Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Disease-Specific Health Risks Reshape Priorities for Rare-Earth Emissions

Bin Wang

binwang@pku.edu.cn

Peking University https://orcid.org/0000-0002-1164-8430

jianghui YUN

State Key Laboratory of Environmental Chemistry and Ecotoxicology, Research Center for Eco-Environmental Sciences, Chinese Academy of Sciences

Qianyi Wang

Peking University

Shuo Yang

Peking University

Tianxiang Wu

Peking University

Zhenyu Zhang

Peking University https://orcid.org/0000-0002-3327-8637

Guorui Liu

Zhejiang Normal University https://orcid.org/0000-0002-8462-6734

Article

Keywords: Rare-earth elements, industrial emissions, long-term trends, population health

Posted Date: November 5th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11512608/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: There is **NO** Competing Interest.

Title: Disease-Specific Health Risks Reshape Priorities for Rare-Earth Emissions

Author: Jianghui Yun[1,2,3#], Qianyi Wang[4,5,#], Shuo Yang[4,5], Tianxiang Wu[4,5], Zhenyu Zhang[6], Bin Wang[4,5,7,8,*], Guorui Liu[1,3,9,*]

[1] Zhejiang Key Laboratory of Digital Intelligence Monitoring and Restoration of Watershed Environment, College of Geography and Environmental Sciences, Zhejiang Normal University, Jinhua, 321004, China.
[2] State Key Laboratory of Environmental Chemistry and Ecotoxicology, Research Center for Eco-Environmental Sciences, Chinese Academy of Sciences, Beijing, China.
[3] College of Resource and Environment, University of Chinese Academy of Sciences, Beijing, China
[4] Institute of Reproductive and Child Health, Peking University/Key Laboratory of Reproductive Health, National Health Commission of the People's Republic of China, Beijing 100191, China
[5] Department of Epidemiology and Biostatistics, School of Public Health, Peking University, Beijing 100191, China
[6] Department of Global Health, School of Public Health, Peking University, Beijing, China
[7] Laboratory for Earth Surface Processes, College of Urban and Environmental Science, Peking University, Beijing 100871, China
[8] Key Laboratory of Epidemiology of Major Diseases (Peking University), Ministry of Education, Beijing 100191, China
[9] School of Environment, Hangzhou Institute for Advanced Study, UCAS, Hangzhou, China.

[#]These authors have equal contributions to this study
[*]Correspondence should be addressed to Dr. Bin Wang (Email: binwang@pku.edu.cn); Dr. Guorui Liu (Email: liiugr@zjnu.edu.cn)

SUMMARY

Introduction: Rare-earth elements (REEs) are indispensable to renewable energy, electrification, advanced manufacturing, and national security, but the rapid expansion of REE extraction and industrial use may create an emerging environmental-health burden. Current assessments commonly prioritize REEs according to production, environmental concentration, or total emission mass, implicitly treating them as a relatively uniform chemical group. However, individual REEs differ in biological activity, pathway perturbation, and disease relevance, raising the possibility that the elements emitted most abundantly are not necessarily those of greatest public-health importance. China, with several decades of intensive REE development and a broad range of industrial emission sources, provides a uniquely informative setting in which to evaluate this question at population scale.

Rationale: We developed an integrated framework to determine whether disease-specific health evidence reshapes priorities for industrial REE emissions. First, we reconstructed a spatially resolved national inventory of 15 REEs from 13 industrial sectors across China during 1990–2023 at 0.5° × 0.5° resolution, resolving emissions by element, source, year, and location. Second, we quantified biological relevance using complementary Exposure–GO–Disease (EGoD) and Exposure–Protein–Protein–Disease (EPPD) networks and synthesized published human and experimental evidence through a large language mode (LLM)-assisted literature framework. Third, these signals were independently evaluated against multicity health-insurance hospitalization

records. Finally, we developed a disease-specific mixture-prioritization framework that compares conventional mass-based rankings with priorities derived after incorporating biological and population-health evidence, thereby identifying REEs and industrial sources that warrant differentiated control, capture, or recovery.

Results: National REE emissions increased from 11.8 t in 1990 to 123.2 t in 2014, representing an approximately tenfold increase, and subsequently declined by 16.5% by 2023. Emissions were highly heterogeneous across industrial sectors and individual elements, with cement kiln co-processing of solid waste emerging as a major source and Ce contributing the largest elemental share. Spatial reconstruction revealed persistent emission hotspots concentrated primarily across industrialized regions of eastern and central China. Biological-network analyses further demonstrated that REEs were not equivalent in disease relevance: Ce and Gd showed among the highest disease connectivity, linking to cancer, neurological, circulatory, and respiratory outcomes through biological processes including oxidative stress and related signaling pathways. Literature evidence likewise showed substantial heterogeneity across REE–disease pairs rather than a common toxicity profile. Population-level analyses provided an independent line of support. During 2013–2017, each interquartile-range increase in Ce emissions from cement kiln co-processing was associated with a relative risk (RR) of 2.02 for circulatory-disease hospitalization (95% CI, 1.11–3.67). Likewise, each interquartile-range increase in Ce emissions from coal-fired power plants was associated with a RR of 3.12 for respiratory-disease hospitalization (95% CI, 1.29–7.53). Integrating emission magnitude with biological and human-health evidence substantially altered the relative contributions of individual REEs across 13 disease categories, identifying both high-emission/high-health-priority elements and lower-emission elements whose potential importance would be underestimated by emission mass alone.

Conclusion: Our findings show that industrial REEs should not be managed as a toxicologically uniform class and that health relevance can substantially reorder priorities inferred from emission magnitude alone. By integrating a first-of-its-kind long-term national emission inventory with mechanistic evidence, literature synthesis, population validation, and mixture prioritization, this study establishes a transferable framework for disease-specific REE management. Such an approach can support more targeted source control while also identifying opportunities for selective capture and recovery, helping reconcile population-health protection with economic development, critical-mineral security, and more efficient resource use.

Title: Disease-Specific Health Risks Reshape Priorities for Rare-Earth Emissions

Author: Jianghui Yun[1,2,3#], Qianyi Wang[4,5,#], Shuo Yang[4,5], Tianxiang Wu[4,5], Zhenyu Zhang[6], Bin Wang[4,5,7,8,*], Guorui Liu[1,3,9,*]

[1] Zhejiang Key Laboratory of Digital Intelligence Monitoring and Restoration of Watershed Environment, College of Geography and Environmental Sciences, Zhejiang Normal University, Jinhua, 321004, China.
[2] State Key Laboratory of Environmental Chemistry and Ecotoxicology, Research Center for Eco-Environmental Sciences, Chinese Academy of Sciences, Beijing, China.
[3] College of Resource and Environment, University of Chinese Academy of Sciences, Beijing, China
[4] Institute of Reproductive and Child Health, Peking University/Key Laboratory of Reproductive Health, National Health Commission of the People's Republic of China, Beijing 100191, China
[5] Department of Epidemiology and Biostatistics, School of Public Health, Peking University, Beijing 100191, China
[6] Department of Global Health, School of Public Health, Peking University, Beijing, China
[7] Laboratory for Earth Surface Processes, College of Urban and Environmental Science, Peking University, Beijing 100871, China
[8] Key Laboratory of Epidemiology of Major Diseases (Peking University), Ministry of Education, Beijing 100191, China
[9] School of Environment, Hangzhou Institute for Advanced Study, UCAS, Hangzhou, China.

[#]These authors have equal contributions to this study
[*]Correspondence should be addressed to Dr. Bin Wang (Email: binwang@pku.edu.cn); Dr. Guorui Liu (Email: liiugr@zjnu.edu.cn)

Abstract: Rare-earth elements (REEs) support energy and technological transitions, yet the history and health relevance of their industrial release remain poorly resolved. We reconstructed emissions of 15 REEs from 13 industrial sectors across China during 1990–2023 and integrated biological networks, human-study evidence and multicity hospitalization records. Annual REE emissions rose from 11.8 t in 1990 to 123.2 t in 2014, then declined by 16.5% by 2023, with emissions concentrated in cement kiln co-processing and cerium (Ce). Biological network analysis identified Ce and gadolinium (Gd) as the elements with the highest disease connectivity, linking to cancer and neurological diseases through pathways including oxidative stress. During 2013–2017, each interquartile-range increase in Ce emissions from cement kiln co-processing solid waste was associated with a relative risk of 2.02 for circulatory disease hospitalization (95% CI, 1.11–3.67), whereas each interquartile-range increase in Ce emissions from coal-fired power plants was associated with a relative risk of 3.12 for respiratory disease hospitalization (95% CI, 1.29–7.53). This large-scale nationwide evidence strongly supported disease-specific, source-focused control and recovery rather than uniform management based on emissions alone.

Keywords: Rare-earth elements, industrial emissions, long-term trends, population health

One-Sentence Summary: A 34-year national inventory shows that industrial REE emissions rose more than tenfold before declining after 2014 and exhibit source- and disease-specific population-health signals.

Introduction

Rare-earth elements (REEs) have become indispensable to the technologies underpinning electrification, renewable energy, advanced electronics, and national security. Global rare-earth mineral production increased from approximately 124000 metric tons in 2015 to 390000 metric tons in 2023, a 214% rise within a decade(*1*). The REEs market is projected to grow from approximately ~15 billion USD to over ~33 billion USD in early 2030s(*2*). Efforts to secure supply are accelerating the extraction, processing, and recovery of REEs(*2-6*). These activities continuously mobilize REEs from geological reservoirs into mines, processing facilities, industrial products, and wastes, ultimately releasing them into the environment(*7-10*). Importantly, environmental releases of REEs are not confined to dedicated rare-earth mining and processing; conventional industrial activities, including cement production, coal combustion, coking, metal smelting, and waste treatment, can also release REEs from the raw materials and fuels they use(*6, 7*). However, because industrial activities have changed substantially over recent decades, the potential environmental and health impacts of long-term REE emissions remain difficult to evaluate. At the same time, avoiding the transformation of resource-security efforts into new burdens of environmental pollution and human health has become an increasingly important challenge(*11*). Source-resolved measurements in China have shown that industrial particles differ substantially in their concentrations and other biologically active constituents, producing source-specific toxic potencies that can vary by as much as two orders of magnitude(*12, 13*). Therefore, establishing long-term, source- and element-resolved emission histories is essential for identifying when, where, and from which industrial activities potential population exposure may have arisen.

A critical limitation is that REEs are often treated as a chemically coherent group, implicitly assuming that environmental importance scales with total concentration or emission mass. They differ in ionic radius, redox behavior, solubility, protein binding, and interference with calcium homeostasis, producing substantial heterogeneity in biological responses(*14, 15*). Respiratory evidence, for example, indicates distinct effects of La, Ce, Sm, Gd, Dy, and Y on pulmonary toxicity(*16*). Human studies likewise suggest that the elements most strongly associated with one health outcome may not be those implicated in another: La and Ce have been linked to neural-tube defects(*17*), whereas Dy, Tm, Yb, and Y have been associated with hypertension(*18*), and Eu, Gd, Dy, Ho, Er, and Yb with restricted fetal growth(*19*). These observations suggest that using total REE mass alone to characterize environmental pressure may be insufficient to capture the potential health relevance of individual elements. Meanwhile, existing epidemiological evidence remains fragmented across occupational settings, mining communities, individual cities, and relatively small cohorts, leaving a substantial gap between mechanistic evidence and population-scale patterns of human health(*11, 20, 21*).

China provides a uniquely informative setting in which to address this challenge. For several decades, China has played a dominant role in global REE extraction, separation, refining, and manufacturing, while its energy system, industrial structure, and urbanization have also undergone rapid transformation(*4, 22, 23*). At its peak, China supplied the great majority of global rare-earth raw materials and rare-earth oxides, concentrating not only production capacity but also many of the environmental consequences associated with mining and processing(*4*). Tailings, mineral dust, smelting emissions, and industrial use can release REEs into air, water, soils, and residential

environments, creating opportunities for chronic population exposure(*10, 23*). More recent evidence indicates that this problem is no longer confined to mining districts. REE enrichment has been documented in urban air, dust, and soils, while biomonitoring studies have detected REEs in human hair, blood, nails, and placental tissue(*17, 18*). However, these environmental and population measurements have largely been limited to specific locations and time periods, and high-quality population-based studies examining associations between REE exposure and disease risk remain scarce, with the available studies all conducted in China(*17, 18, 20*). China's unusually long history and intensity of REE development therefore provide a real-world record that is difficult to reproduce elsewhere and may offer critical lessons as other countries expand production. Nevertheless, nationwide population-level evidence linking long-term industrial REE emissions to disease patterns remains lacking. Reviews of REEs released from industrial activities have highlighted the limited integration of environmental exposure, mechanistic evidence, and population health, and have called for epidemiological studies to clarify the health implications of long-term REE exposure(*11, 24*). Establishing a long-term, spatially resolved national REE emission inventory differentiated by element and industrial source, and further linking it with mechanistic and population-health evidence, is therefore critical for addressing this gap.

Here, we developed an integrated framework linking industrial sources, biological evidence and population health (Fig. 1). We first reconstructed a national inventory of industrial emissions of 15 REEs from 13 sources for 1990–2023 at 0.5° × 0.5° resolution, allowing the environmental pressure of each element to be distinguished from total REE emissions (Fig. 1A). We then quantified disease-specific biological relevance using REE-protein/biological process–disease networks and complemented these relationships with LLM-assisted synthesis of the published evidence (Fig. 1B). These mechanistic and literature-derived signals were evaluated independently against nationwide health-insurance records by relating city-level residential exposure to disease occurrence (Fig. 1C). Finally, we compared rankings based on emission magnitude with those derived after incorporating disease-specific health evidence (Fig. 1D). The principal aim is to characterize long-term industrial REE release and its associated health signals. The disease-weighted ranking provides a reference for identifying the industrial emissions that require priority control, and when it comes to REE with high strategic value, it helps shift the focus to capture and recovery rather than indiscriminate reduction.

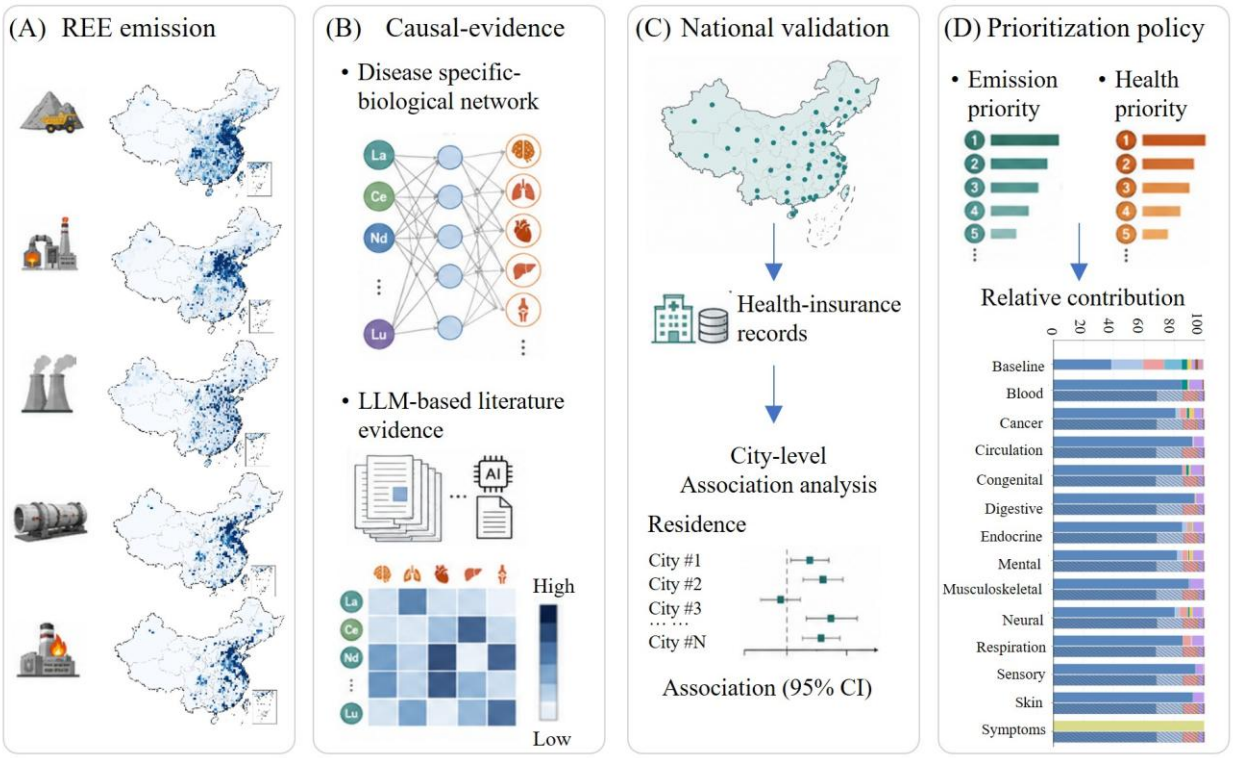


Fig. 1 | Framework for disease-specific prioritization of rare-earth element emissions. (A) A spatially resolved inventory was developed to characterize emissions of individual rare-earth elements (REEs) from major industrial sources across China, e.g., smelting and industrial processing, coal combustion, cement production and waste incineration. The resulting emission maps capture source- and location-specific variation in REE emissions. (B) Disease-specific health relevance was evaluated using two complementary evidence streams. A biological network linked individual REEs to diseases through proteins and biological processes to quantify disease-specific biological relevance, while a large language model (LLM)-assisted literature synthesis summarized existing human and experimental evidence for individual REE–disease associations. (C) Population-level associations were independently evaluated using national health-insurance records. Individual residential locations were linked to city-level REE emissions, and city-level associations between REEs emissions and disease outcomes were estimated with 95% confidence intervals. (D) Emission-based priorities were then compared with health-informed priorities. Relative contributions of individual REEs to different disease categories were used to identify elements whose importance increased or decreased after incorporating disease-specific health evidence, thereby differentiating priorities for environmental control and resource management.

Results

Long term changes and source profiles of industrial REE emissions.

Using an emission factor–based approach, we developed a high-resolution emission inventory of industrial REEs in China for 1990–2023, covering 13 major industrial sources (Extended Data Table 1) and 15 REEs (Table S1). The total REE emissions from all industrial sources from 1990 to 2023 reached 2312 t, with an annual average of 67997 kg. REE emissions were highly concentrated in a small number of industrial sources and elements (Fig. 2A), which is consistent with our previous study(*7*). Cement kiln co-processing solid waste (CK) were the dominant industrial source of total REE emissions, with average annual emissions reaching 58307 kg, accounting for approximately 85.7% of the total. This is mainly attributed to their high production and high emission factors (Tables S2 and S3). Coking plants (COP), coal-fired power plants (CFPP), and blast-furnace ironmaking (BFI) ranked next, with average annual emissions ranging from 672 kg to7700 kg (Tables S4). The above four industrial sources contribute more than 99% of total REE emissions. Other industrial sources, including electric arc furnace steelmaking (EAF), hazardous waste incineration (HWI), iron ore sintering (IOS), municipal solid waste incineration (MSWI), primary copper smelting (PCu), and secondary metal (Al, Cu, Pb, Zn) smelting (SAl, SCu, SPb, SZn), each emitted less than 500 kg of average total REEs annually. At the elemental level, Ce was the most highly emitted REE in every year, with annual emissions ranging from 4560 to 47568 kg and an average of 26234 kg (Tables S5). Ce, La, Nd, and Y together contributed more than 85% of the total average annual REE emissions, which is mainly caused by their high concentrations (Tables S6).

From a temporal perspective, emissions of nearly all REEs exhibited three distinct phases over the entire study period (Fig. 2B): a period of slow growth from 1990 to 2000, a rapid growth stage from 2000 to 2014, and an overall decline from 2014 to 2023. The total annual emissions of REEs increased from 11848 kg in 1990 to a peak of 123237 kg in 2014, and then entered a fluctuating decline stage and dropped to 102893 kg in 2023, which is approximately 16.5% lower than the peak in 2014. The temporal trend of REE emissions was mainly driven by CK (Fig. 2B and Extended Data Fig. 1). For Ce, CK contributed 87.9% of the emission growth during 2000–2014. During 2014–2023, the decline in Ce emissions from CK was equivalent to 106% of the net reduction in total Ce emissions, indicating that the increasing emissions from other industrial sources partially offset the emission reductions brought by CK. It can be clearly seen from the stacked area chart that although CK remained dominant throughout the study period, the source contribution structure became slightly dispersed after 2014, with the relative importance of COP, BFI, and CFPP increasing. COP did not continue to decline after 2014 like CK, but instead entered a relatively stable platform followed by a slight rebound, reaching an emission peak of 4795 kg in 2023. On the contrary, the contributions of BFI and CFPP to Ce emissions continued to increase during the period when CK emissions declined substantially.

Spatial differences of industrial REE emissions.

The spatial distribution of the average annual Ce emissions from various industrial sources from 1990 to 2023 ($0.5° \times 0.5°$) showed that emissions were mainly concentrated in central-eastern China and the eastern coastal region, while the degree of spatial clustering differed markedly among sources (Fig. 2C and Extended Data Fig. 2). The area east of 110°E accounted for approximately 72.5%, 78.7%, 88.0% and 79.7% of Ce emissions from CK, COP, CFPP and BFI, respectively. Emissions were particularly prominent in Shandong, Hebei, Anhui, Jiangsu, Zhejiang and Guangdong regions. Compared with other industrial sources, Ce emissions from CK were more widely distributed, showing a pattern of broad distribution across eastern and south-central

China, and combined with localized high-intensity clusters. In addition, a distinct high-emission hotspot was observed in Sichuan in western China. For COP, its emission hotspots were obviously concentrated in North China, especially in Shanxi, where coke production is intensive. Driven by regional electricity demand and the distribution of energy infrastructure, CFPP exhibited the most dispersed and relatively even spatial pattern, forming a multicenter distribution centered in eastern and central China. BFI emissions showed a much stronger degree of spatial concentration and were highly similar to the distributions of IOS and EAF, probably because they are important stages in steel production. Ce emissions from MSWI displayed strong urbanization and coastal agglomeration characteristics. High-emission areas were mainly distributed along the eastern and southeastern coasts, and obvious continuous or semi-continuous hotspots are formed near the Yangtze River Delta, the Pearl River Delta and some central urban agglomerations. This pattern is likely attributable to the close spatial correspondence among municipal solid waste generation, incineration facilities deployment, and urban population size. By contrast, HWI emissions were more broadly distributed, with clear hotspots in both eastern industrial zones and some industrial clusters in central and western China, reflecting a more multicenter pattern strongly associated with industrial activity. The spatial patterns of emissions from the metal-smelting sector were also closely related to the distribution of metal mineral resources.

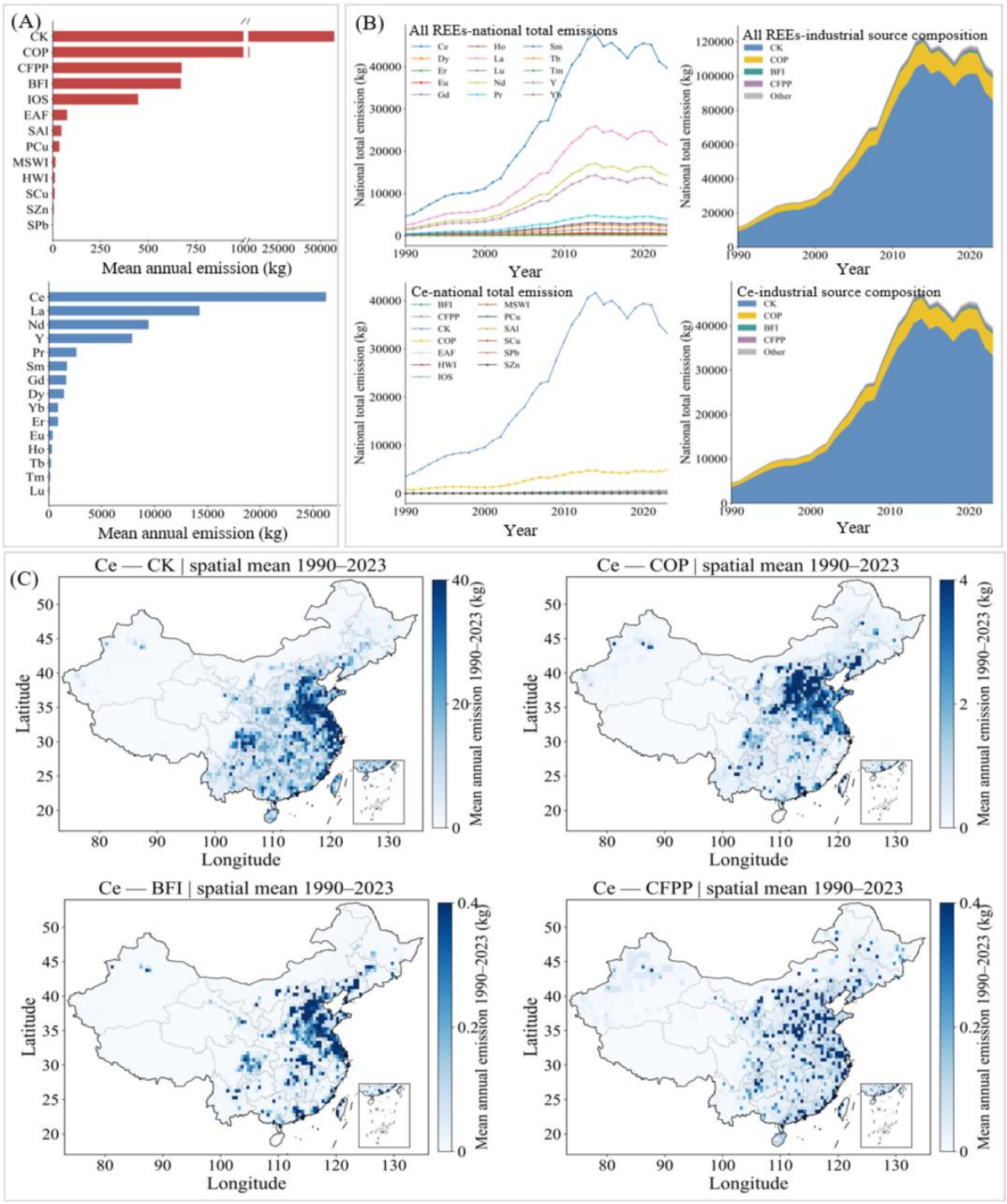


Fig. 2 | Spatiotemporal characteristics of REE emissions in China from 1990 to 2023. (A) Ranking of mean annual emissions by industrial source and by REE. (B) Temporal trends and industrial source composition of REE emissions. The left column shows the national annual total emissions of individual REEs and the annual Ce emissions from different industrial sources, respectively; the right column shows changes in the industrial source composition of total REEs and Ce emissions, respectively. (C) Spatial distributions of mean annual Ce emissions from the four major industrial sources, cement kiln co-processing solid waste (CK), coking plant (COP), blast-furnace pig iron (BFI), and coal-fired power plant (CFPP).

Disease-related biological and literature evidence for REEs

The disease-specific health relevance of individual REEs was evaluated using two complementary evidence streams: biological network analysis and comprehensive literature synthesis assisted by a large language model (LLM). Both analyses revealed pronounced heterogeneity in the health relevance of REEs across elements and disease categories, with evidence concentrated in a limited number of specific REE–disease combinations.

Biological network analysis identified Ce and Gd as the two elements most strongly connected to disease-related biological pathways. In the Exposure–GO–Disease (EGoD) and Exposure–Protein–Protein–Disease (EPPD) networks (Fig. 3A and 3B), Ce accounted for 31.1% and 25.7% of the total weighted path diversity score (PDS), respectively, whereas Gd accounted for 30.4% and 36.9% (Fig. S1; Tables S7 and S8). Cancer represented the dominant disease category in both networks. In the EGoD network, neurological and infectious diseases showed the strongest associations with REEs after cancer, whereas psychiatric, neurological, congenital, and internal diseases were highly connected in the EPPD network. At the level of specific REE–disease pairs, Ce/Gd–cancer, Ce/Gd–neurological disease, and Gd–psychiatric and congenital disease combinations were among the most prominent. The highly connected biological processes linking REEs to disease were mainly related to reactive oxygen species and glutathione metabolism, mitochondrial function, ATP synthesis, apoptosis and cell death, inflammation, and regulation of cell proliferation. Together, these patterns suggest that oxidative stress, mitochondrial dysfunction, and cell-fate regulation form major biological axes linking REE exposure to multiple disease categories.

LLM-assisted literature synthesis provided an empirical complement to the biological network analysis. Existing studies were concentrated primarily on a limited number of elements, particularly Ce and La, and on reproductive, developmental, and musculoskeletal outcomes, whereas many REE–disease combinations still lacked direct investigation (Fig. S2; Table S9). Among 59 quantifiable effect estimates from human studies, 44 showed no statistically significant association, 13 indicated positive risk associations, and 2 suggested protective association (Fig. 3C). Although the overall evidence was dominated by null associations, several REE–disease pairs showed relatively consistent positive signals. It is worth noting that these literature-based signals showed good agreement with the biological network results for several disease outcomes. Ce and La showed clear positive associations with neural tube defects, whereas La and Nd were positively associated with orofacial clefts. Ce also exhibited extensive connections with congenital and developmental diseases in both the EGoD and EPPD networks. Spontaneous miscarriage/miscarriage had a positive risk association with multiple REEs. Gd, Pr, and Yb showed relatively strong health associations with osteoarthritis, with Gd further supported by its high connectivity to musculoskeletal diseases in both biological networks. In contrast, several disease categories that showed strong associations in the biological networks still lacked sufficient direct empirical evidence, most notably cancer, with similar evidence gaps observed for neurological, psychiatric, and endocrine diseases. Conversely, although multiple positive risk associations were reported for reproductive outcomes in the existing literature, the degree of support from the two biological networks was not entirely consistent.

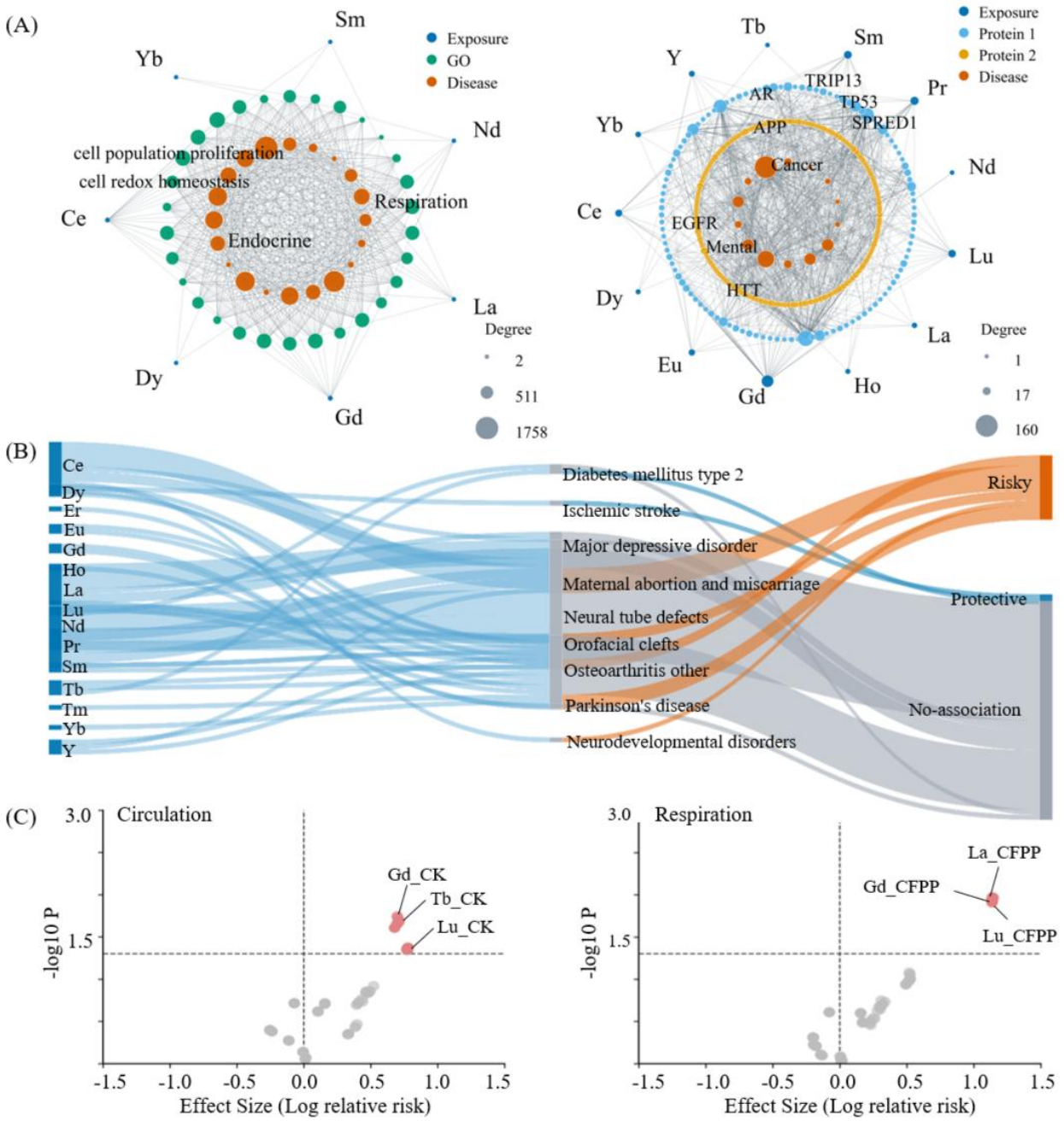


Fig. 3 | Disease-specific health relevance of rare earth elements (REEs)

(A) Exposure–GO–Disease (EGoD) and Exposure–Protein–Protein–Disease (EPPD) networks. Nodes represent REEs, biological processes or proteins, and disease categories. (B) Sankey diagram summarizing reported REE–disease associations and their effect classifications as risky, protective, or no association. Link widths represent the number of literature records corresponding to each REE–disease or disease–effect-type combination. (C) Volcano plots showing associations of industrial REE emissions with circulatory and respiratory hospitalizations in the fully adjusted two-way fixed-effects Poisson model. Effect sizes are expressed as log relative risks (log RRs) per interquartile-range increase in REE emissions. The vertical dashed line indicates the null effect, and the horizontal dashed line indicates the nominal significance threshold. Red points represent nominally significant positive associations, whereas grey points indicate not significance.

Associations between REE emissions and hospitalization outcomes.

Based on city-level REE emissions and medical insurance hospitalization data from 2013–2017 (Table S10), we used a two-way fixed-effects Poisson model to evaluate associations between REE emissions from different industrial sources and hospitalizations for hematologic, circulatory, and respiratory diseases. Exposure effects were expressed as relative risks (RRs) associated with a one-interquartile-range (IQR) increase in REE emissions. The fully adjusted model (M4) controlled for temperature, per capita GDP, the share of secondary industry, the proportion of older adults, and the number of hospital beds per 1000 population, while also including city and year fixed effects (Table S11). In total, the M4 model yielded 495 valid exposure–outcome effect estimates. At the nominal significance threshold of $P < 0.05$, 45 positive associations were identified, whereas no nominally significant negative associations were observed. Significant associations were concentrated in circulatory diseases (30 associations) and respiratory diseases (15 associations), while no associations with hematologic outcomes reached this threshold (Fig. 3D; Table S12).

At the national scale, the positive associations were primarily driven by REE emissions from CFPP and CK. Emissions from both sources were positively associated with hospitalizations for circulatory diseases, with RRs ranging from 2.16 to 2.18 for CFPP and from 1.97 to 2.04 for CK. For Ce, the corresponding RRs were 2.16 (95% CI, 1.02–4.58) and 2.02 (95% CI, 1.11–3.67), respectively. All 15 nominally significant associations for respiratory outcomes were attributable to CFPP, with RRs ranging from 3.11 to 3.15; the RR for Ce was 3.12 (95% CI, 1.29–7.53). Stepwise covariate adjustment showed that the CK–circulatory and CFPP–respiratory associations remained consistently positive and nominally significant from models M1 through M4. In contrast, the CFPP–circulatory association reached nominal significance after adjustment for socioeconomic covariates and remained stable after further adjustment for demographic structure and healthcare resources (Extended Data Fig. 3 and Fig. S3).

Region-stratified analyses further revealed distinct north–south patterns in the associations between REE emissions and hospitalization outcomes (Fig. S4 and S5). In northern China, REE emissions from COP were consistently positively associated with hospitalizations for hematologic, circulatory, and respiratory diseases, with broadly similar effect sizes and median RRs of 2.53, 2.63, and 2.51, respectively. In contrast, the most consistent signal in southern China was an inverse association between REE emissions from IOS and hematologic hospitalizations, with a median RR of approximately 0.366. Further comparisons of effect estimate between northern and southern China indicated that nominal regional differences were mainly concentrated in emissions from COP, IOS, SCu, and PCu, suggesting potential regional heterogeneity in the associations related to these industrial sources (Table S13). Notably, effect estimates for different REEs within the same industrial source were highly similar. Therefore, the current medical insurance analysis more strongly supports population-level associations between source-specific mixtures of REE emissions and hospitalization outcomes, while providing insufficient evidence to distinguish whether any individual REE has a higher or lower independent health effect.

Disease-Specific reshapes priorities for REE Emissions.

To examine how disease-related biological weighting alters emission-control priorities, we integrated disease-specific PDS weights with historical industrial REE emissions and constructed REE integrated emission indices (REEMI) for 20 and 13 disease outcomes under the EGoD and EPPD frameworks, respectively, over 1990–2023 (Table S14). Disease weighting substantially altered the control priorities of individual REEs relative to their original emission levels, whereas the long-term temporal trends in REEMI remained broadly consistent with historical changes in

REE emissions. In 2023, the mass-based REE emission profile was dominated by Ce, La, Nd, and Y, which accounted for approximately 38.6%, 20.9%, 13.9%, and 11.6% of total emissions, respectively. After disease-specific weighting, the relative contribution profiles of individual elements changed markedly, with distinct reordering patterns observed between the EGoD and EPPD pathways (Fig. 4A).

Under the EGoD pathway, the weighted REE contribution profiles were generally stable across diseases. Ce remained the dominant contributor across all 20 outcomes, accounting for 67.5–69.3% of REEMI in 2023. Among the 13 disease categories directly comparable with EPPD, the mean contribution of Ce increased from 38.6% of mass-based emissions to approximately 68.2%, while La and Nd contributed about 17.9% and 8.7%, respectively. Although Gd accounted for only 2.43% of mass-based emissions, its mean contribution under EGoD increased to 3.81%, raising its rank from seventh in the mass-based profile to fourth. When emission magnitude was considered jointly with disease-weighted contribution, Ce, La, and Nd exhibited a "high-emission–high-disease-weighted-contribution" pattern for both circulatory and respiratory diseases (Fig. 4B). In contrast, although Y accounted for 11.6% of mass-based emissions, its disease-weighted contribution declined markedly because corresponding disease links were absent from the current EGoD network.

The EPPD pathway showed greater heterogeneity across diseases. Excluding the Symptoms category, for which Tb was the only element forming a non-zero pathway, Ce contributed an average of 87.3% across the remaining 12 disease categories, further reinforcing its dominant role. Gd also emerged as a major contributor, with a mean contribution of approximately 7.13%, and exhibited a characteristic "low-emission–high-disease-weighted-contribution" pattern, suggesting that its disease-related importance could be underestimated by rankings based solely on emission magnitude. These findings indicate that conventional rankings based solely on emission magnitude may fail to identify some low-emission elements with relatively high disease relevance, and that disease-specific health evidence can provide an additional basis for prioritizing REE emission control. A conceptual management schematic summarizes possible uses of these comparisons for monitoring, emission control and evaluation of capture or recovery (Extended Data Fig. 4).

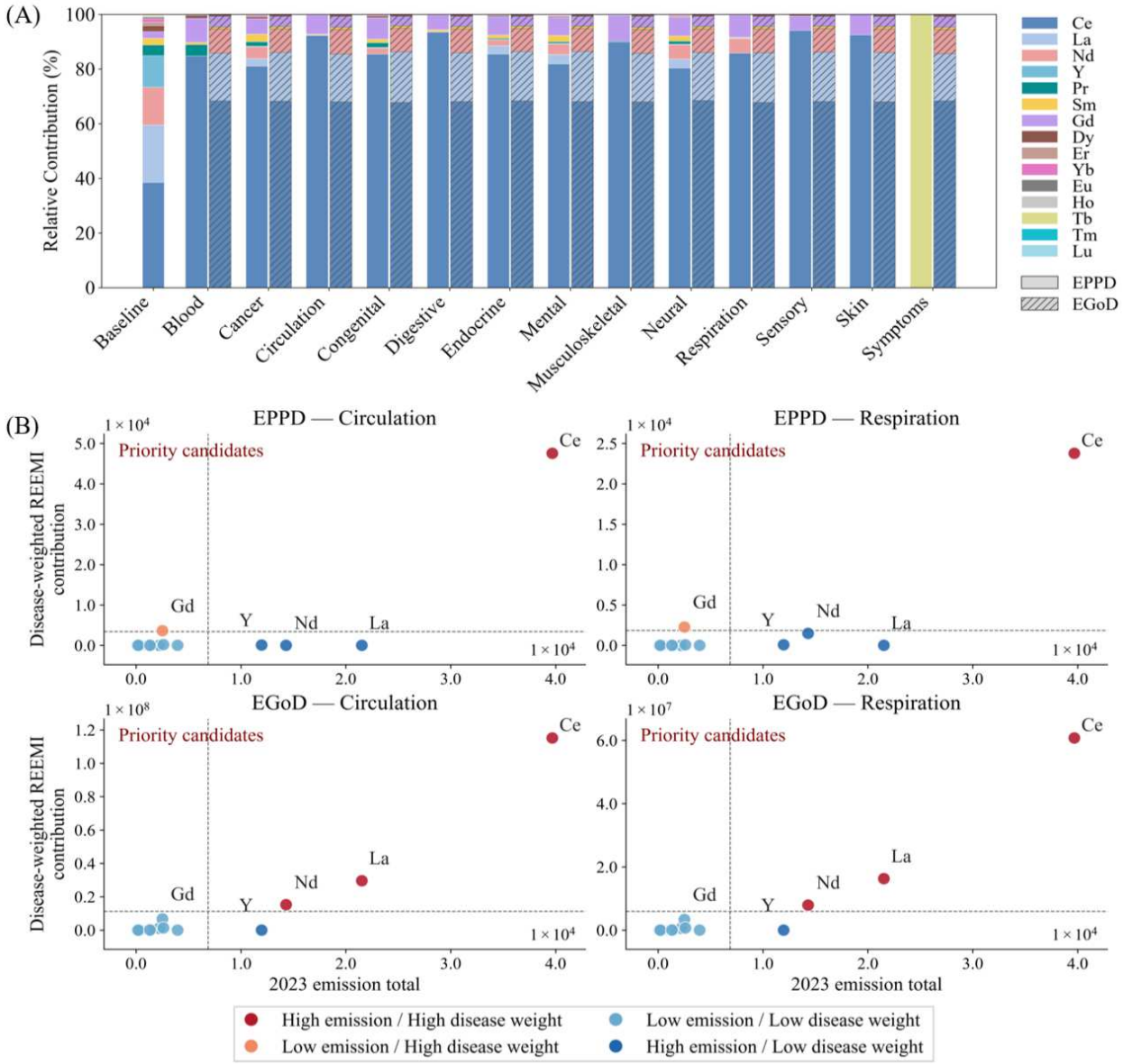


Figure 4 | Disease-specific reweighting reshapes REE emission control priorities.
(A) Relative contributions of individual REEs to the 2023 mass-based emission profile and disease-weighted REEMI across the 13 disease categories shared by the EPPD and EGoD frameworks. Solid and hatched bars denote EPPD and EGoD estimates, respectively; the baseline bar shows the national mass-based REE emission composition. (B) Relationships between national REE emissions and disease-weighted REEMI contributions in 2023 for circulatory and respiratory diseases under the EPPD and EGoD frameworks. Each point represents an individual REE. Dashed lines indicate the mean emission magnitude and mean disease-weighted contribution across the 15 REEs, separating elements into high- and low-emission and high- and low-disease-weighted-contribution groups. Abbreviations: La, lanthanum; Ce, cerium; Pr, praseodymium; Nd, neodymium; Sm, samarium; Eu, europium; Gd, gadolinium; Tb, terbium; Dy, dysprosium; Ho, holmium; Er, erbium; Tm, thulium; Yb, ytterbium; Lu, lutetium; Y, yttrium.

Discussion

This study reconstructed more than three decades of industrial REE emissions and evaluated the associated health evidence. Emissions increased substantially during industrial expansion and declined partially after 2014, but remained well above 1990 levels. This trajectory was dominated by a small number of sources and elements, with CK contributing 85.7% of mean annual emissions and Ce accounting for approximately 38.6%. Evidence from biological networks, human studies, and nationwide hospitalization data consistently indicated that health relevance varied across elements, diseases, and industrial sources. Multicity analyses further identified nominal associations of CK and CFPP emissions with hospitalizations for circulatory and respiratory diseases. Together, these findings provide a long-term nationwide context for evaluating REEs as emerging industrial exposures, rather than viewing them solely as strategically important materials or as a single class of pollutants.

The historical emission inventory extends the earlier static measurements. Previous national measurements across 114 industrial facilities and 13 sectors in China estimated approximately 136,914 kg of particulate-bound REEs released annually, with cement production, coal-fired power generation, and coking accounting for particularly large contributions(*7*). In addition, receptor-side monitoring has revealed REE enrichment and marked spatial variability in atmospheric particles from rare earth mining areas, street dust from industrial cities, and indoor dust from nearby residents(*9, 10, 23*), indicating that REEs released by industrial and mining activities can migrate from production processes into human living environments. Urban particulate matter studies have also reported enrichment of Eu, Tb, and Y in optoelectronic industry clusters along the southeastern coast, suggesting that rare earth applications can generate localized emissions with distinct elemental signatures(*25*). Those measurements established industry as an important source of atmospheric REEs, but the available evidence primarily reflects emissions and environmental conditions in specific regions or during particular time periods. By reconstructing element-specific emissions across multiple industrial sectors over more than three decades and resolving their spatial distributions, our study extends this foundation from source characterization to a historical exposure framework. To our knowledge, this represents the first long-term, spatially resolved, element-specific national inventory capable of linking the evolution of China's industrial REE emissions to population health.

The historical dimension is more than an incremental improvement in emission accounting. REE-related disease outcomes may reflect cumulative or delayed exposures, whereas industrial structure, emission controls, energy systems, and population distributions have changed profoundly during recent decades. A contemporary emission map cannot reconstruct the exposure environment experienced by populations during earlier stages of industrialization. Our inventory therefore creates a temporal bridge between industrial development and health records, while also revealing that REE emissions are not a single national trajectory. Different elements inherit different geological, technological, and sectoral fingerprints. Previous industrial measurements showed one- to two-order-of-magnitude variation in REE concentrations among industrial sources, together with pronounced differences in elemental composition(*7*). Such heterogeneity means that an inventory of total REEs alone can obscure the environmental history of individual elements. Constructing a long-term inventory resolved by both element and industrial source is important not only for improving spatiotemporal resolution, but also for preserving the correspondence among source, element, year, and location. For example, Ce emissions from CK were widely distributed, whereas hotspots from COP were more concentrated in northern China. If these emissions were aggregated into a single national total for REEs, such source-specific spatial

patterns and temporal dynamics would be substantially obscured. Resolving both source and elemental composition is thus necessary before health relevance can be meaningfully assessed.

Our findings also challenge the implicit assumption that REEs can be evaluated as a toxicologically homogeneous class. Their shared trivalent chemistry and coherent geochemical behavior have encouraged the use of total REE concentrations, light-versus-heavy REE categories, or common reference doses in environmental risk assessment. Yet biological responses are markedly more heterogeneous. Experimental studies have shown that REE toxicity rankings depend on element, chemical form, cell type, dose, and endpoint. The most recent high-throughput comparison of 16 rare-earth ions across eight human cell lines found substantial variation among elements and cell systems, with cellular energy metabolism emerging as a sensitive shared endpoint but with no universal toxicity hierarchy across all experimental settings(*21*). Similarly, developmental experiments have shown that La and Pr can perturb neural and cardiovascular development, whereas other REEs exhibit distinct response profiles(*26*). These observations make a single scalar ranking of "REE toxicity" biologically necessary. Likewise, human studies increasingly point in the same direction. In a pilot epidemiological study of hypertension, Dy, Tm, Yb, and Y retained positive associations with hypertension after multivariable adjustment, whereas several more abundant light REEs did not show comparably robust adjusted associations(*18*). For fetal NTDs, maternal serum La and Ce showed much stronger signals, and Bayesian kernel machine regression indicated increasing risk with the overall REE mixture while retaining an independent contribution from La(*17*). These studies differ in design, exposure matrix, population, and susceptibility window and therefore should not be interpreted as establishing a definitive toxicity ranking.

Our biological-network results provide a mechanistic framework for interpreting this heterogeneity. Rather than assigning each element a generic toxicity coefficient, the exposure–biology–disease framework asks how strongly an individual REE connects to a particular disease through proteins and biological processes. This distinction is essential. A compound can have modest general cytotoxicity yet perturb a pathway highly relevant to one disease, whereas an element with stronger nonspecific cellular toxicity may have weaker evidence for another endpoint. The resulting disease-specific biological relevance therefore complements rather than replaces classical toxicology, while helps connect highly fragmented mechanistic evidence into a structure suitable for population-scale evaluation(*27*). For respiratory outcomes, for example, experimental and clinical literature indicates persistent pulmonary deposition, macrophage uptake, oxidative stress, inflammatory dysregulation, calcium-homeostasis disturbance, and fibrotic remodeling after REE exposure(*16*). The convergence of these pathways with the population signals observed for respiratory disease strengthens biological plausibility without requiring the assumption that all REEs act through an identical mechanism.

The nationwide health-insurance analysis adds a second and independent layer of evidence. Previous epidemiological studies have generally measured internal exposure in hundreds or thousands of participants, providing valuable individual-level information but limited geographic and disease coverage. By linking spatially and temporally resolved industrial emissions to nationwide disease records, our analysis evaluates whether the biological signals identified from prior evidence are reproduced at population scale. The concentration of positive signals in specific disease systems, particularly circulatory and respiratory outcomes, is consistent with several strands of earlier evidence. However, evidence regarding cardiovascular effects remains limited, and findings are not fully consistent across experimental models. Therefore, the cardiovascular signals observed in this study should be regarded as hypotheses requiring further validation rather

than as established causal relationships. Such convergence across molecular pathways, earlier epidemiology, and national health records is more informative than any single evidence stream. A major conceptual consequence of these findings is that emission magnitude and health priority are not interchangeable. Conventional pollutant management often begins with the largest mass flows because they are measurable, intuitive, and directly connected to engineering controls. For REEs, this strategy may systematically favor abundant elements while overlooking less-emitted elements with disproportionate disease relevance. Our comparison between emission-based and health-informed rankings demonstrates how incorporation of biological and human evidence can reorder the apparent importance of individual elements. This reordering, rather than the identification of a universally "most toxic" REE, is the central advance of the study. It reframes the management question from "Which REE is emitted most?" to "Which REE contributes most to the health outcome we seek to prevent?"

Our results demonstrate why a mixture cannot be represented faithfully by the sum of single-element toxicities and why a policy framework based exclusively on element-by-element thresholds is inadequate. We address this problem from a complementary decision perspective by introducing a constrained mixture-prioritization framework. Rather than attempting to estimate an absolute attributable burden for every possible combination, the framework decomposes the observed mixture into the relative contributions of its constituent REEs for each disease. Emission pressure provides the environmental dimension, biological-network evidence provides disease-specific plausibility, and nationwide health data provide population relevance. These dimensions are then used to determine which components of a mixture account for the largest share of a disease-specific priority signal. This constrained prioritization has practical advantages. A sector that releases large amounts of relatively low-priority REEs may require broad emission reduction, whereas a source releasing smaller quantities of a highly disease-relevant element may warrant selective capture. For strategically valuable REEs, capture may also create an opportunity for recovery. This distinction connects environmental-health policy with the rapidly evolving critical-mineral agenda. The recent analysis of US mines emphasized that unrecovered critical elements are simultaneously wasted resources and constituents of mining waste, and that by-product recovery can improve supply security while reducing environmental footprints(*3*). Our findings further extend it that an REE identified as both health-relevant and resource-critical should not necessarily be treated simply as material to eliminate. Preventing its release and returning it to productive use could jointly reduce exposure and improve resource efficiency. Thus, the appropriate policy is not uniformly "reduce", but rather "control, capture, and recover", with their relative importance determined by element, source, and disease.

The framework also has broader relevance beyond China. Global REE production is diversifying as countries seek to reduce supply-chain vulnerability, and historically underdeveloped deposits are likely to face renewed exploration and extraction pressure. Earlier assessments warned that widespread expansion of REE mining could reproduce environmental problems already documented around major Chinese production centers. China's experience therefore functions as an anticipatory evidence base for emerging producers. A health-informed framework could be incorporated before new mining and processing capacity is fully established, allowing countries to identify high-priority elements, monitor disease-relevant exposure pathways, and design capture technologies alongside production infrastructure rather than after contamination becomes established. This preventive use may be more consequential globally than retrospective remediation.

Several uncertainties nevertheless define the boundaries of the present work. Emissions represent environmental pressure rather than internal dose. Atmospheric transport, particle size, chemical speciation, solubility, deposition, dietary uptake, and individual behavior all mediate the pathway from industrial release to human exposure. For inhalation in particular, particulate form can strongly influence pulmonary retention and clearance. The biological-network layer is also influenced by the uneven research history of individual REEs, with Ce and La much better studied than many heavy REEs. LLM-assisted evidence synthesis can improve the scale and consistency of literature extraction but cannot remove publication bias or compensate for weak primary studies. Finally, mixture prioritization identifies the most decision-relevant components under available evidence; it does not substitute for experimental characterization of synergy, antagonism, or dose-dependent interaction. These limitations argue for progressively coupling national emission inventories with atmospheric fate modeling, biomonitoring, high-throughput mixture toxicology, and prospective population cohorts. Despite these uncertainties, the study changes the level at which REE environmental-health management can be formulated. Previous work established that industrial activities release substantial quantities of particulate-bound REEs, that long-term exposure can affect multiple biological systems, and that real-world mixtures display complex interactions. We connect these previously separated domains by reconstructing where and when individual REEs were emitted, identifying disease-specific biological relevance, testing those relationships against nationwide population records, and translating the convergent evidence into a finite mixture-prioritization strategy. The result is a shift from treating REEs as a single emerging contaminant class toward recognizing them as a set of strategically valuable elements with unequal health consequences.

As global competition for critical minerals intensifies, especially for REEs, the central policy challenge will not be to choose between economic development and environmental protection. It will be to make increasingly selective decisions about which resources to develop, which emissions to prevent, which materials to capture, and which exposures to monitor. Our findings suggest that disease-specific health evidence can provide the missing criterion for those decisions. REE governance can therefore move beyond mass-based control toward health-informed prioritization, using China's long-term experience to protect populations while preserving the material foundations of the global energy and technological transition.

Methods

Reconstruction of the industrial REE emission inventory in China.

(1) Data source and Pollutant description

We reconstructed annual emissions of 15 rare earth elements (REEs) from 13 major industrial sources in China from 1990 to 2023. The industrial sources included BFI, CFPP, CK, COP, EAF, HWI, IOS, MSWI, PCu, SAl, SCu, SPb, and SZn (Extended Data Table 1). The 15 REEs included Ce, Dy, Er, Eu, Gd, Ho, La, Lu, Nd, Pr, Sm, Tb, Tm, Y, and Yb (Table S1). Elemental concentrations were based on our previous measurements from 114 large-scale industrial facilities covering 13 industrial sectors in China(*7*). These concentrations were combined with source-specific particulate matter emission factors to derive element-specific emission factors. Provincial emissions were estimated using a bottom-up approach based on industrial activity data, particulate matter emission factors, and elemental concentrations. For element i, industrial source s, province r, and year t, emissions were calculated as:

$$E_{i,j,m,p} = A_{j,m,p} \times EF_{j,m,p} \times C_{i,j,m,p} \times 10^{-6} \quad \text{(S1)}$$

where A represents industrial activity (t/year), EF is the particulate matter emission factor (kg/t), and C is the elemental concentration in particulate matter (μg/g). E is expressed in kg/year. Data on industrial activity levels are shown in Table S2. Particulate matter emission factors and elemental concentrations are provided in Tables S3 and S4, respectively.

(2) Reconstruction of historical industrial activity data

Provincial annual activity data for each industrial source were obtained from the National Bureau of Statistics of China and the China Nonferrous Metals Industry Yearbook and harmonized to cover the period from 1990 to 2023 (Table S2). For the iron and steel, cement, and coking sectors, activity levels were represented by the corresponding product outputs. Activity data for coal-fired power plants were derived from thermal power generation using a coal-to-electricity conversion coefficient. For hazardous waste incineration and municipal solid waste incineration, the corresponding disposal or generation indicators and incineration treatment volumes reported in the original data sources were used. Isolated missing years within the same province and industrial source were filled by linear interpolation between adjacent observed years. Time series affected by changes in units, statistical definitions, or administrative boundaries were harmonized before emission calculations. The purpose of activity-data reconstruction was to recover long-term relative changes in industrial activity rather than to replace available observations with interpolated values.

(3) Machine-learning-based spatial downscaling of provincial emissions

To generate continuous gridded emissions suitable for spatial exposure and health analyses, provincial emissions were spatially downscaled to 0.5° × 0.5° grid cells using a machine-learning framework consistent with that used in related industrial-pollutant studies. Sector-matched high-resolution emission fields from the CMIP6 global emission inventory were used as spatial proxy variables, and eXtreme Gradient Boosting (XGBoost) was used to learn the spatial relationship between provincial emissions and gridded proxy features. To account for latitude-dependent differences in grid-cell area, the actual area of each grid cell was calculated using spherical geometry before model fitting. Provincial total emissions and proxy variables were converted to mass densities and standardized using Z scores. The baseline XGBoost parameters were n_estimators = 300, max_depth = 5, and learning_rate = 0.05. Predicted grid-level densities were multiplied by grid-cell area to obtain initial allocated emissions. For each province, year, industrial source, and element, the initial gridded estimates were subsequently calibrated using the ratio of the provincial inventory total to the sum of the gridded predictions:

$$E_{i,s,g,t} = E_{i,s,g,t} \times \frac{E_{i,s,r,t}}{\sum_{h \in G_r} E_{i,s,h,t}} \quad \text{(S2)}$$

where E denotes the initial gridded emission estimate and G_r represents the set of grid cells assigned to province r. This mass-conservation constraint preserves provincial emission totals while using spatial proxy information to resolve within-province variability. Annual gridded emissions were subsequently aggregated by industrial source and element for temporal-trend and spatial-distribution analyses (Tables S5 and S6).

Disease-specific biological networks and Path Diversity Scores.

(1) E-BIO exposure–biology–disease networks

Disease-specific biological associations were constructed using the E-BIO module of the ExposomeX platform(*28*). The 15 REEs were treated as exposure nodes, and two complementary network structures were considered: Exposure–GO–Disease (EGoD) and Exposure–Protein–Protein–Disease (EPPD). The EGoD network links REEs to diseases through Gene Ontology (GO) biological processes and therefore emphasizes mechanistic connections at the functional-process level. In contrast, the EPPD network establishes multistep molecular connections through

exposure-associated proteins, protein–protein interactions (PPI), and disease-associated proteins, thereby emphasizing mechanisms mediated through protein-interaction networks.

(2) EGoD Path Diversity Score

For each element i and disease d, all path records following the structure Exposure → GO → Disease (E→GO→D) were enumerated and denoted as the set $P_i d$. The simple Path Diversity Score (PDS Simple) was defined as the number of path records for a given element–disease pair. The weighted Path Diversity Score (PDS Weighted) was defined as the sum of the degrees of the intermediate GO nodes across all retained paths:

$$PDS_{EGoD,i,d} = \sum_{\pi \in P_i d} Degree[GO(\pi)] \quad \text{(S3)}$$

When no matching degree value was available for an intermediate node, a value of 1 was assigned. After harmonization of disease names, multiple edge records could share identical endpoints; these records were retained and could therefore contribute repeatedly to the EGoD score. Accordingly, PDS Weighted reflects the connectivity-weighted contribution of retained path records and should not be interpreted directly as the number of mutually independent biological mechanisms. Scores for individual REE–disease pairs are provided in Table S7, and element- and disease-level summary results are shown in Fig. S1.

(3) EPPD Path Diversity Score

Within the EPPD network, for each element i and disease d, all length-3 PPI paths following the structure Exposure → Protein1 → Protein2 → Disease were enumerated. PDS Simple was defined as the number of PPI path records satisfying this structure for each element–disease pair. To further account for the network connectivity of mediating proteins, Protein1 and Protein2 from all paths were pooled and deduplicated by protein identifier to generate a set of unique mediating proteins for each element–disease pair (recorded as $U_i d$). PDS Weighted was then defined as the sum of the degrees of these unique mediating proteins:

$$PDS_{EPPD,i,d} = \sum_{v \in U_i d} Degree(v) \quad \text{(S4)}$$

PDS Simple and PDS Weighted values for individual REE–disease pairs are provided in Table S8, and the corresponding summary results are shown in Fig. S1.

Large language model-assisted organization of literature evidence.

(1) Literature search, study screening, and structured information extraction

A combinatorial search strategy was developed in PubMed using predefined lists of REEs and diseases. The search covered 17 REEs and 54 disease categories, yielding 918 element–disease search combinations. REE-related search terms included full English element names, common compound names, and terms related to exposure or concentration. Disease-related terms included standardized disease names, commonly used synonyms, abbreviations, and related clinical terminology. Terms within each element or disease block were connected using the Boolean operator OR, whereas the element and disease blocks were combined using AND. Searches primarily targeted title and abstract fields. To reduce ambiguity associated with elemental symbols, isolated chemical symbols were not used as REE search terms. For selected combinations, additional NOT terms were applied to reduce retrieval of animal studies, review articles, and publications focused on materials, diagnostic applications, or other topics unrelated to the target exposure–health relationships.

Retrieved publications were first screened on the basis of titles and abstracts, followed by full-text review to confirm eligibility. Studies were included if they met all of the following criteria: (i) the study population consisted of humans or human biological samples; (ii) exposure to at least one individual REE was explicitly measured, estimated, or characterized; (iii) a disease, clinical outcome, or other clearly defined human health outcome was reported; and (iv) a specific REE–

health outcome relationship could be established. Animal experiments, in vitro studies, studies reporting only environmental concentrations without human health outcomes, studies in which individual REE exposures could not be distinguished, and reviews, commentaries, or other publications without original human health data were excluded from the quantitative human evidence analysis. When a study reported multiple REEs, multiple outcomes, or multiple independent effect estimates, each corresponding REE–outcome association was retained as a separate record.

After study screening, a large language model (LLM) was used to assist with structured information extraction. Extracted fields included PMID, article title, first author, publication year, study region, study design, sample size, REE, exposure medium or biological sample, disease or health outcome, effect-measure type, effect estimate, and 95% confidence interval (CI). The LLM was used solely to assist with structured information extraction and was not involved in study-inclusion decisions or statistical interpretation. After extraction, researchers manually checked each record against the original publication and corrected or removed extraction errors, ambiguous fields, or information that could not be verified from the source article. Different REEs, disease outcomes, or effect estimates within the same study were retained as separate records.

(2) Classification of effect direction and evidence matrix

For quantitative human evidence, only records for which the effect estimate and 95% CI could be reliably interpreted were retained. For ratio-based effect measures with a null value of 1, associations were classified as "Risky" when the lower bound of the 95% CI was greater than 1, "Protective" when the upper bound of the 95% CI was less than 1, and "No association" when the 95% CI crossed or included 1. Records with missing or unreliable confidence intervals were excluded from effect-direction classification. The final quantitative evidence set contained 59 interpretable human effect estimates, including 44 showing no statistically significant association, 13 showing positive risk associations, and 2 showing protective associations. These classifications were used only to characterize the distribution of existing evidence on REE–disease associations and were not used for effect-size pooling or causal inference. Individual effect estimates and their classifications are provided in Table S9.

Associations between REE emissions and hospitalization outcomes.

(1) Hospitalization data and disease classification

We used nationwide health-insurance hospitalization records from 2013 to 2017. Pharmacy records and records with unspecified diagnoses had been removed before analysis. For records labelled "provincial-level", the treating institution's city was identified using existing institution–city matches, geographic information contained in institution names, and a manually curated institution–city crosswalk. These records were then reassigned to the corresponding city. Special city names were harmonized and city administrative codes were added. Data quality was assessed by checking annual record counts, residual provincial-level labels and missing administrative codes. Using the principal discharge diagnosis ICD-10 code and a unique patient identifier, patients were classified into four prespecified disease groups: respiratory diseases (J00–J99), circulation-related and blood diseases (D50–D89 and I00–I99). These classifications were used for subsequent city-level exposure–hospitalization analyses. The city and temporal coverage of the hospitalization data is summarized in Table S10.

The 1990–2023 gridded emission inventory was spatially overlaid with administrative boundaries and aggregated to the city–year level. Exposure metrics included both element-specific emissions from individual industrial sources and element-specific emissions aggregated across industrial sources. To ensure comparability of effect estimates across elements and emission

sources, each emission metric was scaled by its interquartile range (IQR) based on the city–year distribution. Accordingly, the model-derived relative risk (RR) represents the relative change in hospitalization rate associated with a 1-IQR increase in emissions.

(2) Two-way fixed-effects Poisson models and sequential covariate adjustment

Two-way fixed-effects (TWFE) Poisson regression models were used to evaluate associations between REE emissions and hospitalization outcomes while controlling for time-invariant city-specific characteristics and common nationwide temporal trends. The natural logarithm of the annual resident population was included as an offset with its coefficient constrained to 1. The basic model was specified as:

$$log\left(E\left[Y_{c,t}\right]\right) = \beta_0 + \beta_1 X_{caled,ct} + \beta_2 Z_{c,t} + \alpha_c + \tau_t + log(Pop_{c,t}) \quad \text{(S5)}$$

where $Y_{c,t}$ represents the number of hospitalizations for a given disease in city c and year t, X denotes the IQR-scaled REE emission metric, and $Temp$ represents annual mean temperature. α and τ denote city and year fixed effects, respectively. Population was included as an offset to model hospitalization rates. The exposure coefficient β_1 therefore represents the change in the log relative risk of hospitalization associated with a 1-IQR increase in REE emissions. Exposure variables were scaled by their interquartile ranges (IQRs), such that the estimated relative risk (RR) represented the change in hospitalization rate associated with a 1-IQR increase in REE emissions. City and year fixed effects were included in all models, and standard errors were clustered at the city level to account for within-city temporal correlation.

Four sequential models were fitted to assess the robustness of the associations to covariate adjustment. M1 included annual mean temperature; M2 additionally included per capita gross domestic product (GDP) and the proportion of secondary industry; M3 further adjusted for population aging; and M4 additionally included the number of hospital beds per 1,000 population. The fully adjusted M4 model was prespecified as the primary model, whereas M1–M3 were used to evaluate the stability of effect estimates across progressive covariate adjustment (Table S11 and Fig. S3).

(3) Multiple-testing correction

Because REE emission metrics were strongly correlated across elements and industrial sources, the effective number of independent exposure tests (Me) was estimated from the eigenvalues of the Pearson correlation matrix of the candidate exposure variables. Correlations were calculated using pairwise complete observations(*29*). A nominal threshold of $P < 0.05$ was used to identify preliminary associations, whereas the Me-based significance threshold was used to evaluate statistical significance after accounting for exposure multiplicity. Benjamini–Hochberg false discovery rate (BH-FDR) adjusted P values were additionally calculated as a complementary measure of multiplicity control. RR estimates, 95% confidence intervals (CIs), nominal P values, and multiple-testing-adjusted results are provided in Table S12.

(4) Region-stratified analyses and north–south heterogeneity

To investigate potential geographic heterogeneity in the associations between industrial REE emissions and hospitalization outcomes, analyses were further stratified by northern and southern China. Cities were classified into northern and southern China primarily according to the Qinling–Huaihe geographic boundary, with cities in the Qinghai–Tibet Plateau treated as a separate Plateau stratum. The north–south heterogeneity analysis was restricted to cities classified as North or South, whereas observations assigned to the Plateau stratum were not included in this comparison.

For each source-specific REE emission metric, the fully adjusted M4 model was refitted separately in the northern and southern strata, retaining the same covariates, city and year fixed effects, population offset, and city-clustered standard errors as in the national analysis. Exposure effects

were expressed as RRs per 1-IQR increase in emissions. To ensure direct comparability of regional effect estimates, the exposure IQR was calculated from the pooled analytic distribution before regional stratification and the same scaling factor was applied to both regional models. North–south heterogeneity was evaluated by comparing the regional log-relative-risk estimates. For each exposure–outcome pair, the heterogeneity statistic was calculated as:

$$Z_{diff} = \frac{\beta_{North} - \beta_{South}}{\sqrt{SE_{North}^2 + SE_{South}^2}} \quad \text{(S6)}$$

where β_{North} and β_{South} denote the estimated log RRs in northern and southern China, respectively, and SE_{North} and SE_{South} are their corresponding city-clustered standard errors. A two-sided heterogeneity *P* value was calculated as:

$$P_{diff} = 2\{1 - \Phi(12b, f^1)\} \quad \text{(S7)}$$

where Φ denotes the cumulative distribution function of the standard normal distribution. A nominal P_{diff} was used to identify exploratory evidence of regional heterogeneity. To account for multiple north–south comparisons, P_{diff} values were further adjusted using the Benjamini–Hochberg false discovery rate procedure, with FDR-adjusted $P < 0.05$ considered statistically significant. Region-specific estimates and north–south comparisons are presented in Fig. S4 and Table S13.

Disease-specific REE Emission Mixture Index (REEMI).

(1) Element-level emission aggregation and molar normalization

For each grid cell g and year t, emissions of element i were first summed across all industrial sources:

$$E_{i,g,t} = \sum_s E_{i,s,g,t} \quad \text{(S8)}$$

The aggregated emissions were then normalized by molar mass:

$$Q_{i,g,t} = \frac{E_{i,g,t}}{MW_i} \quad \text{(S9)}$$

where the molar mass of each element is provided in Table S1. This normalization reduces incomparability in mass-based emission scales arising solely from differences in atomic mass among REEs.

(2) Disease- and network-specific REEMI

For network P∈[EGoD, EPPD], disease d, grid cell g, and year t, REEMI was calculated as:

$$REEMI_{p,d,g,t} = \sum_i w_{p,i,d} \times Q_{i,g,t} \quad \text{(S10)}$$

where d denotes the PDS_Weighted value for the corresponding element–disease pair. When no element–disease record was available within a given network, its weight was set to 0. The relative contribution of element i to the national REEMI for a given disease was defined as the weighted contribution of that element divided by the summed weighted contributions of all elements for that disease. Annual element-specific contributions and relative shares of national REEMI under the EGoD and EPPD frameworks are provided in Table S14. Because PDS Weighted values derived from EGoD and EPPD are on different scales, comparisons were restricted to relative contributions, rankings, and changes in priority within each network. Absolute REEMI values were not compared directly between the two network frameworks.

Acknowledgments

This research was supported by the National Natural Science Foundation of China (Grant No. 42477455, 22576189, 22506186), National Key Research & Development Program of Ministry of Science and Technology of China (Grant No. 2023YFC3708305), Beijing Municipal Ecology and Environment Bureau (Grant No. BJST20250209), Strategic Priority Research Program of the

Chinese Academy of Sciences (XDB0750400). We would like to express our gratitude to the working group of environmental exposure and human health of the China Cohort Consortium (http://chinacohort.bjmu.edu.cn/).

Author contributions

B. W. and G. L. conceived the study. J. Y., Q. W., and Z. Z. curated data and performed the analysis. J. Y., B. W., and Q. W. wrote the first draft of the manuscript. J. Y., Z. Z., G. L., and B.W. reviewed and edited the manuscript. B. W., Z. Z., and G. L. managed the program.

Competing interests

Authors declare no competing interests.

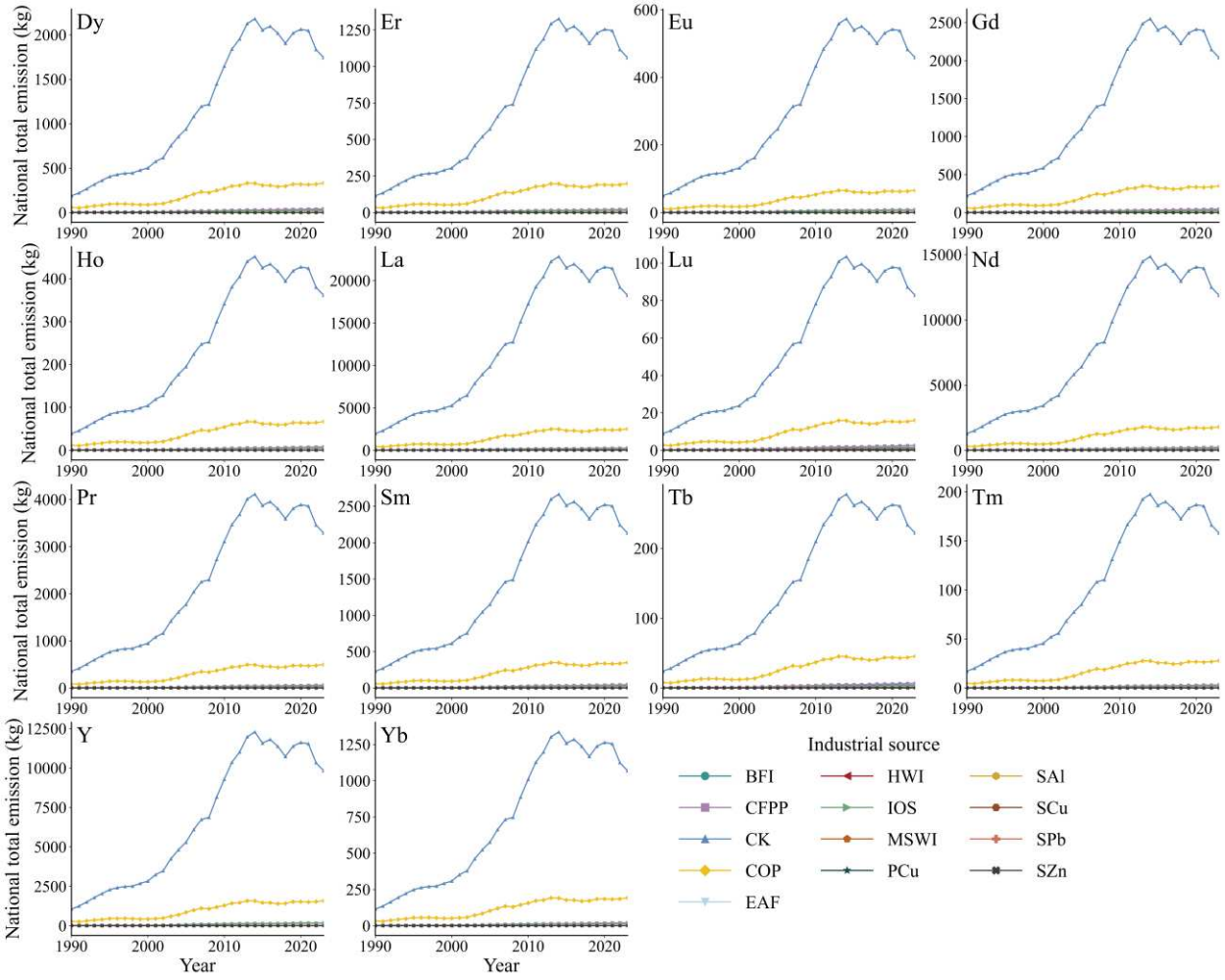
Data and Code materials availability

Readers can contact the corresponding author Dr. Bin Wang (Peking University) for the detailed information, Email: binwang@pku.edu.cn. The data and code used in this study can be freely accessible on Zenodo database (https://doi.org/10.5281/zenodo.23273644).

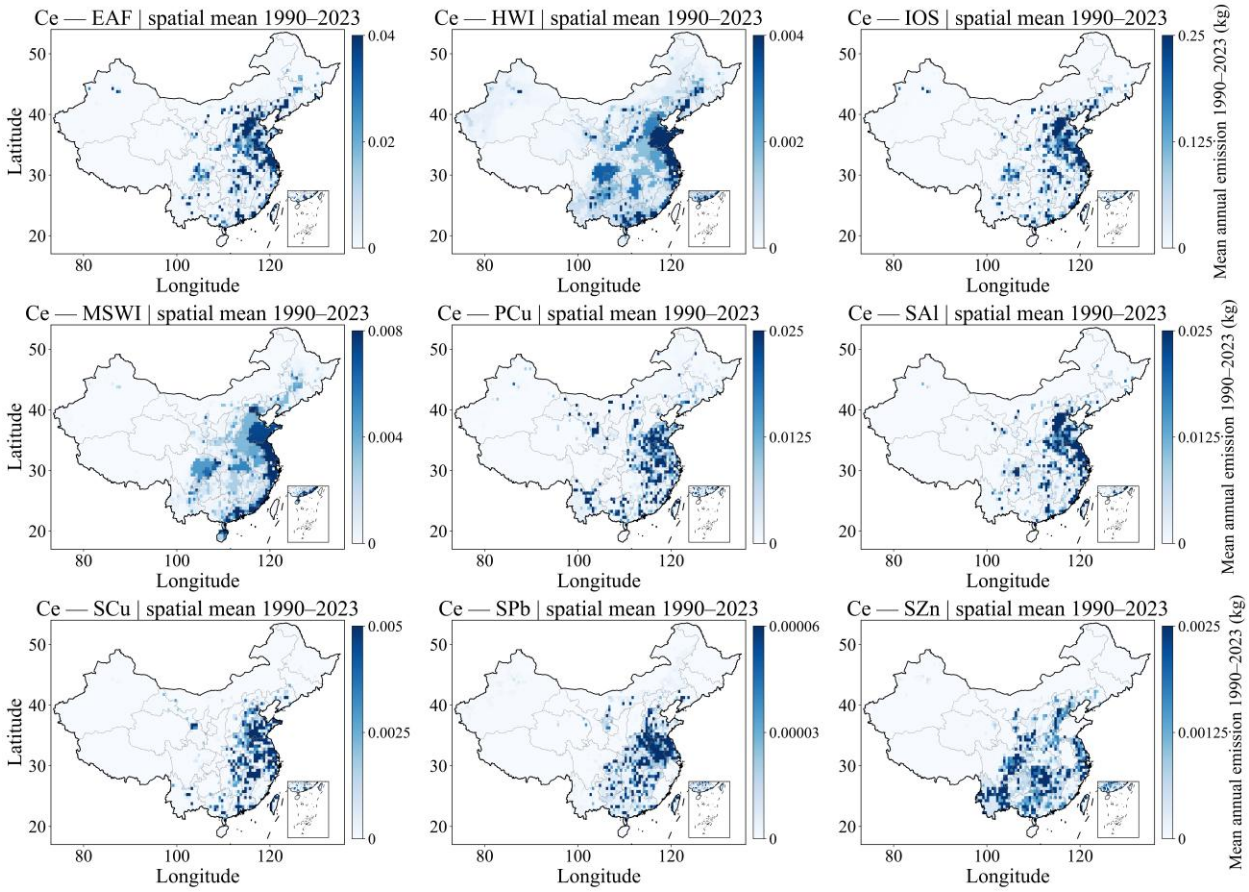
Reference

1. U. S. G. Survey, "Mineral commodity summaries 2024" (2024).
2. K. Joshi, S. Magdouli, S. K. Brar, Bioleaching for the recovery of rare earth elements from industrial waste: A sustainable approach. *Resour Conserv Recy* **215**, (2025).
3. E. A. Holley *et al.*, By-product recovery from US metal mines could reduce import reliance for critical minerals. *Science* **389**, 1325-1331 (2025).
4. X. Huang, G. C. Zhang, A. Pan, F. Y. Chen, C. L. Zheng, Protecting the environment and public health from rare earth mining. *Earths Future* **4**, 532-535 (2016).
5. P. Wang *et al.*, Regional rare-earth element supply and demand balanced with circular economy strategies. *Nat Geosci* **17**, (2024).
6. K. Joshi, S. Magdouli, S. K. Brar, Assessment of the Biogeochemistry of Rare Earth Elements and Distribution of Other Elements in Oil Sands Tailings. *Environmental Science & Technology* **60**, 4840-4853 (2026).
7. J. H. Yun, Q. T. Yang, C. Y. Zhao, C. Z. Chen, G. R. Liu, Atmospheric emissions of fine particle matter bound rare earth elements from industry. *Nature Communications* **15**, (2024).
8. M. T. Niu, Y. Zhang, Y. Wang, M. S. Yao, Differing Ambient Particle Composition and Oxidative Toxicity in 12 Cities in Inner Mongolia: Reference for Regional Optimized Air Pollution Control. *Environmental Science & Technology* **60**, 11353-11365 (2026).
9. K. X. Li, T. Liang, L. Q. Wang, S. H. Tian, Inhalation exposure and potential health risk estimation of lanthanides elements in PM associated with rare earth mining areas: a case of Baotou city, northern China. *Environ Geochem Hlth* **40**, 2795-2805 (2018).
10. G. Y. Sun *et al.*, Rare earth elements in street dust and associated health risk in a municipal industrial base of central China. *Environ Geochem Hlth* **39**, 1469-1486 (2017).
11. J. Ma, F. C. Wu, Urban impact of rare earth elements. *Science* **390**, (2025).
12. H. T. Zheng *et al.*, Control of toxicity of fine particulate matter emissions in China. *Nature* **643**, (2025).
13. D. Wu *et al.*, Achieving health-oriented air pollution control requires integrating unequal toxicities of industrial particles. *Nature Communications* **14**, (2023).
14. S. Hirano, K. T. Suzuki, Exposure, metabolism, and toxicity of rare earths and related compounds. *Environ Health Persp* **104**, 85-95 (1996).

15. A. A. Brouziotis *et al.*, Toxicity of rare earth elements: An overview on human health impact. *Front Env Sci-Switz* **10**, (2022).
16. J. Zhou, P. C. Liang, W. J. Zhu, R. N. Zhang, B. Zhong, Environmental exposure to rare earth elements and respiratory health: an emerging public health concern. *Front Public Health* **14**, (2026).
17. J. Wei *et al.*, Concentrations of rare earth elements in maternal serum during pregnancy and risk for fetal neural tube defects. *Environ Int* **137**, (2020).
18. B. Wang *et al.*, Rare earth elements and hypertension risk among housewives: A pilot study in Shanxi Province, China. *Environ Pollut* **220**, 837-842 (2017).
19. X. J. Fang *et al.*, Associations between maternal urinary rare earth elements during pregnancy and birth weight-for-gestational age: Roles of cord blood vitamin D levels. *Sci Total Environ* **912**, (2024).
20. W. Y. Wang, Y. F. Yang, D. L. Wang, L. H. Huang, Toxic Effects of Rare Earth Elements on Human Health: A Review. *Toxics* **12**, (2024).
21. Y. Q. Wang *et al.*, Health Risk Assessment for Rare Earth Ions: From End Point Identification of Cytotoxicity to Mixed Toxicity Prediction. *Environmental Science & Technology* **60**, 17679-17695 (2026).
22. C. Xu *et al.*, Origin of heavy rare earth mineralization in South China. *Nature Communications* **8**, (2017).
23. X. X. Han *et al.*, Mineral dust deposition drives rare earth element accumulation in indoor dust and associated health risks in the Bayan Obo region. *Environ Pollut* **406**, (2026).
24. J. Yun, Q. Yang, G. Liu, Rare earth elements unintentionally released from global industrial activities. *J Hazard Mater* **480**, 136146 (2024).
25. Y. Yan *et al.*, Characteristics and provenances of rare earth elements in the atmospheric particles of a coastal city with large-scale optoelectronic industries. *Atmospheric Environment* **214**, (2019).
26. Y. B. Zhao *et al.*, Rare Earth Elements Lanthanum and Praseodymium Adversely Affect Neural and Cardiovascular Development in Zebrafish (Danio rerio). *Environmental Science & Technology* **55**, 1155-1166 (2021).
27. Y. Chen, G. Liu, S. Yang, B. Wang, Control Priorities for Persistent Organic Pollutant Mixture Based on Neurodevelopmental Disorders. *Environmental Science & Technology*, (2026).
28. B. Wang *et al.*, ExposomeX: Development of an Integrative Exposomic Platform to Expedite Discovery of the "Exposure-Biology-Disease" Nexus. *Environmental Science & Technology* **59**, 13251-13263 (2025).
29. J. Li, L. Ji, Adjusting multiple testing in multilocus analyses using the eigenvalues of a correlation matrix. *Heredity (Edinb)* **95**, 221-227 (2005).

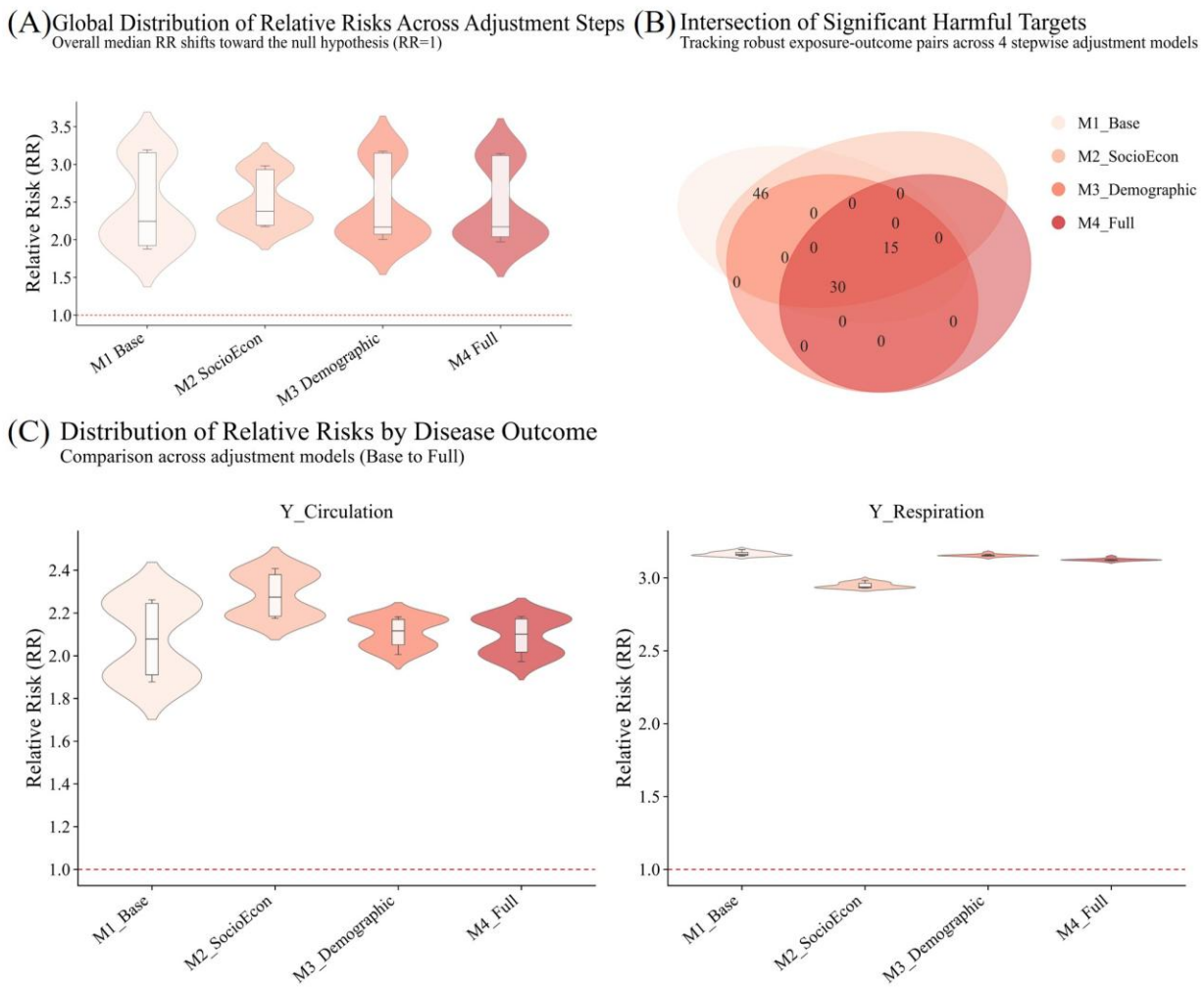


Extended Data Fig. 1 | Annual trends in national emissions of individual REEs, 1990–2023. Annual national emissions of Dy, Er, Eu, Gd, Ho, La, Lu, Nd, Pr, Sm, Tb, Tm, Y, and Yb from 13 industrial source categories during 1990–2023. Emissions were summed across all grid cells for each year and are expressed in kg. Abbreviations: La, lanthanum; Ce, cerium; Pr, praseodymium; Nd, neodymium; Sm, samarium; Eu, europium; Gd, gadolinium; Tb, terbium; Dy, dysprosium; Ho, holmium; Er, erbium; Tm, thulium; Yb, ytterbium; Lu, lutetium; Y, yttrium.



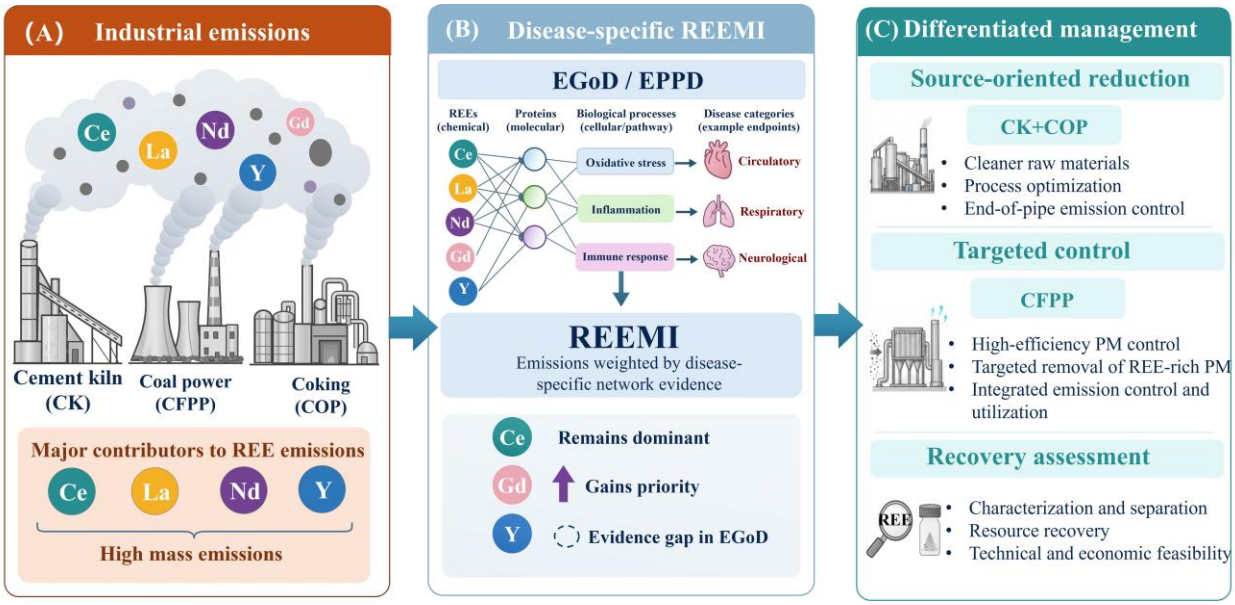
Extended Data Fig. 2 | Spatial distribution of mean annual cerium (Ce) emissions from additional industrial sources, 1990–2023.

Spatial patterns of grid-level mean annual Ce emissions from EAF, HWI, IOS, MSWI, PCu, SAl, SCu, SPb, and SZn during 1990–2023. For each source, annual Ce emissions were averaged across the 34-year study period for each grid cell and are expressed in kg.



Extended Data Fig. 3 | Robustness of REE–hospitalization associations across progressively adjusted models.

(A) Distribution of RRs for all REE emission–hospitalization outcome associations across the four progressively adjusted models (M1–M4); the red dashed line indicates the null value of RR = 1. (B) Overlap of nominally significant positive associations across the four models, illustrating the consistency of significant signals under different covariate adjustment schemes. (C) Changes in the RR distributions for REE emission associations with circulatory and respiratory diseases across models M1–M4. M1 was adjusted for annual mean temperature only; M2 was additionally adjusted for per capita GDP and the proportion of secondary industry; M3 further included population aging; and M4 further included the number of hospital beds per 1000 population. The fully adjusted M4 model was used as the primary model.



Extended Data Fig. 4 | Conceptual framework for disease-specific prioritization of REE emission control and resource recovery.

(A) Industrial sources contribute distinct mixtures of REEs, with Ce, La, Nd, and Y representing major components of mass-based emissions. (B) Disease-specific evidence derived from the EGoD and EPPD pathways is integrated with historical emissions to construct REEMI and reweight element-specific control priorities. Ce remains a prominent priority across both pathways, whereas Gd gains relative priority after disease-specific weighting and Y illustrates an evidence gap in the current EGoD network. (C) The resulting prioritization can inform differentiated management strategies, including source emission reduction, targeted capture and recovery where technically and economically feasible, and continued monitoring and validation of exposure, biological mechanisms, and health associations. The schematic is conceptual and does not represent quantitative thresholds for intervention.

974 **Extended Data Table 1 | Industrial sources and activity indicators**

Abbreviation	Industrial sources	Activity level indicators and data sources
BFI	Blast-furnace pig iron	Iron production; National Bureau of Statistics
CFPP	Coal-fired power plant	Thermal power generation; National Bureau of Statistics
CK	Cement kiln co-processing solid waste	Cement production; National Bureau of Statistics
COP	Coking plant	Coke production; National Bureau of Statistics
EAF	Electric-arc furnace steelmaking	Crude steel production (Electric); National Bureau of Statistics
HWI	Hazardous-waste incineration	Quantity of hazardous waste disposal; National Bureau of Statistics
IOS	Iron-ore sintering	Sintered ore production; National Bureau of Statistics
MSWI	Municipal solid-waste incineration	Volume of waste incinerated for safe disposal; National Bureau of Statistics
PCu	Primary copper smelting	Copper ore production; China Non-ferrous Metals Industry Yearbook
SAl	Secondary aluminium smelting	Aluminium recycling or mixed output; China Non-ferrous Metals Industry Yearbook
SCu	Secondary copper smelting	Copper recycling or mixed output; China Non-ferrous Metals Industry Yearbook
SPb	Secondary lead smelting	Lead recycling or mixed output; China Non-ferrous Metals Industry Yearbook
SZn	Secondary zinc smelting	Zinc recycling or mixed output; China Non-ferrous Metals Industry Yearbook

975
976

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryTableS2.xlsx
- SupplementaryTableS12.xlsx
- SM20261025.docx
- SupplementaryTableS14.xlsx