Received: 13 February 2025 | Revised: 14 April 2025 | Accepted: 22 April 2025

DOI: 10.1002/alz.70282

Alzheimer's & Dementia®
THE JOURNAL OF THE ALZHEIMER'S ASSOCIATION

RESEARCH ARTICLE

NDDRF 2.0: An update and expansion of risk factor knowledge base for personalized prevention of neurodegenerative diseases

Cheng Bi[1] | **Xin Zheng**[1] | **Yuxin Zhang**[1] | **Shengrong Zhou**[2] | **Jie Song**[1] | **Huifang Shang**[3] | **Bairong Shen**[1]

[1]Department of Neurology and Institutes for Systems Genetics, Frontiers Science Center for Disease-related Molecular Network, West China Hospital, Sichuan University, Chengdu, China

[2]Center for Systems Biology, Soochow University, Suzhou, China

[3]Department of Neurology, West China Hospital, Sichuan University, Chengdu, China

Correspondence
Bairong Shen, Department of Neurology and Institutes for Systems Genetics, Frontiers Science Center for Disease-related Molecular Network, West China Hospital, Sichuan University, No. 2222 Xinchuan Road, Gaoxin District, Chengdu 610213, Sichuan Province, China.
Email: bairong.shen@scu.edu.cn

Funding information
National Natural Science Foundation of China, Grant/Award Number: 32270790; Sichuan Science and Technology Program, Grant/Award Number: 2024YFHZ0205

Abstract

INTRODUCTION: Neurodegenerative diseases (NDDs) are chronic diseases caused by brain neuron degeneration, requiring systematic integration of risk factors to address their heterogeneity. Established in 2021, Knowledgebase of Risk Factors for Neurodegenerative Diseases (NDDRF) was the first knowledge base to consolidate NDD risk factors. NDDRF 2.0 expands focus to modifiable lifestyle-related factors, enhancing utility for NDD prevention.

METHODS: Data from the past 4 years were comprehensively updated, while lifestyle factors were manually collected and filtered from 1975 to 2024. Each factor was embedded with International Classification of Diseases codes and clinical stage annotations, and then re-standardized, classified, and annotated in accordance with the Unified Medical Language System Semantic Network.

RESULTS: NDDRF 2.0 encompasses 1971 risk factors classified under 151 subcategories across 20 NDDs, including 536 lifestyle-related factors covering six major categories and is freely accessible at http://sysbio.org.cn/NDDRF/.

DISCUSSION: As the first lifestyle-specific and holistic knowledge base for NDDs, NDDRF 2.0 offers structured and deep phenotype information, enabling personalized prevention strategies and clinical decision support.

KEYWORDS
diagnosis and prevention, knowledge base, lifestyle, neurodegenerative diseases, protective factor, risk factor

Highlights

- An enhanced lifestyle-specific and holistic knowledge base (Knowledgebase of Risk Factors for Neurodegenerative Diseases [NDDRF] 2.0) was built for neurodegenerative diseases (NDDs).
- NDDRF 2.0 provides detailed categorization and deep phenotypes to support targeted NDD prevention.

- NDDRF 2.0 provides a knowledge-driven resource that facilitates personalized risk assessment and proactive health management.
- NDDRF 2.0 provides clinicians, researchers, and at-risk populations with knowledge to develop and implement effective risk prevention strategies.
- NDDRF 2.0 can be used to build chatbots by enhancing large language models in the future.

1 | BACKGROUND

Neurodegenerative diseases (NDDs) represent a series of chronic diseases triggered by the degeneration of neurons in the brain, each with distinct pathophysiological mechanisms.[1–3] In the mid to late stages of these diseases, some patients exhibit cognitive dysfunction (e.g., Alzheimer's disease [AD]),[4,5] while others develop motor impairments (e.g., Parkinson's disease [PD]).[6] Globally, the incidence and mortality rates of NDDs, including AD and PD, are rising.[7,8] In the United States, the number of individuals aged ≥ 65 with AD is expected to increase from 5.8 million in 2021[9] to 6.7 million in 2023[10,11]; in China, the total number of individuals > 65 with PD has already surpassed 2.5 million, with ≈ 100,000 new cases diagnosed annually.[12] The pathological processes of NDDs are often irreversible, and currently, there are no effective treatments to halt disease progression. Available treatments can only delay the progression of these diseases. Therefore, the preclinical phase of NDDs is the most valuable and effective stage for initiating neuroprotective interventions and control.[13,14]

Research into disease risk factors has demonstrated significant potential in preventing and slowing the progression of NDDs. Mounting evidence strongly endorses the Lancet Commission on Dementia Prevention, Intervention, and Care's identification of 14 modifiable risk factors for dementia. These factors have the potential to prevent or delay as many as 40% of cases.[15–17] Currently, the number of research articles on AD risk factors continues to grow annually on PubMed, surpassing 8700 publications, underscoring the importance of AD risk factor research. As the process of screening and querying NDD risk factors is time consuming and prone to inaccuracies for researchers and clinicians, we launched the Knowledgebase of Risk Factors for Neurodegenerative Diseases (NDDRF) in 2021 to facilitate the integration of heterogeneous risk factor data for NDDs. This knowledge base is the inaugural resource solely dedicated to risk factors associated with NDDs, concentrating on biochemical, epidemiological, genetic, and combined factors. It furnishes users with precise and thorough risk factor data meticulously curated and sifted through by our team of experts.[18,19]

Presently, medical research paradigms are evolving toward precision medicine and intelligent health care, emphasizing a shift from the conventional disease-centered model to health-centric approaches.[20,21] The unique variations in the root causes of specific NDDs are shaped not just by genetic components but also by non-genetic factors.[22] Lifestyle factors, unlike genetic and environmental influences, are more malleable, rendering them potent tools in disease prevention and health enhancement.[23,24] Notably, in certain regions, the age-specific occurrence of dementia has decreased, a trend likely attributable to enhanced lifestyle adjustments.[15,16,25] Research indicates that more than one third of dementia cases could be averted through lifestyle modifications.[15] Furthermore, delaying the onset of symptoms by a mere 5 years could potentially decrease the number of cases by 41%.[26] For individuals > 65, maintaining a comprehensive healthy lifestyle score—NEURO (nutrition, exercise, unwind, restorative sleep, optimizing social and mental activity)—can significantly reduce the risk of AD.[27,28]

Enhancing lifestyles not only plays a pivotal role in the prevention of NDDs but also conserves health-care resources, embodying a superior health-care strategy.[23] In this study, we introduce an enhanced iteration, NDDRF 2.0, which integrates categorized and structured modifiable lifestyle risk factors. This version is an extension and update that incorporates newly reported additional factors. The knowledge base has now broadened to include 1971 risk factors for 20 NDDs, comprising a total of 4379 records. The enhancements in NDDRF 2.0 encompass the re-standardization of risk factor types, the inclusion of intricate patient phenotype-related risk factors, and additional details like International Classification of Diseases (ICD) codes, clinical stage annotations, single nucleotide polymorphism (SNP) identifiers (IDs), and preconditions for risk factors. These improvements are complemented by a revamped website, enriched data visualization, and refined search functionalities.

2 | METHODS

2.1 | Data inclusion and exclusion criteria

Expanding on the inclusion and exclusion criteria specified in NDDRF 1.0,[18] we have made enhancements to now integrate lifestyle factors as a key risk factor while refining the standardization and precision of data selection.

Articles were included if they investigated risk factors and provided clear, definitive conclusions regarding their impact. Examples include:

1. Harmful lifestyle factor: Statements such as "is not a healthy lifestyle for AD," "has harmful effects on AD," and so forth.
2. Beneficial lifestyle factor: Statements such as "is a healthy lifestyle for AD," "healthy lifestyle reduces disease progression in AD," "lifestyle protects against AD," and so forth.

Articles were excluded if they did not focus on risk factor studies or lacked definitive conclusions regarding their impact. Examples of exclusion criteria include statements such as "is not a lifestyle factor for AD," "a specific lifestyle has no effect on AD," and so forth.

2.2 | Data sources update

As described in version 1.0,[18] information on risk factors for NDDs was meticulously gathered and curated from published journal articles on PubMed. For the updated version 2.0, we improved our search methodology by using MeSH (medical subject headings) terms to retrieve disease-specific content, such as: Alzheimer's Disease[MeSH Terms] AND (risk factor*[tiab] or genetic factor*[tiab] or risk marker*[tiab]) AND 2021/01/01: 2024/09/30[dp]. This refinement has significantly enhanced the precision of our search results.

In the NDDRF update segment, we accessed 4470 original articles from literature published between January 2021 and September 2024. In the section focusing on the expansion of lifestyle factors, we initiated the search by exploring lifestyle-related content in PubMed using 78 NDDs identified from the MeSH database. Subsequently, a secondary search was carried out based on the acquired lifestyle factors, encompassing literature from January 1975 to September 2024. This process yielded a total of 6415 original articles. After the application of our inclusion and exclusion criteria and the removal of duplicates from the NDDRF update and lifestyle expansion findings, we identified 20 NDDs, 996 risk factors, 2086 records, and a total of 669 articles, as illustrated in Figure 1.

2.3 | Embedding of ICD-11 code

Given that risk factors mentioned in the literature often correspond to specific subtypes or classifications of NDDs, our objective was to offer a more precise description to aid clinicians in their searches. To accomplish this, we leveraged the ICD-11 coding tool[29] provided by the World Health Organization[30] and, with input from clinicians, integrated the relevant disease codes for each risk factor outlined in the articles. In cases in which an NDD was not explicitly identified in the article or was not directly covered by the ICD-11 codes, we categorized it using the "Other specified" and "Unspecified" designations from the coding tool (Table 1).

2.4 | Annotation for clinical phase of risk factors

In the studies on NDD risk factors that we included, each risk factor corresponds to a specific clinical stage, as shown in Table S1 in supporting information. Therefore, in version 2.0, we categorized the clinical stages of risk factors into diagnosis, prognosis, and treatment, and labeled each record accordingly.

RESEARCH IN CONTEXT

1. **Systematic review:** The authors conducted a comprehensive literature review using PubMed to identify modifiable risk factors for neurodegenerative diseases (NDDs) and explore their implications for personalized prevention. Key findings highlighted the importance of lifestyle factors as critical targets for intervention. However, challenges remain in synthesizing heterogeneous risk factors across NDD subtypes and translating these findings into clinical practice.
2. **Interpretation:** The Knowledgebase of Risk Factors for Neurodegenerative Diseases (NDDRF) 2.0 update integrates standardized risk classifications and expanding evidence to include 1971 risk factors across 20 NDDs, with lifestyle factors being a prominent focus. Our study reveals that the risk factors associated with various NDDs extend beyond the 14 factors reported by the Lancet Commission, encompassing a broader range of additional factors such as dietary patterns and occupational types. NDDRF 2.0 serves as a versatile resource for clinicians, researchers, and at-risk populations, supporting diverse applications to develop and implement effective proactive health management.
3. **Future directions:** We plan to develop an ontology for NDD risk factors to standardize terminology and map it to a large language model (LLM) for integration. This will enable the creation of a chatbot designed for high-risk populations, providing personalized risk assessment capabilities.

2.5 | SNP ID

SNPs are common genetic variations. Understanding this information is crucial for genetic research, disease association studies, and personalized medicine. In our collection of genetic factors, many SNPs reported in the articles lacked an SNP ID. During data curation, to ensure consistent descriptions of SNPs, we referenced the dbSNP database (https://www.ncbi.nlm.nih.gov/snp/), obtaining information such as chromosomal location, associated genes, allele variations, and functional impact to annotate the SNP IDs, for example, rs1990622 for the *PTPRT* gene. For variations not retrievable in dbSNP, we retained the original reporting format in NDDRF 2.0, such as *SORL1* gene variant NM_003105.6: c.4519+5G > A.

2.6 | New classification criteria of risk factors

The risk factor categories have been reclassified as lifestyle factors, epidemiological factors, genetic factors, combination factors, and non-

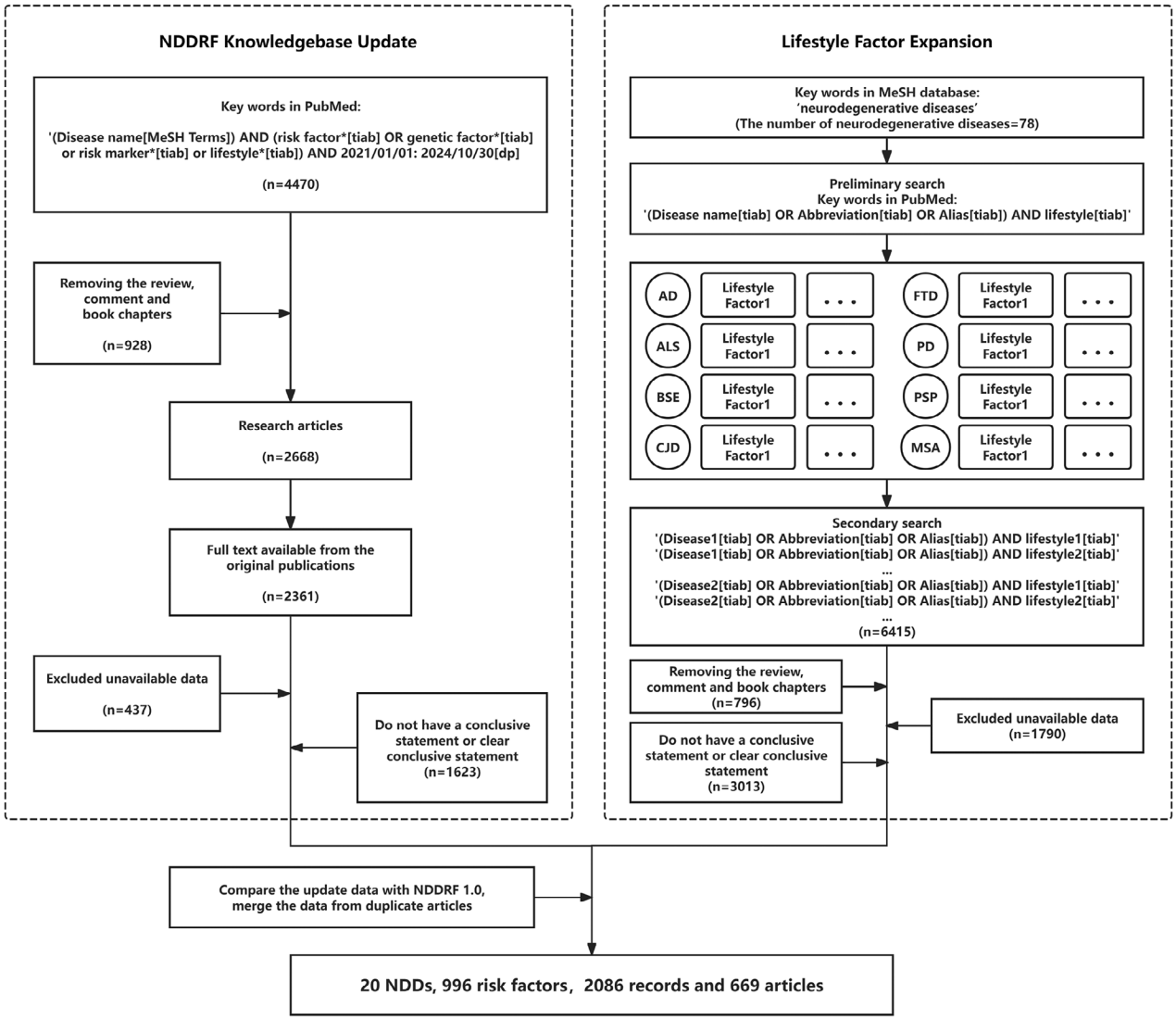

FIGURE 1 Flowchart of NDD lifestyle factor collection and risk factor update. AD, Alzheimer's disease; ALS, amyotrophic lateral sclerosis; BSE, bovine spongiform encephalopathy; CJD, Creutzfeldt–Jakob disease; FTD, frontotemporal dementia; MeSH, medical subject heading; MSA, multiple system atrophy; NDD, neurodegenerative disease; NDDRF, Knowledgebase of Risk Factors for Neurodegenerative Diseases; PD, Parkinson's disease; PSP, progressive supranuclear palsy.

TABLE 1 Instance of ICD code embedding in NDDRF 2.0.

NDDs classification	ICD code	Code name
PD-lateral trunk flexion (LTF)	8A00.Y	Other specified Parkinsonism
Idiopathic Parkinson's disease (IPD)	8A00.Z	Parkinsonism, unspecified

Abbreviations: ICD, International Classification of Diseases; NDDRF, Knowledgebase of Risk Factors for Neurodegenerative Diseases; PD, Parkinson's disease.

risk factors. Using the UMLS Semantic Network[31,32] and ICD-11 codes[31] as references, and in close consultation with clinicians, we performed a detailed categorization of NDD risk factors (Table 2). After this classification, all data in NDDRF 2.0 were systematically labeled to ensure comprehensive and standardized organization.

2.7 | Data model

Based on the data collected, we updated the database tables from NDDRF version 1.0.[18] The newly incorporated table fields are detailed in Table S2 in supporting information.

2.8 | Database architecture

We used the WAMP (Windows + Apache + MySQL + PHP) architecture to update the website. The network topology of risk factors was constructed using the dynamic visualization library Vis.js.

TABLE 2 Detailed risk factor categories in NDDRF 2.0.

Categories	First-class	Second-class	Third-class
Lifestyle factors	Diet	Food	Vegetable/fruit/meat/sausage/ dairy products/egg/fish/ supplementation/condiment/ food additives/grains/ processed foods and other foods/animal feed/others
		Drinks	Alcohol/coffee drinking/milk/ milk replacer/tea drinking/ water intake/beverage
		Nutrient substance	Saccharide/extract of botanical/ plant extracts/lipid and fat/ fungus/lipopolysaccharide/ supplementation/vitamin intake/ fiber/carbohydrate/amino acids/ fatty acids/others
		Microelement	
		Dietary patterns	Fat intake/Mediterranean diet (MeDi)/others
	Education		
	Habit	Behavior	Social behavior/individual behavior/others
		Keeping a pet	
		Reading	
		Sleeping	
		Smoking	
		Travel	
		Attitude	
	Physical activity		
	Physical fitness	Obesity	
		Height	
		BMI	
		Weight	
		Strength	
	Psychology	Anxiety	
		Depression	
		Marriage	
		Meditation	
		Stress	
		Social factors	
		Other characteristics	
Epidemiological factors	Biochemical factors	Tissue	
		Cell	
		Cell component	
		Biomedical or dental material	
		Biologically active substance	Hormone/enzyme/ immunologic factor/receptor/others
	Demographic characteristics	Age	
		Ethnicity	
		Sex	
		Family	
		Marital status	

(Continues)

TABLE 2 (Continued)

Categories	First-class	Second-class	Third-class
	Disease	Disease history	
		Infection	
		Tumor	
		Congenital abnormality	
		Acquired abnormality	
		Mental or behavioral dysfunction	
		Cell or molecular dysfunction	
		Cerebrovascular disease	
		Metabolic syndrome (MetS)	
		Hypertension	
		Dyslipidemia	
		Cardiovascular disease (CVD)	
		Neurodegenerative diseases (NDDs)	
		Ophthalmic diseases	
		Inflammation	
		Diabetes	
		Autoimmune disease	
		Comorbidity	
		Thyroid diseases	
		Osteoarthritis	
		Diseases of the respiratory system	
	Sign or symptom		
	Environment	Air pollution	
		Exposure	
		Occupational characteristics	
		Occupational exposure	
		Occupational types	
		Other	
		Residence	
	Medical methods	Health examination	
		Medicines	Antibiotic/statin/antihypertensive/analgesic and anti-inflammatory drugs/neuroleptics/hypnagogue/ antidiabetic/asthma drugs/ targeted medicine/ gastrointestinal drugs/ anticholinergic drugs/ lipid-lowering drugs/ contraceptive pills/others
		Non-pharmaceutical therapy	
		Surgical therapy	
		Medical history	
		Medication	
		Non-pharmaceutical therapy	
		Other	

(Continues)

TABLE 2 (Continued)

Categories	First-class	Second-class	Third-class
	Physiological indices		
	Reproduction		
	Syndrome		
	Trauma		
	Poisoning		
Genetic factors	Family history		
	Gene or genome	Single nucleotide polymorphisms	
		Copy number variations	
		Other polymorphisms	
	miRNA		
Combination factors			

Abbreviations: BMI, body mass index; miRNA, mitochondrial RNA; NDDRF, knowledgebase of risk factors for neurodegenerative diseases.

3 | RESULTS

3.1 | Advancements in NDDRF 2.0: enriched knowledge base and expanded scope

In contrast to version 1.0, our repository now features a distinct section dedicated to lifestyle factors, encompassing six primary categories: diet, mental health, education, hobbies, exercise, and physical fitness, along with 66 associated subcategories. At present, the database comprises 536 lifestyle factors across 1363 records. After 4 years of ongoing updates, the cumulative count of risk factors has risen to 1971, spanning 4379 records and 151 subcategories.

3.2 | Descriptive statistics

Statistical analysis conducted on the data within NDDRF 2.0 unveiled that AD continues to be the most extensively researched NDD, closely trailed by PD and amyotrophic lateral sclerosis (ALS). Epidemiological factors emerged as prominently featured among the risk factors associated with NDDs (Figure 2A). Within the top five most scrutinized NDDs (AD, PD, ALS, Creutzfeldt–Jakob disease [CJD], frontotemporal lobar degeneration [FTLD]), common risk factors are observed across each disease pair. This implies that a unified prevention strategy could potentially prove efficacious in mitigating the risk of multiple NDDs simultaneously (Figure 2B). Nonetheless, there are no universally common risk factors shared among all five NDDs, underscoring the substantial heterogeneity in NDD risk factor investigations (Figure 2B). Regarding the clinical stages at which risk factors are relevant, diagnostic factors predominate; intriguingly, physical exercise (e.g., yoga[33,34]) and sleep-related factors (e.g., sleep quality[35,36] and efficiency[37,38]) significantly impact diagnosis, prognosis, and treatment stages in NDDs (Figure 2C). This underscores the critical role of modifiable lifestyle factors throughout the clinical continuum of NDDs. Targeted interventions addressing these factors may yield positive outcomes across all clinical phases of these disorders.[25,39] Among the studies sourced from the 65 countries featured in our database, the United States, China, Italy, and the United Kingdom emerge as frontrunners in research contributions, trailed by Spain, France, Japan, and South Korea (Figure 2D). Categorized based on their relevance to NDDs, the risk factors encompass 1803 detrimental factors and 631 protective factors, alongside 463 factors that could act as both protective and harmful under varying conditions (Figure 2E). The distribution of risk factor types differs for each NDD, accentuating the considerable heterogeneity inherent in NDDs (Figure 2E).

In the network diagram showcasing the connections between NDDs and risk factors (Figure 3), AD, PD, and ALS exhibit the highest count of associated risk factors. Noteworthy risk factors linked to multiple NDDs (Degree ≥ 5) encompass education, smoking, apolipoprotein E gene, *MAPT* gene, various occupational types, obesity, diabetes, *TREM2* gene, Mediterranean diet (MeDi), depression, hypertension, and head injury. Among these pivotal risk factors influencing multiple NDDs, lifestyle factors—such as education, smoking, and dietary habits—are prominently featured. Enhancements in these lifestyle factors play a pivotal role in the prevention of diverse NDDs.

3.3 | Enhanced website features and functionality improvements

The website interface has been revamped to offer a more user-friendly experience with enhanced interactivity (Figure 4A, B). The search feature has been enriched with refined filters tailored to match the distinctive attributes of lifestyle, epidemiological, and genetic factors. These updated filters include ethnicity, country of study, risk factor categories (with subcategories), and clinical phases, enabling users to perform customized searches that precisely meet their needs (Figure 4C). A categorized network of risk factors has been integrated for visualization and exploration. Detailed information for each risk factor category is displayed upon hovering over nodes, with clickable

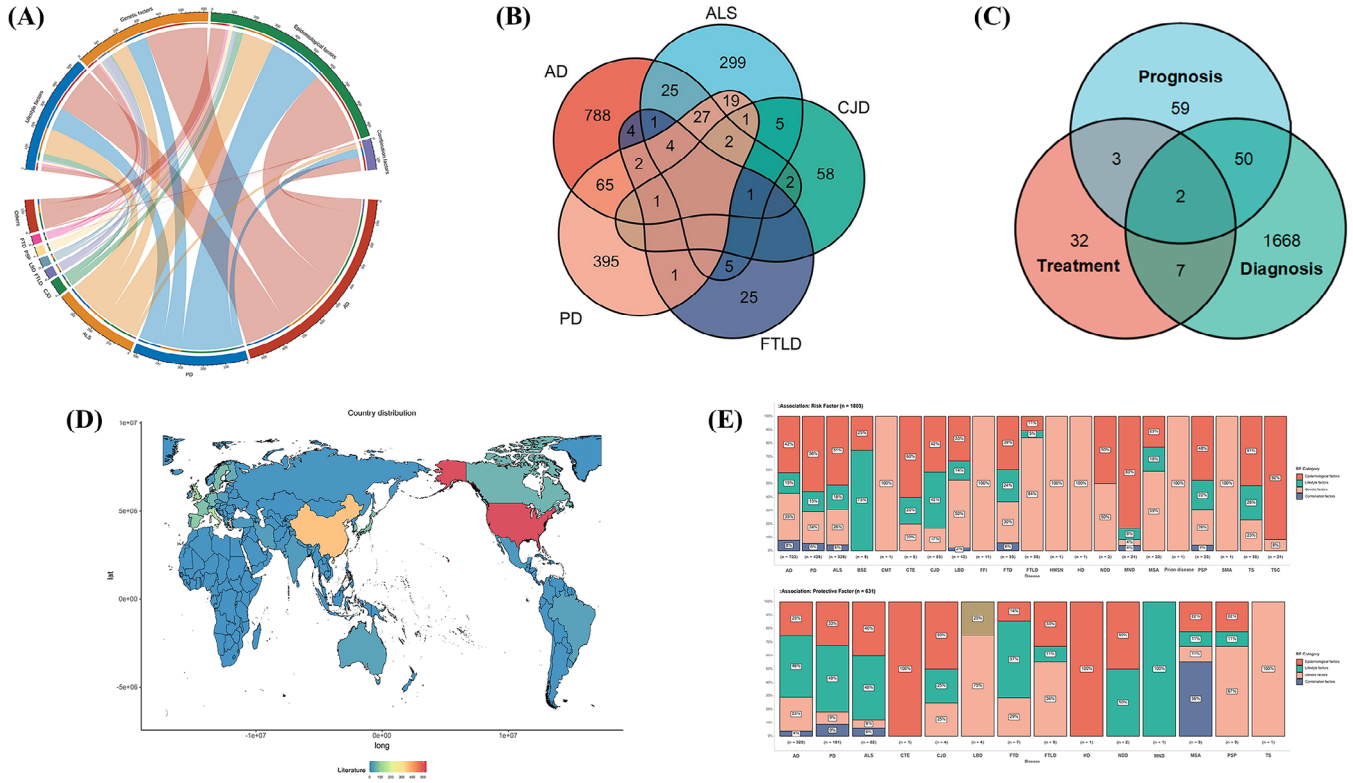

FIGURE 2 Descriptive statistics for NDDRF 2.0. A, Chord diagram shows the quantitative relationships between NDDs and the four categories of risk factors (bidirectional). B, Venn diagram illustrates the distribution of risk factors (single + combined) associated with the five most-studied NDDs. C, Distribution of risk factors (single + combined) according to their clinical phases. D, Heat map showing the distribution of research countries. E, Distribution of NDDs and types of risk factors grouped by relevance. AD, Alzheimer's disease; ALS, amyotrophic lateral sclerosis; BSE, bovine spongiform encephalopathy; CMT, Charcot–Marie–Tooth disease; CTE, chronic traumatic encephalopathy; CJD, Creutzfeldt–Jakob disease; FFI, fatal familial insomnia; FTD, frontotemporal dementia; FTLD, frontotemporal lobar degeneration; HMSN, hereditary motor and sensory neuropathy; HD, Huntington's disease; LBD, Lewy body dementia; MND, motor neuron disease; MSA, multiple system atrophy; NDD, neurodegenerative disease; NDDRF, Knowledgebase of Risk Factors for Neurodegenerative Diseases; PD, Parkinson's disease; PSP, progressive supranuclear palsy; RF, risk factor; SMA, spinal muscular atrophy; TS, Tourette syndrome; TSC, tuberous sclerosis complex.

nodes enabling further exploration of subclass searches (Figure 4D). Considering the availability of the data, a preview version is provided on the download page. Users may contact us to request approval for accessing the full dataset for academic use.

4 | DISCUSSION

The diagnosis of NDDs continues to pose a significant challenge, often relying heavily on the clinical expertise of neurologists and neurosurgeons.[40] While substantial global funding has been allocated to drug development for treatment, most options are expensive, may lead to adverse effects, and primarily aim to slow disease progression rather than provide a cure.[41,42] During the preclinical stages of NDDs, patients may exhibit no symptoms or signs, despite likely biological changes in the brain.[10,43,44] Therefore, for high-risk populations, proactive lifestyle adjustments offer a more effective health-care approach in preventing or delaying disease onset compared to traditional clinical interventions.[23]

Alfalahi et al. proposed a shift from data-based to knowledge-based models in research on digital phenotyping of neurodegenerative manifestations, emphasizing patient-specific phenotypes.[45] The inherent personalization and heterogeneity of medical data make it challenging to find identical individuals in clinical settings. For more accurate NDD risk identification and assessment, collecting in-depth information on risk factors plays a crucial role in matching patient-specific phenotypes.

Currently, lifestyle-related risk factors for NDDs are reported with broad variability, and the populations affected by each lifestyle factor are often not clearly defined. This heterogeneity makes it difficult to provide precise risk assessments and individualized prevention strategies for specific at-risk individuals. In the NDDRF 2.0 update, we added comprehensive lifestyle factors encompassing six main categories and > 60 subcategories, which include detailed quantitative metrics, baseline information of study populations, and preconditions affecting disease impact.

Compared to the first version of NDDRF, our data volume has more than doubled in size, with significant expansions across five dimensions: the number of references, categories of NDDs, categories of

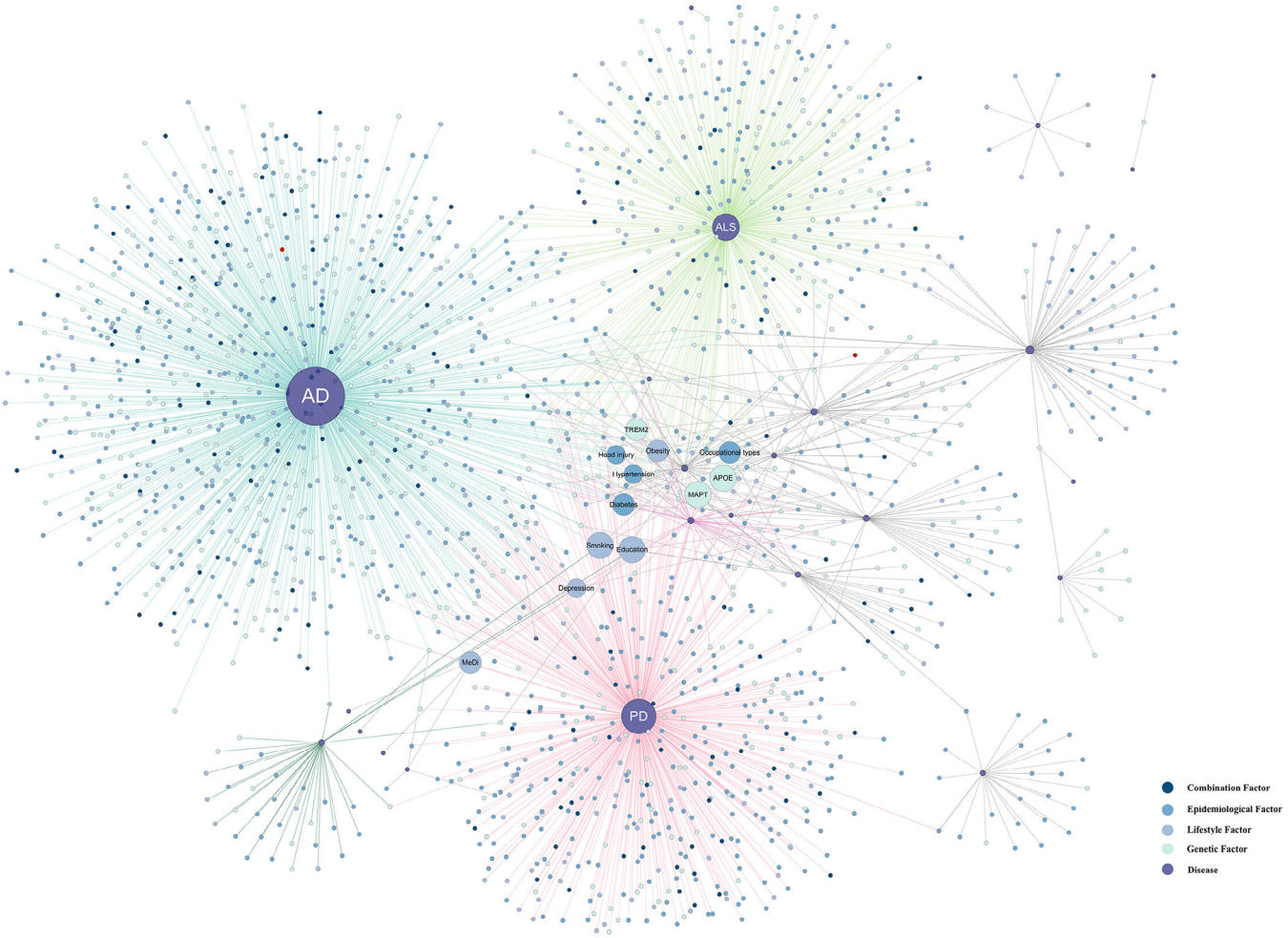

FIGURE 3 Risk factor network of NDDRF. Network diagram shows correlation between neurodegenerative diseases and risk factors. AD, Alzheimer's disease; ALS, amyotrophic lateral sclerosis; PD, Parkinson's disease; *APOE*, apolipoprotein E; *MAPT*, microtubule-associated protein tau gene; *TREM2*, triggering receptor expressed on myeloid cells 2; MeDi, mediterranean diet; NDDRF, Knowledgebase of Risk Factors for Neurodegenerative Diseases.

risk factors, the number of risk factors, and the overall records of risk factors (Figure 5).

To showcase the impact of NDDRF 2.0 in mitigating the risk of NDDs, we present three cases that exemplify real-world applications of the knowledge base. These instances underscore the effectiveness of NDDRF 2.0 and demonstrate how it provides crucial assistance to users from diverse backgrounds with a range of requirements.

4.1 | Enhancing multi-disease genetic risk analysis for bioinformaticians

By combining NDDRF 2.0 with the PheWAS Catalog, bioinformaticians can adopt a robust approach to discern how specific genetic variations associated with NDDs influence the risk across various diseases (Figure 6A). The PheWAS Catalog (https://phewascatalog.org/phewas), a comprehensive database and tool for conducting phenome-wide association studies (PheWAS),[46] facilitates the exploration of relationships between a single genotype and a multitude of phenotypes. For instance, the genetic risk factors cataloged in NDDRF 2.0 encompass numerous genes linked to risk, which can be queried in the PheWAS Catalog using gene names or SNP IDs. By amalgamating the effect sizes detailed in our knowledge base with PheWAS data, researchers can refine disease risk assessment models and devise personalized therapeutic strategies, thereby anticipating the likelihood of developing multiple associated conditions in the future.

4.2 | Empowering clinicians to recommend optimal health care for NDD patients

For example, consider Patient A, diagnosed with a specific subtype of AD. By using the advanced search functionality in NDDRF 2.0, clinicians can apply filters such as "AD—protective factors—lifestyle factors—prognosis/treatment." By considering factors like ethnicity and age that align with the patient's profile, clinicians can leverage

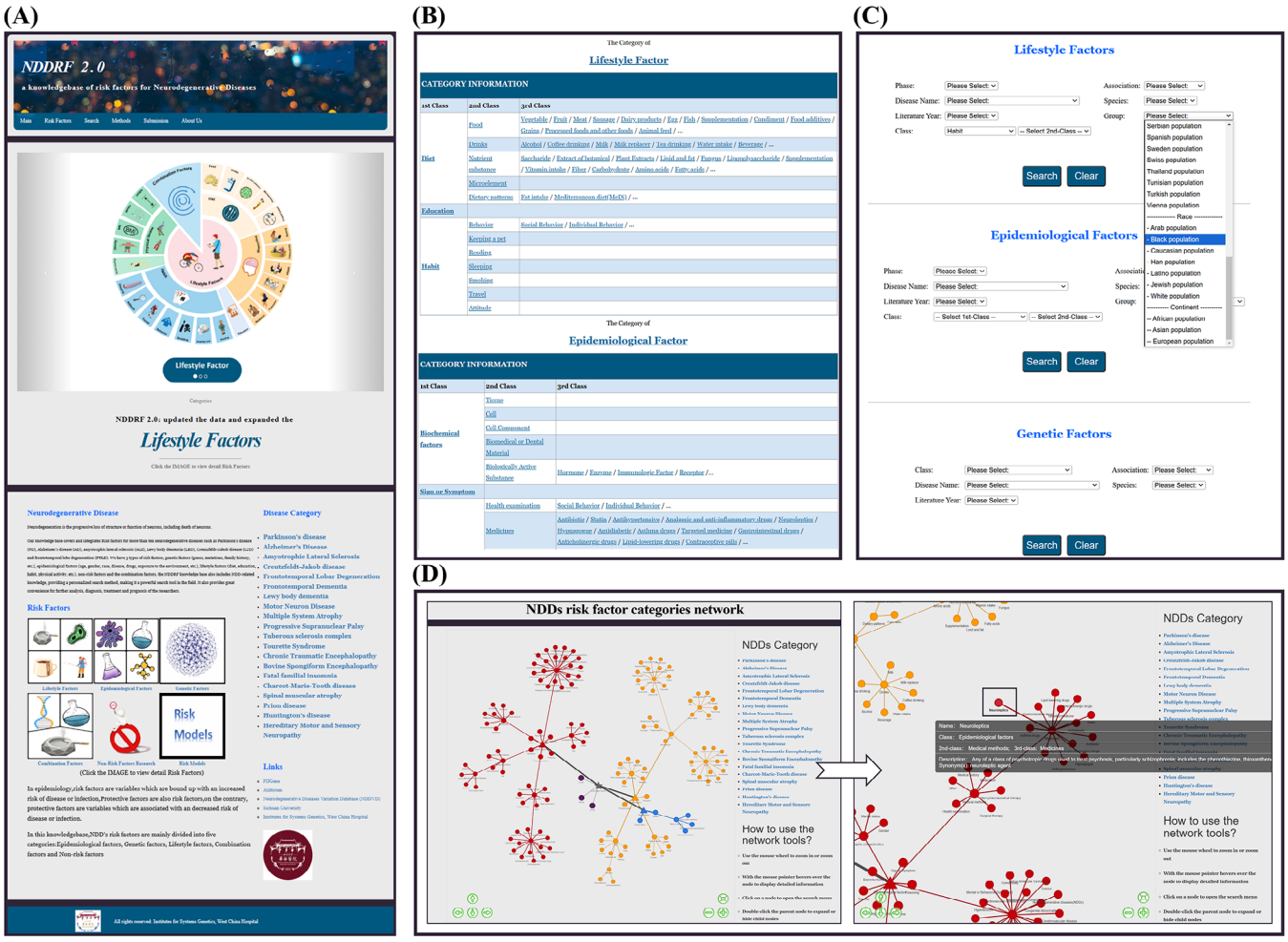

FIGURE 4 Website update of NDDRF. A, Main interface of website. B, Web page with risk factor subcategory information. C, Advanced search. D, Categorized network of risk factors. NDD, neurodegenerative disease; NDDRF, Knowledgebase of Risk Factors for Neurodegenerative Diseases.

effect sizes and quantitative indicators from NDDRF 2.0, in conjunction with their clinical expertise, to counsel the patient on lifestyle adjustments designed to slow disease advancement and enhance the patient's overall quality of life (Figure 6B).

4.3 | In-depth analysis of lifestyle factors in specific risk groups

By leveraging the detailed phenotypic data on risk factors within NDDRF 2.0, users can conduct precise risk assessments, thereby improving the precision of risk screening (Figure 6C). For instance, User B, who engages in lifestyle practices like high sugar consumption, smoking, and alcohol consumption, can use NDDRF 2.0 to investigate the potential risks associated with these habits for various NDDs. By examining modifiable risk factors and using the quantitative metrics available, proactive interventions can be implemented to mitigate the risk of developing diseases. In essence, NDDRF 2.0 empowers users across diverse roles to access personalized information with precision, thereby making a substantial contribution to the systematic prevention of NDDs.

While the NDDRF stands as the sole knowledge repository exclusively focused on NDD risk factors, numerous extensive databases also encompass information on disease risk factors. Notable among these are longitudinal cohort study databases like the Alzheimer's Disease Neuroimaging Initiative (ADNI),[47,48] the Parkinson's Progression Markers Initiative (PPMI),[49] UK Biobank,[50] and the Australian Imaging, Biomarkers & Lifestyle Flagship Study of Ageing (AIBL).[51,52] These aging cohort studies provide crucial insights into the impact of diverse risk factors on NDDs. Nevertheless, the quantitative assessment of risk factors for NDDs reveals substantial heterogeneity, leading to discrepancies in current research outcomes.[53] This underscores the critical need for an integrated disease risk factor database to facilitate systematic analysis.

When it comes to the integration of disease risk factors, several databases exist, each with notable limitations. For example, AlzRisk is a curated database dedicated to AD and related dementia risk factors, covering 15 lifestyle factors such as alcohol consumption,

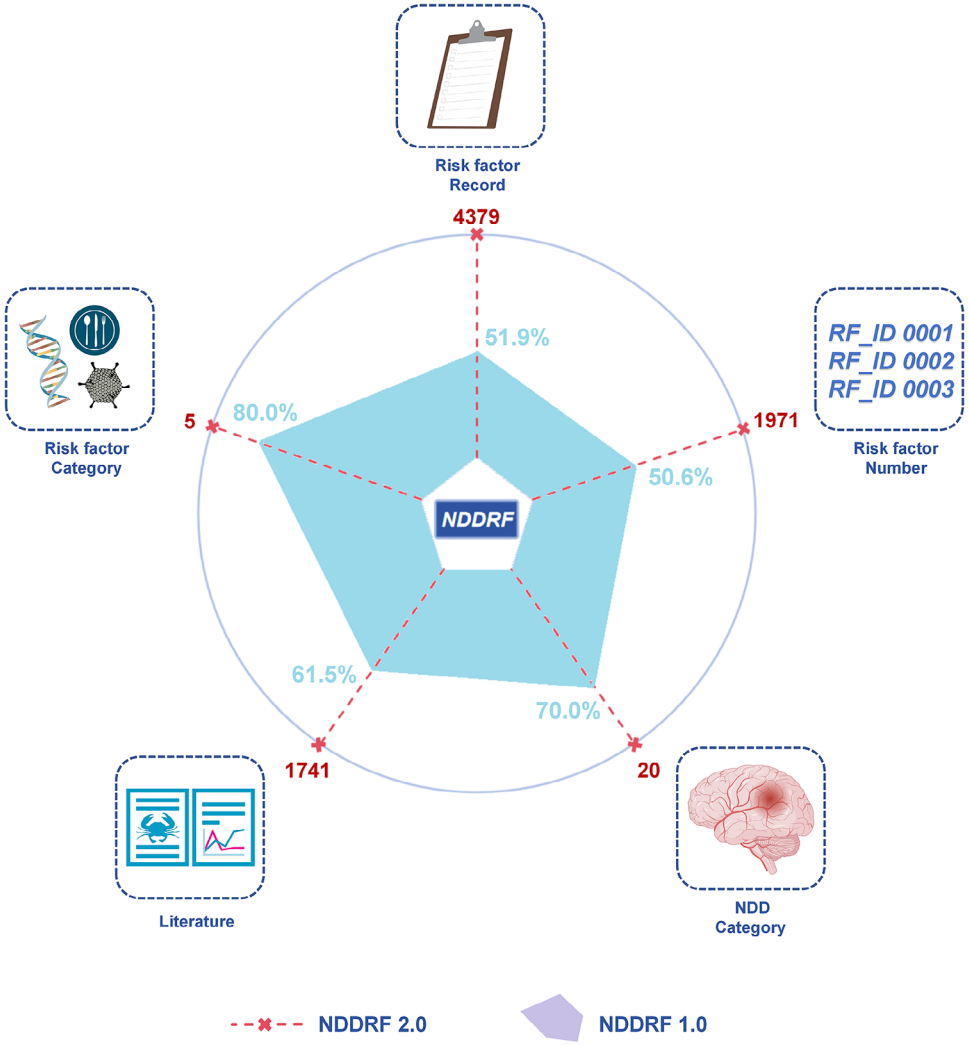

FIGURE 5 Comparison between NDDRF Knowledge Base 2.0 and 1.0. NDD, neurodegenerative disease; NDDRF, Knowledgebase of Risk Factors for Neurodegenerative Diseases.

blood pressure, cognitive activity, and diabetes mellitus. However, it lacks coverage of genetic and combination factors.[53] Similarly, CTD,[54] the Comparative Toxicogenomics Database, focuses on relationships among chemicals, genes, and diseases, with a specific emphasis on integrating environmental exposures and disease associations, yet it does not include lifestyle factors. HuGE Navigator serves as a genomics and epidemiology database for multiple diseases, primarily focused on genetic information, with only limited data on non-genetic risk factors.[55] Last, MIKB (Myocardial Infarction Knowledge Base),[56] a knowledge base developed by our research group, integrates cardiovascular disease risk factors, encompassing both genetic and non-genetic factors. While MIKB serves as an important resource for cardiovascular disease prevention and management, its coverage of lifestyle factors remains relatively limited and requires further updating and expansion. In summary, integrative databases lay the foundation for understanding the connections and distinctions among various NDDs. The earliest benchmark database in this field, INDD[57], the Integrated Neurodegenerative Disease Database, facilitates efficient and comprehensive querying of biomarker research results by integrating multiple upstream databases, thereby enhancing research efficiency.

These databases collectively underscore the fragmented nature of existing resources, underscoring the necessity for a comprehensive and balanced database that systematically integrates genetic, epidemiological, and lifestyle factors. Compared to these resources, NDDRF 2.0 offers several distinct advantages:

1. Comprehensive inclusion of diverse risk factor types. Particularly noteworthy is the incorporation of modifiable lifestyle factors, with detailed subcategories crucial for disease prevention.
2. Specialized focus on NDDs. Encompassing risk factors for 20 distinct NDDs, NDDRF 2.0 delves deeply into this specific area of research.
3. In-depth and systematically curated risk factor information. Users can benefit from meticulously curated data that is both thorough and reliable.

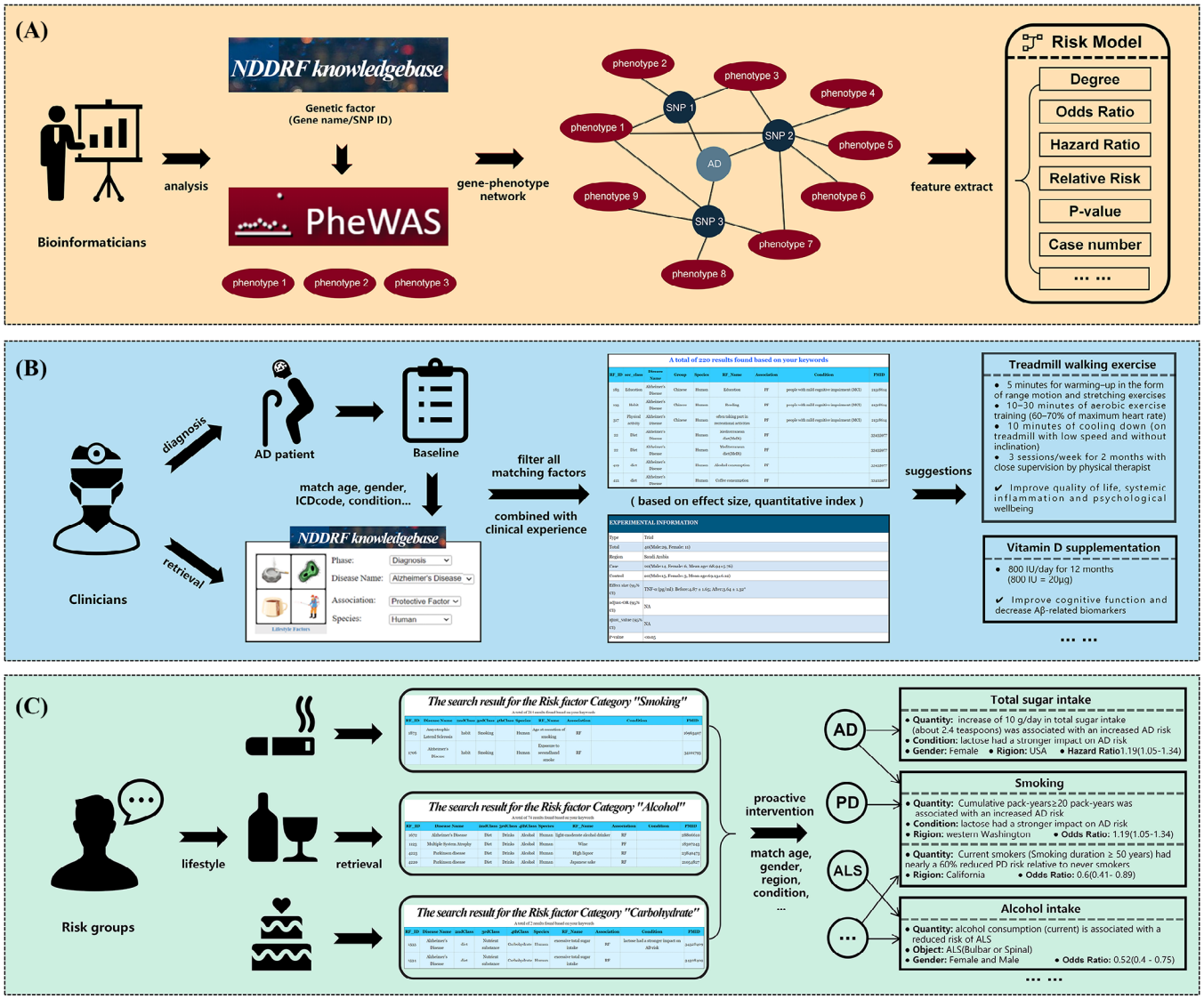

FIGURE 6 Framework diagram of NDDRF 2.0 application instance: (A) bioinformaticians, (B) clinicians, and (C) risk group. AD, Alzheimer's disease; ALS, amyotrophic lateral sclerosis; ICD, International Classification of Diseases; NDDRF, Knowledgebase of Risk Factors for Neurodegenerative Diseases; PD, Parkinson's disease; PheWAS, phenome-wide association studies; SNP ID, single nucleotide polymorphism identifier.

4. Broad applicability. NDDRF 2.0 caters to a wide array of users with diverse professional and research needs, making it a versatile resource.
5. Enhanced and user-friendly search functionality. The database boasts improved search capabilities, enhancing accessibility and usability for all users.

Currently, NDDRF 2.0 effectively addresses the query requirements of clinicians and researchers concerning risk factors. However, for the public and at-risk populations, a more intuitive and accessible interactive approach is essential. Contemporary large language models (LLMs) have achieved substantial advancements, demonstrating exceptional efficacy in handling domain-specific inquiries across professional fields.[58,59] Nevertheless, general-purpose LLMs trained on non-specialized data may generate hallucinated or inaccurate medical responses due to insufficient domain-specific training.[60] Such errors in applying generalized knowledge could pose significant safety risks to at-risk populations. To mitigate this, NDDRF 2.0—a manually curated, evidence-based knowledge base—will serve as a foundation for fine-tuning LLMs, enabling them to deliver precise and reliable answers to specialized queries.[61] In the future, we plan to develop an ontology for NDD risk factors by comprehensively collecting and standardizing relevant terminology. Domain-specific ontology is instrumental in enhancing the semantic organization and representation of knowledge related to NDDs.[62] Ultimately, NDDRF 2.0 will be mapped to the ontology and integrated with LLMs to create a chatbot tailored to at-risk individuals, offering personalized risk assessment functionalities.[63,64]

In this study, we have significantly enhanced the NDDRF knowledge base by broadening its scope to include a comprehensive range of modifiable lifestyle factors, in addition to a substantial update and

refinement of existing data. The enhanced NDDRF 2.0 knowledge base now encompasses a detailed categorization of lifestyle-related risk factors across six major domains, enabling a proactive approach to health management. By providing in-depth phenotypic data on NDD risk factors, NDDRF 2.0 facilitates precision-based risk modeling for more precise and individualized risk assessment. With this enriched resource, we empower both clinicians and researchers to formulate and implement more effective risk prevention strategies. This thorough and comprehensive approach enhances the versatility and relevance of the knowledge base, establishing it as a valuable tool for in-depth analysis, aiding clinicians in decision-making, and advancing public health initiatives in NDD prevention and risk mitigation.

AUTHOR CONTRIBUTIONS

Cheng Bi: writing—original draft, methodology, data curation, visualization, software, and formal analysis. Xin Zheng: methodology, data curation, and validation. Yuxin Zhang: investigation, data curation, and validation. Shengrong Zhou: mmethodology, data curation, and validation. Jie Song: data curation and validation. Huifang Shang: validation, investigation, and data curation. Bairong Shen: writing—review and editing, supervision, and conceptualization.

ACKNOWLEDGMENTS

The symbols for making the graphical abstract come from University of Maryland Center for Environmental Science (ian.umces.edu/imagelibrary/). This work was supported by the National Natural Science Foundation of China (32270790) as well as the Sichuan Science and Technology Program (grant no. 2024YFHZ?>0205).

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest. Author disclosures are available in the supporting information.

CONSENT STATEMENT

Consent was not necessary because this study uses online public resources.

REFERENCES

1. Dugger BN, Dickson DW. Pathology of neurodegenerative diseases. *Cold Spring Harb Perspect Biol*. 2017;9:a028035.
2. Brettschneider J, Del Tredici K, Lee VM, Trojanowski JQ. Spreading of pathology in neurodegenerative diseases: a focus on human studies. *Nat Rev Neurosci*. 2015;16:109-120.
3. Agnello L, Ciaccio M. Neurodegenerative diseases: from molecular basis to therapy. *Int J Mol Sci*. 2022;23:12854.
4. Bateman RJ, Xiong C, Benzinger TL, et al. Clinical and biomarker changes in dominantly inherited Alzheimer's disease. *N Engl J Med*. 2012;367:795-804.
5. Ravikumar S, Denning AE, Lim S, et al. Postmortem imaging reveals patterns of medial temporal lobe vulnerability to tau pathology in Alzheimer's disease. *Nat Commun*. 2024;15:4803.
6. Poewe W, Seppi K, Tanner CM, et al. Parkinson disease. *Nat Rev Dis Primers*. 2017;3:17013.
7. Hoyert DL, Xu J. Deaths: preliminary data for 2011. *Natl Vital Stat Rep*. 2012;61:1-51.
8. Feigin VL, Vos T, Nichols E, et al. The global burden of neurological disorders: translating evidence into policy. *Lancet Neuro*. 2020;19:255-265.
9. He W, Goodkind D, Kowal P, Bureau UC. *International Population Reports: An Aging World: 2015*. U.S. Government Publishing Office; 2016:P95/16-11.
10. 2023 Alzheimer's disease facts and figures. *Alzheimer Dement*. 2023;19:1598-1695.
11. Rajan KB, Weuve J, Barnes LL, McAninch EA, Wilson RS, Evans DA. Population estimate of people with clinical Alzheimer's disease and mild cognitive impairment in the United States (2020-2060). *Alzheimers Dement*. 2021;17:1966-1975.
12. Zheng Z, Zhu Z, Zhou C, Cao L, Zhao G. Burden of Parkinson disease in China, 1990-2019: findings from the 2019 global burden of disease study. *Neuroepidemiology*. 2023;57:51-64.
13. Hardy J. Pathways to primary neurodegenerative disease. *Ann NY Acad Sci*. 2000;924:29-34.
14. Appel SH, Smith RG, Le WD. Immune-mediated cell death in neurodegenerative disease. *Adv Neurol*. 1996;69:153-159.
15. Livingston G, Sommerlad A, Orgeta V, et al. Dementia prevention, intervention, and care. *Lancet*. 2017;390:2673-2734.
16. Livingston G, Huntley J, Sommerlad A, et al. Dementia prevention, intervention, and care: 2020 report of the Lancet Commission. *Lancet*. 2020;396:413-446.
17. Livingston G, Huntley J, Liu KY, et al. Dementia prevention, intervention, and care: 2024 report of the Lancet standing Commission. *Lancet*. 2024;404:572-628.
18. Bi C, Zhou S, Liu X, et al. NDDRF: a risk factor knowledgebase for personalized prevention of neurodegenerative diseases. *J Adv Res*. 2022;40:223-231.
19. Yang Y, Xu C, Liu X, et al. NDDVD: an integrated and manually curated neurodegenerative diseases variation database. *Database*. 2018;2018:bay018.
20. Yang Y, Xu S, Hong Y, et al. Computational modeling for medical data: from data collection to knowledge discovery. *Innov Life*. 2024;2:100079.
21. Shen L, Bai J, Wang J, Shen B. The fourth scientific discovery paradigm for precision medicine and healthcare: challenges ahead. *Precis Clin Med*. 2021;4:80-84.
22. Wilson DM, 3rd, Cookson MR, Van Den Bosch L, Zetterberg H, Holtzman DM, Dewachter I. Hallmarks of neurodegenerative diseases. *Cell*. 2023;186:693-714.
23. Shen B, Lin Y, Bi C, et al. Translational informatics for Parkinson's disease: from big biomedical data to small actionable alterations. *Genom Proteom Bioinform*. 2019;17:415-429.
24. Chen Y, Liu X, Yu Y, et al. PCaLiStDB: a lifestyle database for precision prevention of prostate cancer. *Database*. 2020;2020.
25. Kivipelto M, Mangialasche F, Ngandu T. Lifestyle interventions to prevent cognitive impairment, dementia and Alzheimer disease. *Nat Rev Neurol*. 2018;14:653-666.
26. Dhana K, Evans DA, Rajan KB, Bennett DA, MC Morris. Healthy lifestyle and the risk of Alzheimer dementia: findings from 2 longitudinal studies. *Neurology*. 2020;95:e374-e83.
27. Zissimopoulos J, Crimmins E, St ClairP. The value of delaying Alzheimer's disease onset. *Forum Health Econ Policy*. 2014;18:25-39.
28. Sherzai D, Sherzai A, Sherzai A. Lifestyle intervention and Alzheimer disease. *J Fam Pract*. 2022;71:eS83-eS9.
29. ICD-11 Coding Tool. World Health Organization.
30. The Lancet. ICD-11. *Lancet*. 2019;393:2275.
31. Humphreys BL, Lindberg DA, Schoolman HM, Barnett GO. The Unified Medical Language System: an informatics research collaboration. *J Am Med Inform Assoc*. 1998;5:1-11.
32. UMLS Knowledge Sources. Release 2024AA. National Library of Medicine (US).

33. Ni M, Signorile JF, Mooney K, et al. Comparative effect of power training and high-speed yoga on motor function in older patients with Parkinson disease. *Arch Phys Med Rehabil*. 2016;97:345-354.e15.
34. Krause-Sorio B, Siddarth P, Kilpatrick L, et al. Yoga prevents gray matter atrophy in women at risk for Alzheimer's disease: a randomized controlled trial. *J Alzheimers Dis*. 2022;87:569-581.
35. Sinha N, Fausto BA, Mander B, Gluck MA. High-quality sleep mitigates abca7-related generalization deficits in healthy older African Americans. *J Alzheimers Dis*. 2023;94:281-290.
36. Schrag A, Bohlken J, Dammertz L, et al. Widening the spectrum of risk factors, comorbidities, and prodromal features of Parkinson disease. *JAMA Neurol*. 2023;80:161-171.
37. Yang JJ, Keohane LM, Pan XF, et al. Association of healthy lifestyles with risk of Alzheimer disease and related dementias in low-income Black and White Americans. *Neurology*. 2022;99:e944-e53.
38. Cullell N, Cárcel-Márquez J, Gallego-Fábrega C, et al. Sleep/wake cycle alterations as a cause of neurodegenerative diseases: a Mendelian randomization study. *Neurobiol Aging*. 2021;106:320.e1-.e12.
39. Dhana K, Franco OH, Ritz EM, et al. Healthy lifestyle and life expectancy with and without Alzheimer's dementia: population based cohort study. *BMJ*. 2022;377:e068390.
40. Hansson O. Biomarkers for neurodegenerative diseases. *Nat Med*. 2021;27:954-963.
41. Knox C, Wilson M, Klinger CM, et al. DrugBank 6.0: the DrugBank knowledgebase for 2024. *Nucleic Acids Res*. 2024;52:D1265-D1275.
42. Mahase E. FDA allows drugs without proven clinical benefit to languish for years on accelerated pathway. *BMJ*. 2021;374:n1898.
43. Berg D, Postuma RB, Adler CH, et al. MDS research criteria for prodromal Parkinson's disease. *Mov Disord*. 2015;30:1600-1611.
44. Heinzel S, Berg D, Gasser T, Chen H, Yao C, Postuma RB. Update of the MDS research criteria for prodromal Parkinson's disease. *Mov Disord*. 2019;34:1464-1470.
45. Alfalahi H, Dias SB, Khandoker AH, Chaudhuri KR, Hadjileontiadis LJ. A scoping review of neurodegenerative manifestations in explainable digital phenotyping. *NPJ Parkinson's Dis*. 2023;9:49.
46. Denny JC, Bastarache L, Ritchie MD, et al. Systematic comparison of phenome-wide association study of electronic medical record data and genome-wide association study data. *Nat Biotechnol*. 2013;31:1102-1110.
47. Petersen RC, Aisen PS, Beckett LA, et al. Alzheimer's disease neuroimaging initiative (ADNI): clinical characterization. *Neurology*. 2010;74:201-209.
48. jr Jack CR, Bernstein MA, Fox NC, et al. The Alzheimer's disease neuroimaging initiative (ADNI): MRI methods. *J Magn Reson Imaging*. 2008;27:685-691.
49. The Parkinson progression marker initiative (PPMI). *Prog Neurobiol*. 2011;95:629-635.
50. Sudlow C, Gallacher J, Allen N, et al. UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med*. 2015;12:e1001779.
51. Ellis KA, Bush AI, Darby D, et al. The Australian imaging, biomarkers and lifestyle (AIBL) study of aging: methodology and baseline characteristics of 1112 individuals recruited for a longitudinal study of Alzheimer's disease. *Int Psychogeriatr*. 2009;21:672-687.
52. Fowler C, Rainey-Smith SR, Bird S, et al. Fifteen years of the Australian imaging, biomarkers and lifestyle (AIBL) study: progress and observations from 2,359 older adults spanning the spectrum from cognitive normality to Alzheimer's disease. *J Alzheimers Dis Rep*. 2021;5:443-468.
53. Kinoshita J, Clark T. Alzforum. *Methods Mol Biol*. 2007;401:365-381.
54. Davis AP, Wiegers TC, Sciaky D, et al. Comparative toxicogenomics database's 20th anniversary: update 2025. *Nucleic Acids Res*. 2024.
55. Yu W, Gwinn M, Clyne M, Yesupriya A, Khoury MJ. A navigator for human genome epidemiology. *Nat Genet*. 2008;40:124-125.
56. Zhan C, Zhang Y, Liu X, et al. MIKB: a manually curated and comprehensive knowledge base for myocardial infarction. *Comput Struct Biotechnol J*. 2021;19:6098-6107.
57. Xie SX, Baek Y, Grossman M, et al. Building an integrated neurodegenerative disease database at an academic health center. *Alzheimers Dement*. 2011;7:e84-93.
58. Shool S, Adimi S, Saboori Amleshi et al. A systematic review of large language model (LLM) evaluations in clinical medicine. *BMC Med Inf Decis Making*. 2025;25:117.
59. Zong H, Wu R, Cha J, et al. Large language models in worldwide medical exams: platform development and comprehensive analysis. *J Med Internet Res*. 2024;26:e66114.
60. Ullah E, Parwani A, Baig MM, Singh R. Challenges and barriers of using large language models (LLM) such as ChatGPT for diagnostic medicine with a focus on digital pathology—a recent scoping review. *Diagn Pathol*. 2024;19:43.
61. Ge J, Sun S, Owens J, et al. Development of a liver disease-specific large language model chat interface using retrieval-augmented generation. *Hepatology*. 2024;80:1158-1168.
62. Yu C, Zong H, Chen Y, et al. PCAO2: an ontology for integration of prostate cancer associated genotypic, phenotypic and lifestyle data. *Briefings Bioinf*. 2024;25:bbae136.
63. Li J, Tang T, Wu E, et al. RARPKB: a knowledge-guide decision support platform for personalized robot-assisted surgery in prostate cancer. *Int J Surg*. 2024;110:3412-3424.
64. Ren S, Jin Y, Chen Y, Shen B. CRPMKB: a knowledge base of cancer risk prediction models for systematic comparison and personalized applications. *Bioinformatics*. 2022;38:1669-1676.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Bi C, Zheng X, Zhang Y, et al. NDDRF 2.0: An update and expansion of risk factor knowledge base for personalized prevention of neurodegenerative diseases. *Alzheimer's Dement*. 2025;21:e70282.
https://doi.org/10.1002/alz.70282