

CHINA CDC WEEKLY



Vol. 8 No. 36 Sept. 4, 2026

中国疾病预防控制中心周报 (英文)

KASHIN-BECK DISEASE



Preplanned Studies

Putative Adverse Outcome Pathway in Kashin-Beck Disease Suggested by Lipidomics and Network Toxicology — Xi'an City, Shaanxi Province, China, 2023–2025 1117

Network Analysis of Visual Behaviors, Family Factors, and Refractive Status in Chinese Schoolchildren — 10 PLADs, China, 2020–2024 1123

Occurrence and Risk Assessment of Typical Endocrine-Disrupting Chemicals in Surface and Drinking Water Across Four River Basins — Henan Province, China, 2025 1129

Methods and Applications

A Hierarchical Multi-Stage Machine-Learning Study on Predicting Oil and Salt in Chinese Restaurant Dishes for Nutrition Surveillance — 6 Survey Sites, China, 2019–2020 1135



ISSN 2096-7071



Editorial Board

Honorary Editor-in-Chief Hongbing Shen

Founding Editor-in-Chief George F. Gao

Advisory Board Member Jianguo Xu Liming Li Yu Wang Gabriel M Leung Zijian Feng

Editor-in-Chief Jianwei Wang

Deputy Editor-in-Chief

Zhuo Chen (USA) Zhibin Hu Qun Li Zhengliang Li Xiaoming Shi Yan Sun

Changjun Wang Tangchun Wu Yongning Wu Ningshao Xia Chihong Zhao

Editorial Board Member

Jianping Cao Guobing Chen Xi Chen (USA) Gong Cheng Gangqiang Ding

Xiaoping Dong Pei Gao Xin Guo Jun Han Mengjie Han

Weidong Hao Na He Yuping He Guoqing Hu Cunrui Huang

John S. Ji (USA) Na Jia Weihua Jia Zhongwei Jia Biao Kan

Haidong Kan Jianqiang Lai Lance Rodewald (USA) Ni Li Shizhu Li

Ying Li Zhenjun Li Zhongjie Li Geyu Liang Yuan Lin

Aidong Liu Min Liu Qiyong Liu Qingjie Liu Yawen Liu

Jinxing Lu Xiangfeng Lu David LYE Chien Boon (Singapore) Fan Lyu

Jun Lyu Huilai Ma Jiaqi Ma Chen Mao Xiaoping Miao

An Pan Jie Pan Lili Ren Guoqing Shi Yuelong Shu

Chengye Sun Quanfu Sun Xin Sun Hua Wang Huaqing Wang

Hui Wang Jianming Wang Junling Wang Lin Wang Tong Wang

Shenghui Wu (USA) Min Xia Lin Xiao Dongqun Xu Hongyan Yao

Guojing Yang Zundong Yin Dianke Yu Hongjie Yu Siyan Zhan

Jianzhong Zhang Jun Zhang Liubo Zhang Tao Zhang Yanping Zhang

Wei Zhao Yanlin Zhao Maigeng Zhou Xiaonong Zhou Yongqun Zhu

Guihua Zhuang

Editorial Office

Directing Editor Chihong Zhao

Managing Editor Yu Chen

Senior Scientific Editors

Xuejun Ma Daxin Ni Ning Wang Wenwu Yin Shicheng Yu Jianzhong Zhang Qian Zhu

Scientific Editors

Weihong Chen Jie Dong Tao Jiang Dongmin Li Xudong Li Nankun Liu

Liwei Shi Meng Wang Zhihui Wang Qi Yang Qing Yue Lijie Zhang

Ying Zhang Lishu Zheng

Editor in Charge Meng Wang

Preplanned Studies

Putative Adverse Outcome Pathway in Kashin-Beck Disease Suggested by Lipidomics and Network Toxicology — Xi'an City, Shaanxi Province, China, 2023–2025

Liangyu Cui^{1,2}; Xi Wang^{1,2,3}; Chaowei Wang^{2,3}; Yang Liu^{1,2}; Lulu Bai^{1,2}; Lian Liu^{1,2}; Qi Jing^{1,2}; Yuyang Ma^{1,2}; Pengyan Jiao^{2,3}; Zhouyun Chen^{2,3}; Yirong Qin^{2,3}; Hui Wang^{2,3}; Yuequan Yuan^{2,3}; Xi Lyu^{2,3}; Yu Zhang^{1,2}; Xiong Guo^{2,3}; Yujie Ning^{2,3}; Feng Zhang^{2,3,#}

Summary

What is already known about this topic?

Lipids are key mediators in cartilage degeneration. While T-2 toxin exposure drives Kashin-Beck disease (KBD), systematic metabolic comparisons between KBD and osteoarthritis (OA) are lacking.

What is added by this report?

KBD and OA share cartilage lipid alterations linked to ferroptosis, but KBD exhibits unique changes indicating severe oxidative stress. Furthermore, prostaglandin-endoperoxide synthase 2 (PTGS2) was identified as the candidate molecular target potentially involved in T-2 toxin-induced cartilage damage.

What are the implications for public health practice?

Uncovering the adverse outcome pathway of T-2 toxin exposure clarifies KBD pathogenesis. Targeting the PTGS2-mediated signaling pathway may provide a potential strategy for clinical intervention and disease management.

and OA when compared with FNF controls. These included reductions in sphingomyelins, ceramides, and 5-hydroxyeicosatetraenoic acid and increases in lysophospholipids and vitamin D3 sulfate. These lipids are predominantly involved in arachidonic acid and linoleic acid metabolism and ferroptosis. Plasma lipidomics revealed alterations in triglyceride and diglyceride levels. KBD was associated with elevated levels of cartilage ether triglycerides and plasma polyunsaturated fatty acids. Within the proposed AOP network, prostaglandin-endoperoxide synthase 2 (PTGS2) was identified as a potential molecular initiating event (MIE).

Conclusion: This study proposes a hypothetical AOP framework that links exposure to T-2 toxin with cartilage injury in KBD and identifies PTGS2 as a potential target for further biochemical and functional validation.

ABSTRACT

Introduction: This study compared the metabolic characteristics of Kashin-Beck disease (KBD) and osteoarthritis (OA) to establish an adverse outcome pathway (AOP) framework linking T-2 toxin exposure with cartilage damage.

Methods: Knee cartilage and plasma samples were collected from 17 patients with KBD, 22 patients with OA, and 5 controls with femoral neck fracture (FNF). Targeted lipidomics, partial least squares discriminant analysis (PLS-DA), pathway analysis, network toxicology, and molecular docking analyses were performed.

Results: Lipidomic analysis of the cartilage identified 158 differential lipids (DLs) common to both KBD

Kashin-Beck disease (KBD) is an endemic, chronic osteoarthropathy characterized by multifocal hyaline cartilage necrosis and significant disabilities. Despite the availability of preventive measures, the burden remains substantial (1). Osteoarthritis (OA) is a degenerative joint disease caused by aging and mechanical stress (2). In contrast to OA, KBD lesions manifest during childhood and are closely associated with environmental exposure, indicating a distinct pathogenic mechanism (3). Although lipids mediate cartilage degeneration (4), comprehensive lipidomic comparisons are lacking.

Environmental T-2 toxin exposure is a primary risk factor for KBD. Besides inducing chondrocyte apoptosis, the T-2 toxin disrupts lipid homeostasis, causing systemic lipid alterations, oxidative stress, and ferroptosis (5). Nevertheless, how the T-2 toxin

coordinates multitissue lipidomic alterations across the cartilage and plasma remains unclear.

To address this gap, this study integrated multitissue lipidomics with an adverse outcome pathway (AOP). Instead of healthy controls, patients with femoral neck fractures (FNF) served as the orthopedic controls. Differential lipids (DLs) were correlated with metabolic enzymes, integrated with KBD-related genes, and T-2 toxin targets were predicted to construct a hypothetical AOP network.

Patients were recruited at Xi'an Honghui Hospital between April 2023 and March 2025. Cartilage and plasma samples were obtained from 17 patients with KBD, 22 patients with OA, and five with FNF (Supplementary Table S1, available at <https://weekly.chinacdc.cn>). Patient selection was based on the WS/T207–2010 criteria for KBD, the Outerbridge grading system for OA, and clinical assessments for FNF. All patients with KBD were Grade II or III and underwent knee arthroplasty, during which cartilage samples were collected. Detailed patient inclusion and exclusion criteria, sample collection, and information processing are provided in the Supplementary Material (available at <https://weekly.chinacdc.cn>).

Targeted lipidomic analysis utilizing liquid chromatography-mass spectrometry (LC-MS) was conducted with a Vanquish Flex ultra-high-performance liquid chromatography (UPLC) system interfaced with a Q-Exactive Plus high-resolution tandem mass spectrometer (Thermo Fisher Scientific, USA). Raw data were processed using the XCMS software (version 4.2.3, Bioconductor, CA, USA), and lipid annotation was performed using the Kyoto Encyclopedia of Genes and Genomes (KEGG) and Human Metabolome Database (HMDB). Partial least squares discriminant analysis (PLS-DA) models were evaluated using permutation tests. DLs were defined by t -test $P < 0.05$, fold change (FC) > 1.20 , and variable importance in projection (VIP) ≥ 1.00 . Pearson's correlation assessed the cross-tissue associations. KEGG pathway analysis used hypergeometric testing ($P < 0.05$). Metabolite gene set enrichment analysis identified altered pathways using the absolute value of normalized enrichment score (NES) > 1.00 , $P < 0.05$, and false discovery rate (FDR) < 0.25 .

The KEGGREST R package was used to map the DLs to human genes encoding the corresponding metabolic enzymes. KBD-related genes were sourced from GeneCards and OMIM databases, whereas T-2 toxin targets were obtained from the Comparative Toxicogenomics Database (CTD). The intersection of

the mapped genes, KBD-related genes, and T-2 toxin targets was analyzed using the Search Tool for the Retrieval of Interacting Genes/Proteins (STRING) and Cytoscape software (version 3.10.2, ISB, WA, USA). Additionally, AutoDock Vina (version 1.1.2, TSRI, CA, USA) was used to predict binding between T-2 toxin and human prostaglandin-endoperoxide synthase 2 (PTGS2) using the Protein Data Bank (PDB ID: 5KIR). The docking poses were prioritized by binding energy. Here, 117 intersecting genes were considered potential molecular initiating events (MIEs) for enrichment analysis. After excluding terms unrelated to cartilage biology, eight gene ontology (GO) phenotypes and seven KEGG pathways were selected as candidate key events (KEs) within the proposed AOP network (Supplementary Table S2, available at <https://weekly.chinacdc.cn>).

Data analysis used R (version 4.5.1, R Foundation, Vienna, Austria). Statistical tests, including Student's t -test, ANOVA, Mann–Whitney U test, and Kruskal–Wallis test, were applied as appropriate, with $P < 0.05$ indicating significance.

Cartilage lipidomics identified 214, 204, and 47 DLs for KBD versus FNF, OA versus FNF, and KBD versus OA, respectively (Figure 1A–B). The corresponding plasma analyses revealed 212, 124, and 223 DLs, respectively (Figure 1C–D). PLS-DA permutation tests confirmed model stability without significant overfitting (Supplementary Figure S1, available at <https://weekly.chinacdc.cn>). In the cartilage, KBD versus FNF and OA versus FNF showed reduced 5-hydroxyeicosatetraenoic acid (5-HETE) and increased vitamin D3 sulfate levels, with methyl stearate and elaidic acid additionally elevated in KBD versus FNF (Figure 1E–H). Plasma changes were more heterogeneous and driven by triglyceride and diglyceride shifts versus FNF, along with polyunsaturated fatty acid (PUFA) and membrane lipid variations between KBD and OA. Bredemolonic acid, isoacitretin, cinnacsiol E, and 3-palmitoyl-sn-glycerol were elevated in KBD plasma, whereas isotretinoin levels were higher in OA (Figure 1I–L). KBD versus FNF and OA versus FNF shared 158 cartilage DLs (phospholipids, lysophospholipids, free fatty acids, and sphingolipids) and 80 downregulated plasma DLs. A comparison of KBD and OA revealed 33 cross-tissue DLs with higher levels of fatty acids, ether triglycerides, and glycerolipids in KBD (Supplementary Figure S2, available at <https://weekly.chinacdc.cn>).

Cartilage DLs in KBD versus FNF were enriched in

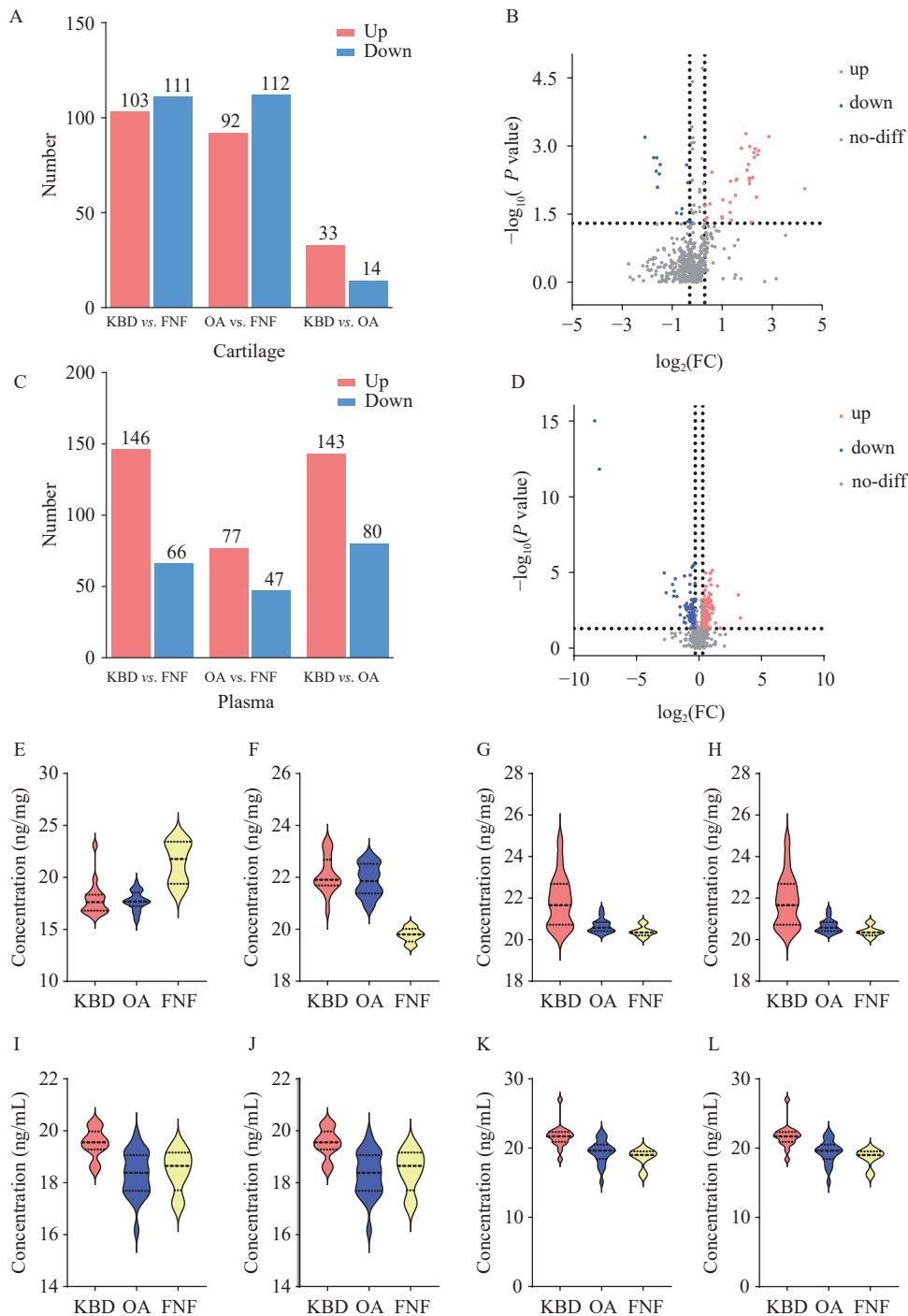


FIGURE 1. DLs in cartilage and plasma. Samples were obtained from patients with KBD ($n=17$), OA ($n=22$), and FNF ($n=5$) at Xi'an Honghui Hospital, Shaanxi Province, China, April 2023–March 2025. (A) Numbers of upregulated and downregulated DLs in cartilage across the three pairwise comparisons. (B) Volcano plot of cartilage DLs in KBD versus OA. (C) Numbers of upregulated and downregulated DLs in plasma across the three pairwise comparisons. (D) Volcano plot of plasma DLs in KBD versus OA. (E) 5-HETE levels in cartilage. (F) Vitamin D3 sulfate levels in cartilage. (G) Methyl stearate levels in cartilage. (H) Elaidic acid levels in cartilage. (I) Bredemolic acid levels in plasma. (J) Isoacitrein levels in plasma. (K) Cinnassiol E levels in plasma. (L) 3-palmitoyl-sn-glycerol levels in plasma.

Note: In panels A and B, red, blue and gray indicate upregulated, downregulated, and non-differential lipids, respectively. In panels C and D, red, blue, and gray indicate upregulated, downregulated, and non-differential lipids, respectively. In panels E–L, red, blue, and yellow represent KBD, OA, and FNF, respectively.

Abbreviation: DLs=differential lipids; KBD=Kashin-Beck disease; OA=osteoarthritis; FNF=femoral neck fracture; FC=fold change; 5-HETE=5-hydroxyeicosatetraenoic acid.

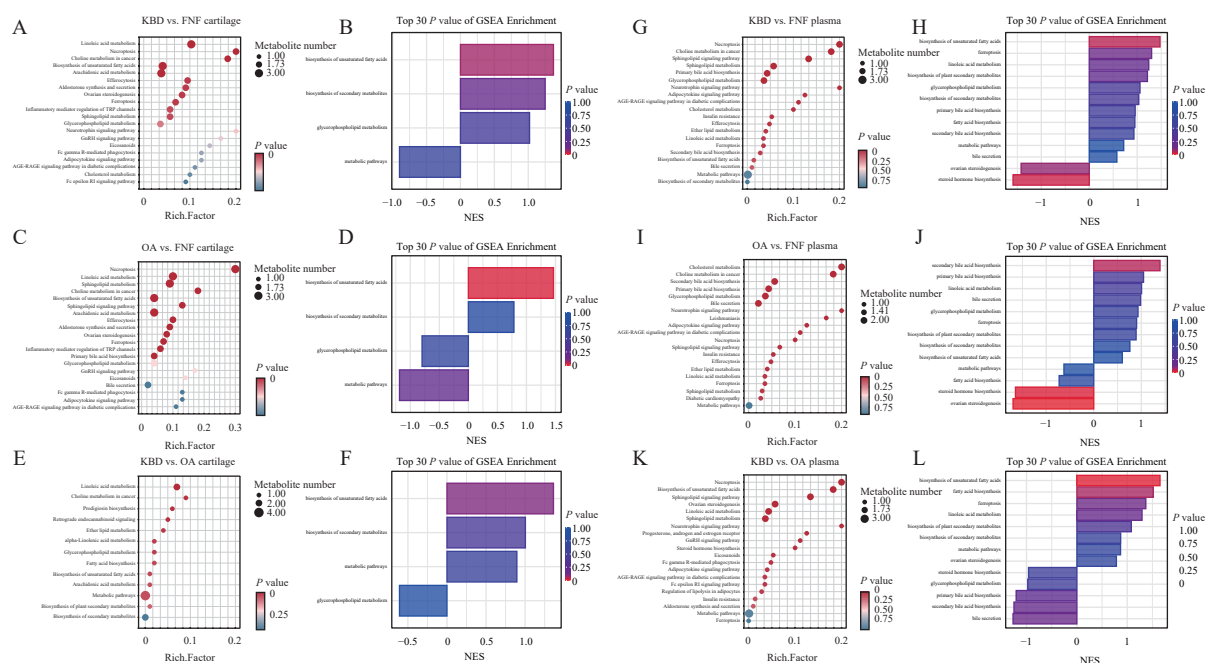


FIGURE 2. KEGG and GSEA of cartilage and plasma DLs. (A) KEGG enrichment of cartilage DLs in KBD versus FNF. (B) GSEA of cartilage DLs in KBD versus FNF. (C) KEGG enrichment of cartilage DLs in OA versus FNF. (D) GSEA of cartilage DLs in OA versus FNF. (E) KEGG enrichment of cartilage DLs in KBD versus OA. (F) GSEA of cartilage DLs in KBD versus OA. (G) KEGG enrichment of plasma DLs in KBD versus FNF. (H) GSEA of plasma DLs in KBD versus FNF. (I) KEGG enrichment of plasma DLs in OA versus FNF. (J) GSEA of plasma DLs in OA versus FNF. (K) KEGG enrichment of plasma DLs in KBD versus OA. (L) GSEA of plasma DLs in KBD versus OA.

Note: In panels A, C, E, G, I and K, bubble size represents the number of metabolites and color represents the P value. In panels B, D, F, H, J and L, the x-axis represents the NES, and bar color represents the P value.

Abbreviation: DLs=differential lipids; KBD=Kashin-Beck disease; OA=osteoarthritis; FNF=femoral neck fracture; KEGG=Kyoto Encyclopedia of Genes and Genomes; GSEA=gene set enrichment analysis; NES=normalized enrichment score.

linoleic acid, arachidonic acid, unsaturated fatty acid biosynthesis, ferroptosis, necroptosis, and apoptosis pathways. OA versus FNF showed similar enrichment but lacked apoptosis. KBD versus OA showed enrichment in linoleic acid, arachidonic acid, and glycerophospholipid metabolism without cellular process enrichment (Figure 2A–F). Plasma enrichment paralleled cartilage, highlighting the sphingolipid, glycerophospholipid, and ferroptosis pathways, with metabolite gene set enrichment suggesting the involvement of fatty acid and sterol pathways (Figure 2G–L).

Systematic mapping identified 165 common genes between KBD and OA DLs (Figure 3A). The protein-protein interaction (PPI) network analysis revealed that core hub genes, notably PTGS2, arachidonate 5-lipoxygenase (ALOX5), and cytochrome P450 family 4 subfamily F (CYP4F), were involved in lipid metabolism (Figure 3B–C). Intersecting the mapped genes with KBD disease genes and T-2 toxin targets identified 117 genes (Figure 3D). Further intersection

with the core network genes highlighted PTGS2 as the singular critical overlapping gene associated with the tumor necrosis factor (TNF) and interleukin-17 (IL-17) signaling pathways (Figure 3E–F). Molecular docking predicted a favorable interaction between T-2 toxin and the active site of PTGS2, which was mediated through hydrogen bonds and hydrophobic interactions (Figure 3G).

A putative AOP network was constructed from 117 candidate MIEs (Figure 3H). These genes were computationally associated with eight GO phenotypes and seven KEGG pathways potentially involved in cartilage injury, including the TNF signaling pathway (Supplementary Figure S1, available at <https://weekly.chinacdc.cn/>). Frequency analysis of these pathways identified PTGS2 as the only consistently overlapping candidate MIE (Figure 3I).

DISCUSSION

This study combined multitissue lipidomics and

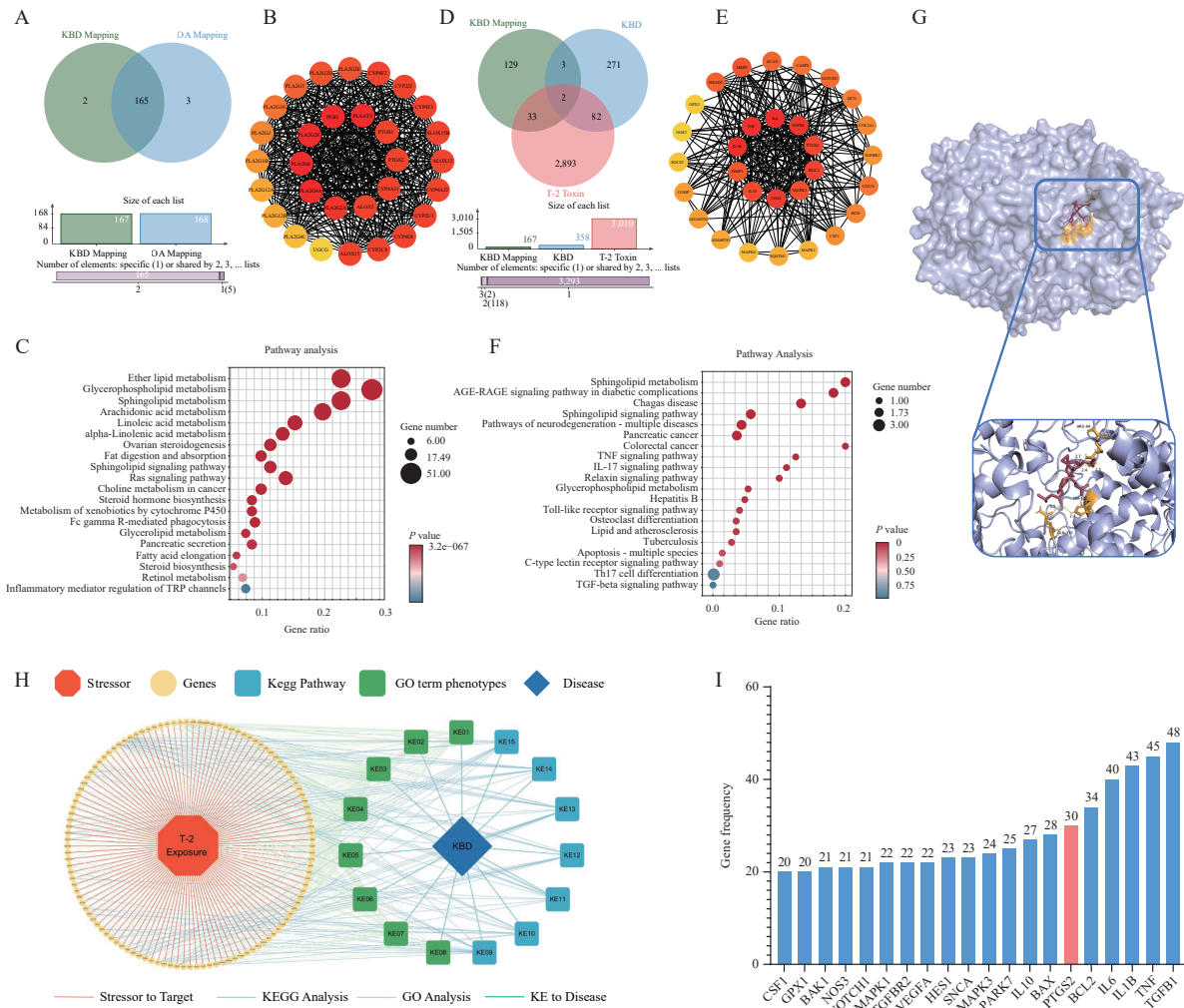


FIGURE 3. Network toxicology analysis of DL mapping genes. (A) Venn diagram of common mapped genes between KBD and OA. (B) PPI network of shared mapped genes between KBD and OA. (C) KEGG enrichment of shared mapped genes. (D) Venn diagram of KBD mapped genes, KBD disease genes, and T-2 toxin targets. (E) PPI network of genes intersecting between KBD and T-2 toxin targets. (F) KEGG enrichment of the intersecting genes. (G) Molecular docking of PTGS2 and T-2 toxin. (H) Integrated network linking T-2 toxin exposure to candidate genes, Gene Ontology (GO) phenotypes, KEGG pathways, and KBD. (I) Frequency of genes across the selected phenotypes and pathways; PTGS2 is highlighted in red. Abbreviation: DLs=differential lipids; KBD=Kashin-Beck disease; OA=osteoarthritis; PPI=protein-protein interaction; KEGG=Kyoto Encyclopedia of Genes and Genomes; GO=Gene Ontology; PTGS2=prostaglandin-endoperoxide synthase 2.

network toxicology to compare KBD and OA and proposed a hypothetical AOP linking T-2 toxin with cartilage damage. Both diseases share alterations in lysophospholipids (LPLs), 5-HETE, vitamin D3 sulfate, arachidonic and linoleic acid metabolism, and ferroptosis, which are consistent with joint lipid remodeling (2,6–7). PTGS2, ALOX5, and CYP4F participate in arachidonic acid metabolism and inflammatory lipid mediator production (8). Conversely, KBD-specific shifts in glycerides, ether lipids, and plasma PUFAs suggest distinct metabolic responses (5). Reduced 5-HETE levels may indicate altered ALOX5-dependent arachidonic acid flux through substrate redistribution toward the

cyclooxygenase (COX)/PTGS2 or alternative lipoxygenase (LOX) pathways (9). Elevated vitamin D3 sulfate suggests altered vitamin D metabolism, although its role remains unclear.

The 117 intersecting genes were associated with signaling pathways related to TNF, IL-17, mitogen-activated protein kinase (MAPK), phosphoinositide 3-kinase/protein kinase B (PI3K-Akt), and sphingolipids. The TNF and IL-17 signaling may activate nuclear factor kappa B (NF- κ B) and MAPK pathways, potentially converging on PTGS2 regulation and prostanoid metabolism, linking inflammatory signaling to lipid dysregulation (10). Nonetheless, this model is derived from network relationships rather than

empirical validation. Consequently, PTGS2 should be considered a candidate rather than a confirmed causal MIE.

This study has two limitations that merit attention. First, the small FNF group reduced the statistical power and limited the generalizability. This was due to stringent inclusion criteria and practical challenges in obtaining matched human cartilage specimens. Owing to the ethical constraints associated with the collection of completely healthy articular cartilage, FNF cartilage is often used as a relatively normal orthopedic control. However, it does not entirely represent healthy tissue, and the acute inflammatory and metabolic responses may complicate plasma comparisons. Second, key lipids and PTGS2 lacked orthogonal validation, and the study did not include PTGS2 knockdown or inhibition experiments, or ferroptosis assays. Larger cohorts, targeted biochemical validation, and functional experiments are required to test the proposed AOP.

In summary, KBD and OA share substantial lipid remodeling, whereas KBD shows distinct glyceride and PUFA features. The proposed T-2 toxin-PTGS2 AOP provides a testable framework for mechanistic validation.

Conflicts of interest: No conflicts of interest.

Acknowledgements: We thank all sample donors and study participants for their participation and support.

Ethical statement: This study was conducted in accordance with the Declaration of Helsinki. This study was approved by the Ethics Committee of Xi'an Jiaotong University (Approval Date: March 1, 2022; Reference Number: 2022SLCJ0421).

Funding: This study was supported by the National Key Research and Development Program of China (grant number 2022YFC2503100).

doi: 10.46234/ccdcw2026.182

* Corresponding author: Feng Zhang, fzhxjtu@mail.xjtu.edu.cn.

¹ Department of Occupational and Environmental Health, School of Public Health, Xi'an Jiaotong University Health Science Center, Xi'an City, Shaanxi Province, China; ² Collaborative Innovation Center of Endemic Diseases and Health Promotion for Silk Road Region, School of Public Health, Xi'an Jiaotong University Health Science Center, Xi'an City, Shaanxi Province, China; ³ Key Laboratory of Trace Elements and Endemic Diseases, National Health and Family Planning

Commission, School of Public Health, Xi'an Jiaotong University Health Science Center, Xi'an City, Shaanxi Province, China.

Copyright © 2026 by Chinese Center for Disease Control and Prevention & Chinese Academy of Preventive Medicine. All content is distributed under a Creative Commons Attribution Non Commercial License 4.0 (CC BY-NC).

Submitted: June 10, 2026

Accepted: August 12, 2026

Issued: September 04, 2026

REFERENCES

1. Cui SL, Liu H, Pei JR, Li JX, Jiao Z, Deng Q, et al. Summary analysis of national surveillance on Kashin-beck disease from 1990 to 2023. *Biomed Environ Sci* 2024;37(9):1056 – 66. <https://doi.org/10.3967/bes2024.109>.
2. Mei ZX, Yilamu K, Ni WY, Shen PY, Pan N, Chen HS, et al. Chondrocyte fatty acid oxidation drives osteoarthritis via SOX9 degradation and epigenetic regulation. *Nat Commun* 2025;16(1):4892. <https://doi.org/10.1038/s41467-025-60037-4>.
3. Han LX, Cheng BL, Wei WM, Liu L, Cheng SQ, Liu H, et al. Whole-transcriptome sequencing of knee joint cartilage from Kashin-beck disease and osteoarthritis patients. *Int J Mol Sci* 2024;25(8):4348. <https://doi.org/10.3390/ijms25084348>.
4. Chang H, Liu L, Zhang QP, Xu GY, Wang JP, Chen P, et al. A comparative metabolomic analysis reveals the metabolic variations among cartilage of Kashin-Beck disease and osteoarthritis. *Bone Joint Res* 2024;13(7):362 – 71. <https://doi.org/10.1302/2046-3758.137.Bjr-2023-0403.R1>.
5. An KY, Shi BZ, Lv XZ, Liu YH, Xia ZF. T-2 toxin triggers lipid metabolism disorder and oxidative stress in liver of ducks. *Ecotoxicol Environ Saf* 2024;286:117169. <https://doi.org/10.1016/j.ecoenv.2024.117169>.
6. Wang GY, Qin SL, Zheng YN, Xia C, Zhang P, Zhang LX, et al. T-2 toxin induces ferroptosis by increasing lipid reactive oxygen species (ROS) and downregulating solute carrier family 7 member 11 (SLC7A11). *J Agric Food Chem* 2021;69(51):15716 – 27. <https://doi.org/10.1021/acs.jafc.1c05393>.
7. Miao Y, Chen YW, Xue F, Liu KX, Zhu B, Gao JJ, et al. Contribution of ferroptosis and GPX4's dual functions to osteoarthritis progression. *EBioMedicine* 2022;76:103847. <https://doi.org/10.1016/j.ebiom.2022.103847>.
8. Wu XX, Sun AR, Crawford R, Xiao Y, Wang YP, Prasadam I, et al. Inhibition of leukotriene A₄ hydrolase suppressed cartilage degradation and synovial inflammation in a mouse model of experimental osteoarthritis. *Cartilage* 2024;15(2):184 – 94. <https://doi.org/10.1177/19476035231169940>.
9. Jónasdóttir HS, Brouwers H, Kwekkeboom JC, van der Linden HMJ, Huizinga T, Kloppenburg M, et al. Targeted lipidomics reveals activation of resolution pathways in knee osteoarthritis in humans. *Osteoarthritis Cartilage* 2017;25(7):1150 – 60. <https://doi.org/10.1016/j.joca.2017.01.018>.
10. Mimpfen JY, Baldwin MJ, Cribbs AP, Philpott M, Carr AJ, Dakin SG, et al. Interleukin-17A causes osteoarthritis-like transcriptional changes in human osteoarthritis-derived chondrocytes and synovial fibroblasts *in vitro*. *Front Immunol* 2021;12:676173. <https://doi.org/10.3389/fimmu.2021.676173>.

SUPPLEMENTARY MATERIAL

Section 1: Patient Sources and Inclusion and Exclusion Criteria

Patient Sources

A total of 44 patients were recruited at Xi'an Honghui Hospital between April 10, 2023, and March 25, 2025, and all patients signed informed consent forms. The 17 patients with Kashin-Beck disease (KBD) were primarily from KBD-endemic areas in Shaanxi Province, such as Xianyang, Yan'an, and Baoji. The 22 patients with osteoarthritis (OA) were mainly from Xi'an and Weinan in Shaanxi Province, and the 5 patients with femoral neck fracture (FNF) were all from Shaanxi Province.

Criteria for Identification of Patients with KBD

According to the National Diagnostic Criteria for KBD (WS/T207-2010), the diagnosis of KBD in patients and controls is made by combining clinical signs and right-hand orthopantomograms, and the diagnostic basis is as follows.

1) Epidemiological investigation: Patients with KBD have a history of contact with the endemic area of the disease.

2) Clinical signs and symptoms: Patients with KBD often have multiple deformed symmetrical epiphyseal plates, and articular cartilage damage, which is mainly manifested as thickening of finger joints, pain and limitation of flexion and extension, and other clinical signs. According to the severity of clinical symptoms and signs, patients are classified into I, II and III degrees.

3) X-ray manifestations: Before the epiphyseal-plate isometric stage, the metacarpophalangeal epiphyses are first indented and sclerotic with calcified bands, and some of them are accompanied by deformational changes of the bone ends or epiphyses and carpal bones. KBD can be classified as mild, moderate, or severe according to the degree of lesions on the X-ray film, or as active, inactive, and obsolete according to the changes on the X-ray film, which can be used to evaluate the degree of activity of KBD.

4) Differential diagnosis: KBD is distinguished from bone and cartilage-related diseases such as osteoarthritis, rheumatoid arthritis, hereditary osteoarthropathy, chondrodysplasia, and epiphyseal dysplasia based on its spatial, temporal, and population distribution.

Criteria for Identification of Patients with OA

OA was diagnosed based on a combination of clinical symptoms, physical examinations, and imaging and arthroscopic findings according to the Outerbridge classification system for cartilage damage.

1) Clinical history: Patients with OA usually had a long-term history of chronic joint pain, most commonly involving weight-bearing joints such as the knee or hip.

2) Clinical signs and symptoms: Patients with OA often showed localized joint tenderness, bony enlargement, crepitus during joint movement, restricted flexion and extension, joint deformity, or varus/valgus malalignment in advanced cases.

3) Imaging and cartilage manifestations: Radiographic examination, computed tomography (CT), magnetic resonance imaging (MRI), or intraoperative observation was used to evaluate degenerative changes of the affected

SUPPLEMENTARY TABLE S1. Baseline information of patients with KBD, OA and FNF recruited at Xi'an Honghui Hospital, Shaanxi Province, China, April 2023–March 2025.

Primary diagnosis	Subjects (n)	Sex (male/female)	Age (years)	P	BMI (kg/m ²)	P
KBD	17	2/15	61.59 (5.78)	0.100	24.52 (4.56)	0.327
OA	22	3/19	63.50 (4.83)		25.43 (3.41)	
FNF	5	2/3	68.80 (13.01)		21.85 (1.63)	

Note: Age and BMI values are mean standard deviation.

Abbreviations: KBD=Kashin-Beck disease; OA=osteoarthritis; FNF=femoral neck fracture; BMI=body mass index.

SUPPLEMENTARY TABLE S2. Classification of phenotypes and pathways.

Classification	Ancestor terms	Phenotypes	Gene count
GO terms	Homeostatic process	Amide metabolic process	19
		Olefinic compound metabolic process	12
		Homeostasis of number of cells	15
		Tissue homeostasis	13
		Anatomical structure homeostasis	13
		Regulation of cellular catabolic process	16
		Maintenance of location	13
		Membrane organization	14
		Carboxylic acid biosynthetic process	12
		Organic acid biosynthetic process	12
		Regulation of membrane potential	14
		Regulation of metal ion transport	13
		Regulation of monoatomic ion transport	14
		Regulation of protein-containing complex assembly	12
		Calcium ion transport	12
		Negative regulation of catalytic activity	12
		Regulation of transmembrane transport	12
		Positive regulation of protein localization	12
	Immune system process	Response to lipopolysaccharide	18
		Leukocyte proliferation	16
		Regulation of leukocyte proliferation	14
		Mononuclear cell proliferation	14
		T cell proliferation	12
		Lymphocyte proliferation	13
		Regulation of mononuclear cell proliferation	12
		Cellular response to biotic stimulus	12
		Regulation of leukocyte cell-cell adhesion	13
		Leukocyte cell-cell adhesion	13
		Regulation of inflammatory response	13
		Chemotaxis	13
		Taxis	13
		Lymphocyte differentiation	12
	Anatomical structure development	Mesenchyme development	18
		Ossification	20
		Connective tissue development	15
		Glial cell differentiation	14
		Mesenchymal cell differentiation	14
		Positive regulation of cell development	17
		Gliogenesis	15
		Cartilage development	12
		Skeletal system morphogenesis	12
		Epithelial cell proliferation	16
		Morphogenesis of an epithelium	13

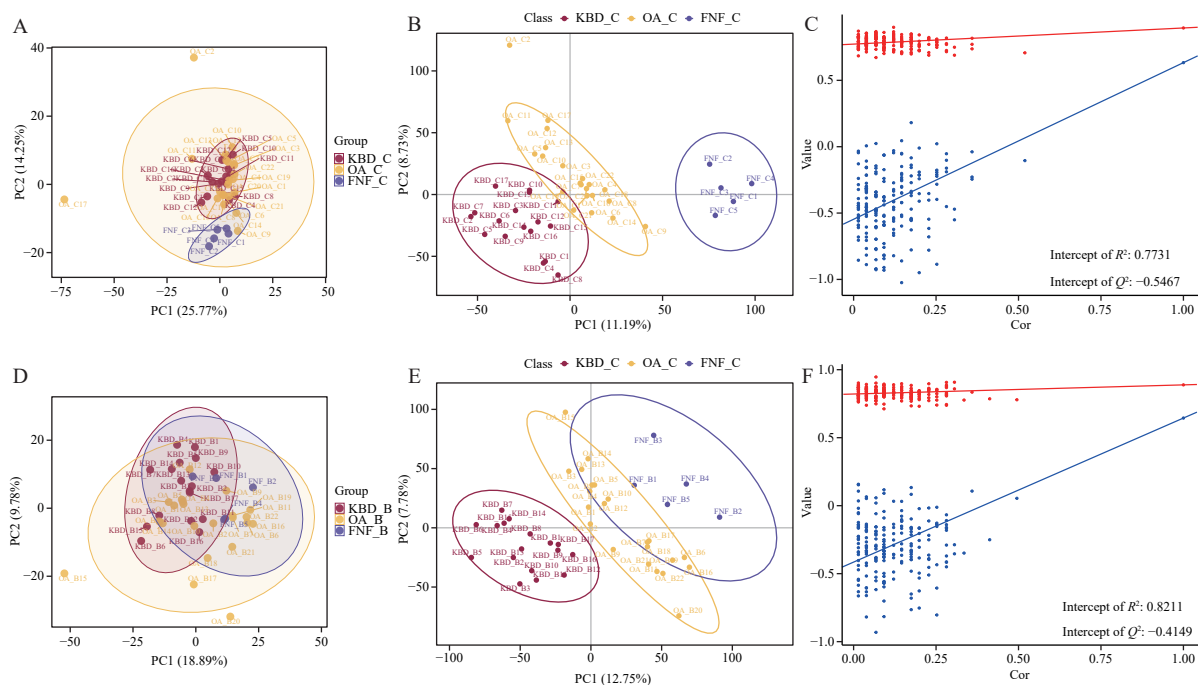
Continued

Classification	Ancestor terms	Phenotypes	Gene count
GO terms	Response to stress	Response to xenobiotic stimulus	25
		Reactive oxygen species metabolic process	16
		Response to oxidative stress	18
		Response to toxic substance	15
		Response to reactive oxygen species	13
		Response to mechanical stimulus	13
		Cellular response to chemical stress	14
		Cellular response to abiotic stimulus	14
		Cellular response to environmental stimulus	14
		Cellular response to oxidative stress	12
		Response to steroid hormone	12
	Lipid metabolic process	Sphingolipid metabolic process	15
		Ceramide metabolic process	13
		Membrane lipid metabolic process	16
		Lipid catabolic process	17
		Fatty acid metabolic process	17
		Steroid metabolic process	14
	Extracellular matrix organization	Phospholipid metabolic process	13
		Extracellular matrix organization	15
		Extracellular structure organization	15
		External encapsulating structure organization	15
		Positive regulation of cell-cell adhesion	12
		Positive regulation of cell adhesion	14
	Programmed cell death	Negative regulation of apoptotic signaling pathway	17
		Regulation of apoptotic signaling pathway	20
		Regulation of intrinsic apoptotic signaling pathway	13
		Extrinsic apoptotic signaling pathway	14
		Intrinsic apoptotic signaling pathway	15
	Signaling	Positive regulation of phosphorylation	14
		Positive regulation of protein phosphorylation	13
		Positive regulation of MAPK cascade	13
KEGG pathways	Not further categorized	Sphingolipid metabolism	14
		Sphingolipid signaling pathway	13
		TNF signaling pathway	11
		Osteoclast differentiation	11
		MAPK signaling pathway	13
		Ras signaling pathway	11
		PI3K-Akt signaling pathway	13

Note: GO terms were grouped into eight functional categories for clarity, whereas KEGG pathways were not further categorized due to the limited number of enriched pathways.

Abbreviations: DLs=differential lipids; KBD=Kashin-Beck disease; OA=osteoarthritis; FNF=femoral neck fracture.

joint. Typical related changes included joint space narrowing, osteophyte formation, subchondral sclerosis, subchondral cysts, cartilage thinning, cartilage defects, and deformity of the articular surface. The severity of



SUPPLEMENTARY FIGURE S1. Quality control of cartilage and plasma lipidomics. (A) PCA of cartilage lipids. (B) PLS-DA of cartilage lipids. (C) Permutation test of the PLS-DA model for cartilage lipids. (D) PCA of plasma lipids. (E) PLS-DA of plasma lipids. (F) Permutation test of the PLS-DA model for plasma lipids.

Note: In the PCA and PLS-DA plots, red, yellow, and blue represent KBD, OA, and FNF samples, respectively. In the permutation tests, red and blue represent R^2 and Q^2 , respectively.

Abbreviation: KBD=Kashin-Beck disease; OA=osteoarthritis; FNF=femoral neck fracture; PCA=principal component analysis; PLS-DA=partial least squares-discriminant analysis; R^2 =coefficient of determination; Q^2 =cross-validated predictive ability.

cartilage degeneration was assessed according to the Outerbridge classification:

- Grade 0: normal articular cartilage;
- Grade I: cartilage softening and swelling;
- Grade II: superficial fragmentation or fissuring with partial-thickness cartilage damage;
- Grade III: deep fissuring or fragmentation extending toward the subchondral bone;
- Grade IV: full-thickness cartilage erosion with exposed subchondral bone.

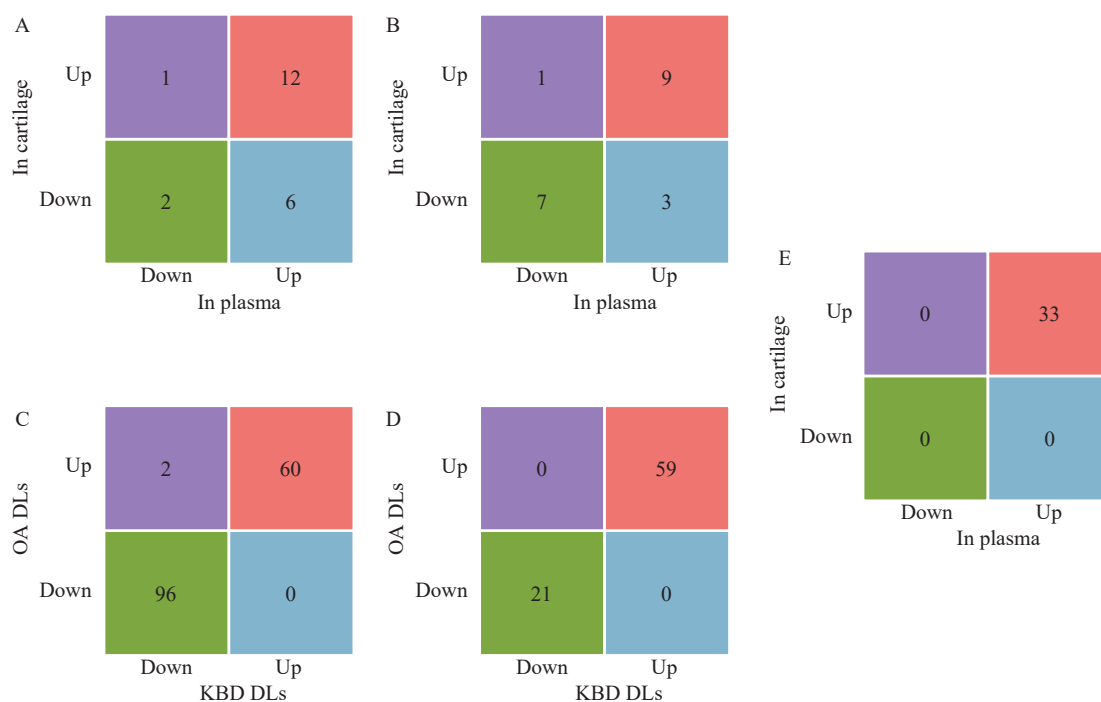
In this study, patients with OA were identified when degenerative cartilage lesions consistent with OA were observed, particularly moderate-to-severe cartilage degeneration corresponding to advanced Outerbridge grades.

4) Patients were excluded if they had evidence of KBD, rheumatoid arthritis, gouty arthritis, septic arthritis, ankylosing spondylitis, psoriatic arthritis, systemic lupus erythematosus-related arthritis, osteonecrosis of the femoral head, congenital or hereditary skeletal dysplasia, traumatic cartilage injury, tumor-related joint destruction, or other secondary causes of joint degeneration. Patients with a clear history of residence in or exposure to KBD-endemic areas combined with typical symmetrical hand joint deformities or characteristic hand X-ray changes were not classified as OA cases.

Criteria for Identification of Patients with FNF

The diagnosis of FNF was based on clinical assessment and imaging confirmation. The clinical evaluation included medical history, trauma history, physical examination, and radiological examination of the hip joint. The diagnostic criteria were as follows.

1) Clinical history: Patients with FNF usually had a definite history of trauma, such as a fall from standing height in elderly patients or high-energy trauma in younger patients. Some patients, especially elderly individuals with osteoporosis, could develop FNF after minor trauma.



SUPPLEMENTARY FIGURE S2. Common DLs across tissues and groups. (A) Common DLs in KBD cartilage and plasma (Consistency rate-66.7%). (B) Common DLs in OA cartilage and plasma (Consistency rate-80.0%). (C) Common DLs in KBD versus OA cartilage (Consistency rate-98.7%). (D) Common DLs in KBD versus OA plasma (Consistency rate-100.0%). (E) Common DLs in cartilage and plasma of KBD and OA (Consistency rate-100.0%).

Note: Red and green indicate concordant upregulation and downregulation, respectively. Purple indicates upregulation in the dataset represented on the y-axis and downregulation in the dataset represented on the x-axis, whereas blue indicates the opposite directional pattern. The consistency rate represents the proportion of common DLs showing concordant directional changes in the two compared datasets.

Abbreviation: DLs=differential lipids; KBD=Kashin-Beck disease; OA=osteoarthritis; FNF=femoral neck fracture.

2) Clinical signs and symptoms: Patients with FNF often presented with pain around the hip or groin, reduced range of motion of the hip joint, tenderness over the anterior hip or groin region, pain during passive hip rotation, and inability to perform straight leg raising. Displaced FNF could be accompanied by shortening, abduction, and external rotation of the affected lower limb.

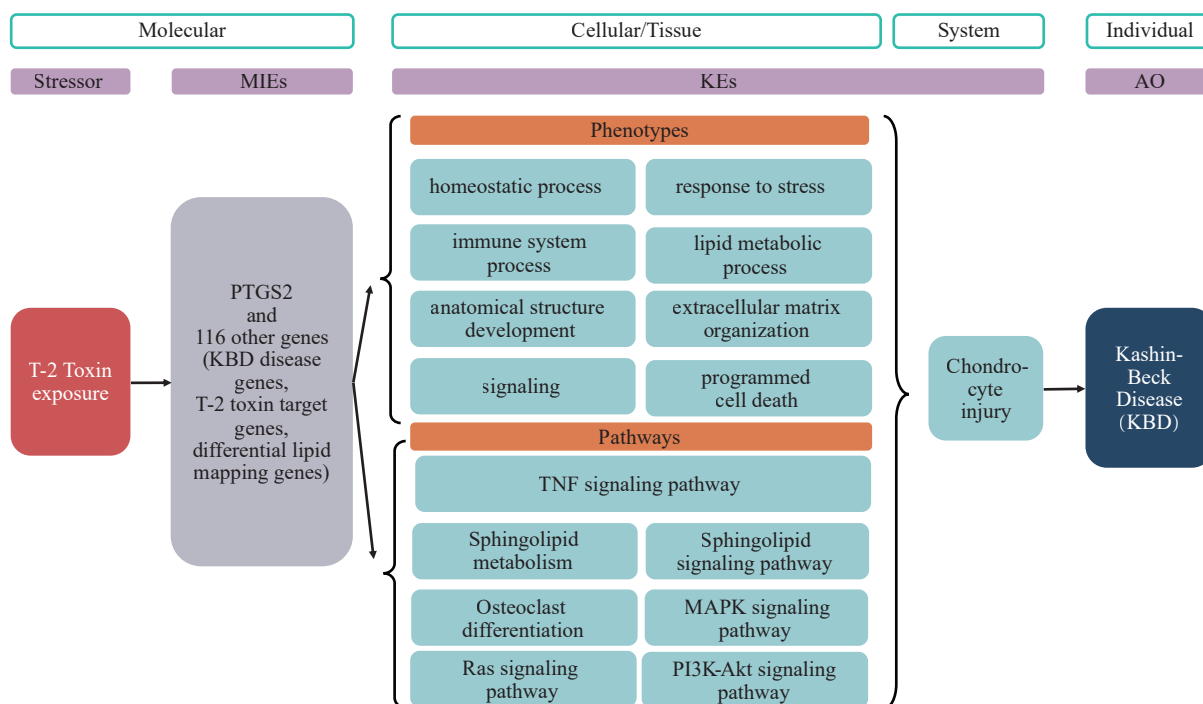
3) Imaging manifestations: The diagnosis of FNF was confirmed by imaging examination. FNF was identified by the presence of a fracture line through the femoral neck, cortical interruption, trabecular discontinuity, impaction, angulation, or displacement of the femoral head-neck fragment. When plain radiographs were negative or equivocal but clinical suspicion remained high, CT or MRI was used to detect occult, non-displaced, impacted, or stress-related FNF.

4) Patients were excluded if the lesion was not located in the femoral neck or if imaging suggested intertrochanteric fracture, subtrochanteric fracture, femoral shaft fracture, acetabular fracture, hip dislocation, isolated greater trochanteric fracture, pelvic fracture, osteonecrosis of the femoral head, hip osteoarthritis, inflammatory arthritis, bone tumor, metastatic bone disease, osteomyelitis, or pathological fracture unrelated to ordinary trauma or osteoporosis. Patients with chronic degenerative joint disease without an acute FNF line were not classified as FNF cases.

Section 2: Cartilage and Plasma Sample Collection and Preprocessing Procedures

Cartilage and Plasma Sample Collection Methods

Peripheral venous blood was collected from all patients and centrifuged within 2 h of collection. After centrifugation at 3,000 rpm for 15 min at 4 °C, the supernatant plasma was carefully aspirated, aliquoted into



SUPPLEMENTARY FIGURE S3. Putative adverse outcome pathway (AOP) framework linking T-2 toxin exposure to Kashin-Beck disease.

Note: T-2 toxin exposure is shown as the stressor; PTGS2 and 116 other intersecting genes are presented as candidate molecular initiating events (MIEs). Candidate key events (KEs) include eight Gene Ontology phenotype categories and seven pathways. These KEs converge on chondrocyte injury at the system level and KBD as the adverse outcome (AO).

nuclease-free cryovials, and immediately stored at -80°C for subsequent lipidomic analysis. Samples from the femoral condyles were collected within 1 h following joint replacement surgery, ensuring the inclusion of an intact cartilage layer and a 0.5–1 cm layer of subchondral bone, in preparation for subsequent lipidomic analysis.

Cartilage and Plasma Sample Preprocessing Methods

A plasma sample (200 μL) was extracted using methyl tert-butyl ether (MTBE, 400 μL) and ice-cold methanol (80 μL). Cartilage samples (50 mg) were homogenized in 50% cold methanol (200 μL) and extracted with cold MTBE (600 μL). Both mixtures were centrifuged and the supernatants were dried and reconstituted with dichloromethane-methanol (v/v, 1:1) for instrumental analysis.

Section 3: Lipidomics Data Acquisition Process and Chromatographic Conditions

Lipidomics Data Acquisition Process

Lipidomics data were acquired using a Thermo Fisher Scientific Q-Exactive Plus high-resolution tandem mass spectrometer coupled to a Thermo Vanquish Flex UPLC system. Data acquisition employed the full mass spectrometry (MS) and data-dependent acquisition (DDA) modes. In each acquisition cycle, the full scan range was 100–1,500 Da, with a resolution of 70,000, an automatic gain control (AGC) target of $3\text{E}6$, and a maximum ion injection time (IT) of 100 ms. The top six ions with response intensities exceeding 100,000 were selected from the full scan for MS/MS analysis, with an MS/MS resolution of 17,500 and a maximum IT of 50 ms. The dynamic exclusion time was set to 6 s.

Chromatographic and Mass Spectrometric Conditions

Lipidomic separation was performed using an ACQUITY UPLC CSH C18 column (100 mm $\times$ 2.1 mm, 1.7 μm ; Waters). The mobile phase consisted of solvent A (acetonitrile:water, 6:4 v/v, containing 10 mM ammonium

formate and 0.1% formic acid) and solvent B (isopropanol:acetonitrile, 9:1, v/v, containing 10 mM ammonium formate and 0.1% formic acid). The gradient elution program was as follows: 0–0.5 min, 30% B; 0.5–2.0 min, 30%–75% B; 2.0–5.2 min, 75%–100% B; 5.2–6.8 min, 100% B; 6.8–6.81 min, 100%–30% B; and 6.81–10.0 min, 30% B. The flow rate was maintained at 0.35 mL/min, the injection volume was 4 μ L, and the column temperature was kept at 40 °C. Mass spectrometry (MS) was performed using a Q-Exactive Plus high-resolution tandem mass spectrometer (Thermo Fisher Scientific) equipped with an electrospray ionization (ESI) source. Each sample was analyzed in both the positive and negative ion modes. The ESI source temperature was set to 350 °C. The spray voltage was +3,800 V in the positive ion mode and –3,500 V in the negative ion mode. The pressures of sheath, auxiliary, and sweep gases were maintained at 45, 12, and 0 arbitrary units (Arb), respectively. Data acquisition was performed using the full scan-data-dependent acquisition (full scan-DDA) mode. The mass-to-charge ratio (m/z) range for the full scan was 100–1,500 Da, with a resolution of 70,000, an automatic gain control (AGC) target of 3E6, and a maximum ion injection time (IT) of 100 ms. Subsequently, the six most intense precursor ions with signal intensities exceeding 100,000 were selected from the full-scan results for DDA MS/MS fragmentation analysis. The DDA scan parameters included a resolution of 17,500, a maximum ion injection time of 50 ms, and dynamic exclusion time of 6 s.

Section 4: Data Processing and Analysis Process

Data Preprocessing Process

The acquired MS data were preprocessed using XCMS software, involving steps such as peak picking, peak grouping, retention time correction, secondary peak grouping, and annotation of isotopes and adducts. Raw LC-MS data files were converted to mzXML format and subsequently processed using R-based toolkits, including XCMS, CAMERA, and metaX. Individual ions were identified based on the retention time (RT) and mass-to-charge ratio (m/z). Peak intensities were recorded to generate a three-dimensional matrix containing specific peak indices (RT- m/z combinations), sample names (observations), and ion intensity information (variables).

Information Analysis Process

Metabolites were annotated using online KEGG and HMDB databases by matching the samples' accurate molecular mass data (m/z) against database entries. Annotations were assigned when the mass deviation between the observed and database values was less than 10 ppm; molecular formulas were further determined and verified through isotopic distribution analysis. Additionally, metabolite identification results were validated using an in-house metabolite fragmentation spectral library. Metabolite data underwent three key processing steps: data filtering (removing samples with >80% missing values or quality control [QC] samples with >50% missing data), data imputation using the K-nearest neighbors (KNN) method, and data normalization via probabilistic quotient normalization (PQN). Principal component analysis (PCA) and differential metabolite analysis were performed using the 'metaX' package. Partial least squares-discriminant analysis (PLS-DA) was conducted using the 'ropls' package, and variable importance in projection (VIP) values were calculated. Finally, significantly differential metabolites were selected based on three criteria: a t -test $P < 0.05$, $FC > 1.20$, and a VIP value ≥ 1.00 in the PLS-DA analysis. Hypergeometric testing was employed for KEGG pathway enrichment analysis, with a $P < 0.05$ indicating significant enrichment. GSEA (v4.1.0) was used for metabolite set enrichment analysis; KEGG pathways meeting the criteria of $|NES| > 1.00$, nominal $P < 0.05$, and FDR < 0.25 were considered significantly different between the two groups.

Preplanned Studies

Network Analysis of Visual Behaviors, Family Factors, and Refractive Status in Chinese Schoolchildren — 10 PLADs, China, 2020–2024

Keke Liu^{1,2}; Huijuan Luo²; Yan Chen^{1,2}; Boran E^{1,2}; Huining Kuang^{1,2}; Chenyu Zhang^{1,2}; Minglu Wang³; Xin Guo^{1,3,#}

Summary

What is already known about this topic?

Myopia is a growing public health concern, with established risk factors including near-work activities and family-related factors. However, the interactions between visual behaviors and family factors remain poorly understood.

What is added by this report?

This study uses network analysis to examine the relationships among visual behaviors, family factors, and refractive status in 29,525 children. It identifies key central and bridging nodes, highlighting the need for age-specific interventions.

What are the implications for public health practice?

Targeting postural behaviors, reducing screen time, and promoting family-based interventions may help prevent myopia.

children than among older children (0.76 *vs.* 0.67), whereas older children exhibited fewer but stronger connections. Based on cycloplegic refraction in the study sample, the detection rate of myopia was also higher among older children than among younger children [51.85% (4,395/8,477) *vs.* 15.64% (3,291/21,048)].

Conclusion: This study provides a comprehensive understanding of the interrelationships among visual behaviors, family factors, and refractive status, highlighting postural behaviors and the bridging roles of screen time, homework/reading, and tutoring activities. Age-related differences in network characteristics may help inform future prevention strategies.

ABSTRACT

Introduction: Myopia is a growing public health concern, particularly in East Asia, where it contributes substantially to pediatric healthcare expenditures. The development of myopia is influenced by complex interactions between visual behaviors, family factors, and refractive status.

Methods: This study used network analysis to examine the interrelationships among visual behaviors, family factors, and refractive status in 29,525 Chinese schoolchildren aged 6–14 years. Central and bridging nodes were identified within the myopia development network.

Results: Postural behaviors dominated the network, with eye-to-book distance (strength=1.20) and head support (1.16) emerging as the most central nodes. In contrast, break location (bridge=0.30) was the strongest behavioral bridge node, followed by non-study screen time (0.24) and homework/reading time (0.22). Network density was higher among younger

The high prevalence of myopia and its irreversible ocular complications underscore the urgency of early prevention and control strategies. In China, the annual economic burden of myopia in urban areas was estimated at 173.6 billion Chinese Yuan (CNY) [approximately 26.3 billion United States dollars (USD)], with rapid urbanization and increasing educational pressure identified as key drivers of this growing burden (1). Myopia develops through complex interactions between genetic predisposition and environmental influences, including lifestyle and behavioral exposures (2). Whereas traditional studies typically focus on isolated risk factors, network analysis provides a complementary approach by modeling interrelationships and identifying central nodes (highly influential factors) and bridging nodes (connectors between domains). In contrast to conventional regression models that estimate independent associations, network analysis enables simultaneous examination of interconnections among behavioral, family, and physiological factors, making it well suited to the multifactorial nature of myopia. Accordingly, this study applies network analysis to examine the

interrelationships among visual behaviors, family factors, and refractive status in Chinese schoolchildren, with the aims of identifying central and bridging factors, comparing network structures across age groups, and informing targeted prevention strategies.

This study analyzed data from a national cross-sectional survey on eye development and related behavioral factors among children and adolescents conducted by the Chinese Center for Disease Control and Prevention between 2020 and 2024. The selected provincial-level administrative divisions (Beijing, Shanxi, Liaoning, Zhejiang, Shandong, Henan, Hunan, Guangdong, Chongqing, and Shaanxi) represent eastern, central, and western China and encompass varying levels of economic development and urbanization. A multistage cluster sampling design was used to recruit students from kindergarten to high school, with an initial sample of 48,646 participants. Although the original survey included a broad age range, the present analysis was restricted to children aged 6–14 years. Exclusions comprised participants outside this age range ($n=1,947$); missing or abnormal axial length values (10–40 mm) ($n=662$); missing spherical equivalent data (669); abnormal spherical equivalent values outside $Q1 - 3 \times \text{interquartile range (IQR)}$ or $Q3 + 3 \times \text{IQR}$ ($n=1,144$); missing or invalid screen time data ($n=8,244$); missing parental myopia data ($n=947$); and invalid responses for daily outdoor activity time ($n=3,901$), homework/reading time ($n=709$), tutoring duration ($n=881$), break-related variables ($n=14$), and posture variables ($n=3$). The final analytic sample comprised 29,525 children aged 6–14 years.

All participants underwent standardized ophthalmic examinations following a unified protocol. Noncycloplegic and cycloplegic autorefractor were performed using identical autorefractor models across all sites. Cycloplegia was induced using 0.5% tropicamide administered four times at 5-minute intervals. Due to high inter-eye correlation, only right-eye data were analyzed. Spherical equivalent (SE) was calculated as sphere plus half cylinder. Myopia was defined as cycloplegic SE ≤ -0.50 diopters (D), and insufficient hyperopic reserve as age-specific cycloplegic SE below established normative thresholds (3). Visual behaviors were assessed using standardized questionnaires, including daily outdoor activity time, homework/reading time, tutoring duration, non-study screen time, break location and frequency, and six posture-related items during reading and writing (chest-to-desk distance, eye-to-book distance, finger-

to-pen distance, lying posture, head support, and sitting posture). Parental myopia and covariates (age, sex, province, residence, and parental education) were also collected.

Continuous variables are presented as mean $\pm$ standard deviation and compared using the Wilcoxon rank-sum test. Categorical variables are presented as frequencies (percentages) and compared using Pearson's chi-square test. All tests were two-sided, with statistical significance set at $P < 0.05$. Networks were estimated using EBICglasso (Extended Bayesian Information Criterion graphical Least Absolute Shrinkage and Selection Operator) with LASSO (Least Absolute Shrinkage and Selection Operator) regularization. Edge weights represent partial correlations between nodes, density reflects the proportion of observed connections, strength centrality is the sum of absolute edge weights, and bridge centrality quantifies cross-module connectivity. To reduce confounding, each node was residualized for prespecified covariates (age, sex, province, residence, and parental education) prior to network estimation. These variables were selected based on prior evidence as established confounders of the visual behavior–myopia associations. To account for clustering at the school level, cluster-robust standard errors were applied, and province fixed effects were included. Stability was evaluated using 1,000 bootstrap resamples, with a correlation stability (CS) coefficient > 0.50 indicating acceptable reliability. Participants were stratified into younger (6–9 years) and older (10–14 years) groups to examine developmental differences while maintaining network stability estimation. The Network Comparison Test (NCT) assessed differences in global structure and edge weights between groups. All analyses were performed in R software (version 4.5.2; R Foundation for Statistical Computing, Vienna, Austria) using the q graph, bootnet, and NCT packages.

A total of 29,525 children were included, of whom 21,048 [71.29% (21,048/29,525)] were aged 6–9 years and 8,477 [28.71% (8,477/29,525)] were aged 10–14 years. Compared with younger children, older children had lower spherical equivalent (0.46 ± 1.10 D *vs.* -0.81 ± 1.63 D, $P < 0.001$) and higher axial length-to-corneal radius ratio (2.96 ± 0.09 *vs.* 3.06 ± 0.11 , $P < 0.001$). Based on cycloplegic refraction in our study sample, the detection rate of myopia was substantially higher in older children [51.85% (4,395/8,477) *vs.* 15.64% (3,291/21,048), $P < 0.001$]. Significant differences were also observed in residence, parental

TABLE 1. Characteristics of study participants.

Characteristic	Total (N=29,525)	6–9 years (n=21,048)	10–14 years (n=8,477)	P
Age, years, mean±SD	8.66±1.87	7.66±0.87	11.17±1.21	<0.001
Sex, n (%)				0.178
Male	15,127 (51.23)	10,731 (50.98)	4,396 (51.86)	
Female	14,398 (48.77)	10,317 (49.02)	4,081 (48.14)	
Residence, n (%)				<0.001
Urban	22,270 (75.43)	15,544 (73.85)	6,726 (79.34)	
Rural	7,255 (24.57)	5,504 (26.15)	1,751 (20.66)	
Parental education, n (%)				<0.001
≤ High school	12,447 (42.16)	8,274 (39.31)	4,173 (49.23)	
Associate/Bachelor's degree	15,769 (53.41)	11,742 (55.79)	4,027 (47.51)	
Graduate degree	1,208 (4.09)	945 (4.49)	263 (3.10)	
Parental myopia, n (%)				<0.001
Neither parent myopic	4,513 (15.29)	3,289 (15.63)	1,224 (14.44)	
One parent myopic	12,634 (42.79)	9,541 (45.33)	3,093 (36.49)	
Both parents myopic	12,378 (41.92)	8,218 (39.04)	4,160 (49.07)	
AL/CR ratio, mean±SD	2.99±0.11	2.96±0.09	3.06±0.11	<0.001
Spherical equivalent, D, mean±SD	0.10±1.40	0.46±1.10	−0.81±1.63	<0.001
Myopia, n (%)	7,686 (26.03)	3,291 (15.64)	4,395 (51.85)	<0.001
Pre-myopia, n (%)	12,284 (41.61)	9,291 (44.14)	2,993 (35.31)	<0.001
Emmetropia, n (%)	9,555 (32.36)	8,466 (40.22)	1,089 (12.85)	<0.001

Abbreviation: AL/CR ratio=axial length-to-corneal radius ratio; D=Diopters; SD=standard deviation; SE=spherical equivalent.

TABLE 2. Network structure metrics.

Network	Nodes	Edges	Density	Strength
Full sample	15	80	0.76	5.55
Age 6–9 years	15	80	0.76	4.64
Age 10–14 years	15	71	0.67	4.74

Note: Density=proportion of possible edges present; Strength=sum of absolute edge weights. Network Comparison Test showed no significant difference in global structure between age groups ($P=0.802$), although edge weights differed ($P=0.010$).

education, and parental myopia (all $P<0.001$) (Table 1). Distributions of behavioral variables are presented in Supplementary Table S1(available at <https://weekly.chinacdc.cn/>).

The full-sample network included 15 nodes and 80 edges, with a density of 0.76 and mean strength of 5.55 (Table 2, Figure 1A). Behavioral variables were densely interconnected, particularly within near-work activities (homework/reading time, tutoring duration, and non-study screen time). Posture variables formed a distinct cluster, with eye-to-book distance and head support showing strong links to physiological outcomes. Age-stratified analysis revealed differences in network characteristics (Table 2, Figure 1B–C). The younger group showed higher density (0.76 *vs.* 0.67), whereas the older group showed slightly greater overall

strength (4.74 *vs.* 4.64). However, the NCT indicated no significant difference in global structure between groups ($P=0.802$), although specific edge weights differed ($P=0.010$). Bootstrap analysis supported network stability, with a CS coefficient of 0.75 for strength centrality and generally narrow confidence intervals for edge estimates.

Strength centrality indicated that posture-related nodes dominated the network. Eye-to-book distance (1.20) and head support (1.16) were the most central nodes, followed by lying posture (0.97) and chest-to-desk distance (0.92) (Figure 2A, Supplementary Table S2, available at <https://weekly.chinacdc.cn/>). Among behavioral variables, break location ranked fifth (0.89). In contrast, non-study screen time (0.42), tutoring duration (0.44), homework/reading time

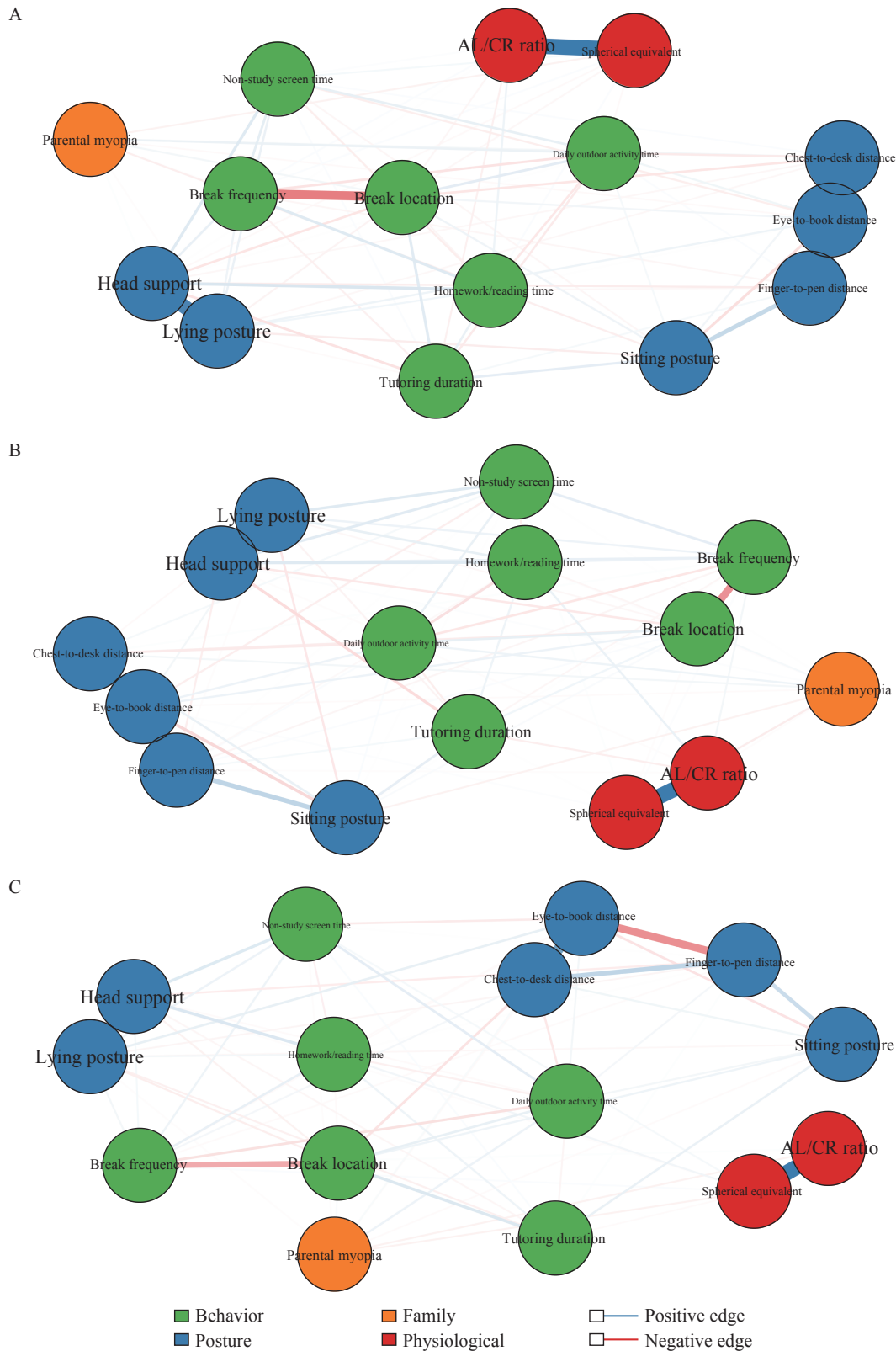


FIGURE 1. Network structures of the full sample and age-stratified groups. Networks for (A) full sample ($N=29,525$), (B) children aged 6–9 years ($n=21,048$), and (C) children aged 10–14 years ($n=8,477$). Nodes are colored by type: green=Behavior (6 nodes), blue=Posture (6 nodes), orange=Family (1 node), red=Physiological (2 nodes). Edges represent partial correlations: blue=positive, red=negative. Edge thickness indicates correlation strength. Abbreviation: AL/CR ratio=axiallength-to-corneal radius ratio.

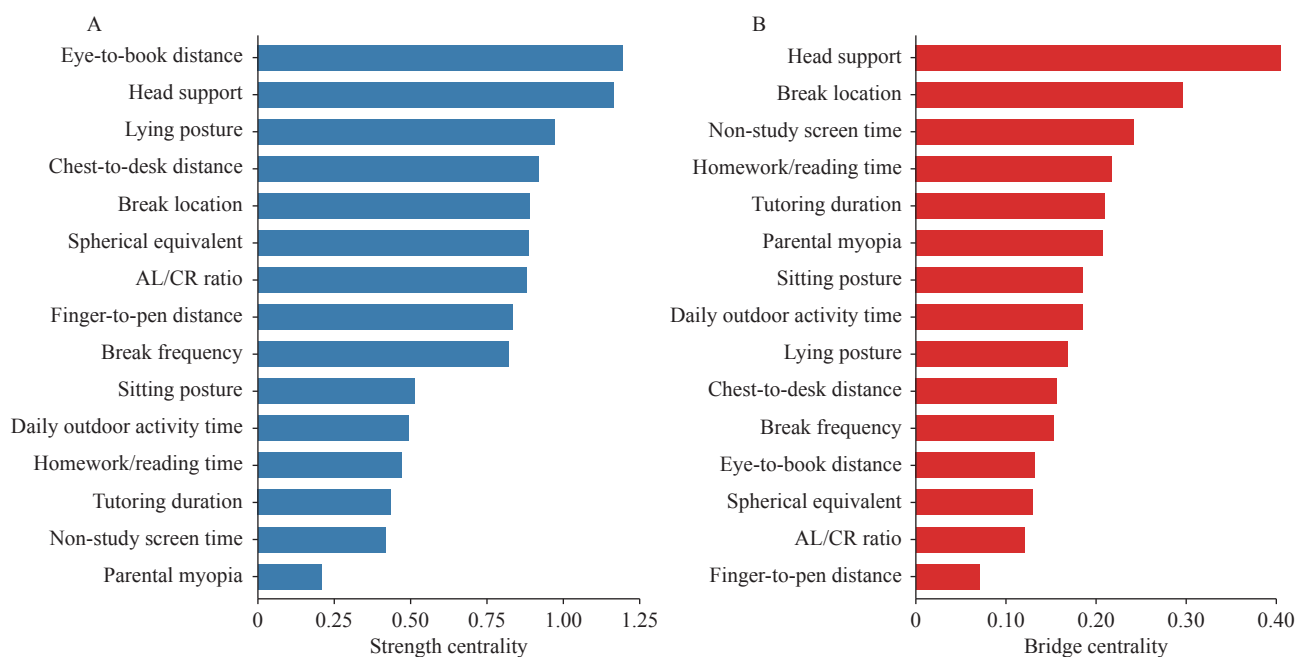


FIGURE 2. Node centrality analysis. (A) Strength centrality rankings (15 nodes). (B) Bridge centrality rankings (15 nodes). Note: In panel A, higher strength indicates greater influence in the network. Nodes are arranged in descending order of strength. In panel B, higher bridge strength indicates greater importance in connecting different modules (Behavior, Posture, Family Factors, Physiological). Error bars represent 95% bootstrap confidence intervals. Abbreviation: AL/CR ratio=axiallength-to-corneal radius ratio.

(0.47), and outdoor time (0.50) showed low centrality, while parental myopia (0.21) exhibited the lowest value. Bridge centrality identified key connectors between modules (Figure 2B, Supplementary Table S3, available at <https://weekly.chinacdc.cn/>). Head support (0.41) was the strongest bridge node linking posture and physiological domains, followed by break location (0.30). Non-study screen time (0.24), homework/reading time (0.22), and tutoring duration (0.21) also showed moderate bridging roles. Parental myopia (0.21) additionally demonstrated cross-module connectivity, suggesting a potential role of family factors within the broader network structure.

DISCUSSION

This study applied network analysis to examine the interrelationships among visual behaviors, family factors, and refractive status in Chinese schoolchildren. In contrast to traditional approaches focusing on isolated risk factors, the findings highlight a structured interdependence among behavioral, familial, and physiological domains, offering a systems-level perspective on myopia development.

Postural behaviors, particularly eye-to-book distance and head support, occupied central positions in the

network, indicating stronger connectivity with other variables than most behavioral and familial factors. Their centrality underscores their relevance within the broader behavioral–physiological system and supports further longitudinal investigation. These findings align with previous evidence linking near-work posture to retinal defocus, ocular responses, and myopia-related structural changes (4).

Screen time and homework, although less central, functioned as key bridging nodes connecting behavioral and physiological domains. This suggests that near-work activities may exert influence through contextual pathways involving posture and refractive outcomes, consistent with prior research (5–6). The results emphasize the importance of considering both duration and context of near-work behaviors in prevention strategies.

Parental myopia emerged as a bridging factor between family and behavioral domains, consistent with established evidence of familial aggregation in childhood myopia (7). Its cross-domain connectivity suggests that family-related factors may interact with behavioral patterns within the broader network, although causal direction cannot be inferred from the cross-sectional data (8).

Age-stratified analyses revealed broadly similar

network structures but differences in specific connections. Younger children showed more densely interconnected behavioral and postural relationships, whereas older children exhibited more selective connectivity. These findings suggest developmental variation in the relative importance of specific pathways, supporting the potential value of age-tailored prevention strategies.

A key strength of this study is the application of network analysis to capture interdependencies among multiple risk domains. This approach moves beyond isolated associations by identifying central and bridging nodes, offering a structured view of myopia-related factors (9). The observed age-related differences further support a dynamic, developmentally sensitive framework for understanding myopia risk.

Several limitations should be noted. First, the cross-sectional design precludes causal inference. Second, residual confounding from unmeasured factors (e.g., household income, school policies, light exposure) may remain. Third, behavioral measures were self-reported and subject to recall bias. Fourth, findings are limited to children aged 6–14 years. Fifth, although missing data led to exclusions, comparisons suggested limited selection bias. Future studies should include broader age ranges and longitudinal designs to strengthen generalizability and causal interpretation.

In conclusion, network analysis revealed that posture-related behaviors occupy central positions in the myopia-related system, while screen time, homework, and parental myopia function as key bridging factors. Age-related variations in network structure further suggest developmental differences in the organization of myopia risk. These findings provide a systems-level perspective that may inform future prevention strategies.

Acknowledgements: Gratitude goes to all staff members who participated in this study, including those from Beijing, Zhejiang, Chongqing, Shaanxi, Liaoning, Henan, Guangdong, Shandong, Hunan, and Shanxi Provinces.

Ethical statements: This study was approved by the Institutional Review Board of the Beijing Center for Disease Prevention and Control (Approval No. 2022 No.24).

Conflicts of interest: No conflicts of interest.

Funding: This work was supported by the Operation

of Public Health Emergency Response Mechanism (grant number 102393220020010000017).

doi: 10.46234/ccdcw2026.183

Corresponding author: Xin Guo, guoxin@chinacdc.cn.

¹ School of Public Health, Capital Medical University, Beijing, China;

² Institute for School Health, Beijing Center for Disease Prevention and Control, Beijing, China; ³ Center for School Health, Chinese Center for Disease Control and Prevention & Chinese Academy of Preventive Medicine, Beijing, China.

Copyright © 2026 by Chinese Center for Disease Control and Prevention & Chinese Academy of Preventive Medicine. All content is distributed under a Creative Commons Attribution Non Commercial License 4.0 (CC BY-NC).

Submitted: April 04, 2026

Accepted: June 14, 2026

Issued: September 04, 2026

REFERENCES

1. Ma YY, Wen YC, Zhong H, Lin SL, Liang L, Yang YF, et al. Healthcare utilization and economic burden of myopia in urban China: a nationwide cost-of-illness study. *J Glob Health* 2022;12:11003. <https://doi.org/10.7189/jogh.12.11003>.
2. Baird PN, Saw SM, Lanca C, Guggenheim JA, Smith III EL, Zhou XT, et al. Myopia. *Nat Rev Dis Primers* 2020;6(1):99. <https://doi.org/10.1038/s41572-020-00231-4>.
3. National Disease Control and Prevention Administration. WS/T 10039-2025 Specification for evaluation of uncorrected distant visual acuity and refractive status in children and adolescents. Beijing: Standards Press of China, 2025. <http://www.csres.com/detail/440903.html>. (In Chinese).
4. Jiang DJ, Shi BJ, Gao H, Guo YB, Zhou SY, Zhang Y. Associations between reading and writing postures and myopia among school students in Ningbo, China. *Front Public Health* 2022;10:713377. <https://doi.org/10.3389/fpubh.2022.713377>.
5. Zong ZQ, Zhang YX, Qiao JC, Tian Y, Xu SJ. The association between screen time exposure and myopia in children and adolescents: a meta-analysis. *BMC Public Health* 2024;24(1):1625. <https://doi.org/10.1186/s12889-024-19113-5>.
6. Gajjar S, Ostrin LA. A systematic review of near work and myopia: measurement, relationships, mechanisms and clinical corollaries. *Acta Ophthalmol* 2022;100(4):376 – 87. <https://doi.org/10.1111/aos.15043>.
7. Guo L, Zhang JY, Guo YF, Hu Y, Wang LL, Yang X, et al. Associations of incident myopia and its development with related factors among school-aged children across different grades: a four-year longitudinal study. *Eye (Lond)* 2025;39(4):725 – 33. <https://doi.org/10.1038/s41433-024-03467-6>.
8. Enthoven CA, Tideman JWL, Polling JR, Tedja MS, Raat H, Iglesias AI, et al. Interaction between lifestyle and genetic susceptibility in myopia: the Generation R study. *Eur J Epidemiol* 2019;34(8):777 – 84. <https://doi.org/10.1007/s10654-019-00512-7>.
9. Xiang YJ, Lu CJ, Zhu HM, Sun XY, Mu JF. Exploring the interrelationship between myopia, psychological resilience, and emotion regulation in adolescents using a network analysis. *Sci Rep* 2025;15(1):25703. <https://doi.org/10.1038/s41598-025-10390-7>.

SUPPLEMENTARY MATERIAL

SUPPLEMENTARY TABLE S1. Distribution of visual behavioral variables by age group.

Variable	Category	Total (N=29,525)	Aged 6–9 years (n=21,048)	Aged 10–14 years (n=8,477)
Daily outdoor activity time (hours/day)				
	<1	5,828 (19.74)	3,964 (18.83)	1,864 (21.99)
	1–2	11,444 (38.76)	8,193 (38.93)	3,251 (38.35)
	2–3	5,901 (19.99)	4,239 (20.14)	1,662 (19.61)
	≥3	6,352 (21.51)	4,652 (22.10)	1,700 (20.05)
Homework/reading time (hours/day)				
	No homework	1,091 (3.70)	940 (4.47)	151 (1.78)
	<1	8,557 (28.98)	6,612 (31.41)	1,945 (22.94)
	1–2	12,312 (41.70)	8,834 (41.97)	3,478 (41.03)
	2–3	5,135 (17.39)	3,233 (15.36)	1,902 (22.44)
	≥3	2,430 (8.23)	1,429 (6.79)	1,001 (11.81)
Tutoring duration (hours/week)				
	None	14,187 (48.05)	10,936 (51.96)	3,251 (38.35)
	<1	3,079 (10.43)	2,368 (11.25)	711 (8.39)
	1–2	6,231 (21.10)	4,249 (20.19)	1,982 (23.38)
	2–3	3,186 (10.79)	1,939 (9.21)	1,247 (14.71)
	≥3	2,842 (9.63)	1,556 (7.39)	1,286 (15.17)
Non-study screen time (hours/day)				
	None	5,908 (20.01)	4,591 (21.81)	1,317 (15.54)
	<0.5	11,285 (38.22)	8,155 (38.74)	3,130 (36.92)
	0.5–1	6,874 (23.28)	4,687 (22.27)	2,187 (25.80)
	1–2	3,876 (13.13)	2,635 (12.52)	1,241 (14.64)
	2–3	939 (3.18)	604 (2.87)	335 (3.95)
	≥3	643 (2.18)	376 (1.79)	267 (3.15)
Break location				
	Indoor	17,082 (57.86)	11,997 (57.00)	5,085 (59.99)
	Outdoor	12,443 (42.14)	9,051 (43.00)	3,392 (40.01)
Break frequency (outdoor)				
	Every break	16,306 (55.23)	11,876 (56.42)	4,430 (52.26)
	Every 2 breaks	8,252 (27.95)	5,690 (27.03)	2,562 (30.22)
	Rarely	4,400 (14.90)	3,089 (14.68)	1,311 (15.47)
	Never	567 (1.92)	393 (1.87)	174 (2.05)
Chest-to-desk distance				
	Never	2,283 (7.73)	1,545 (7.34)	738 (8.71)
	Occasionally	15,381 (52.09)	11,231 (53.36)	4,150 (48.96)
	Often	8,477 (28.71)	6,006 (28.53)	2,471 (29.15)
	Always	3,384 (11.46)	2,266 (10.77)	1,118 (13.19)
Eye-to-book distance				
	Never	2,079 (7.04)	1,465 (6.96)	614 (7.24)
	Occasionally	15,507 (52.52)	11,148 (52.96)	4,359 (51.42)
	Often	9,012 (30.52)	6,511 (30.93)	2,501 (29.50)
	Always	2,927 (9.91)	1,924 (9.14)	1,003 (11.83)

Continued

Variable	Category	Total (N=29,525)	Aged 6–9 years (n=21,048)	Aged 10–14 years (n=8,477)
Finger-to-pen distance	Never	2,268 (7.68)	1,504 (7.15)	764 (9.01)
	Occasionally	12,422 (42.07)	8,962 (42.58)	3,460 (40.82)
	Often	10,206 (34.57)	7,475 (35.51)	2,731 (32.22)
	Always	4,629 (15.68)	3,107 (14.76)	1,522 (17.95)
Lying posture	Never	14,010 (47.45)	10,181 (48.37)	3,829 (45.17)
	Occasionally	12,237 (41.45)	8,644 (41.07)	3,593 (42.39)
	Often	2,731 (9.25)	1,852 (8.80)	879 (10.37)
	Always	547 (1.85)	371 (1.76)	176 (2.08)
Head support	Never	15,590 (52.80)	11,593 (55.08)	3,997 (47.15)
	Occasionally	11,382 (38.55)	7,775 (36.94)	3,607 (42.55)
	Often	2,201 (7.45)	1,434 (6.81)	767 (9.05)
	Always	352 (1.19)	246 (1.17)	106 (1.25)
Sitting posture	Never	3,915 (13.26)	2,791 (13.26)	1,124 (13.26)
	Occasionally	10,337 (35.01)	7,417 (35.24)	2,920 (34.45)
	Often	10,266 (34.77)	7,413 (35.22)	2,853 (33.66)
	Always	5,007 (16.96)	3,427 (16.28)	1,580 (18.64)

Note: Data are presented as *n* (%). The percentages may not sum to 100 because of rounding.

SUPPLEMENTARY TABLE S2. Strength centrality rankings.

Rank	Node name	Node type	Strength
1	Eye-to-book distance	Posture	1.20
2	Head support	Posture	1.16
3	Lying posture	Posture	0.97
4	Chest-to-desk distance	Posture	0.92
5	Break location	Behavior	0.89
6	Spherical equivalent	Physiological	0.89
7	AL/CR ratio	Physiological	0.88
8	Finger-to-pen distance	Posture	0.83
9	Break frequency	Behavior	0.82
10	Sitting posture	Posture	0.51
11	Daily outdoor activity time	Behavior	0.50
12	Homework/reading time	Behavior	0.47
13	Tutoring duration	Behavior	0.44
14	Non-study screen time	Behavior	0.42
15	Parental myopia	Family	0.21

Note: Strength centrality indicates the sum of absolute edge weights connected to each node. Higher values denote greater influence in the network.

Abbreviation: AL/CR ratio=axial length-to-corneal radius ratio.

SUPPLEMENTARY TABLE S3. Bridge centrality rankings.

Rank	Node name	Node type	Bridge strength	Connected modules
1	Head support	Posture	0.41	Posture ↔ Physiological
2	Break location	Behavior	0.30	Behavior ↔ Physiological
3	Non-study screen time	Behavior	0.24	Behavior ↔ Physiological
4	Homework/reading time	Behavior	0.22	Behavior ↔ Physiological
5	Tutoring duration	Behavior	0.21	Behavior ↔ Physiological
6	Parental myopia	Family	0.21	Family ↔ Others
7	Sitting posture	Posture	0.19	Posture ↔ Physiological
8	Daily outdoor activity time	Behavior	0.19	Behavior ↔ Physiological
9	Lying posture	Posture	0.17	Posture ↔ Physiological
10	Chest-to-desk distance	Posture	0.16	Posture ↔ Physiological
11	Break frequency	Behavior	0.15	Behavior ↔ Physiological
12	Eye-to-book distance	Posture	0.13	Posture ↔ Physiological
13	Spherical equivalent	Physiological	0.13	Physiological ↔ Behavior/Posture
14	AL/CR ratio	Physiological	0.12	Physiological ↔ Behavior/Posture
15	Finger-to-pen distance	Posture	0.07	Posture ↔ Physiological

Note: Bridge centrality quantifies the sum of absolute edge weights connecting a given node to nodes in other modules. Higher values indicate stronger cross-module connectivity

Abbreviation:AL/CR ratio=axial length-to-corneal radius ratio.

Preplanned Studies

Occurrence and Risk Assessment of Typical Endocrine-Disrupting Chemicals in Surface and Drinking Water Across Four River Basins — Henan Province, China, 2025

Xie Wang¹; Qingqing Ma¹; Jie Zhang¹; Ruiling Liang¹; Yaoguang Zhang¹; Yongheng Su¹; Congke Zhang^{1,†}

Summary

What is already known about this topic?

Bisphenols and per- and polyfluoroalkyl substances (PFASs) are endocrine-disrupting chemicals found in aquatic ecosystems. Henan, a critical water conveyance region within the South-to-North Water Diversion Project, encompasses four river basins characterized by complex hydrological conditions. However, systematic monitoring of contaminant levels remains insufficient.

What is added by this report?

This study identified key priority pollutants and highlighted the dominance of short-chain PFASs in the region.

What are the implications for public health practice?

These baseline data address gaps in existing knowledge on emerging pollutants omitted from GB 5749-2022, thereby providing critical scientific evidence to guide the revision of national drinking water standards, strengthen watershed pollution control, and safeguard public health and ecotoxicological integrity.

ABSTRACT

Introduction: Bisphenols (BPs) and per- and polyfluoroalkyl substances (PFASs) in aquatic environments have attracted significant research attention. Assessing their abundance in surface and drinking water across the Yangtze, Yellow, Huai, and Hai River Basins in Henan Province is crucial for exploring exposure and toxicity-related risks, thereby guiding future mitigation strategies.

Methods: Surface and drinking water samples ($n=191$) were collected and analyzed in October 2025 for health and ecotoxicological risk assessments.

Results: Four BPs and 11 PFASs were detected in Henan Province, and bisphenol A (BPA) and short-chain PFASs were identified as the predominant pollutants. The Yangtze River Basin had the highest

average total concentrations of endocrine-disrupting chemicals (Σ EDCs). Hazard quotients were consistently higher for children than for adults. BPA poses a high aquatic ecological risk, whereas PFBS demonstrates a stronger potential for human non-carcinogenic risks through drinking water exposure.

Conclusion: Continuous monitoring of EDCs, such as BPA, in the Yangtze River Basin is essential for comprehensively evaluating their associated impacts on aquatic environments. Enhanced surveillance of priority pollutants in major river basins is recommended to safeguard drinking water for vulnerable populations, particularly children.

Endocrine-disrupting chemicals (EDCs), such as bisphenols (BPs) and per- and polyfluoroalkyl substances (PFASs), are exogenous organic compounds that interfere with hormonal physiology and endocrine system functions (1). Previous studies have highlighted varying EDC levels in drinking water across countries (2). However, existing research on BPs and PFASs in major river basins, such as the Yangtze (3) and Haihe (4) Rivers, mainly focuses on specific basins or compounds, without accounting for the water supply chain.

Henan Province spans the Yellow, Huai, Yangtze, and Hai River Basins. It is both an important water source and receiving region for the South-to-North Water Diversion Project. Nevertheless, no studies have documented the changing BP and PFAS concentrations during drinking water treatment across these basins. This study explored the concentrations of 4 BPs and 11 PFASs, and their associated risks, in 4 major river basins of Henan Province.

A total of 191 samples, including 46 source, 46 treated, 51 tap, and 48 surface water samples, were collected from 16 counties (cities and districts) across 4 river basins (Supplementary Table S1, available at

<https://weekly.chinacdc.cn>) in September–November 2025. Sampling sites were selected based on the hydrological characteristics of Henan Province, covering the upper, middle, and lower reaches of the basins. Sampling covered water sources, water treatment plants, pipe network terminals, and densely populated areas. Three surface, three source, three treated drinking water, and three tap water samples were collected from each administrative region.

Samples were analyzed at the Henan Center for Disease Control and Prevention using liquid chromatography–mass spectrometry, following the internal standard method. Blank and recovery experiments were performed for quality control. The Wilcoxon rank-sum test was used to assess differences among river basins. All plots were generated and analyzed using CNSknowall (<https://cnsknowall.com>), a web service for data analysis and visualization. EDC analysis plots were generated using this platform for data analysis and visualization.

Non-carcinogenic risks associated with EDCs were evaluated in both adults and children using the health risk assessment model recommended by the U.S. Environmental Protection Agency (USEPA) (5). The risk quotient (*RQ*) approach has been used to evaluate

the potential impacts of EDCs on aquatic organisms (6). Reference doses, acute toxicity parameters, and data are presented in Supplementary Tables S2 and S3 (available at <https://weekly.chinacdc.cn>).

Bisphenol AF (BPAF), bisphenol A (BPA), and perfluorobutane sulfonate (PFBS) were the most frequently detected EDCs, followed by perfluorobutanoic acid (PFBA) and bisphenol S (BPS). Detection rates for other EDCs remained below 20%. BPAF, BPA, PFBS, PFBA, and BPS concentrations ranged from not detected (nd) to 6.4 ng/L, nd to 646.0 ng/L, nd to 57.1 ng/L, nd to 27.0 ng/L, and nd to 17.8 ng/L, respectively. The Anderson–Darling test used to evaluate the normality of the concentration datasets for the 15 EDCs demonstrated that the data followed non-normal distributions (Table 1). In contrast, perfluorohexanoic acid, perfluorononanoic acid, perfluorodecanoic acid, and perfluoroheptanoic acid were not detected in any samples. BPA was the predominant pollutant among BPs, whereas PFBS and PFBA, both short-chain PFASs, were the dominant congeners.

The Yangtze River Basin exhibited the highest average Σ EDCs values, followed by the Huai and Hai River Basins. The lowest average Σ EDC concentration

TABLE 1. Detection rates and concentrations of endocrine-disrupting chemicals in the collected water samples in China, 2025.

EDCs		Detection rate (%)	Mean (ng/L)	Max (ng/L)	P25 (ng/L)	Median (ng/L)	P75 (ng/L)	P90 (ng/L)
BPs	BPA	52.2	20.5	646.0	0.45	1.15	8.28	43.0
	BPS	25.6	0.7	17.8	0.1	0.1	0.153	1.046
	BPAF	57.0	0.3	6.4	0.008	0.056	0.267	0.822
	TBBPA	9.7	0.6	49.2	0.05	0.05	0.05	0.05
PFASs	PFBA	28.0	2.0	27.0	0.9	0.9	3.36	9.62
	PFPeA	0.5	1.1	41.1	0.9	0.9	0.9	0.9
	PFHxA	0	nd	nd	nd	nd	nd	nd
	PFHpA	0	nd	nd	nd	nd	nd	nd
	PFOA	3.9	2.5	53.1	2.5	2.5	2.5	2.5
	PFNA	0	nd	nd	nd	nd	nd	nd
	PFDA	0	nd	nd	nd	nd	nd	nd
	PFBS	37.2	1.5	57.1	0.1	0.1	1.02	3.14
	PFHxS	6.8	0.2	3.4	0.1	0.1	0.1	0.1
	PFOS	15.5	0.6	6.6	0.35	0.35	0.35	1.23
	PFHpS	1.4	0.1	0.8	0.05	0.05	0.05	0.05

Note: Concentrations of undetected samples were assigned as half the limit of detection.

Abbreviation: BPs=bisphenols; PFASs=per - and polyfluoroalkyl substances; BPA=bisphenol A; BPS=bisphenol S; BPAF=bisphenol AF; TBBPA=tetrabromobisphenol A; PFBA=perfluorobutanoic acid; PFPeA=perfluoropentanoic acid; PFHxA=perfluorohexanoic acid; PFHpA=perfluoroheptanoic acid; PFOA=perfluorooctanoic acid; PFNA=perfluorononanoic acid; PFDA=perfluorodecanoic acid; PFBS=perfluorobutane sulfonate; PFHxS=perfluorohexanesulfonic acid; PFOS=perfluorooctanesulfonic acid; PFHpS=perfluoroheptanesulfonic acid; nd=not detected.

was observed in the Yellow River Basin (Figure 1A). The BPA, BPS, BPAF, and PFBA concentrations in the Yangtze River basin were significantly higher than those in the Hai River Basin ($P<0.05$). Moreover, significantly higher concentrations of BPA and PFBA were observed in the Yangtze River basin than in the Huai River basin ($P<0.05$). In contrast, PFBS concentrations did not vary significantly among the

four basins ($P>0.05$) (Figure 1B).

The Σ BP concentrations in surface water samples showed the widest range (nd–687.0 ng/L), exceeding all other samples. The Σ BP concentrations were markedly lower in both treated and tap water samples. The Σ PFAS concentrations in surface water samples ranged from nd to 85.4 ng/L. The concentrations of BPs and PFASs in surface water were significantly

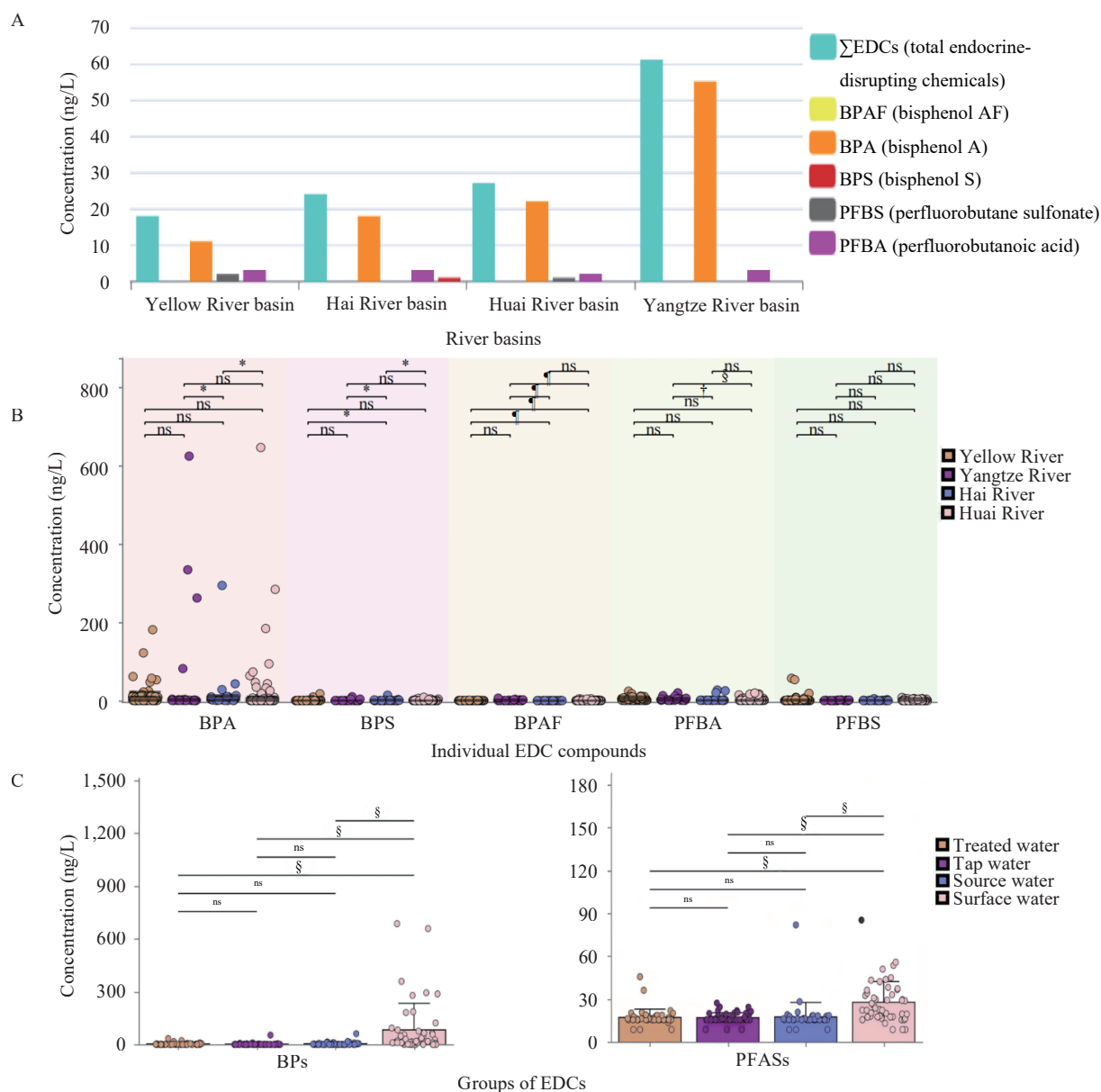


FIGURE 1. EDCs concentrations in four major river basins in Henan Province, 2025. (A) Concentrations of EDCs in different river basins; (B) Distribution of individual EDCs among river basins; (C) Concentrations of different EDC groups in different water sample types.

Note: ns represents no statistically significant differences.

Abbreviation: EDCs=endocrine-disrupting chemicals; BPA=bisphenol A; BPS=bisphenol S; BPAF=bisphenol AF; PFBA=perfluorobutanoic acid; PFBS=perfluorobutane sulfonate.

* $P<0.05$, † $P<0.01$, § $P<0.005$, ¶ $P<0.001$; ($P>0.05$).

higher than those in source, treated, and tap water ($P<0.005$) (Figure 1C). These findings suggest that BPs and PFASs exhibit pronounced environmental persistence, reducing the efficacy of conventional drinking water treatment.

Due to the absence of USEPA cancer slope factors, this investigation focused exclusively on non-carcinogenic risk assessments. The hazard quotients (HQs) of the five EDCs (BPA, BPS, BPAF, PFBA, and PFBS) with detection frequencies $>20\%$ were calculated (Figure 2A). The HQs of all EDCs were below 1, with a maximum of 1.9×10^{-2} , suggesting that EDCs are unlikely to pose non-carcinogenic risks. The most serious non-carcinogenic risk was observed in the Yellow River Basin, where the maximum HQs reached

5.4×10^{-3} for adults and 1.9×10^{-2} for children, corresponding to a PFBS concentration of 57.1 ng/L. Moreover, the maximum HQs in the Yangtze River Basin reached 6.0×10^{-4} and 2.0×10^{-3} for adults and children, respectively, corresponding to a PFBS concentration of 20.0 ng/L. Conversely, the associated human health risks were relatively low in the Huai and Hai River Basins. Across all basins, HQs were consistently higher in children than adults, indicating greater vulnerability to five EDCs. Importantly, all HQs for adults and children were far below the threshold of 1.0, suggesting negligible health risks associated with EDCs in the studied basins.

The distributions of the ecotoxicological RQ values associated with the five dominant EDCs in the surface

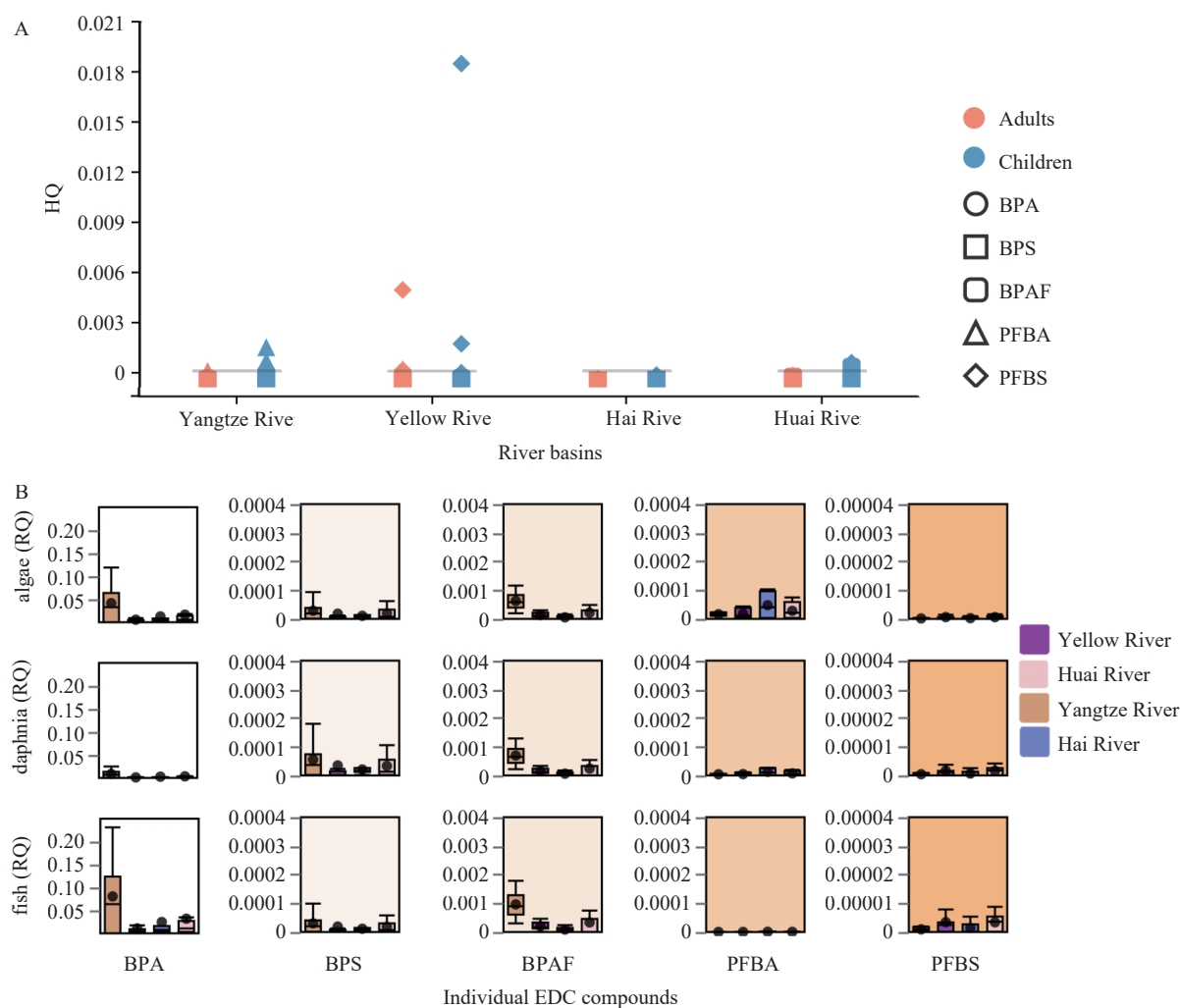


FIGURE 2. Risk assessment results of five EDCs in China, 2025. (A) Non-carcinogenic risk assessment results of five EDCs in drinking water for children and adults in different river basins; (B) Ecotoxicological risk assessment results of five EDCs in surface water at three trophic levels.

Abbreviation: EDCs=endocrine-disrupting chemicals; BPA=bisphenol A; BPS=bisphenol S; BPAF=bisphenol AF; PFBA=perfluorobutanoic acid; PFBS=perfluorobutane sulfonate.

waters of four river basins were determined across three trophic levels (algae, *Daphnia*, and fish) (Figure 2B). *RQs* associated with BPs exceeded those associated with PFASs. Within the Yangtze River, the highest average *RQs* for BPA, BPS, and BPAF reached 4.3×10^{-2} , 3.8×10^{-5} , and 7.6×10^{-4} , respectively. Across all trophic levels, PFBA consistently exhibited higher ecotoxicological *RQs* than PFBS. The Haihe River Basin exhibited the highest average PFBA-associated *RQ*, reaching 1.0×10^{-4} , surpassing other basins. Tanghe County (a city in the Yangtze River Basin) had the highest BPA-associated *RQs*, 0.010 to 0.231. The remaining EDCs showed relatively low *RQs* (<1) suggesting that the ecotoxicological risks of surface water EDCs are acceptable.

DISCUSSION

This is the first comprehensive study of 15 EDCs coupled with a dual health–ecological assessment across Henan Province. The data were highly representative and covered the entire study area. The investigated BPs and PFASs represent the major pollutants threatening the source waters of the South-to-North Water Diversion Project. The findings provide critical evidence for developing and strengthening protective measures to ensure the Project's long-term sustainability.

The results identified short-chain PFAS (e.g., PFBS and PFBA) and BPA as the dominant pollutants in the study region. Notably, BPA presents higher ecological risks than PFBS, whereas PFBS poses greater human non-carcinogenic health risks through drinking water exposure. Industrial activities along river basins generate copious BPs and PFASs (7). The open burning of dumped waste and landfill leachate can release EDCs contained in plastic waste into the surrounding environment (8), whereas food packaging and domestic sewage are stable sources of EDCs (9).

PFASs are chemically stable, highly recalcitrant to degradation, and highly mobile over long distances, making them difficult to eliminate with conventional water treatment (10). Although conventional treatments can effectively eliminate BPs, source reduction remains essential for reducing treatment loads because source control is inherently more efficient than terminal remediation. Therefore, the advancement of green material substitution coupled with the enforcement of stricter emission controls in EDC-related industries is strongly recommended to alleviate ecological and public health risks. As a critical

water-supply basin for the South-to-North Water Diversion Project, local authorities should strictly supervise industrial emissions and incorporate short-chain PFAS and BPA into routine monitoring protocols.

This study confirmed that children face significantly greater health risks than adults. Henan Province should prioritize addressing the hazards of prolonged low-dose exposure in children. Moreover, efforts to disseminate water safety knowledge should be intensified, with an emphasis on educating residents and safeguarding children. Inspecting water quality in schools and kindergartens should be prioritized.

Under Drinking Water Quality Standards (GB 5749-2022), the maximum permissible concentrations are 10.0 µg/L for BPA, 0.08 µg/L for PFOA, and 0.04 µg/L for PFOS. However, no domestic regulatory thresholds have been established for other EDCs. By contrast, in April 2024 the USEPA promulgated markedly stricter limits at 4.0 ng/L for both PFOA and PFOS, and 1.0 ng/L for PFBS. However, no federal limits have been established for BPA levels in drinking water.

The findings in this report are subject to one limitation. Because sampling was restricted to a single season, data from different periods could not be obtained, preventing the assessment of seasonal fluctuations in pollutant concentrations and potentially introducing bias.

As for recommendations, stricter pollution control measures are imperative to safeguard drinking water sources in the Yangtze River Basin and long-term monitoring of emerging pollutants is necessary. Additional protection should be prioritized for vulnerable groups, particularly children. Collectively, these measures can substantially improve drinking water safety across provinces.

Conflicts of interest: No conflicts of interest.

Acknowledgements: The authors thank all researchers involved in this study and all participants who assisted in completing this investigation.

Ethical statement: This study did not involve human subjects, animal experiments, informed consent, private data, or any other content requiring an ethical review. The authors declare that the research was conducted in compliance with academic norms and that the data were authentic and reliable.

Funding: This work was supported by the Autonomous Research Project of the Chinese Center for Disease Control and Prevention (No.gs2025002) and the Medical Scientific Research Project Plan of Henan Province (No.LHGJ20250134).

doi: 10.46234/ccdcw2026.181

Corresponding author: Congke Zhang, zhangckhenancdc@163.com.

¹ Henan Center for Disease Control and Prevention (Henan Academy of Preventive Medicine), Zhengzhou City, Henan Province, China.

Copyright © 2026 by Chinese Center for Disease Control and Prevention & Chinese Academy of Preventive Medicine. All content is distributed under a Creative Commons Attribution Non Commercial License 4.0 (CC BY-NC).

Submitted: May 28, 2026

Accepted: August 12, 2026

Issued: September 04, 2026

REFERENCES

1. Zaborowska M, Wyszowska J, Borowik A, Kucharski J. Bisphenols-A threat to the natural environment. *Materials (Basel)* 2023;16(19):6500. <https://doi.org/10.3390/ma16196500>.
2. Lazofsky A, Buckley B. Recent trends in multiclass analysis of emerging endocrine disrupting contaminants (EDCs) in drinking water. *Molecules* 2022;27(24):8835. <https://doi.org/10.3390/molecules27248835>.
3. Li WY, Ma WX, Chang JJ, Xiao Y, Wang R, Zhu ZL, et al. Distribution characteristics and environmental risk assessment of bisphenol analogues in the lower reaches of the Yangtze River. *Environ Sci Process Impacts* 2025;27(8):2298 – 306. <https://doi.org/10.1039/d5em00294j>.
4. Li YN, Niu ZG, Zhang Y. Occurrence of legacy and emerging poly- and perfluoroalkyl substances in water: A case study in Tianjin (China). *Chemosphere* 2022;287(Pt 4):132409. <http://dx.doi.org/10.1016/j.chemosphere.2021.132409>.
5. USEPA. Risk assessment guidance for superfund. Volume I : Human health evaluation manual (Part A), interim final. Washington (WA): Office of Emergency and Remedial Response; 1989 Dec. Report No.: EPA/540/1-89/002. https://www.epa.gov/sites/production/files/2015-09/documents/rags_a.pdf.
6. van der Hoeven N. Current issues in statistics and models for ecotoxicological risk assessment. *Acta Biotheor* 2004;52(3):201 – 17. <https://doi.org/10.1023/B:ACBI.0000043442.05899.8c>.
7. Carlson GL, Andersen M. Tracking environmental contamination from multiple sources of per- and polyfluoroalkyl substances. *Environ Res* 2025;276:121470. <https://doi.org/10.1016/j.envres.2025.121470>.
8. Abraham A, Chakraborty P. A review on sources and health impacts of bisphenol A. *Rev Environ Health* 2020;35(2):201 – 10. <https://doi.org/10.1515/reveh-2019-0034>.
9. Zhang YJ, Ling Y, Zhang Y, Zhang F. Research progress on pretreatment and detection of bisphenols in food and environmental samples. *Chin J Chromatogr* 2019;37(12):1268 – 74. <https://doi.org/10.3724/SP.J.1123.2019.08002>.
10. Qian ZX, Feng C, Chen YH, Lin YJ, Liang ZW, Qian HL, et al. Occurrence, sources, and prioritization of per- and polyfluoroalkyl substances (PFASs) in drinking water from Yangtze River Delta, China: focusing on emerging PFASs. *Molecules* 2025;30(11):2313. <https://doi.org/10.3390/molecules3011231>.

SUPPLEMENTARY MATERIALS

Instrument Parameters

Mass spectral analysis was conducted using a QTRAP 6500 linear ion trap quadrupole LC-MS/MS spectrometer (AB SCIEX, USA), equipped with an ESI probe. The LC-MS/MS system equipped with a 30A system (SHIMADZU, JAPAN) used a Waters ACQUITY UPLCTM BEH C₁₈ column (100 mm × 2.1 mm, 1.7 μm).

Formulas for Risk Assessment

Human Health Risk Assessment: The carcinogenic and non-carcinogenic risks of EDCs for adults and children were quantified using Equations 1 and 2 based on the health risk assessment model recommended by the USEPA (1). The hazard quotient (*HQ*) was applied to assess the non-carcinogenic risk to humans via the direct ingestion of surface water used as drinking water, whereas alternative exposure pathways (e.g., dermal contact) were not considered.

$$HQ = CDI / RfD \quad (1)$$

where *CDI* denotes chronic daily intake, representing the estimated human exposure to contaminants through ingestion [mg/(kg.day)] and was calculated using Equation 2; *RfD* is the reference dose [mg/(kg.day)], defined as the oral exposure threshold for a given contaminant, and was obtained from the Regional Screening Level Summary Table of the USEPA:

$$CDI_i = \frac{C \times IR_i \times EF_i \times ED_i}{BW_i \times AT} \quad (2)$$

where *C* denotes the measured concentration of contaminants in each water sample (mg/L). For samples with concentrations below the limit of detection (LOD), values were assigned as one-half LOD for subsequent calculation; *IR_i* represents the water ingestion rate (1 L/day and 2 L/day for children and adults, respectively) (L/day); *EF_i* is the exposure frequency (365 days/year for both ages); *ED_i* represents exposure duration (6 years for children and 70 years for adults), *BW_i* denotes body weight of the exposed person (10 kg for children and 70 kg for adults), and *AT* indicates average lifespan (2,190 and 25,550 days for children and adults, respectively). An *HQ* greater than 1 indicates a non-tolerable exposure level, suggesting that the compound in question is likely to cause adverse effects. Conversely, when the *HQ* < 1, adverse effects are unlikely.

Ecotoxicological Risk Assessment: The risk quotient (*RQ*) was employed to assess the potential impacts of EDCs on aquatic organisms and to predict the extent of the harm they pose to both aquatic organisms and humans (2). Ecotoxicological risk assessments were conducted using Equations 3 and 4. In this study, the ecological *RQs* of 15 EDCs were calculated across 3 categories of aquatic organisms (algae, *Daphnia*, and fish) based on 48 surface water samples.

$$RQ = MEC / PNEC \quad (3)$$

$$PNEC = EC_{50} / AF \quad (4)$$

where *MEC* denotes the measured environmental concentration, mg/L; *PNEC* represents the predicted no-effect concentration of the substance, mg/L; *EC₅₀* denotes the median effect concentration of the substance on organisms, obtained through a literature review, mg/L; and *AF* is the assessment factor, adopting the recommended value of 1,000 specified in the EU Water Framework Directive. The risk assessment was divided into four levels: minor risk (*RQ* ≤ 0.01), low risk (0.01 < *RQ* ≤ 0.1), moderate risk (0.1 < *RQ* ≤ 0.5), and high risk (*RQ* > 0.5).

SUPPLEMENTARY TABLE S1. Average concentrations of endocrine-disrupting chemicals in the different administrative regions of the study area in China, 2025.

River Basins	Administrative Region	Mean concentration (ng/L)
Yellow River Basin	Lingbao	27.0
	Gongyi	16.6
	Wuzhi	4.5
	Luoning	5.4
	Pingyuan	10.2
	Taiqian	39.2
Huai River Basin	Lankao	5.7
	Suiping	33.5
	Chuanhui	18.2
	Shenqiu	17.4
	Wuyang	2.5
	Liangyuan	78.6
Yangtze River Basin	Xichuan	10.2
	Tanghe	103.2
Hai River Basin	Linzhou	34.0
	Junxian	11.0

SUPPLEMENTARY TABLE S2. Reference doses (*RfD*s) of five endocrine-disrupting chemicals for human health risk calculation.

EDCs	CAS	<i>RfD</i> [mg/(kg·d)]	Sources
BPAF	1478-61-1	0.0006	EPA Database-Calibrated Assessment Product (DCAP)
BPA	80-05-7	0.05	EPA Integrated Risk Information System (IRIS)
PFBS	375-73-5	0.0003	EPA Provisional Peer-Reviewed Toxicity Values (PPRTVs)
PFBA	375-22-4	0.001	EPA IRIS
BPS	80-09-1	0.0059	EPA DCAP

Note: All *RfD* values were retrieved from the publicly available USEPA IRIS, DCAP, and PPRTVs online databases for 2026.

Abbreviation: BPA=bisphenol A; BPS=bisphenol S; BPAF=bisphenol AF; PFBA=perfluorobutanoic acid; PFBS=perfluorobutane sulfonate; *RfD*=reference dose; EDCs=endocrine-disrupting chemicals.

SUPPLEMENTARY TABLE S3. Acute toxicity data (*EC*₅₀) of five EDCs and corresponding references.

EDCs	Algae	<i>EC</i> ₅₀ (mg/L) <i>Daphnia</i>	Fish	References
BPA	5.25	25.1	2.7	(3–5)
BPS	106	55	100.1	(6–8)
BPAF	3.0	2.7	1.95	(9–11)
PFBA	262.2	1,006	13,795	(12–14)
PFBS	5,023	1937	921.9	(15–16,12)

Abbreviation: *EC*₅₀=half-maximal effective concentration; BPA=bisphenol A; BPS=bisphenol S; BPAF=bisphenol AF; PFBA=perfluorobutanoic acid; PFBS=perfluorobutane sulfonate.

REFERENCES

- USEPA. Risk Assessment Guidance for Superfund. Volume I : Human Health Evaluation Manual (Part A), Interim Final, Office of Emergency and Remedial Response. USEPA, Washington, D.C. 20450. https://www.epa.gov/sites/production/files/2015-09/documents/rags_a.pdf.
- Van der HN. Current issues in statistics and models for ecotoxicological risk assessment. *Acta Biotheor* 2004;52(3):201-217. <https://doi.org/10.1023/B:ACBI.0000043442.05899.8c>.
- Alexander HC, Dill DC, Smith LW, Guiney PD, Dorn P. Bisphenol A: acute aquatic toxicity. *Environ Toxicol Chem* 1988;7(1):19 – 26. <https://doi.org/10.1002/etc.5620070104>.

4. Plahuta M, Tişler T, Pintar A, Toman MJ. Adverse effects of Bisphenol A on water louse (*Asellus aquaticus*). *Ecotoxicol Environ Saf* 2015;117:81 – 8. <https://doi.org/10.1016/j.ecoenv.2015.03.031>.
5. Chan WK, Chan KM. Disruption of the hypothalamic-pituitary-thyroid axis in zebrafish embryo-larvae following waterborne exposure to BDE-47, TBBPA and BPA. *Aquat Toxicol* 2012;108:106 – 11. <https://doi.org/10.1016/j.aquatox.2011.10.013>.
6. European Chemicals Agency (ECHA). Disseminated REACH registration dossier: Section 6.1.5 Toxicity to aquatic algae and cyanobacteria(OECD Test Guideline 201); ECHA CHEM Database, European Chemicals Agency, Helsinki, Finland. CAS number 80-09-1 https://chem.echa.europa.eu/100.001.137/dossier-view/2c405b30-c0bd-497a-944a-1754da992ca0/IUC5-3cd624e4-cd2b-4b41-b35d-8d3a668f0756_17930586-6ec6-404a-94e8-7b1afb013123?searchText=80-09-1.
7. European Chemicals Agency (ECHA). Disseminated REACH registration dossier:Section 6.1.3 Short-term toxicity to aquatic invertebrates(OECD Test Guideline 202); ECHA CHEM Database, European Chemicals Agency, Helsinki, Finland. CAS number 80-09-1 https://chem.echa.europa.eu/100.001.137/dossier-view/2c405b30-c0bd-497a-944a-1754da992ca0/IUC5-82d1cab7-3555-43e6-87ea-606c36cc8951_17930586-6ec6-404a-94e8-7b1afb013123?searchText=80-09-1.
8. Blanc M, Rüegg J, Scherbak N, Keiter SH. Environmental chemicals differentially affect epigenetic-related mechanisms in the zebrafish liver (ZF-L) cell line and in zebrafish embryos. *Aquat Toxicol* 2019;215:105272. <https://doi.org/10.1016/j.aquatox.2019.105272>.
9. European Chemicals Agency (ECHA). Disseminated REACH registration dossier:Section 6.1.5 Toxicity to aquatic algae and cyanobacteria(ISO 8692); ECHA CHEM Database, European Chemicals Agency, Helsinki, Finland. CAS number 1478-61-1 https://chem.echa.europa.eu/100.014.579/dossier-view/d20a1a9a-5686-4996-9180-0c9d1aa0af6c/c5099731-8d40-4165-addd-4bae56e313a3_6feef51f-ed0d-4bc7-91c3-9928993fc28c?searchText=1478-61-1.
10. European Chemicals Agency (ECHA). Disseminated REACH registration dossier:Section 6.1.3 Short-term toxicity to aquatic invertebrates(ISO 6341); ECHA CHEM Database, European Chemicals Agency, Helsinki, Finland. CAS number 1478-61-1 https://chem.echa.europa.eu/100.014.579/dossier-view/d20a1a9a-5686-4996-9180-0c9d1aa0af6c/a850c831-08ee-40cb-81bd-33869409a461_6feef51f-ed0d-4bc7-91c3-9928993fc28c?searchText=1478-61-1.
11. Mu XY, Huang Y, Li XX, Lei YL, Teng MM, Li XF, et al. Developmental effects and Estrogenicity of Bisphenol A alternatives in a zebrafish embryo model. *Environ Sci Technol* 2018;52(5):3222 – 31. <https://doi.org/10.1021/acs.est.7b06255>.
12. Zheng ZY, Zhou JX, Peng ZX, Meng Y, Yuan WB, Sun Q, et al. Conflicting results of ecological and health risk assessment of perfluorinated compounds in major river basins in China. *J Environ Sci* 2026;162:112 – 20. <https://doi.org/10.1016/j.jes.2025.06.013>.
13. Boudreau TM. Toxicity of Perfluorinated organic acids to selected freshwater organisms under laboratory and field conditions [dissertation]. Ontario, Canada: University of Guelph; 2002. <https://atrium.lib.uoguelph.ca/items/b2d039c9-f298-4276-8992-6666f10b0ef7>.
14. Godfrey A, Abdel-Moneim A, Sepúlveda MS. Acute mixture toxicity of halogenated chemicals and their next generation counterparts on zebrafish embryos. *Chemosphere* 2017;181:710 – 2. <https://doi.org/10.1016/j.chemosphere.2017.04.146>.
15. European Chemicals Agency (ECHA). Disseminated REACH registration dossier:Section 6.1.5 Toxicity to aquatic algae and cyanobacteria(OECD Test Guideline 201); ECHA CHEM Database, European Chemicals Agency, Helsinki, Finland. CAS number375-73-5 https://chem.echa.europa.eu/100.006.176/dossier-view/f6fdbdac-3a36-4f42-b79e-a0877dc562d2/eee89050-c737-42ba-9ac9-091b2deb0dfd_145202ac-df61-4c7b-8052-c730d0084be8?searchText=375-73-5.
16. European Chemicals Agency (ECHA). Disseminated REACH registration dossier:Section 6.1.3 Short-term toxicity to aquatic invertebrates(OECD Test Guideline 202); ECHA CHEM Database, European Chemicals Agency, Helsinki, Finland. CAS number375-73-5 https://chem.echa.europa.eu/100.006.176/dossier-view/f6fdbdac-3a36-4f42-b79e-a0877dc562d2/ab1f445b-3622-4369-a4be-7a0076bcd763_145202ac-df61-4c7b-8052-c730d0084be8?searchText=375-73-5.

Methods and Applications

A Hierarchical Multi-Stage Machine-Learning Study on Predicting Oil and Salt in Chinese Restaurant Dishes for Nutrition Surveillance — 6 Survey Sites, China, 2019–2020

Ying Xing¹; Jiguo Zhang¹; Xiaofang Jia¹; Chang Su¹; Wenwen Du^{1,†}; Huijun Wang¹; Aidong Liu¹

ABSTRACT

Introduction: The rise in dining out has created a demand for scalable methods for estimating the oil, salt, and sodium content in restaurant meals. These preparation-dependent components are challenging to measure consistently.

Methods: A hierarchical multi-stage prediction model was developed to estimate five dish-level targets: added cooking oil, salt, salt-containing seasonings, sodium from salt-containing seasonings, and total sodium from salt and seasonings in 8,455 Chinese dishes from 192 restaurants across six survey sites in China. The model performance was evaluated using standard regression metrics and task-specific indicators, including zero-use identification, tolerance-based accuracy, multiplicative consistency, symmetric mean absolute percentage error (SMAPE), Spearman's rank correlation, and discrete intake-level classification.

Results: The stage-wise light gradient-boosting machine (LightGBM) model outperformed the linear regression, random forest, and multi-task LightGBM models for all five targets. The log-scale coefficient of determination (R^2) values ranged from 0.224 for salt to 0.570 for salt-containing seasonings. On the original scale, strict tolerance-based accuracy was limited, with 22%–37% of the predictions meeting the tolerance band. Multiplicative and rank-based performances were stronger: 40% to 73% of the predictions were within two-fold of the observed value, and the Spearman correlations ranged from 0.486 to 0.771. The predicted means were lower than the observed means, indicating a conservative estimation and limited suitability for uncalibrated mean estimation.

Conclusion: The model can support the preliminary estimation, relative comparison, and screening of restaurant dishes for nutrition surveillance. Therefore, it should not be interpreted as a precise dish-level measurement tool without further calibration.

Oil, salt, and sodium are key targets for dietary surveillance because their excess intake is associated with obesity, hypertension, and cardiovascular disease (1–3). Surveillance increasingly requires scalable estimates from restaurant dishes; however, direct weighing and laboratory analysis are difficult to scale (4). Existing nutrition estimation methods use food images, labels, recipes, or ingredient data to infer intrinsic nutrients (5–8). The added cooking oil, added salt, and salt-containing seasonings are different; they are preparation-dependent quantities that vary with cooking methods, chef habits, customer requests, recipe standardization, pre-prepared sauces, condiment brands, batch preparation, and kitchen practices. This study developed a hierarchical multistage machine learning framework to evaluate whether routinely collected dish information can support the preliminary estimation and comparison of these targets for nutrition surveillance.

METHODS

Data Sources and Study Targets

Data were obtained from 8,455 restaurant dishes collected during 2019–2020 from 192 restaurants across 6 survey sites selected to cover geographic and culinary diversity: Shijiazhuang, Hebei Province, North China; Xining, Qinghai Province, Northwest China; Daqing, Heilongjiang Province, Northeast China; Changsha, Hunan Province, Central China; Chengdu, Sichuan Province, Southwest China; and Nanchang, Jiangxi Province, East/Southeast China. A total of 2 districts/counties and 32 restaurants were surveyed in each province.

Trained staff collected dish names, ingredients, ingredient weights, cooking methods, and restaurant characteristics through interviews with chefs and on-

site investigations. Identically named dishes from different restaurants were retained as separate records. Five targets were predicted: added cooking oil, salt-containing seasonings, sodium in salt-containing seasonings, salt, and total sodium from added salt and salt-containing seasonings

Data Preprocessing and Feature Description

All targets were log1p-transformed before model fitting, and observations above the 99.5th percentile for each target were excluded. Predictors included three feature blocks: ingredient composition, including total ingredient weight, main ingredient weight, and 65 food-category weight variables; cooking method, including 6 major categories and 22 subtypes; and restaurant-level characteristics, including survey site, north-south region, and restaurant size. The feature definitions are shown in Table 1, and the food category and cooking method classifications are shown in Supplementary Table S1 (available at <https://weekly.chinacdc.cn/>).

Model Development and Design

The primary aim of this study was to predict five dish-level targets: added cooking oil, salt-containing seasonings, sodium in salt-containing seasonings, salt, and total sodium from seasonings and salt. These outcomes are related in both culinary practice and

empirical data, and the pairwise Spearman correlations among them suggested that upstream information could be informative for downstream prediction (Figure 1A). We therefore used a stage-wise framework with five sequential stages (Figure 1B). Stage 1 predicted added cooking oil from the original dish features. Stage 2 predicted salt-containing seasonings using the same features plus the stage 1 output. Stage 3 predicted sodium in salt-containing seasonings using the updated feature set. Stage 4 predicted salt using the original features plus upstream stage outputs. Stage 5 predicted total sodium from seasonings and salt using all previously accumulated stage outputs.

Four models were evaluated to compare different modeling strategies. Linear regression and random forest were implemented as alternative learners using the same stage-wise prediction logic. The multi-task light gradient-boosting machine (LightGBM) model predicted all five targets simultaneously from the same original input features without stage-to-stage feedback. The stage-wise LightGBM model predicted the five targets sequentially and passed predictions forward from one stage to the next.

Model Evaluation and Performance Metrics

Model performance was evaluated using the coefficient of determination (R^2), mean absolute error (MAE), and root mean squared error (RMSE) on a

TABLE 1. Feature dictionary for model input variables.

Feature block	Variable name(s)	Definition	Coding	Source/derivation
Ingredient composition	Total ingredient weight	Total recorded weight of the dish	Continuous (g)	Calculated by summing all recorded ingredient weights
	Main ingredient weight	Weight of the dominant ingredient in the dish	Continuous (g)	Calculated from raw ingredient-level records
	65 food-category weight variables	Dish-level total weight assigned to each mapped food category after manual ingredient-to-category mapping	Continuous (g), 0 if absent	Raw item-level ingredients were mapped to the 65-category system and aggregated at the dish level; complete list shown in Table S1
Cooking method	6 major cooking-category indicators	Major cooking category of the dish	Binary indicator (0/1)	Derived from the original coarse cooking-method records and standardized according to the study classification
	22 cooking-subtype indicators	Finer-grained cooking subtype of the dish	Binary indicator (0/1)	Manually refined from original cooking-method information according to the study classification scheme
Restaurant/site level	6 survey-site indicator variables	Survey-site location of the restaurant	Binary indicator (0/1)	Directly derived from recorded restaurant location
	North-south regional indicator	Whether the restaurant was located in southern China	Binary indicator (0/1)	Derived from survey-site location
	Restaurant size indicators	Restaurant scale category (large, medium, small)	Binary indicator (0/1)	Directly recorded in field survey

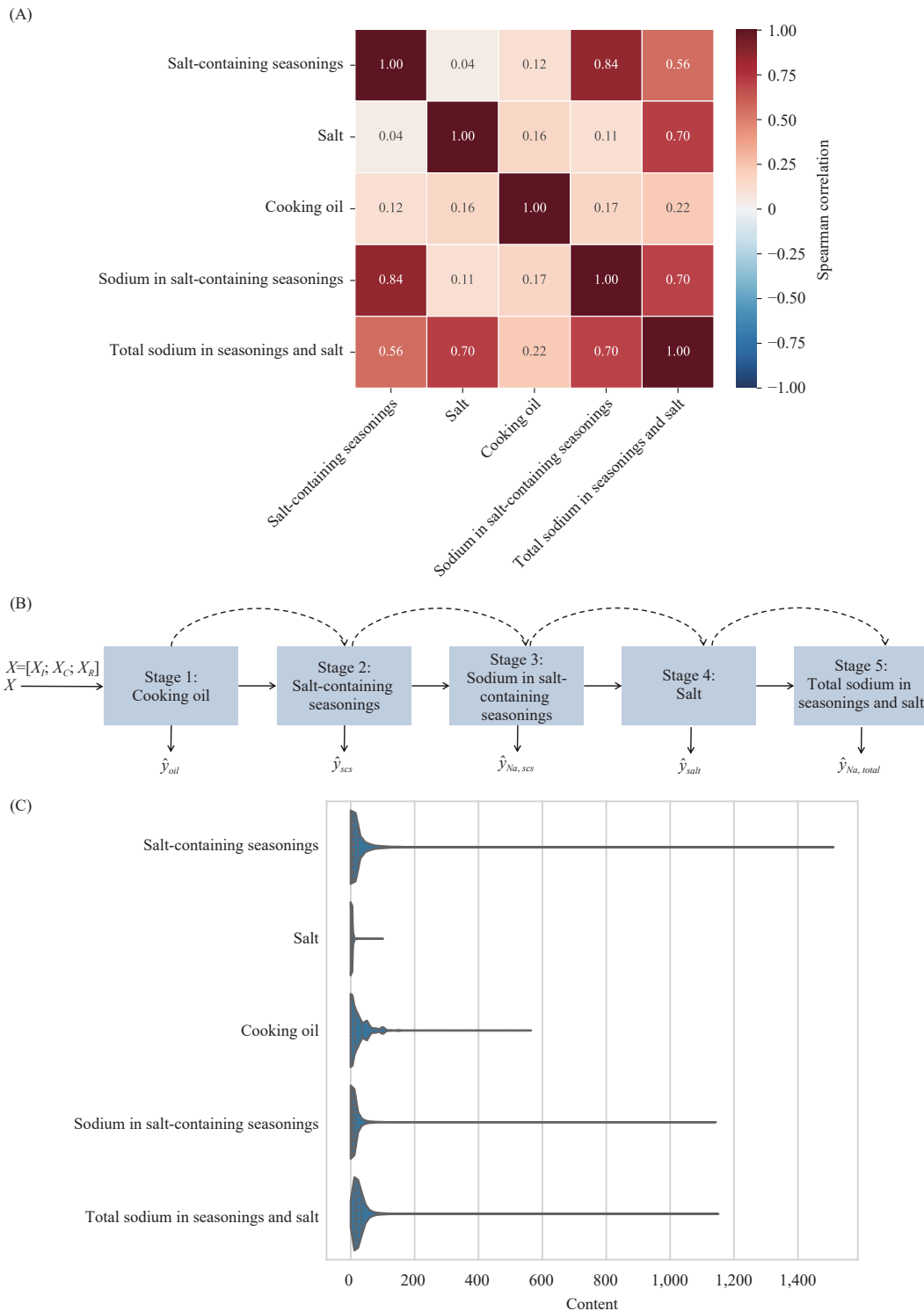


FIGURE 1. Target variable correlations, model framework, and distribution characteristics for predicting oil, salt, and sodium in Chinese restaurant dishes. (A) Spearman correlation heatmap of the five target variables. (B) Hierarchical multi-stage prediction framework. (C) Distribution characteristics of target variables.

Note: For (A), correlations were calculated among added cooking oil, salt-containing seasonings, sodium in salt-containing seasonings, added salt, and total sodium from seasonings and salt. For (B), original dish features from ingredient composition, cooking method, and restaurant/site-level feature blocks were used as initial inputs. The model predicted the five targets sequentially, and each upstream prediction was added to the input of the next stage. Solid arrows indicate forward prediction flow, and dashed arrows indicate out-of-fold predictions used during training to prevent information leakage. For (C), distributions of the five target variables are shown.

log-transformed scale. As the primary goal of this study was to support the rapid estimation of oil, salt, and sodium intake in large-scale nutrition surveillance or surveys, task-specific metrics were also calculated on the original scale, including zero-use identification, tolerance-based accuracy, multiplicative consistency, symmetric mean absolute percentage error (SMAPE), Spearman rank correlation, and discrete intake-level classification. Definitions of these metrics are provided in Supplementary Table S2 (available at <https://weekly.chinacdc.cn/>).

Statistical Analysis

The dataset was split 80/20 into training and test sets. Five-fold cross-validation was applied to the training set. To prevent information leakage, out-of-fold upstream predictions were used as downstream training inputs and the test-set upstream predictions were passed forward during the final evaluation. Subgroup performance was evaluated according to survey site, north-south region, and major cooking-method category. Feature-group contributions were assessed using grouped Shapley additive explanations (SHAP) values for ingredient composition, cooking method, and restaurant-level features.

All analyses were conducted in Python 3.11 using pandas, Numpy, scikit-learn, and LightGBM.

RESULTS

Data Description

The full dataset covered six major cooking categories (Supplementary Table S1). Mean values per dish were 24.72 g for cooking oil, 21.79 g for salt-containing seasonings, 3.46 g for salt, 1,275.57 mg for sodium in salt-containing seasonings, and 2,663.71 mg for total sodium. The targets were right-skewed and heavy-tailed, consistent with Figure 1C. Of the 8,455 records, 3,729 (44.1%) shared dish names, including 987 repeated dish names. Same-name dishes still varied substantially: median within-name ranges were 27.0 g for cooking oil, 13.5 g for salt-containing seasonings, and 3.0 g for added salt.

Model Results

Statistical metrics on log scale On a log-transformed scale, the stage-wise LightGBM model performed the best for all five targets (Table 2). The R^2 values were 0.550 for cooking oil, 0.570 for salt-containing seasonings, 0.454 for sodium in salt-containing

seasonings, 0.224 for added salt, and 0.394 for total sodium from seasonings and salt. Subgroup analyses by survey site, north-south region, and major cooking-method category generally supported the stability of the results, although added salt remained the weakest target (Supplementary Table S3, available at <https://weekly.chinacdc.cn/>).

Grouped SHAP analysis showed that ingredient-composition, cooking-method, and restaurant-level features contributed 49.8%, 37.3%, and 12.9% for cooking oil; 52.2%, 37.1%, and 10.7% for salt-containing seasonings; 58.8%, 17.9%, and 23.3% for sodium in salt-containing seasonings; 40.8%, 14.8%, and 44.4% for added salt; and 69.8%, 16.4%, and 13.8% for total sodium, respectively.

Task-specific metrics on original scale The predicted means and standard deviations were lower than the observed values for all five targets (Table 3), indicating a conservative shrinkage of high values. Thus, raw predictions should be interpreted as preliminary reference values rather than exact dish-level measurements, and calibration is required for absolute mean intake estimation. On the original scale, BandAcc ranged from 0.217 to 0.374, WIF@2× from 0.403 to 0.729, and Spearman's ρ from 0.486 to 0.771, indicating better relative ranking than precise gram-level agreement (Supplementary Table S2).

DISCUSSION

This study evaluated whether routinely collected restaurant dish information could support the estimation of added oil, salt, salt-containing seasonings, and sodium-related outcomes for nutrition surveillance. Stage-wise LightGBM outperformed the comparison models and showed generally stable performance across survey sites, regions, and cooking-method categories. However, the model should be interpreted as a surveillance support tool for preliminary estimation and comparison, and not as a substitute for direct measurement.

This task differs from conventional nutrient estimation. Intrinsic nutrients can often be estimated from ingredient weights and food composition tables (5–10), whereas added oil, salt, and salt-containing seasonings depend on preparation practices. Chef habits, customer requests, recipe standardization, pre-prepared sauces, condiment brands, batch preparation, and kitchen practices are difficult to measure consistently. The repeated-dish analysis supports this

TABLE 2. Test-set performance of four regression models*.

Target	Model	R ²	MAE	RMSE
Cooking oil	Linear regression	0.432	0.878	1.226
	Random forest	0.438	0.989	1.225
	LightGBM multi-task	0.527	0.860	1.124
	LightGBM stage-wise (proposed) [†]	0.550	0.831	1.069
Salt-containing seasonings	Linear regression	0.527	0.664	0.870
	Random forest	0.393	0.762	0.994
	LightGBM multi-task	0.544	0.656	0.862
	LightGBM stage-wise (proposed) [†]	0.570	0.632	0.813
Sodium in salt-containing seasonings	Linear regression	0.310	0.662	0.925
	Random forest	0.323	0.727	0.931
	LightGBM multi-task	0.426	0.655	0.858
	LightGBM stage-wise (proposed) [†]	0.454	0.624	0.806
Salt	Linear regression	0.142	0.557	0.711
	Random forest	0.175	0.547	0.704
	LightGBM multi-task	0.165	0.539	0.708
	LightGBM stage-wise (proposed) [†]	0.224	0.518	0.666
Total sodium in seasonings and salt	Linear regression	0.303	0.528	0.746
	Random forest	0.288	0.553	0.790
	LightGBM multi-task	0.327	0.545	0.768
	LightGBM stage-wise (proposed) [†]	0.394	0.506	0.707

Abbreviation: R²=coefficient of determination; MAE=mean absolute error; RMSE=root mean squared error.

* All outcomes were log-transformed.

[†] indicates the primary model used in this study.

TABLE 3. Observed and predicted mean values and standard deviations of the five target variables in the test set.

Target	Observed mean	Observed SD	Predicted mean	Predicted SD	Mean difference (Pred - Obs)
Cooking oil (g)	24.84	34.84	17.13	20.18	-7.70
Salt-containing seasonings (g)	21.63	41.35	14.31	16.01	-7.32
Sodium in salt-containing seasonings (mg)	1,301.05	2,487.01	888.15	886.21	-412.90
Salt (g)	3.55	4.59	2.65	1.35	-0.89
Total sodium in seasonings and salt (mg)	2,745.68	3,633.57	2,162.92	1,267.24	-582.75

Abbreviation: SD=standard deviation; Pred=predicted; Obs=observed.

point: identical dish names do not necessarily represent standardized recipes or stable seasoning amounts.

The feature-contribution results also support the use of multiple information sources. Ingredient composition captured much of the dish-level structure, cooking methods captured preparation-related variation, and restaurant-level variables captured part of geographic or operational variation, especially for added salt. Restaurant size was included as a predictor; however, it is only a coarse indicator and cannot

directly measure recipe standardization or chef-level practices.

The model's public health value lies in its screening and prioritization. In large-scale monitoring, direct measurement of restaurant dishes is often impractical. Approximate model outputs can help field investigators identify dishes, cooking methods, regions, or restaurant types that are likely to have high oil or sodium content and prioritize them for further assessment or intervention. Raw predictions should not be used for

menu labeling, regulatory enforcement, individual dietary assessment, or uncalibrated estimation of the population mean intake.

The findings in this report are subject to some limitations. First, although the dataset covers diverse survey sites, it was not designed to generate province-specific estimates. Second, the target values derived from chef interviews and on-site investigations may contain measurement errors. Third, routinely collected predictors did not include condiment brands, chef-specific habits, customer requests, or detailed cooking process information. Fourth, external validation and calibration are required before broader application.

In conclusion, routinely collected dish information can support the preliminary estimation and relative comparison of oil, salt, and sodium in Chinese restaurant dishes, while revealing clear limits for precise gram-level prediction.

Conflicts of interest: No conflicts of interest.

Funding: Supported by Public Health Emergency Project Nutrition Health and Healthy Diet Campaign (No. 102393220020070000012).

doi: 10.46234/ccdcw2026.184

* Corresponding author: Wenwen Du, duww@ninh.chinacdc.cn.

¹ The National Institute for Nutrition and Health, Chinese Center for Disease Control and Prevention & Chinese Academy of Preventive Medicine, Beijing, China.

Copyright © 2026 by Chinese Center for Disease Control and Prevention & Chinese Academy of Preventive Medicine. All content is distributed under a Creative Commons Attribution Non Commercial License 4.0 (CC BY-NC).

Submitted: January 27, 2026

Accepted: May 22, 2026

Issued: September 04, 2026

REFERENCES

1. Allison SJ. High salt intake as a driver of obesity. *Nat Rev Nephrol* 2018;14(5):285. <https://doi.org/10.1038/nrneph.2018.23>.
2. Watso JC, Fancher IS, Gomez DH, Hutchison ZJ, Gutiérrez OM, Robinson AT. The damaging duo: obesity and excess dietary salt contribute to hypertension and cardiovascular disease. *Obes Rev* 2023;24(8):e13589. <https://doi.org/10.1111/obr.13589>.
3. Zhou L, Stamler J, Chan Q, Van Horn L, Daviglus ML, Dyer AR, et al. Salt intake and prevalence of overweight/obesity in Japan, China, the United Kingdom, and the United States: the INTERMAP Study. *Am J Clin Nutr* 2019;110(1):34 – 40. <https://doi.org/10.1093/ajcn/nqz067>.
4. McLean RM. Measuring population sodium intake: a review of methods. *Nutrients* 2014;6(11):4651 – 62. <https://doi.org/10.3390/nu6114651>.
5. Yunus R, Arif O, Afzal H, Amjad MF, Abbas H, Bokhari HN, et al. A framework to estimate the nutritional value of food in real time using deep learning techniques. *IEEE Access* 2019;7:2643 – 52. <https://doi.org/10.1109/ACCESS.2018.2879117>.
6. Lei ZF, ul Haq A, Dorrahi M, Zhang DF, Abbott D. Composing recipes based on nutrients in food in a machine learning context. *Neurocomputing* 2020;415:382 – 96. <https://doi.org/10.1016/j.neucom.2020.08.071>.
7. Shen ZD, Shehzad A, Chen S, Sun H, Liu J. Machine learning based approach on food recognition and nutrition estimation. *Procedia Comput Sci* 2020;174:448 – 53. <https://doi.org/10.1016/j.procs.2020.06.113>.
8. Al-Saffar M, Baiee W. Nutrition information estimation from food photos using machine learning based on multiple datasets. *Bull Electr Eng Inform* 2022;11(5):2922 – 9. <https://doi.org/10.11591/eei.v11i5.4007>.
9. Ispirova G, Eftimov T, Koroušić Seljak B. P-NUT: predicting nutrient content from short text descriptions. *Mathematics* 2020;8(10):1811. <https://doi.org/10.3390/math8101811>.
10. Li JT, Han FD, Guerrero R, Pavlovic V. Picture-to-amount (PITA): predicting relative ingredient amounts from food images. In: *Proceedings of the 2020 25th international conference on pattern recognition*. Milan, Italy: IEEE. 2021; p. 10343-50. <http://dx.doi.org/10.1109/icpr48806.2021.9412828>.

SUPPLEMENTARY MATERIAL

SUPPLEMENTARY TABLE S1. Food-category mapping and hierarchical cooking-method classification used in feature construction.

Classification domain	Category	Subcategories	Frequency in the dataset (n)*
Cooking categories	Cold dishes and cold mixed noodles	Light cold dish type, Heavily sauced cold dish type, Spicy and oily cold dish type, Oily cold dish type, Mildly seasoned staple dish type, Mildly seasoned cold dish type	1,539
	Braised, marinated, and sugar-based preparations	Traditional braised type, Composite sauce-braised type, Grilled and sauced braised type	407
	Steamed and boiled dishes	Plain steamed type, Sauced steamed type	662
	High-heat stir-frying	Home-style stir-fried type, Dry-pot stir-fried type, Sauce-rich stir-fried type, Spicy stir-fried type	3,353
	Moist-heat cooking	Clear-soup braised type, Richly seasoned braised type, Hotpot base-braised type, Heavy-oil braised type	1,689
	Deep-fried, pan-fried, and grilled/roasted dishes	Plain-fried or grilled type, Marinated fried or grilled type, Sauced fried or grilled type	805
Food categories	Eggs and egg products	Quail eggs, goose eggs, chicken eggs, duck eggs	610
	Beverages	Popsicles and ice cream, tea and tea beverages, powdered beverages, fruit juices and fruit drinks, milk beverages, carbonated beverages, plant-protein beverages	69
	Fish, shrimp, crab and shellfish	Shellfish, shrimp, crab, fish	1,173
	Dried legumes and legume products	Broad beans, adzuki beans, soybeans, mung beans, kidney beans	770
	Vegetables and vegetable products	Allium vegetables, root vegetables, tender-stem-leafy and flowering vegetables, fruiting and gourd vegetables, aquatic vegetables, fresh legumes, wild vegetables	14,473
	Cereals and cereal products	Barley, rice, wheat, millet and yellow millet, corn	576
	Snacks and sweet baked goods	Cakes and desserts, snacks	93
	Tubers, starches and related products	Starches	985
	Fats and oils (excluding added cooking oil)	Animal fats	362
	Poultry and poultry products	Goose meat, chicken, duck	666
	Alcoholic beverages	Fermented alcoholic beverages, distilled spirits	117
	Sugars, preserved fruits, candied fruits and honey	Honey, preserved and candied fruits, candy	74
	Fruits and fruit products	Citrus fruits, melons and stone fruits, berries, tropical and subtropical fruits, pome fruits	231
	Livestock meat and meat products	Donkey meat, beef, lamb and mutton, pork	3,790
	Milk and dairy products	Milk powder, cheese, cream, yogurt, liquid milk	28
	Nuts and seeds	Tree nuts, seeds	445
	Condiments and seasonings (seasonings not included in this study)	Pickled vegetables, herbs and spices	1,751

* For cooking categories, *n* denotes the number of dishes assigned to that category; for food categories, *n* denotes the number of occurrences of that category across dishes.

Interpretation of Task-Specific Performance Metrics in Supplementary Table S2

The task-specific metrics in Supplementary Table S2 were used to complement the standard log-scale regression metrics by evaluating model performance on the original scale. These metrics were intended to assess whether the predictions could support preliminary estimation, relative comparison, and screening in nutrition surveillance rather

SUPPLEMENTARY TABLE S2. Task-specific performance metrics of the stage-wise LightGBM model*.

Target	ZeroAcc@0.1	ZeroPrec@0.1	ZeroRec@0.1	ZeroF1@0.1	BandAcc@abs 0.5+20%†	EpsMAE@abs 0.5+20%†	WIF@1.5×	WIF@2×	SMAPE	Spearman ρ	ClsAcc (bins)	MacroF1 (bins)
Cooking oil	0.086	0.867	0.086	0.157	0.217	8.141	0.285	0.403	1.010	0.767	0.286	0.254
Salt- containing seasonings	0.021	1.000	0.021	0.041	0.258	6.596	0.370	0.555	0.739	0.771	0.350	0.257
Sodium in salt- containing seasonings	0.040	1.000	0.040	0.077	0.279	3.666	0.350	0.539	0.770	0.667	0.325	0.253
Salt	0.016	0.750	0.016	0.031	0.374	0.893	0.366	0.547	0.841	0.486	0.412	0.241
Total sodium in seasonings and salt	0.034	1.000	0.034	0.067	0.324	5.591	0.508	0.729	0.517	0.595	0.469	0.239

Abbreviation: ZeroAcc@0.1=zero-use accuracy at threshold 0.1; ZeroPrec@0.1=zero-use precision at threshold 0.1; ZeroRec@0.1=zero-use recall at threshold 0.1; ZeroF1@0.1=zero-use F1 score at threshold 0.1; BandAcc=tolerance-based band accuracy; EpsMAE=epsilon mean absolute error beyond tolerance band; WIF=within interval factor; SMAPE=symmetric mean absolute percentage error; ClsAcc=classification accuracy; MacroF1=macro F1 score.

* All metrics were calculated using the original scale. The values for cooking oil, salt-containing seasonings, and added salt are in grams, and sodium-related quantities are in milligrams, except for dimensionless metrics.

† For the abs0.5+20% criterion, the average tolerance bands on the original scale were 5.44 g for salt-containing seasonings, 305 mg for sodium in salt-containing seasonings, 1.19 g for added salt, and 583 mg for total sodium in seasonings and salt. The EpsMAE reports the mean excess absolute error beyond this band.

SUPPLEMENTARY TABLE S3. Subgroup performance of the stage-wise LightGBM model by survey site, cooking method, and north-south region.

Subgroup	Category	n	Target	R ² *	MAE*	RMSE*	Spearman's ρ
site	Shijiazhuang	258	Cooking oil (g)	0.490	0.766	0.975	0.743
			Salt-containing seasonings (g)	0.584	0.630	0.816	0.778
			Sodium from salt-containing seasonings (mg)	0.428	0.682	0.848	0.616
			Salt (g)	0.128	0.438	0.586	0.319
	Xining	282	Total sodium in seasonings and salt (mg)	0.471	0.450	0.591	0.548
			Cooking oil (g)	0.386	0.939	1.141	0.555
			Salt-containing seasonings (g)	0.496	0.574	0.765	0.727
			Sodium in salt-containing seasonings (mg)	0.357	0.621	0.817	0.559
	Nanchang	277	Salt (g)	0.081	0.617	0.782	0.274
			Total sodium in seasonings and salt (mg)	0.323	0.655	0.880	0.456
			Cooking oil (g)	0.465	0.805	1.063	0.765
			Salt-containing seasonings (g)	0.446	0.616	0.853	0.709
	Chengdu	300	Sodium in salt-containing seasonings (mg)	0.412	0.544	0.765	0.643
			Salt (g)	0.033	0.356	0.543	0.135
			Total sodium in seasonings and salt (mg)	0.177	0.336	0.528	0.573
			Cooking oil (g)	0.519	1.040	1.303	0.680
	Daqing	294	Salt-containing seasonings (g)	0.523	0.664	0.872	0.709
			Sodium in salt-containing seasonings (mg)	0.351	0.683	0.900	0.562
			Salt (g)	0.078	0.663	0.792	0.335
			Total sodium in seasonings and salt (mg)	0.253	0.661	0.901	0.461
	Changsha	280	Cooking oil (g)	0.523	0.828	1.087	0.739
			Salt-containing seasonings (g)	0.623	0.650	0.825	0.798
			Sodium in salt-containing seasonings (mg)	0.510	0.632	0.822	0.705
			Salt (g)	0.160	0.562	0.717	0.435
Region	North	834	Total sodium in seasonings and salt (mg)	0.380	0.511	0.739	0.594
			Cooking oil (g)	0.491	0.815	1.068	0.738
			Salt-containing seasonings (g)	0.473	0.784	0.986	0.678
			Sodium in salt-containing seasonings (mg)	0.368	0.754	0.938	0.604
	South	857	Salt (g)	0.101	0.479	0.633	0.369
			Total sodium in seasonings and salt (mg)	0.321	0.509	0.720	0.588
			Cooking oil (g)	0.541	0.846	1.073	0.757
			Salt-containing seasonings (g)	0.584	0.618	0.803	0.780
			Sodium in salt-containing seasonings (mg)	0.441	0.644	0.829	0.640
			Salt (g)	0.146	0.542	0.703	0.385
			Total sodium in seasonings and salt (mg)	0.387	0.541	0.750	0.546

Continued

Subgroup	Category	n	Target	R ² *	MAE*	RMSE*	Spearman's ρ
Cooking method	High-heat stir-frying	689	Cooking oil (g)	0.554	0.792	1.003	0.821
			Salt-containing seasonings (g)	0.485	0.657	0.860	0.709
			Sodium in salt-containing seasonings (mg)	0.400	0.638	0.821	0.603
			Salt (g)	0.243	0.466	0.631	0.518
	Moist-heat cooking	328	Total sodium in seasonings and salt (mg)	0.307	0.464	0.658	0.567
			Cooking oil (g)	0.504	0.892	1.166	0.710
			Salt-containing seasonings (g)	0.655	0.517	0.706	0.857
			Sodium in salt-containing seasonings (mg)	0.525	0.523	0.722	0.743
	Cold dishes and cold mixed noodles	292	Salt (g)	0.199	0.529	0.690	0.490
			Total sodium in seasonings and salt (mg)	0.255	0.472	0.665	0.558
			Cooking oil (g)	0.582	0.780	0.984	0.672
			Salt-containing seasonings (g)	0.475	0.735	0.974	0.695
	Deep-Fried, Pan-Fried, and Grilled/Roasted Dishes	148	Sodium in salt-containing seasonings (mg)	0.325	0.732	0.959	0.563
			Salt (g)	0.200	0.556	0.732	0.479
			Total sodium in seasonings and salt (mg)	0.421	0.610	0.873	0.543
			Cooking oil (g)	0.245	1.079	1.435	0.567
	Steamed and Boiled Dishes	140	Salt-containing seasonings (g)	0.617	0.741	0.899	0.802
			Sodium in salt-containing seasonings (mg)	0.495	0.742	0.916	0.718
			Salt (g)	0.215	0.598	0.740	0.503
			Total sodium in seasonings and salt (mg)	0.422	0.659	0.889	0.609
	Braised, Marinated, and Sugar-Based Preparations	94	Cooking oil (g)	0.185	1.036	1.307	0.455
			Salt-containing seasonings (g)	0.622	0.655	0.810	0.789
			Sodium in salt-containing seasonings (mg)	0.499	0.622	0.803	0.706
			Salt (g)	0.058	0.558	0.701	0.346
			Total sodium in seasonings and salt (mg)	0.271	0.554	0.742	0.581
			Cooking oil (g)	0.502	1.043	1.202	0.733
			Salt-containing seasonings (g)	0.485	0.713	0.907	0.680
			Sodium in salt-containing seasonings (mg)	0.302	0.868	1.057	0.515
			Salt (g)	0.217	0.641	0.781	0.498
			Total sodium in seasonings and salt (mg)	0.352	0.609	0.894	0.604

Abbreviation: R²=coefficient of determination; MAE=mean absolute error; RMSE=root mean squared error; ρ=Spearman's rank correlation coefficient.

* All outcomes are on the log-transformed scale.

than to claim precise dish-level measurements.

Zero-use detection (ZeroAcc, ZeroPrec, ZeroRec, ZeroF1). Zero-use metrics evaluate whether the model correctly identifies dishes with negligible or no use of the target ingredient or seasoning. Across the five targets, zero-use recall was low, but precision was high, indicating that the model rarely predicted exact zero values; however, when it did, such predictions were usually correct. This suggests that the model tends to assign small positive values, rather than strictly zero values. Therefore, predictions should not be used as definitive indicators of nonuse.

Tolerance-based accuracy (BandAcc and EpsMAE). BandAcc was calculated using a strict tolerance band defined as absolute error ≤ 0.5 plus 20% of the true value on the original scale. The BandAcc ranged from 0.217 to 0.374 across the targets, indicating limited agreement under this strict criterion. The EpsMAE further quantified the

average excess error beyond the tolerance band. These results show that precise gram- or milligram-level prediction remains difficult, especially for dishes with large variabilities in oil, salt, or seasoning use.

Multiplicative agreement and ranking (WIF@1.5×/WIF@2× and Spearman's ρ). WIF@1.5× and WIF@2× assessed whether the predicted values remained within a specified multiplicative range of the observed values. Spearman's ρ assessed whether the model preserved rank ordering across dishes. Compared with strict tolerance-based accuracy, multiplicative and rank-based performance was stronger, with WIF@2× ranging from 0.403 to 0.729 and Spearman's ρ ranging from 0.486 to 0.771. These findings indicate that the model captured broad relative differences and approximated magnitudes better than the exact dish-level amounts.

Bin-based classification metrics (ClsAcc and MacroF1). ClsAcc and MacroF1 were used to evaluate whether the predicted and observed values fell within the same predefined intake categories. The moderate classification results suggest that the model partly preserved the intake level categories, but the category agreement was not sufficient for the precise classification of individual dishes.

Overall interpretation. Taken together, these metrics indicate that the model provides useful preliminary reference estimates for nutrition surveillance, particularly for comparing dishes and identifying those likely to have relatively high oil, salt, or sodium content. However, this model should not be interpreted as a substitute for direct weighing, recipe dissection, or laboratory analysis. Raw predictions are more appropriate for surveillance-oriented screening and prioritization than for precise dish-level gram measurements or uncalibrated estimations of absolute mean intake.

Youth Editorial Board

Director Lei Zhou

Vice Directors Jue Liu Tiantian Li Tianmu Chen

Members of Youth Editorial Board

Jingwen Ai	Li Bai	Yuhai Bi	Yunlong Cao
Liangliang Cui	Meng Gao	Jie Gong	Yuehua Hu
Jia Huang	Xiang Huo	Xiaolin Jiang	Yu Ju
Min Kang	Huihui Kong	Lingcai Kong	Shengjie Lai
Fangfang Li	Jingxin Li	Huigang Liang	Di Liu
Jun Liu	Li Liu	Yang Liu	Chao Ma
Yang Pan	Zhixing Peng	Menbao Qian	Tian Qin
Shuhui Song	Kun Su	Song Tang	Bin Wang
Jingyuan Wang	Linghang Wang	Qihui Wang	Xiaoli Wang
Xin Wang	Feixue Wei	Yongyue Wei	Zhiqiang Wu
Meng Xiao	Tian Xiao	Wuxiang Xie	Lei Xu
Lin Yang	Canqing Yu	Lin Zeng	Yi Zhang
Yang Zhao	Hong Zhou		

Indexed by Science Citation Index Expanded (SCIE), Social Sciences Citation Index (SSCI), PubMed Central (PMC), Scopus, Chinese Scientific and Technical Papers and Citations, and Chinese Science Citation Database (CSCD)

Copyright © 2026 by Chinese Center for Disease Control and Prevention & Chinese Academy of Preventive Medicine

Under the terms of the Creative Commons Attribution-Non Commercial License 4.0 (CC BY-NC), it is permissible to download, share, remix, transform, and build upon the work provided it is properly cited. The work cannot be used commercially without permission from the journal.

References to non-China-CDC sites on the Internet are provided as a service to *CCDC Weekly* readers and do not constitute or imply endorsement of these organizations or their programs by China CDC or National Health Commission of the People's Republic of China. China CDC is not responsible for the content of non-China-CDC sites.

The inauguration of *China CDC Weekly* is in part supported by Project for Enhancing International Impact of China STM Journals Category D (PIIJ2-D-04-(2018)) of China Association for Science and Technology (CAST).

CHINA CDC WEEKLY



中国疾病预防控制中心周报 (英文)

Responsible Authority

National Disease Control and Prevention Administration

Sponsor

Chinese Center for Disease Control and Prevention &
Chinese Academy of Preventive Medicine

Editor-in-Chief

Jianwei Wang

Editing and Publishing

China CDC Weekly Editorial Office
No.155 Changbai Road, Changping District, Beijing, China
Tel: 86-10-63150501, 63150701
Email: weekly@chinacdc.cn

Printing: Beijing Kexin Printing Co., Ltd

Complimentary Access

CSSN

ISSN 2096-7071 (Print)

ISSN 2097-3101 (Online)

CN 10-1629/R1