories for testing.

We determined anaemia for WRA based on Hb level <12.0 g dl^{-1} after adjustment for smoking and altitude in enumeration areas above 1,000 m. District-level anaemia prevalence and average Hb concentration (measured and reported in g l^{-1}) are considered primary outcomes for the ecological analysis. For the individual-level analysis, we estimated individual exposure to ambient $PM_{2.5}$ by clustering the exposure using geocode information of each PSU in the NFHS-4 data.

Ambient $PM_{2.5}$ exposure assessment. Our primary exposure metric is ten-year (2007–2016) average ambient $PM_{2.5}$ exposure, modelled as a continuous variable. We estimated long-term exposure for all 636 districts from the national $PM_{2.5}$ database created for India[22]. This database was created by converting MODIS-MAIAC aerosol optical depth to surface $PM_{2.5}$ using a dynamic scaling factor from MERRA-2 reanalysis data. The instantaneous $PM_{2.5}$ (representing the satellite overpass time) was then converted to a 24-hr average using the diurnal scaling factor from MERRA-2. Both these scaling factors were calibrated against the data from the existing ground-based network of the Central Pollution Control Board of India (CPCB). At the annual scale, satellite-derived $PM_{2.5}$ concentration shows a correlation coefficient of 0.97 and a root mean square error of 7.2 μg m^{-3} with the coincident ground-based measurements from the CPCB network. We used the population distribution from the Indian census to estimate the exposure at a 1-km resolution for 2007–2016. For the individual-level analysis, individual exposure to ambient $PM_{2.5}$ was calculated by clustering the exposure using geocode information of each PSU in the NFHS-4 data.

$PM_{2.5}$ composition. India lacks long-term $PM_{2.5}$ composition data. Hence, we used MERRA-2 reanalysis data for annual-averaged BC, sulfate, OC, dust and sea salt information, bilinearly re-gridded from 0.5°×0.6° to 1 km×1 km to match the resolution of the satellite-based $PM_{2.5}$ dataset. In our study, the MERRA-2 total $PM_{2.5}$ was estimated using the methodology of Bali et al.[23] and Buchard et al.[45].

Bali et al.[23] presented a detailed account for validation and bias correction of the MERRA-2 data for use in air quality studies in India. We downscaled the MERRA-2 $PM_{2.5}$ exposure to 1 km^2 to capture the exposure information of districts (smaller in size) that are missed at coarser resolution. MERRA-2 data shows a significant correlation with the satellite-derived $PM_{2.5}$ from the same time, with an R^2 value of 0.88 (Supplementary Fig. 8).

The spatial patterns of chronic exposure to individual $PM_{2.5}$ species from MERRA-2 are shown in Supplementary Fig. 3. We used the $PM_{2.5}$ composition data measured in Delhi and Varanasi[46] as well as other available data from the literature (Supplementary Table 4) during the period 2007–2016 to evaluate the MERRA-2 $PM_{2.5}$ composition. The R^2 values were 0.63 for $PM_{2.5}$, BC, OC and SO_4^{2-} (Supplementary Fig. 6). We also investigated the region-specific correlation between ground-based and MERRA-2 measurements for Delhi from 2013 to 2016 (Supplementary Fig. 7). The R^2 value range (0.53–0.88) demonstrated the robustness of the MERRA-2 data used for the analysis.

A generalized linear model with a binomial family was used to assess the effect of single air pollutants on anaemia prevalence. In a single-pollutant model, we introduced one pollutant at a time as a linear term and adjusted for all risk factors.

WRF-CMAQ model. The model set-up of Weather Research Forecasting (WRF) v.3.9.1 (ref. [47]) and Community Multi-scale Air Quality Modelling System (CMAQ) v.5.3.1 (ref. [48]) was used to estimate the ambient $PM_{2.5}$ concentration for the year 2016 over the study domain[49]. The ERA5 dataset of ECMWF was used as an input to the WRF model. The inputs to the CMAQ model included multipollutant emissions at 36 km×36 km resolution and pollutants from transboundary sources along with the meteorological input generated by the WRF model. The national-scale emissions were generated using the Greenhouse Gas and Air Pollution Interactions and Synergies (GAINS)-ASIA model (https://gains.iiasa.ac.at)[50] in which actual consumption data for the different sectors was used as an input for the year 2016. The ECLIPSE (v.5) database of IIASA (2014) was used for the national ammonia emissions, ship emissions and the emissions emitted from those neighbouring countries that fall within the study domain, including Nepal, Myanmar, Bangladesh, Sri Lanka and Pakistan. Additionally, transboundary pollutants coming from outside the study domain were taken from boundary conditions developed by the Community Atmosphere Model with Chemistry (CAM-chem) model[51] (https://www.acom.ucar.edu/cam-chem/cam-chem.shtml). Sectoral data from the CMAQ model (2016) are spatially represented in Supplementary Fig. 4.

GEOS-chem model. The contribution of source sectors to $PM_{2.5}$ concentrations was assessed through simulations using the Goddard Earth Observing System (GEOS)-Chem model v.10.01, with an Indian region emission inventory generated using the Speciated-Multipollutant Generator (SMoG-India-v1; https://ncapcoalesce.iitb.ac.in/resources/smog-india-emission-inventory/). Gridded emissions for 2015 were produced over India[52] for primary particulate matter constituents (that is, BC and OC) and precursor gases (SO_2, NOx, NMVOCs and NH_3). SMoG-India uses an engineering model approach of technology-emissions modelling, using parameters for process and emissions control technologies and technology-linked emission factors (g of pollutant per kg of fuel) to estimate emissions from sectors including energy, industry, transport, residential, agricultural residue burning and distributed diesel (diesel generator sets, agricultural tractors and irrigation pumps). To improve spatiotemporal distributions, SMoG-India uses sector-specific spatial proxies, along with seasonality, to generate monthly emissions at a resolution of 0.25°×0.25°. Other features include emissions of anthropogenic mineral matter (or coal fly-ash), a key primary pollutant missing from global inventories, speciated emissions of NMVOCs and relevant sectoral grouping for the input of emissions at both surface and higher model levels. Additionally, emissions of forest fires were taken from the global GFED-4s database[53], while NH_3 emissions were from the MIX emission inventory[54]. The South Asia nested version of GEOS-chem was used to resolve the South Asia domain at a resolution of 0.5°×0.67° (approximately 56 km×74 km at the Equator) with dynamic boundary conditions using meteorological fields from the NASA Goddard Earth Observation System (GEOS-5). Simulations used a three-month initialization period with the year 2012 meteorology to best represent the mean $PM_{2.5}$ concentrations. The results were consistent with the findings of Phillip et al.[55]. Sectoral contributions to $PM_{2.5}$ were estimated from the difference between simulated $PM_{2.5}$ with all emissions and $PM_{2.5}$ simulated with all emissions except the particular sector. The outputs at relative humidity of 50% were considered.

Confounders and effect modifiers. We obtained all the information about covariates from the NFHS-4 questionnaires administered to WRA. We used daily iron intake derived from NSSO-68[6]. Exposure to indoor air pollution from biomass burning has been linked to anaemia in children,[56,57] so we adjusted our model for the type of cooking fuel in the analysis. Biogas, liquefied petroleum gas and electricity were considered as clean fuel, while kerosene or biomass fuel (wood, straw/shrubs/grass, agricultural crops, animal dung, coal/lignite, charcoal, other) were considered as polluting fuel. The education level of WRA was covered in the analysis and categorized information about levels of education as primary, secondary, higher and no education. BMI in the NFHS-4 questionnaire was categorized as underweight (BMI <18.5), normal (BMI 18.5–24.9), overweight (BMI 25.0–29.9), obese (BMI ≥30.0). SES was quantified using the individual wealth index, with levels of 'poorest', 'poorer', 'middle', 'richer' and 'richest.' Smoking status was also captured for WRA, where a person smoking bidi, cigarettes or tobacco was categorized as a smoker. Other covariates used as dichotomous variables in the analysis were second-hand smoke exposure (yes/no) and type of residence (urban/rural). The effect modification by relevant covariates on anaemia as well as Hb level for every 10 μg m^{-3} increase in ambient $PM_{2.5}$ exposure was further examined by stratified analysis. Additional risk factors such as infection[58] can act as effect modifiers, but the required information was not available in the dataset to be included in the analysis.

Statistical modelling. We used R v.4.1.0 to conduct all statistical analyses and created the figures using R and ArcGIS v.10.3. We first estimated the impact of long-term (ten-year) ambient $PM_{2.5}$ exposure modelled as a continuous variable per district on average anaemia prevalence and average Hb levels using the standard multiple linear regression model. We then performed individual-level analysis using the generalized linear mixed-effects models accounting for clustering by PSUs with Gaussian and logit links for Hb and anaemia, respectively. We developed an exposure–response curve to assess the nonlinearity between decade-long ambient $PM_{2.5}$ exposure and the probability of anaemia prevalence and Hb levels using a generalized addition model with penalized cubic splines. Penalized splines are used to determine the potential nonlinearity in the association of ambient $PM_{2.5}$ exposure (μg m^{-3}) with Hb level (g L^{-1}) and anaemia status at an individual level.

The possible effect modifications by wealth index (SES), residence type (urban/rural), education level, cooking fuel, second-hand smoke and daily iron intake were examined by stratified estimates of OR with 95% UI at the levels of respective factors using multiplicative interaction terms. Mutually disjointed UIs across the levels are an indicator of potential effect modification by the factors. To estimate the simultaneous effects of $PM_{2.5}$ components, we replaced $PM_{2.5}$ in the above

models with its components one at a time, along with residual $PM_{2.5}$ of respective components derived by regressing $PM_{2.5}$ on each component. The residual-based approach was used to estimate the effect of $PM_{2.5}$ components while controlling for risk factors by the total amount of $PM_{2.5}$. We regressed each component of interest on the total $PM_{2.5}$ in a linear model and used the residual to estimate the effect of each individual component while holding $PM_{2.5}$ constant.

Health benefit analysis. We examined the expected benefits of the three recent clean-air and social policies: (1) the NCAP focusing on ambient air pollution; (2) the PMUY aiming to supply clean fuel to 80 million poor households that use solid fuel for cooking; and (3) the DDUGJY aiming to electrify rural areas and eradicate kerosene lighting. The NCAP aimed to reduce the annual ambient $PM_{2.5}$ exposure by up to 30% by targeting the reduction of emissions from outdoor sources, focusing on the non-attainment cities.

We first estimated the reduction in ambient $PM_{2.5}$ exposure ($\Delta PM_{2.5}$) at the district level if India meets the NCAP target (that is, complete success of the NCAP) relative to the long-term exposure level. Similarly, the expected reduction in $PM_{2.5}$ exposure was calculated for the complete mitigation of household emissions by successfully implementing the PMUY and DDUGJY. For this, we used the relative contributions of household emissions to ambient $PM_{2.5}$ derived from the model simulation (Supplementary Fig. 4). If the household emissions are mitigated completely, the ambient $PM_{2.5}$ (averaged over 2007–2016) exposure is expected to reduce by around 30% from 62.3 $\mu g\,m^{-3}$ to 43.6 $\mu g\,m.^{-3}$ These estimates are at par with those in the literature, suggesting a very similar proportion of household emissions to ambient $PM_{2.5}$ concentrations in India[25,28].

Next, we derived the attributable fraction by the following formula:

$$AF = \frac{(RR - 1)}{RR} \quad (1)$$

where $RR = \exp(\log(OR) \times \Delta PM_{2.5})$ and OR is the estimated odds ratio for each unit increase of $PM_{2.5}$ exposure. Then we calculated the expected reduction in district level anaemia prevalence (E) due to the reduction in long-term $PM_{2.5}$ exposure as:

$$E = AF \times \text{District level anaemia prevalence} \quad (2)$$

In addition to the GoI's clean-air policies, we also estimated the expected health benefits of meeting the previous (now interim target 4) WHO air quality guideline of 5 $\mu g\,m^{-3}$ following a similar methodology.

Assumptions and uncertainty. Our analysis made several key assumptions. We assumed that the participants did not leave their residence during the exposure period of ten years. Due to the paucity of ground-based $PM_{2.5}$ composition measurements across India, we could only use datasets from a few cities for the validation of MERRA-2 species data. Even in the literature, composition data were available for only 13 cities. We assumed that the validation of MERRA-2 reanalysis and model data with the limited in-situ data was adequate. While doing the NCAP policy analysis, we assumed a uniform 30% reduction throughout the districts of the country to calculate the resultant anaemia prevalence. Since our study utilizes NFHS-4 data for health outcomes and the survey period was 2015–2016, we considered both models (WRF-CMAQ and GEOS-chem) to be representative of the study period. This also allowed us to assess the robustness of the sectoral priority. Also, there was no specific national emission inventory available to carry out the task for 2015–2016, so we used both the models in our study. We matched the sectors to generate similar and comparable sectoral categories in the two model outputs. Also, the 'other' sector in the GEOS-chem model was generated by subtracting all the contributing sources from the total $PM_{2.5}$ concentration. This 'other' category was a bigger fraction that was not separately accounted for, including anthropogenic mineral matter, natural dust, residential lighting (kerosene lamps), informal industry (food and agro-product processing), waste burning and fugitive dust. It contributed to a 30–50% fraction of different grids.

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

The satellite-based $PM_{2.5}$ data that support the findings of this study are available in the Code Ocean database (https://codeocean.com/capsule/5860976/tree). The MERRA-2 reanalysis data were retrieved from https://gmao.gsfc.nasa.gov/reanalysis/MERRA-2/. The model data used in this study are available from the corresponding author on reasonable request. DHS data source: https://www.dhsprogram.com/data/available-datasets.cfm.

Code availability

The code that supports the findings of this study is available in the Code Ocean database (https://codeocean.com/capsule/5860976/tree) and from the corresponding author on reasonable request.

Received: 6 March 2022; Accepted: 11 July 2022;
Published online: 25 August 2022

References

1. Kassebaum, N. J. The global burden of anemia. *Hematol. Oncol. Clin. North Am.* **30**, 247–308 (2016).
2. Kinyoki, D., Osgood-Zimmerman, A. E., Bhattacharjee, N. V., Kassebaum, N. J. & Hay, S. I. Anemia prevalence in women of reproductive age in low- and middle-income countries between 2000 and 2018. *Nat. Med.* **27**, 1761–1782 (2021).
3. *Global Nutrition Targets 2025: Policy Brief Series* (WHO, 2014); https://www.who.int/publications-detail-redirect/WHO-NMH-NHD-14.2
4. Wirth, J. P. et al. Predictors of anemia in women of reproductive age: Biomarkers Reflecting Inflammation and Nutritional Determinants of Anemia (BRINDA) project. *Am. J. Clin. Nutr.* **106**, 416S–427S (2017).
5. Swaminathan, S. et al. The burden of child and maternal malnutrition and trends in its indicators in the states of India: the Global Burden of Disease Study 1990–2017. *Lancet Child Adolesc. Health* **3**, 855–870 (2019).
6. *National Family Health Survey (NFHS-4), 2015–16* (IIPS and ICF, 2017); https://dhsprogram.com/pubs/pdf/FR339/FR339.pdf
7. Kapil, U., Kapil, R. & Gupta, A. National iron plus initiative: current status & future strategy. *Indian J. Med. Res.* **150**, 239–247 (2019).
8. Ghosh, S., Sinha, S., Thomas, T., Sachdev, H. S. & Kurpad, A. V. Revisiting dietary iron requirement and deficiency in Indian women: implications for food iron fortification and supplementation. *J. Nutr.* **149**, 366–371 (2019).
9. Morales-Ancajima, V. C. et al. Increased outdoor $PM_{2.5}$ concentration is associated with moderate/severe anemia in children aged 6–59 months in Lima, Peru. *J. Environ. Public Health* **2019**, e6127845 (2019).
10. Stauder, R., Valent, P. & Theurl, I. Anemia at older age: etiologies, clinical implications, and management. *Blood* **131**, 505–514 (2018).
11. Honda, T., Pun, V. C., Manjourides, J. & Suh, H. Anemia prevalence and hemoglobin levels are associated with long-term exposure to air pollution in an older population. *Environ. Int.* **101**, 125–132 (2017).
12. Elbarbary, M. et al. Ambient air pollution exposure association with anaemia prevalence and haemoglobin levels in Chinese older adults. *Int. J. Environ. Res. Public Health* **17**, 3209 (2020).
13. Cohen, A. J. et al. Estimates and 25-year trends of the global burden of disease attributable to ambient air pollution: an analysis of data from the Global Burden of Diseases Study 2015. *Lancet* **389**, 1907–1918 (2017).
14. Dandona, L. et al. Nations within a nation: variations in epidemiological transition across the states of India, 1990–2016 in the Global Burden of Disease Study. *Lancet* **390**, 2437–2460 (2017).
15. Spears, D. et al. The association of early-life exposure to ambient $PM_{2.5}$ and later-childhood height-for-age in India: an observational study. *Environ. Health* **18**, 62 (2019).
16. Balakrishnan, K. et al. Exposures to fine particulate matter ($PM_{2.5}$) and birthweight in a rural–urban, mother–child cohort in Tamil Nadu, India. *Environ. Res.* **161**, 524–531 (2018).
17. Mehta, U. et al. The association between ambient $PM_{2.5}$ exposure and anemia outcomes among children under five years of age in India. *Environ. Epidemiol.* **5**, e125 (2021).
18. Ostro, B., Feng, W.-Y., Broadwin, R., Green, S. & Lipsett, M. The effects of components of fine particulate air pollution on mortality in California: results from CALFINE. *Environ. Health Perspect.* **115**, 13–19 (2007).
19. Janssen, N. A. H. et al. Black carbon as an additional indicator of the adverse health effects of airborne particles compared with PM_{10} and $PM_{2.5}$. *Environ. Health Perspect.* **119**, 1691–1699 (2011).
20. Lippmann, M., Chen, L.-C., Gordon, T., Ito, K. & Thurston, G. D. National Particle Component Toxicity (NPACT) Initiative: integrated epidemiologic and toxicologic studies of the health effects of particulate matter components. *Res. Rep. Health Eff. Inst.* **177**, 5–13 (2013).
21. Swaminathan, S. et al. Dietary iron intake and anemia are weakly associated, limiting effective iron fortification strategies in India. *J. Nutr.* **149**, 831–839 (2019).
22. Dey, S. et al. A satellite-based high-resolution (1-km) ambient $PM_{2.5}$ database for India over two decades (2000–2019): applications for air quality management. *Remote Sens.* **12**, 3872 (2020).
23. Bali, K., Dey, S. & Ganguly, D. Diurnal patterns in ambient $PM_{2.5}$ exposure over India using MERRA-2 reanalysis data. *Atmos. Environ.* **248**, 118180 (2021).
24. Villalobos, A. M. et al. Source apportionment of carbonaceous fine particulate matter ($PM_{2.5}$) in two contrasting cities across the Indo–Gangetic Plain. *Atmos. Pollut. Res.* **6**, 398–405 (2015).
25. GBD MAPS Working Group. *Burden of Disease Attributable to Major Air Pollution Sources in India* (HEI, 2018).
26. Ganguly, T., Selvaraj, K. L. & Guttikunda, S. K. National Clean Air Programme (NCAP) for Indian cities: review and outlook of clean air action plans. *Atmos. Environ.: X* **8**, 100096 (2020).

27. Ranjan, R. & Singh, S. Household cooking fuel patterns in rural India: pre- and post-Pradhan Mantri Ujjwala Yojana. *Indian J. Hum. Dev.* **14**, 518–526 (2020).
28. Chowdhury, S. et al. Indian annual ambient air quality standard is achievable by completely mitigating emissions from household sources. *Proc. Natl Acad. Sci. USA* **116**, 10711–10716 (2019).
29. Hajat, A., MacLehose, R. F., Rosofsky, A., Walker, K. D. & Clougherty, J. E. Confounding by socioeconomic status in epidemiological studies of air pollution and health: challenges and opportunities. *Environ. Health Perspect.* **129**, 065001 (2021).
30. Krewski, D. et al. Extended follow-up and spatial analysis of the American Cancer Society study linking particulate air pollution and mortality. *Res. Rep. Health Eff. Inst.* **140**, 5–114 (2009).
31. Clougherty, J. E. et al. Synergistic effects of traffic-related air pollution and exposure to violence on urban asthma etiology. *Environ. Health Perspect.* **115**, 1140–1146 (2007).
32. Fuller, C. H., Feeser, K. R., Sarnat, J. A. & O'Neill, M. S. Air pollution, cardiovascular endpoints and susceptibility by stress and material resources: a systematic review of the evidence. *Environ. Health* **16**, 58 (2017).
33. Hajat, A., Hsia, C. & O'Neill, M. S. Socioeconomic disparities and air pollution exposure: a global review. *Curr. Environ. Health Rep.* **2**, 440–450 (2015).
34. Sun, X. et al. The associations between birth weight and exposure to fine particulate matter ($PM_{2.5}$) and its chemical constituents during pregnancy: a meta-analysis. *Environ. Pollut.* **211**, 38–47 (2016).
35. Reiss, R. et al. Evidence of health impacts of sulfate-and nitrate-containing particles in ambient air. *Inhal. Toxicol.* **19**, 419–449 (2007).
36. Garçon, G. et al. Dunkerque City air pollution particulate matter-induced cytotoxicity, oxidative stress and inflammation in human epithelial lung cells (L132) in culture. *Toxicol. In Vitro* **20**, 519–528 (2006).
37. Das, A., Habib, G., Vivekanandan, P. & Kumar, A. Reactive oxygen species production and inflammatory effects of ambient $PM_{2.5}$ -associated metals on human lung epithelial A549 cells 'one year-long study': the Delhi chapter. *Chemosphere* **262**, 128305 (2021).
38. Viehmann, A. et al. Long-term residential exposure to urban air pollution, and repeated measures of systemic blood markers of inflammation and coagulation. *Occup. Environ. Med* **72**, 656–663 (2015).
39. Trayhurn, P. & Wood, I. S. Signalling role of adipose tissue: adipokines and inflammation in obesity. *Biochem. Soc. Trans.* **33**, 1078–1081 (2005).
40. Herter-Aeberli, I., Thankachan, P., Bose, B. & Kurpad, A. V. Increased risk of iron deficiency and reduced iron absorption but no difference in zinc, vitamin A or B-vitamin status in obese women in India. *Eur. J. Nutr.* **55**, 2411–2421 (2016).
41. Dasi, T. et al. Point-of-care haemoglobin measurement in pooled capillary blood by a portable autoanalyser: comparison with venous blood haemoglobin measured by reference methods in cross-sectional and longitudinal studies. B. *J. Nutr.* **10**, 1 (2021).
42. Petry, N. et al. The proportion of anemia associated with iron deficiency in low, medium, and high human development index countries: a systematic analysis of national surveys. *Nutrients* **8**, 693 (2016).
43. *National Nutrition Monitoring Bureau (NNMB) Diet and Nutritional Survey of Rural Population, Prevalence of Hypertension and Diabetes among Adults and Infant and Young Child Feeding Practices. Report of the Third Repeat Survey* (NNMB, 2012).
44. Sanchis-Gomar, F., Cortell-Ballester, J., Pareja-Galeano, H., Banfi, G. & Lippi, G. Hemoglobin point-of-care testing: the HemoCue system. *J. Lab. Autom.* **18**, 198–205 (2013).
45. Buchard, V. et al. The MERRA-2 aerosol reanalysis, 1980 onward. Part II: evaluation and case studies. *J. Clim.* **30**, 6851–6872 (2017).
46. Jain, S., Sharma, S. K., Vijayan, N. & Mandal, T. K. Seasonal characteristics of aerosols ($PM_{2.5}$ and PM_{10}) and their source apportionment using PMF: a four year study over Delhi, India. *Environ. Pollut.* **262**, 114337 (2020).
47. Skamarock, W. C. et al. *A Description of the Advanced Research WRF Version 3* (No. NCAR/TN-475+STR) (University Corporation for Atmospheric Research, 2008).
48. Appel, K. W. et al. The Community Multiscale Air Quality (CMAQ) model versions 5.3 and 5.3.1: system updates and evaluation. *Geosci. Model Dev.* **14**, 2867–2897 (2021).
49. Singh, N., Agarwal, S., Sharma, S., Chatani, S. & Ramanathan, V. Air pollution over India: causal factors for the high pollution with implications for mitigation. *ACS Earth Space Chem.* **5**, 3297–3312 (2021).
50. Amann, M. et al. Cost-effective control of air quality and greenhouse gases in Europe: modeling and policy applications. *Environ. Model. Softw.* **26**, 1489–1501 (2011).
51. Buchholz, R. R., Emmons, L. K., Tilmes, S. & the CESM2 Development Team. *CESM2.1/CAM-chem Instantaneous Output for Boundary Conditions* (UCAR/NCAR – Atmospheric Chemistry Observations and Modeling Laboratory, 2019).
52. Venkataraman, C. et al. Source influence on emission pathways and ambient $PM_{2.5}$ pollution over India (2015–2050). *Atmos. Chem. Phys.* **18**, 8017–8039 (2018).
53. Giglio, L., Randerson, J. T. & van der Werf, G. R. Analysis of daily, monthly, and annual burned area using the fourth-generation global fire emissions database (GFED4). *J. Geophys. Res.: Biogeosci.* **118**, 317–328 (2013).
54. Li, M. et al. MIX: a mosaic Asian anthropogenic emission inventory under the international collaboration framework of the MICS-Asia and HTAP. *Atmos. Chem. Phys.* **17**, 935–963 (2017).
55. Philip, S. et al. Spatially and seasonally resolved estimate of the ratio of organic mass to organic carbon. *Atmos. Environ.* **87**, 34–40 (2014).
56. Kankaria, A., Nongkynrih, B. & Gupta, S. K. Indoor air pollution in India: implications on health and its control. *Indian J. Community Med* **39**, 203–207 (2014).
57. Kyu, H. H., Georgiades, K. & Boyle, M. H. Biofuel smoke and child anemia in 29 developing countries: a multilevel analysis. *Ann. Epidemiol.* **20**, 811–817 (2010).
58. Viana, M. B. Anemia and infection: a complex relationship. *Rev. Bras. Hematol. Hemoter.* **33**, 90–92 (2011).

Acknowledgements

The work is supported by a research grant under the SUPRA scheme from the SERB, Department of Science and Technology, GoI. The exposure database was generated as part of a project under the NCAP funded by the Ministry of Environment, Forest and Climate Change, GoI. S.D. acknowledges financial support for the Institute Chair fellowship and DST-FIST programme (SR/FST/ESII-016/2014) for computing support. S.J. acknowledges the Institute Postdoctoral fellowship from IIT Delhi. S.W. acknowledges the National Natural Science Foundation of China (grant no. 22188102).

Author contributions

S.D. and E.C. conceived the study and wrote the initial article. E.C. analysed the data with the help of S.G. S.S., N.S., S.A., S.W., C.V. and K.T. did the model simulations and analysis of the model outputs. A.V.K. and A.J.C. provided a critical review of aspects of the epidemiological analysis and aetiological frameworks. S.J. provided the $PM_{2.5}$ composition data for validating the model data. All authors provided comments and contributed to the final version of the article.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41893-022-00944-2.

Correspondence and requests for materials should be addressed to Sagnik Dey or Santu Ghosh.

Peer review information *Nature Sustainability* thanks Trenton Honda, Andrea Pozzer and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

nature portfolio

Corresponding author(s): DBPR NATSUSTAIN-22030524A

Last updated by author(s): Jun 10, 2022

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	The satellite based PM2.5 data that support the findings of this study is available Code Ocean database (https://codeocean.com/capsule/5860976/tree). The MERRA2 reanalysis data is retrieved from https://gmao.gsfc.nasa.gov/reanalysis/MERRA-2/. NFHS-4 data can be accessed from: https://www.dhsprogram.com/data/available-datasets.cfm. It requires to submit a short proposal for an approval to access the data from DHS portal.
Data analysis	The code that support the data analysis of this study is available in the Code Ocean database (https://codeocean.com/capsule/5860976/tree); oftwares used: R 4.1.0, OriginLab Pro 2020b, ArcGIS 10.7

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

The satellite based PM2.5 data that support the findings of this study is available Code Ocean database (https://codeocean.com/capsule/5860976/tree). The MERRA2 reanalysis data is retrieved from https://gmao.gsfc.nasa.gov/reanalysis/MERRA-2/. The model data used in study will be made available from the corresponding author upon reasonable request. DHS data source: https://www.dhsprogram.com/data/available-datasets.cfm.

Human research participants

Policy information about studies involving human research participants and Sex and Gender in Research.

Reporting on sex and gender	The study includes information on women of reproductive age (15-49 years). Authors have not collected this data directly. Its a secondary data procured from DHS website through approval application submission.
Population characteristics	Characteristics of the WRA included in the analysis are: daily iron intake, education level, BMI, Smoking status. Household information involved: socio-economic status, second hand smoke, cooking fuel type
Recruitment	Our study involved 699,686 observations; of these, 59,129 observations had missing PM2.5 exposure, anemia status, and covariate data. We finally included 640557 WRA in the individual analysis across 636 districts of India.
Ethics oversight	NFHS is a nationally representative survey, where multiple organizations were involved. International Institute for Population Sciences (IIPS), Mumbai, is the nodal agency for all of the surveys

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☐ Life sciences ☒ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Behavioural & social sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	We examine the impacts of long-term exposure to ambient PM2.5 and its components on anemia prevalence amongst Women of Reproductive Age (WRA 15-49 years) in India through a cross-sectional study. We show that a 10 ug/m3 increase in long-term exposure to ambient PM2.5 is associated with a 7.2% higher anemia prevalence in India, with approximately four times higher risk due to exposure to secondary in-organics and black carbon compared to other species. We further assess the association with sectoral PM2.5 to characterize the relative importance of sectoral interventions to provide a better policy guidance. Using the indigenous exposure-risk functions, we assess the health benefits of meeting recent clean air targets in the context of anemia burden in India.
Research sample	We retrieved NFHS-4 data that gives information on household and individual sociodemographic characteristics and blood biochemistry, including hemoglobin from 699,686 women of Reproductive Age (WRA) residing in 601,509 households across 640 districts covering 29 states and 6 union territories of India. The sample covers WRA population that is chosen on the basis of the highest anemia burden globally, with nearly 53% of the Indian WRA being anemic.
Sampling strategy	The survey dataset provides information on population, health, and nutrition for India. It is a stratified two-stage sample data. Our study utilizes women of reproductive age (15-49 years) population with 699686 observations. Out of these, 59,129 observations had missing PM2.5 exposure, anemia status, and covariate data, therefore the final observation used was 640557.
Data collection	The health data was accessed from the DHS Program portal (https://www.dhsprogram.com/data/available-datasets.cfm). It requires to submit a short proposal for an approval to access the data from DHS portal.
Timing	The NFHS-4 survey was performed during 2015-2016, this survey data has been utilized in the cross-sectional study

Data exclusions	We made some exclusions on the sample size based on missing PM2.5 exposure, anemia status, and covariate data. Similar exclusions have been established in literature (Mehta et al., 2021; https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7939416/)
Non-participation	59,129 observations/participants had missing PM2.5 exposure, anemia status, and covariate data, therefore the final observation used was 640557 out of 699686.
Randomization	Survey sampling was based on random selection

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging