

Postprint: Construction and Application of the China-US-Korea Health and Nutrition Survey Database

Authors: Wei KuiKui, Gong Haowen, Liu Yuxiu, Wang Yan, Liu Mengdan, Zhang Manting, Yu Xihui, Liu Yuxiu

Date: 2023-08-01T00:00:00+00:00

Abstract

The health and nutrition survey databases of China, the United States, and South Korea are all established for monitoring the health and nutritional status of their respective national populations, and have become important public resources for data mining in recent years. This paper conducts a comparative analysis of the construction status, historical background, sampling methods, and survey content of the three national health and nutrition survey databases. The primary distinction lies in their research design types: China employs a cohort study design, whereas both the United States and South Korea utilize serial cross-sectional surveys. Subsequently, a visual analysis of publication volume and research hotspots based on these databases reveals a more pronounced upward trend in publications from the United States, with research hotspots across all three databases aligning with health and nutrition survey themes. Finally, the construction and application of these databases are discussed, and recommendations are proposed, aiming to provide pathways for researchers to conduct related studies and to better facilitate the generation of scientific evidence.

Full Text

Research on the Construction and Application of Health and Nutrition Survey Databases in China, the United States, and Korea

WEI Kuikui¹, **GONG** Haowen², **LIU** Yuxiu^{1,2,3*}, **WANG** Yan¹, **LIU** Mengdan¹, **ZHANG** Manting³, **YU** Xihui²

¹Department of Critical Care Medicine, Affiliated Jinling Hospital of Nanjing Medical University, Nanjing 210002, China

²Department of Biostatistics, School of Public Health, Southern Medical University, Guangdong 510515, China

³Department of Biostatistics, School of Public Health, Nanjing Medical University, Nanjing 211166, China

*Corresponding author: LIU Yuxiu, Professor; E-mail: liu_{yuxiu}@163.com

Abstract

The health and nutrition survey databases of China, the United States, and Korea are all focused on national health and nutrition surveillance, which have become important public resources for data mining in recent years. This paper compares and analyzes the construction status, brief histories, sampling methods, and survey contents of the three health and nutrition survey databases. The main difference lies in the types of study design, with cohort study used in China and cross-sectional surveys used in both the United States and Korea. We then conduct a visual analysis of the publication volume and research hotspots based on these databases, finding that the upward trend in publications is more prominent in the United States, and that the research hotspots of the three databases are consistent with health and nutrition survey themes. Finally, we discuss the construction and application of these databases and provide recommendations, aiming to offer researchers a pathway for relevant studies and better promote the generation of scientific evidence.

Keywords: Database; Health and nutrition examination survey; China Health and Nutrition Survey; National Health and Nutrition Examination Survey; The Korea National Health and Nutrition Examination Survey; Database

National health and nutrition status are important indicators for measuring a country's healthcare, socioeconomic development, and population quality. Since the 1960s, many countries have regularly conducted national health and nutrition surveys, building corresponding public databases to timely release reports on national health and nutrition status and to formulate and modify social policies and standards to improve population health and nutrition [1-3]. Among these, China, the United States, and Korea have developed relatively systematic and standardized health and nutrition survey databases. This paper introduces and compares the construction and development of health and nutrition survey databases in these three countries, with a focus on visual analysis of publications based on these databases, providing methods and pathways to promote the broad application of medical public data resources.

1.1 Database Construction

The China Health and Nutrition Survey (CHNS) is a cohort study jointly conducted by the Carolina Population Center at the University of North Carolina at Chapel Hill and the National Institute for Nutrition and Health at the Chinese

Center for Disease Control and Prevention [4]. CHNS is a continuously updated dynamic cohort [5] aimed at understanding China's social, economic, and demographic changes and their impact on health across the life cycle, providing scientific evidence for formulating, revising, and improving relevant policies in the new era [6]. CHNS covers information at the individual, household, and community levels for Chinese residents, providing comprehensive data support for research [7]. As China's first large-scale survey database based on individuals and households [8], it is freely accessible to scholars worldwide after registration, with the official website at <https://www.cpc.unc.edu/projects/china>.

The National Health and Nutrition Examination Survey (NHANES) is a continuous, population-based cross-sectional survey that uniquely combines health interviews with health examinations [9]. Its primary purpose is to assess the health and nutrition status of adults and children in the United States through integrated health interviews, examinations, and nutrition surveys [10]. The NHANES database is freely available to researchers worldwide for addressing future medical, environmental, and public health issues, and provides scientific evidence for government public health policy and health service planning [11]. The official NHANES website is <https://www.cdc.gov/nchs/nhanes/index.htm>.

The Korea National Health and Nutrition Examination Survey (KNHANES) is a publicly available, continuous cross-sectional monitoring system designed to conduct nationwide statistics on the health status, health behaviors, and food and nutrition consumption of the Korean population, thereby assessing the population's health and nutrition status. The KNHANES project obtains representative national information through three components: health interviews, health examinations, and nutrition surveys, providing a data foundation for developing and modifying national health policies and plans [12-13]. The official KNHANES website is <https://knhanes.kdca.go.kr/>.

1.2 Historical Background

All three national health and nutrition survey projects have long histories. China's CHNS project began in 1989 but is not a continuous annual survey, with intervals of 2-4 years between waves. To date, 11 rounds have been conducted, with data from the 2019 survey still being processed and not yet publicly released. Since the early 1960s, the Division of Health and Nutrition Examination Surveys at the US Centers for Disease Control and Prevention has conducted a series of health and nutrition surveys. From 1971 to 1994, NHANES was conducted periodically. To better understand national health and nutrition status, it was adjusted in 1999 to a continuous two-year cycle survey [14]. The KNHANES project began in 1998, with the first three phases (1998, 2001, 2005) not being continuous annual surveys. To improve the timeliness of statistical data, it was changed to a continuous annual survey starting in 2007, generating timely and reliable national annual statistics to provide a scientific basis for the country's annual health reports [15]. The NHANES survey was suspended from March 2020 to June 2021 due to COVID-19 but has since resumed, with

new COVID-19-related survey content added (such as previous infection status, severity, vaccination, antigen testing, etc.) [24]. The 2021-2022 survey data are still being processed and are not yet publicly available.

1.3 Sampling Methods

CHNS employs multistage stratified cluster random sampling due to differences in geographic environment, economic development level, population structure, health conditions, dietary patterns, and public resources among participants from different regions. In the 2015 survey, there were 360 primary sampling units, approximately 7,200 households, and over 30,000 individuals [2,16-17]. Both NHANES and KNHANES use complex multistage probability sampling methods to select participants representative of their national populations. NHANES consists of four sampling stages, with approximately 5,000 individuals planned for selection annually. Additionally, NHANES oversamples certain population subgroups (such as Hispanic, non-Hispanic Black, and non-Black Asian individuals) to improve the reliability and accuracy of health status estimates for specific subgroups [18-19]. KNHANES uses a three-stage sampling design, with approximately 10,000 individuals planned for selection annually and a response rate of about 80%. Table 1 compares the sampling methods of CHNS, NHANES, and KNHANES.

1.4 Survey Content

While the survey contents of the three countries are not identical, their overall frameworks are similar, generally divided into three components: health interview, health examination, and nutrition survey [20]. Considering that disease prevalence and risk factor exposure vary across age groups, health interview questionnaire content is adjusted according to different age ranges. Health examinations are conducted by trained professional medical personnel. Dietary surveys use 24-hour dietary recall to collect relevant dietary information [13]. Table 2 summarizes the survey contents of CHNS, NHANES, and KNHANES. To overcome the limitations of cross-sectional studies and improve the utilization of health and nutrition data, the United States and Korea have linked their databases with national cause-of-death data from their statistical offices, providing data support for studying risk factors for morbidity and mortality. The official websites of NHANES and KNHANES provide more research details and are dynamically updated, with relevant survey datasets readily accessible [21-23].

2.1 Publication Volume Analysis

We conducted a topic search in the Web of Science (WOS) database using the following search terms: TS=(“National Health and Nutrition Examination Survey” OR “NHANES”), TS=(“China Health and Nutrition Survey” OR “CHNS”), and TS=(“Korea National Health and Nutrition Examination Survey” OR “KNHANES”) to obtain annual publication volumes based on each

database. Figure 1 [Figure 1: see original paper] shows the publication trends for CHNS, NHANES, and KNHANES from 2012 to 2021. Over the past decade, publications based on CHNS, NHANES, and KNHANES have shown significant year-over-year growth. Publications based on CHNS and KNHANES are similar in volume, reaching over 300 in 2021, while publications based on NHANES are more prominent—more than five times those of CHNS and KNHANES—reaching 1,700 in 2021.

2.2 Research Hotspot Analysis

To understand research hotspots in publications using CHNS, NHANES, and KNHANES, we used the bibliometric analysis software CiteSpace [25-26] to conduct visual analysis of keywords in papers based on these databases. We searched the WOS Core Collection for publications from 2018 to October 2022 using the same search strategy, excluded irrelevant literature, and integrated all documents [27]. From 2018 to October 2022, there were 1,133, 6,912, and 1,330 publications based on CHNS, NHANES, and KNHANES, respectively. Using CiteSpace for keyword visualization, we obtained keyword co-occurrence maps for the three databases (Figures 2 [Figure 2: see original paper]~4 [Figure 4: see original paper]). The maps show similar key research content across the three databases, with large nodes for keywords such as “obesity,” “risk,” “prevalence,” “cardiovascular disease,” “physical activity,” “hypertension,” “association,” “health,” “adult,” “depression,” and “mortality”—larger nodes indicate higher frequency. For visual clarity, we used colored circles representing keywords, where circle size reflects keyword frequency and line thickness represents association strength between keywords. These results reflect current research hotspots in health and nutrition surveys and broader public health fields, providing ideas for researchers conducting database-based studies.

Of course, there are some differences in research hotspots among CHNS, NHANES, and KNHANES. NHANES and KNHANES can be linked to national cause-of-death data, enabling more extensive research on mortality [28-29]. All three databases include mental health surveys; the US and Korean surveys contain depression rating scales, so “depression” appears as a keyword with a large node, indicating substantial research in this area [30-31]. China’s CHNS includes surveys on stress and psychological status but lacks depression scale scores, so no “depression” node appears in the CHNS keyword co-occurrence map.

Discussion

This study systematically organizes health and nutrition survey databases at home and abroad, providing methods and pathways for researchers to use public databases from different countries individually or jointly. Previous scholars have introduced single health and nutrition survey databases, but this paper provides the first systematic and comprehensive comparison of multiple national health and nutrition public databases. Secondary research based on the three countries’

health and nutrition survey databases is limited, mostly comprising researchers' introductions to basic database conditions and changes. Therefore, in analyzing publication trends over the past decade, we did not separate trends for original versus secondary research but combined them to obtain publication trends for each database from 2012 to 2021.

The CHNS project has unique characteristics. It is a prospective cohort study that collects sociological and economic information and better reflects China's basic conditions. The latest CHNS has made many new breakthroughs, including upgrades and improvements to computer-assisted nutrition survey systems and richer research content, such as adding cognitive function screening for middle-aged and elderly individuals. However, it still needs to learn from NHANES and KNHANES. First, NHANES and KNHANES surveys are conducted by trained full-time investigators, with most survey items performed in Mobile Examination Centers (MECs), allowing better error control. Second, NHANES and KNHANES are continuous annual surveys with timely data release. Finally, NHANES and KNHANES can be linked to cause-of-death data from national health statistics centers, overcoming their limitations as cross-sectional studies.

The annual publication volumes for CHNS and KNHANES are similar but substantially lower than NHANES. Reasons include: (1) NHANES database information is released more timely and standardized; (2) NHANES has a longer history with continuously enriched survey content, providing researchers with more comprehensive data resources spanning longer time periods; and (3) NHANES is conducted in the US where English is the native language, and as the world's most widely used language, English enables researchers from various countries to more easily understand NHANES content. Meanwhile, CHNS and KNHANES continue to adjust and optimize their survey content and procedures to facilitate broader researcher application.

This study first systematically organizes and compares the construction status, historical background, sampling methods, and survey contents of CHNS, NHANES, and KNHANES, then conducts bibliometric analysis of publications based on these three countries' health and nutrition survey databases to help more researchers systematically understand and widely apply these public databases. Finally, using CiteSpace software for visual analysis of hotspot research content based on the three databases helps researchers understand current focal areas, aiming to provide pathways for relevant research and better promote scientific evidence generation.

This study has limitations. It only includes health and nutrition survey public databases from China, the US, and Korea; other countries were not included due to lack of public availability or absence of such surveys. Additionally, publication volumes were obtained through WOS Core Collection searches and do not represent all publications using these databases, though they reflect publication trends and differences between countries' databases and hold important reference value.

Recommendations

We propose the following recommendations for future creation and utilization of public databases:

1. **Conduct researcher training and education.** While many internationally influential achievements have been made using the three countries' health and nutrition survey databases, their application remains insufficient. Future researcher training programs could expand understanding of these databases, stimulating more researchers to mine public databases and conduct database-based scientific research.
2. **Create a global health and nutrition public platform.** The University of Southern California has already established the Global Aging and Health website (<https://g2aging.org/>), which aggregates health and aging-related data from multiple countries, providing convenience for global aging research. We hope more countries will open their health and nutrition survey data to establish a global health and nutrition survey public platform, facilitating multidimensional application of national data and providing stronger evidence for scientific research and decision-making.
3. **Adjust survey content and optimize survey processes.** Survey content should be adjusted according to disease prevalence and health hazards, with databases learning from each other's strengths. Survey processes should become more scientific and technologies more advanced to obtain more authentic and reliable data.

References

- [1] HUANG L N, WANG Z H, WANG H J, et al. Nutrition transition and related health challenges over decades in China[J]. *Eur J Clin Nutr*, 2021, 75(2): 247-252. DOI:10.1038/s41430-020-00725-4.
- [2] POPKIN B M, DU S F, ZHAI F Y, et al. Cohort Profile: the China Health and Nutrition Survey—monitoring and understanding socio-economic and health change in China, 1989-2011[J]. *Int J Epidemiol*, 2010, 39(6): 1435-1440. DOI:10.1093/ije/dyp322.
- [3] ZHAO L Y, GUO Q Y, LI S J, et al. Strengthening nutrition surveys and monitoring to improve the nutrition and health status of Chinese residents[J]. *Journal of Hygiene Research*, 2019, 48(4): 517-522. DOI:10.19813/j.cnki.weishengyanjiu.2019.04.027.
- [4] ZHOU C, LIU C Z, ZHANG Z X, et al. Variety and quantity of dietary protein intake from different sources and risk of new-onset diabetes: a Nationwide Cohort Study in China[J]. *BMC Med*, 2022, 20(1): 6. DOI:10.1186/s12916-021-02199-8.
- [5] LI J, YANG Q L, AN R, et al. Famine and trajectories of body mass index, waist circumference, and blood pressure in two generations: results

from the CHNS from 1993-2015[J]. *Hypertension*, 2022, 79(3): 518-531. DOI:10.1161/HYPERTENSIONAHA.121.18022.

[6] CHANG H, TIAN G X, LIU Q Q, et al. CHNS database architecture and data application extraction methods and procedures[J]. *Chinese Journal of Evidence-Based Cardiovascular Medicine*, 2018, 10(9): 1043-1047. DOI:10.3969/j.issn.1674-4055.2018.09.04.

[7] YANG J, LI Y J, LIU Q Q, et al. Brief introduction of medical database and data mining technology in big data era[J]. *J Evid Based Med*, 2020, 13(1): 57-69. DOI:10.1111/jebm.12373.

[8] ZHANG B, ZHAI F Y, DU S F, et al. The China health and nutrition survey, 1989-2011[J]. *Obes Rev*, 2014, 15 Suppl 1(01): 2-7. DOI:10.1111/obr.12119.

[9] GUO X J, TIAN G X, PAN Z Y, et al. Introduction to NHANES project and data extraction process[J]. *Chinese Journal of Evidence-Based Cardiovascular Medicine*, 2019, 11(6): 654-657. DOI:10.3969/j.issn.1674-4055.2019.06.03.

[10] PAULOSE-RAM R, BURT V, BROITMAN L, et al. Overview of Asian American data collection, release, and analysis: national health and nutrition examination survey 2011-2018[J]. *Am J Public Health*, 2017, 107(6): 916-921. DOI:10.2105/AJPH.2017.303815.

[11] AHLUWALIA N, DWYER J, TERRY A, et al. Update on NHANES dietary data: focus on collection, release, analytical considerations, and uses to inform public policy[J]. *Adv Nutr*, 2016, 7(1): 121-134. DOI:10.3945/an.115.009258.

[12] KIM Y. The Korea National Health and Nutrition Examination Survey (KNHANES): current status and challenges[J]. *Epidemiol Health*, 2014, 36: e2014002. DOI:10.4178/epih/e2014002.

[13] KWEON S, KIM Y, JANG M J, et al. Data resource profile: the Korea national health and nutrition examination survey (KNHANES)[J]. *Int J Epidemiol*, 2014, 43(1): 69-77. DOI:10.1093/ije/dyt228.

[14] TU J, HE H J, HU Y D, et al. Application and implications of information systems in the US National Health and Nutrition Examination Survey[J]. *Chinese Journal of Epidemiology*, 2022, 43(7): 1127-1133. DOI:10.3760/cma.j.cn112338-20211109-00871.

[15] OH K, KIM Y, KWEON S, et al. Korea National Health and Nutrition Examination Survey, 20th anniversary: accomplishments and future directions[J]. *Epidemiol Health*, 2021, 43: e2021025. DOI:10.4178/epih.e2021025.

[16] LI Y Z. Study on the relationship between skinfold thickness trajectory in early adulthood and new-onset hypertension—based on CHNS data[D]. Jinan: Shandong University, 2021.

[17] China Health and Nutrition Survey Project Group. Dietary nutrient intake status and trends among residents in nine provinces of China from 1989-2009 (I):

Overall scheme of the health and nutrition survey project[J]. *Acta Nutrimenta Sinica*, 2011, 33(3): 234-236. DOI:10.13325/j.cnki.acta.nutr.sin.2011.03.018.

[18] WANG L, LI X G, WANG Z X, et al. Trends in prevalence of diabetes and control of risk factors in diabetes among US adults, 1999-2018[J]. *JAMA*, 2021, 326(8): 1-13. DOI:10.1001/jama.2021.9883.

[19] AKINBAMI L J, CHEN T C, DAVY O, et al. National health and nutrition examination survey, 2017-march 2020 prepandemic file: sample design, estimation, and analytic guidelines[J]. *Vital Health Stat 1*, 2022(190): 1-36.

[20] Centers for Disease Control and Prevention. National Health and Nutrition Examination Survey: 1999–2022 Survey Content Brochure[EB/OL]. [2023-1-11]. <https://www.cdc.gov/nchs/data/nhanes/survey-content>.

[21] China Health and Nutrition Survey[EB/OL]. [2023-01-11]. <http://www.cpc.unc.edu/projects/china>.

[22] National Health and Nutrition Examination Survey[EB/OL]. [2023-01-11]. <https://www.cdc.gov/nchs/nhanes/index.htm>.

[23] Korea National Health and Nutrition Examination Survey[EB/OL]. [2023-01-11]. <https://knhanes.kdca.go.kr>.

[24] PAULOSE-RAM R, GRABER J E, WOODWELL D, et al. The national health and nutrition examination survey (NHANES) in the COVID-19 environment[J]. *Am J Public Health*, 2021, 111(12): 2149-2156. DOI:10.2105/AJPH.2021.306517.

[25] CHEN Y, CHEN C M, LIU Z Y, et al. Methodological functions of CiteSpace knowledge mapping[J]. *Studies in Science of Science*, 2015, 33(2): 242-253. DOI:10.16192/j.cnki.1003-2053.2015.02.009.

[26] LI J, CHEN C M. CiteSpace: Scientific text mining and visualization[M]. Beijing: Capital University of Economics and Business Press, 2016.

[27] HUANG X X, HUANG T, YANG R, et al. Introduction to the application of UK Biobank data[J]. *Chinese Journal of Evidence-Based Medicine*, 2022, 22(9): 1099-1107.

[28] ZHANG Y B, CHEN C, PAN X F, et al. Associations of healthy lifestyle and socioeconomic status with mortality and incident cardiovascular disease: two prospective cohort studies[J]. *BMJ*, 2021, 373: n604. DOI:10.1136/bmj.n604.

[29] GU K M, MIN S H, CHO J. Sleep duration and mortality in patients with diabetes: results from the 2007-2015 Korea national health and nutrition examination survey[J]. *Diabetes Metab*, 2022, 48(3): 101312. DOI:10.1016/j.diabet.2021.101312.

[30] PARK G N, KIM J O, OH J W, et al. Association between anemia and depression: the 2014, 2016, and 2018 Korea national health and nutrition examination survey[J]. *J Affect Disord*, 2022, 312: 86-91. DOI:10.1016/j.jad.2022.06.015.

[31] DONG L, XIE Y W, ZOU X H. Association between sleep duration and depression in US adults: a cross-sectional study[J]. J Affect Disord, 2022, 296: 183-188. DOI:10.1016/j.jad.2021.09.075.

Author Contributions: WEI Kuikui was responsible for conceptualization, literature collection, and drafting the manuscript; GONG Haowen conducted result analysis, interpretation, and visualization; LIU Yuxiu proposed the research idea, designed the framework, and took overall responsibility for the manuscript; WANG Yan and LIU Mengdan reviewed and revised the article; ZHANG Manting and YU Xihui provided software technical support.

Conflict of Interest Statement: The authors declare no conflicts of interest.

Received: 2023-02-25; **Revised:** 2023-07-12

Edited by: ZHAO Yuecui

Note: Figure translations are in progress. See original paper for figures.

Source: ChinaXiv — Machine translation. Verify with original.