

Dietary diversity and healthy weight management in the population: a data-driven analytical approach

Wenwen Zhang, Jianting Ye *, Haodi Zhang

Department of Electronic Business, South China university of Technology, Guangdong, 510641, China

* Corresponding Author

Abstract. Cardiovascular and cerebrovascular diseases (CVDs), diabetes mellitus, malignant tumours and chronic obstructive pulmonary disease (COPD), as typical chronic non-communicable diseases (NCDs), have increasingly become the major factors threatening the health of Chinese residents. In view of this, this study aimed to deeply analyse the key indicators and their interrelationships affecting the health status of the population, using the eight dietary balance criteria proposed in the newly revised Dietary Guidelines for Chinese Residents of the Chinese Society of Nutrition to screen and process the relevant health indicators. By constructing a comprehensive evaluation model and visualising the data, this study reveals the irrational factors in residents' dietary habits, such as high-fat diet and excessive alcohol consumption, which pose potential risks to residents' health. In addition, this paper explores the relationship between residents' dietary habits, exercise frequency and healthy body weight, aiming to provide a scientific basis and practical guidance for improving public health.

Keywords: Dietary Diversity; Healthy Weight Management; Chronic Non-Communicable Diseases; Data-Driven Analysis; Dietary Guidelines for Chinese Residents.

1. Introduction

In the health agenda of the 21st century, public health awareness has increased in tandem with rapid economic growth and significant improvement in social well-being. However, along with lifestyle changes, chronic non-communicable diseases (NCDs) such as cardiovascular and cerebrovascular diseases, diabetes mellitus, malignant neoplasms and chronic obstructive pulmonary disease (COPD) have gradually become the major obstacles limiting the health of our population [1]. Modern medical research shows that health is not only affected by genetic factors, but also closely related to individual dietary habits, physical activity, occupational life and other social life factors. Therefore, how to improve the overall health of residents through scientific dietary regulation, reasonable physical activity and the promotion of healthy lifestyles has become the focus of current public health research.

In this context, this study analyses the results of a questionnaire and data on the epidemiology of chronic non-communicable diseases and their related influencing factors conducted by the health research department of a city, and combines them with the eight dietary balance standards put forward in the latest revision of the Dietary Guidelines for Chinese Residents of the Chinese Academy of Nutrition, with the aim of discussing the key factors affecting the health of the population [2]. The study aims to explore the key factors affecting the health of the population. This study focuses on the nutritional balance of residents' daily diet, the diversity of food intake and the healthiness of their lifestyles, and seeks to reveal the intrinsic links between the incidence of chronic diseases and lifestyles, and to propose scientific and practical health promotion strategies based on them [3].

By comprehensively analysing the dietary habits of the population, this study will not only focus on whether the nutritional intake is balanced and whether there is excessive intake of certain food groups, but will also explore whether the intake of key food groups such as vegetables, fruits, grains, meat and dairy products meets the recommended standards of the Dietary Guidelines for Chinese Residents. By comprehensively assessing the problems in the dietary habits of the population, this study aims to provide a basis for improving public health and suggesting targeted nutritional interventions.



2. Related Work

With the rapid rise in obesity rates worldwide, healthy weight management has become an important topic in public health research. Especially in China, the incidence of chronic non-communicable diseases (NCDs), including obesity, cardiovascular diseases, and diabetes, has increased dramatically due to rapid economic development and lifestyle changes. It is well known that dietary habits have an important impact on weight management and prevention of chronic diseases [4]. Therefore, this study aimed to explore the relationship between dietary diversity and healthy weight management in the population, using data-driven analyses to provide insights into the associated factors.

This study is based on the results of a questionnaire and data from the Epidemiology of Chronic Non-Communicable Diseases and Related Influencing Factors survey conducted by the Health Research Department of a city on some residents. The study first analysed the dietary habits of the residents with reference to the eight dietary balance criteria proposed in the Dietary Guidelines for Chinese Residents, focusing on the relationship between dietary diversity and weight management [5]. Through the construction of a comprehensive evaluation model and visualisation, the study revealed irrational factors in the residents' dietary habits, such as high-fat diet and excessive alcohol consumption, which pose potential threats to healthy weight management.

The study found that the majority of the population consumed less than four food groups per day, resulting in inadequate dietary balance. In addition, the diversity of food intake of the population was generally low, which might lead to inadequate nutrient intake, thus affecting weight management and health status [6]. Meanwhile, the BMI of the overall residents was high, which indicates that obesity is more prevalent among the study population and requires intervention through improved dietary habits and increased physical activity.

It is worth noting that although the frequency and intensity of exercise was generally good among the residents, there is a need to further improve the quality of diet to achieve better healthy weight management. The study also pointed out that the residents' intake of vegetables, fruits and dairy products was generally insufficient, and this unbalanced dietary structure may lead to insufficient nutrient intake, which in turn may affect health status [7].

Based on these findings, this study emphasises the importance of increasing dietary diversity in healthy weight management. By increasing the intake of food variety, a more complete range of nutrients can be ensured, thereby contributing to the maintenance of a healthy body weight and the prevention of chronic diseases. In addition, this study suggests that in order to effectively manage weight and improve health, public health policies should place more emphasis on raising public awareness of healthy eating and encouraging residents to follow scientific dietary guidelines and adopt balanced eating habits.

In conclusion, this study provides insights into the relationship between dietary diversity and healthy weight management in the population through data-driven analyses, providing a scientific basis for the development of effective public health strategies. Future studies should further explore the specific mechanisms of the impact of dietary diversity on health in different populations and how public health interventions can effectively promote the adoption of healthier eating habits among residents.

3. Modelling and Solving

Firstly the data is preprocessed. Normal data can guarantee the reliability of the model and is a prerequisite for solving the problem. It can be found that there are a lot of duplicates and missing values in the given data, in this paper, we use python to delete the abnormal values and duplicates by calling `drop_duplicates` in the pandas library, and then fill in the empty values.

Based on the data as well as the dietary guidelines, several evaluation indicators were first defined, which included: daily intake of food variety with a reasonable combination, eat more, move more, healthy weight, eat more fruits and vegetables, dairy, whole grains, soybeans, moderate amount of

fish, poultry, eggs, lean meat, less salt and less oil, and sugar control and alcohol restriction, and so on [8].

While constructing the evaluation indicators, attention was paid to the conversion between different units to ensure the consistency of the data outline. The relevant data involved in the evaluation indexes are summarised on a daily, weekly and monthly basis in order to carry out the analysis of the reasonableness of the residents' eating habits. For this indicator, this paper considers two aspects: whether the daily dietary intake is up to the standard or not; whether the daily and weekly food intake is up to the standard or not.

(1) Whether the daily diet includes cereals and potatoes, vegetables and fruits, livestock, poultry, fish, eggs and milk, and legumes. i is 1,2,3,4 for each of these four dietary categories in turn, and x_i denotes the daily intake of dietary category i (intake of $x_i = 1$, no intake of $x_i = 0$).

Based on common sense, we agreed:

Cereals and Potatoes include Edible Rice, Edible Wheat Flour, Edible Mixed Grains, Edible Potatoes, Edible Fried Pasta;

Vegetables and fruits include fresh vegetables, fresh fruits;

Livestock, poultry, fish, eggs and milk include pork, beef, lamb, poultry, offal, aquatic products, eggs, fresh milk, milk powder, yoghurt;

Soya beans include tofu, bean curd, dried bean curd, soya milk, dried soya beans (soybean, black bean, green bean).

Calculation of the total daily intake of each food group by the population:

$$\text{Sum of the population's daily intake of each food group} = \sum_{i=1}^4 x_i \quad (1)$$

(2) Daily food types and weekly food types, when counting these two quantities, we consider that the treatment of the daily and weekly intake of each of the above food types is realised by borrowing the idea of the sign function.

$$\text{sign}(x) = \begin{cases} -1 & x < 0 \\ 0 & x = 0 \\ 1 & x > 0 \end{cases} \quad (2)$$

$$\text{Daily food variety} = \sum \text{sign}(\text{daily consumption frequency of food}) \quad (3)$$

$$\text{Weekly food variety} = \sum \text{sign}(\text{weekly consumption frequency of food}) \quad (4)$$

Body mass index (BMI) was used as an indicator of healthy weight. BMI is calculated by squaring weight (kg)/height (m) [9]. This paper also considers the number of days of exercise per week for residents, and the intensity of exercise per day for residents.

(1) Calculation of BMI:

$$\text{BMI} = \text{Weight (kg)} / (\text{Height (m)})^2 \quad (5)$$

(2) Calculation of the number of days of exercise per week for the population:

For the number of days of physical activity per week (average record), this paper considers the following treatment of the values in the column of "Whether or not to participate in physical activity":

0, 1 corresponds to almost no exercise, i.e. 0 days per week;

2 Consider 2 days of exercise per week;

3 Consider 4 days of exercise per week;

4 Consider 6 days of exercise per week

(3) Calculation of daily exercise intensity:

$$\begin{aligned}
 & \text{Intensity of daily exercise} \\
 &= \text{intensity of average exercise in sports} \\
 & * \text{average daily duration of physical exercise}
 \end{aligned} \tag{6}$$

This indicator is measured by calculating the total daily intake of vegetables, fruits, and dairy foods per day [10]. According to common sense, vegetables, fruits and dairy products are rich in vitamins and proteins, which play a vital role in human health, growth and development.

Calculation of food intake:

$$\begin{aligned}
 & \text{Daily Vegetable Intake of the population} \\
 &= \text{Frequency of consumption of fresh} \frac{\text{vegetables}}{\text{day}} \\
 & * \text{Average quantity per serving} * \text{Conversion factor}
 \end{aligned} \tag{7}$$

$$\begin{aligned}
 & \text{Daily fruit intake of the population} \\
 &= \text{frequency of consumption of} \frac{\text{fruits}}{\text{day}} * \text{average quantity per serving} \\
 & * \text{conversion factor}
 \end{aligned} \tag{8}$$

$$\begin{aligned}
 & \text{Daily milk intake of the population} \\
 &= \left(\text{frequency of consumption of fresh} \frac{\text{milk}}{\text{day}} \right. \\
 & + \text{frequency of consumption of powdered} \frac{\text{milk}}{\text{day}} \\
 & \left. + \text{frequency of consumption of} \frac{\text{yoghurt}}{\text{day}} \right) \\
 & * \text{average quantity per consumption} * \text{conversion factor}
 \end{aligned} \tag{9}$$

(Conversion factor: 1 tael = 50g for mass unit conversion)

This article calculates the total daily intake of fish, poultry, eggs and lean meat, which are important sources of high-quality protein and beneficial to people's health.

$$\begin{aligned}
 & \text{Daily meat intake of the population} \\
 &= \left(\text{frequency of consuming} \frac{\text{poultry}}{\text{day}} \right. \\
 & + \text{frequency of using aquatic} \frac{\text{products}}{\text{day}} + \text{frequency of consuming} \frac{\text{eggs}}{\text{day}} \\
 & + \text{frequency of consuming beef and} \frac{\text{mutton}}{\text{day}} \\
 & \left. + \text{frequency of consuming} \frac{\text{pork}}{\text{day}} \right) \\
 & * \text{average amount per serving} * \text{conversion factor}
 \end{aligned} \tag{10}$$

Calculate the total daily intake of oil, salt and alcohol. Excessive oil, salt intake may increase the risk of cardiovascular disease, while various studies have shown that excessive intake of alcohol is harmful to health.

Noting that oil and salt intake was counted for the last year on a household basis, the number of people eating at home in a household is considered below:

The number of people eating at home in a household is divided into weekdays and weekends and analysed separately for breakfast, Chinese food and dinner attendance.

On weekends, the percentage of breakfast versus Chinese food or breakfast versus dinner with an absolute difference of 0 or 1 is above 0.86 on weekends, and the percentage of Chinese food and dinner with an absolute difference of 0 or 1 is 0.96 on weekends.

Therefore, the data distribution is relatively concentrated, and we can consider dividing three meals a day by three directly as the number of people eating at home on weekends.

Weekdays: the data distribution on weekdays is relatively dispersed, and here we consider the number of people eating at home in the morning, midday and evening on weekdays to be processed by data dimensionality reduction using SPSSPRO [11].

The result of KMO test shows that the value of KMO is 0.721, at the same time, the result of Bartlett's spherical test shows that the significance P-value is 10⁻⁴ orders of magnitude, presenting significance at the level of rejection of the original hypothesis, there is a correlation between the variables, and the principal component analysis is valid to the extent of a relatively appropriate, so we consider the use of PCA on the weekday morning, midday and evening to downsize the three indicators into a corresponding indicators to be analysed.

The dimensionality reduction indicator is $F = 0.3554070865344561 \times \text{number of people eating breakfast at home on weekdays} + 0.3554070865344561 \times \text{number of people eating midday meal at home on weekdays} + 0.3765298553923662 \times \text{number of people eating dinner at home on weekdays}$, and it consists of $(0.3554070865344561 + 0.3554070865344561 + 0.3765298553923662) = 1.0873440284612$ the sum of the coefficients is slightly greater than 1, so round down to the value F.

So, the PEOPLE expression can be derived as:

$$\begin{aligned} \text{people} = & ([F] \times 5 \\ & + [\text{number of people eating breakfast at home on weekends} \\ & + \text{number of people eating lunch at home on weekends} \\ & + \text{number of people eating dinner at home on weekends}] \div 3 \times 2) \\ & \div 7 \end{aligned} \quad (11)$$

At the same time to people for 0 data to set 1 processing, to ensure that the data is real at the same time to meet the needs of its follow-up as a divisor, to ensure that the follow-up work is carried out normally.

(1) Calculation of oil, salt and intake:

$$\text{Daily oil intake} = (\text{vegetable oil intake} + \text{animal oil intake}) * \text{conversion factor} \quad (12)$$

$$\text{Daily salt intake} = \text{salt intake} * \text{conversion factor} \quad (13)$$

The conversion factor is specified here:

In this paper, it is assumed that the oil and salt intake of the people sitting at home eating in a household is the same. Now what needs to be estimated is: the average daily intake of the population.

This paper is now known data is the family monthly salt and oil content, here to make the assumption that the amount of oil and salt intake is equal to each day of the month, then the conversion coefficient is divided into three parts: the amount of units of conversion, the conversion of units of time, the conversion of units of the family.

Where: 1 tael = 500 grams 1 month = 30 days Average number of people eating at home per day people.

$$\text{Conversion factor} = 500 \div 30 \div \text{people} \quad (14)$$

(2) Calculation of alcohol intake:

Daily alcohol intake

$$= \sum (\text{frequency of drinking} * \text{amount of alcohol per drink}) * \text{conversion factor} \quad (15)$$

(Conversion factor: 1 tael = 50 grams for mass unit conversion)

All the above data are integrated into a pandas DataFrame, with each indicator calculated above as a column, and the rows corresponding to the data of each resident. This allows for each indicator to be analysed either individually or together with other indicators so that we get a more accurate assessment of healthy eating.

The main objective of the model is to provide a more comprehensive quantitative assessment of each individual's dietary habits and to identify possible problems in the dietary habits of the population, e.g., is the population's diet varied in terms of the variety of foods consumed and the amount of foods consumed per day up to the standard? Does the resident's daily physical activity meet the standard? Is the daily intake of oil, salt and alcohol excessive? And so on. In this way, we can give specific advice on how to improve the dietary habits of each individual.

By drawing histograms, the data for each indicator is visualised and the distribution of the data is understood graphically. The datavisualised presentation of each indicator is shown in Figure 1 below:

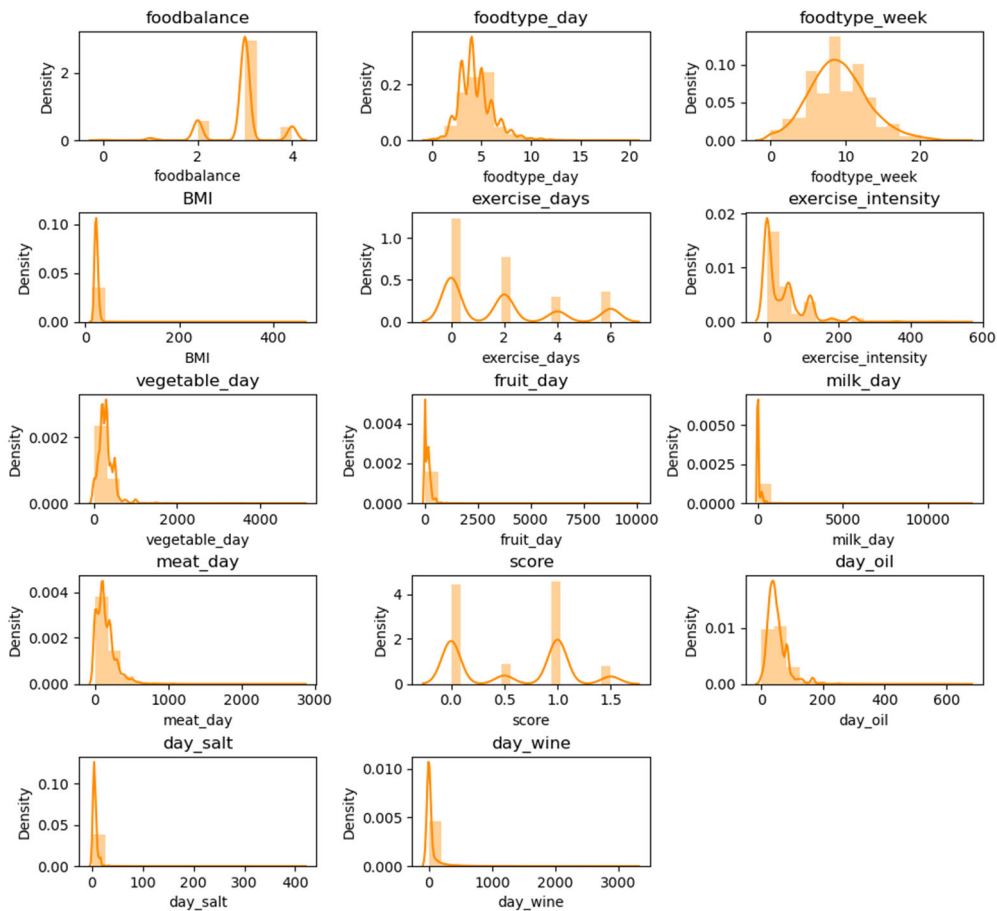


Figure 1. Visualisation of the data for each indicator

It can be seen that most of the data have small quantitative indicators, and it can also be seen that there are very small differences in the indicators for the sampled residents in terms of Body Mass (BMI), Daily Fruit Intake, Daily Milk Intake, Daily Salt and Alcohol Intake. Indicators with larger differences include the total daily intake of all food groups, daily food group intake, weekly food group intake, number of days of exercise per week, daily intensity of exercise, the degree of

satisfaction with the indicators of eating aquatic products twice a week and one egg per day, and the daily intake of cooking oil.

Scatter plots were drawn for each indicator, and standard lines in line with the Dietary Guidelines for Chinese Residents were added to the plots so that we could compare the values taken for different indicators. The standard line can include thresholds or minimum/maximum standards, which are used to judge whether the data are reasonable or not, so as to infer the reasonableness of the residents' diet. The scatter plot is shown in Figure 2 below:

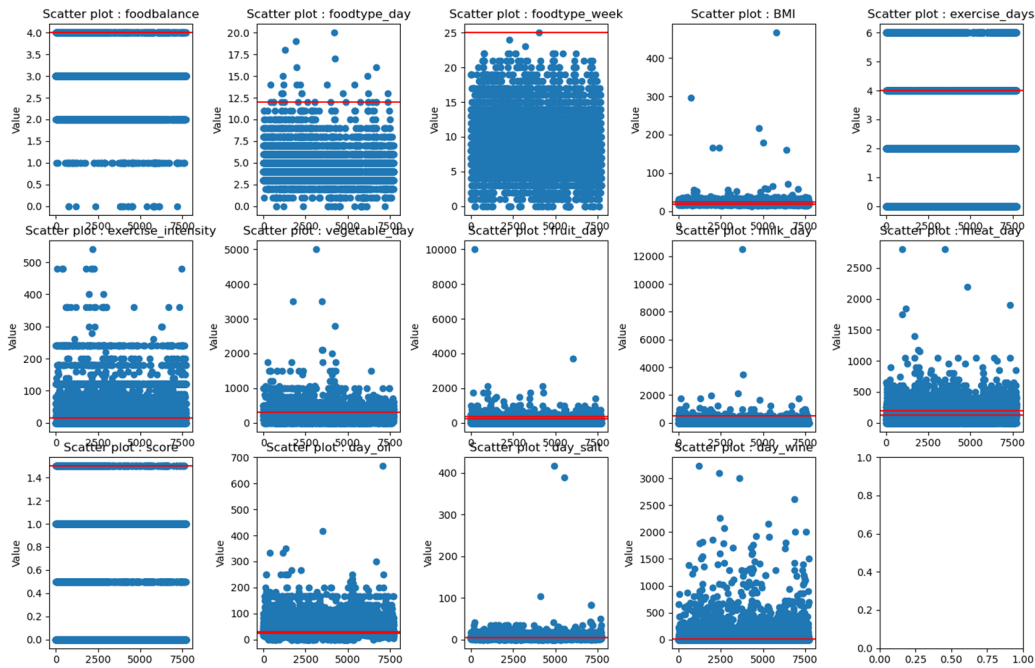


Figure 2. Scatterplot of indicators

In Figure 2 above, the scatter points in the scatter plot represent the values of the corresponding indicators for each sampled urban resident, and the red horizontal lines represent the thresholds of the indicators recommended by the Dietary Guidelines for Chinese Residents.

4. Conclusion

This study reveals several key findings through in-depth analyses of visual data on the dietary habits of urban residents and their impact on health. Firstly, the daily dietary structure of the vast majority of residents failed to achieve the balance recommended by the Dietary Guidelines for Chinese Residents, as evidenced by an intake of less than four food groups. This result points out that public awareness and practice of dietary balance still need to be strengthened. Secondly, residents' intake of food diversity was significantly insufficient, which may be related to factors such as partiality and solidified dietary habits, and needs to be further improved through public health education.

In addition, the study found that the residents' body mass index (BMI) was generally high, which not only reflected the prevalence of the obesity problem, but also hinted at the risk of a variety of chronic diseases that may arise as a result. It is worth noting that although residents' exercise frequency and intensity were generally good, indicating high levels of health awareness and physical activity, there is still a need to further improve the quality of their diets to match their healthy lifestyles.

The study also pointed out that although most residents' vegetable intake was up to standard, their intake of meat, fruit and dairy products was generally inadequate, with significant gaps in high-quality protein and vitamin supplementation in particular. This unbalanced dietary structure may lead to insufficient intake of nutrients, which in turn affects the overall health of the residents.

Finally, residents' daily intake of oil, salt, and alcohol exceeded the recommended health thresholds, especially the excessive intake of oil and alcohol, which poses a significant risk to cardiovascular

health. Therefore, this study highlights the need to promote healthy eating habits through public health policies and individual behavioural changes, especially limiting the intake of high-fat, high-salt and high-alcohol diets.

In summary, this study highlights the importance of balanced diets, increasing food diversity, controlling BMI, improving the quality of physical activity, and adjusting oil, salt and alcohol intake through in-depth analyses of the dietary habits of the population. Future public health intervention strategies should focus more on improving residents' knowledge of healthy eating and promoting healthier eating habits and lifestyles through comprehensive health education and environmental support.

5. Discussion

This study explored the relationship between dietary diversity and healthy weight management among residents through a data-driven analytical approach, revealing the importance of dietary habits for healthy weight maintenance. The results of the study showed that the majority of residents failed to meet the food variety intake recommended by the Dietary Guidelines for Chinese Residents, resulting in dietary imbalance. In addition, residents' low food intake diversity was associated with high BMI, suggesting that dietary habits are strongly associated with the risk of obesity and other chronic diseases.

Dietary diversity is one of the key factors for nutritional adequacy, ensuring the intake of multiple nutrients from different foods. However, the present study found that lack of diversity in food groups may lead to nutritional imbalance and increase the risk of chronic diseases. This is consistent with findings in the existing literature that dietary diversity is associated with lower rates of obesity and chronic disease. Therefore, increasing dietary diversity is one of the most important strategies to improve the health status of the population.

In addition, this study highlights the importance of combining exercise and dietary habits for healthy weight management. Although the frequency and intensity of exercise was generally good, it failed to match with healthy eating habits, resulting in a generally high BMI. This suggests that a single health behaviour change (e.g. increased exercise) may not be sufficient to achieve the goal of healthy weight management, and that a combination of measures, including improved diet quality and increased physical activity, is needed to maintain a healthy weight.

Notably, this study also identified deficiencies in residents' intake of vegetables, fruits and dairy products, which may lead to inadequate intake of vitamins and minerals and affect health. Therefore, encouraging residents to increase their intake of these foods is key to improving diet quality and promoting healthy weight management.

A limitation of this study is that the data were derived from residents of specific regions and may not be fully representative of the dietary habits and health status of the entire population. Therefore, future studies need to validate these findings in a wider range of regions and populations. In addition, the assessment of dietary habits was mainly based on self-reported questionnaires, which may be subject to reporting bias.

In summary, this study highlights the importance of improving dietary diversity in the prevention of obesity and chronic diseases by analysing the relationship between dietary habits and healthy weight management in the population. Future public health strategies should place greater emphasis on improving diet quality in combination with increased physical activity to promote healthy weight management and overall health in the population. In addition, interventions targeting dietary habits need to be individually designed, taking into account the specific needs and preferences of different populations, in order to increase the effectiveness of interventions.

References

- [1] Castelló, A., Buijsse, B., Martín, M., Ruiz, A., Casas, A. M., Baena-Cañada, J. M., ... & Pollán, M. (2016). Evaluating the applicability of data-driven dietary patterns to independent samples with a focus on measurement tools for pattern similarity. *Journal of the Academy of Nutrition and Dietetics*, 116(12), 1914-1924.
- [2] Teng, Y., Jing, H., Chacha, S., Wang, Z., Huang, Y., Yang, J., ... & Dang, S. (2023). Maternal Dietary Diversity and Birth Weight in Offspring: Evidence from a Chinese Population-Based Study. *International Journal of Environmental Research and Public Health*, 20(4), 3228.
- [3] Makate, M., & Nyamuranga, C. (2023). The long-term impact of education on dietary diversity among women in Zimbabwe. *Review of Development Economics*, 27(2), 897-923.
- [4] Dunlop, E., Boorman, J. L., Hambridge, T. L., McNeill, J., James, A. P., Kiely, M., ... & Black, L. J. (2023). Evidence of low vitamin D intakes in the Australian population points to a need for data-driven nutrition policy for improving population vitamin D status. *Journal of Human Nutrition and Dietetics*, 36(1), 203-215.
- [5] Zang, C., Zhang, Y., Xu, J., Bian, J., Morozyuk, D., Schenck, E. J., ... & Kaushal, R. (2023). Data-driven analysis to understand long COVID using electronic health records from the RECOVER initiative. *Nature Communications*, 14(1), 1948.
- [6] Dhingra, L. S., Shen, M., Mangla, A., & Khera, R. (2023). Cardiovascular care innovation through data-driven discoveries in the electronic health record. *The American Journal of Cardiology*, 203, 136-148.
- [7] Yang, Y., Piao, W., Cai, S., Huang, K., Yuan, C., Cheng, X., ... & Yu, D. (2023). Comparison of data-driven identified hypertension-protective dietary patterns among Chinese adults: based on a nationwide study. *European Journal of Nutrition*, 1-21.
- [8] Machado, P., McNaughton, S. A., Livingstone, K. M., Hadjidakou, M., Russell, C., Wingrove, K., ... & Lawrence, M. (2023). Measuring adherence to sustainable healthy diets: a scoping review of dietary metrics. *Advances in Nutrition*, 14(1), 147-160.
- [9] Musanase, C., Vodacek, A., Hanyurwimfura, D., Uwitonze, A., & Kabandana, I. (2023). Data-Driven Analysis and Machine Learning-Based Crop and Fertilizer Recommendation System for Revolutionizing Farming Practices. *Agriculture*, 13(11), 2141.
- [10] Lee, U., Jung, G., Ma, E. Y., San Kim, J., Kim, H., Alikhanov, J., ... & Kim, H. (2023). Toward data-driven digital therapeutics analytics: Literature review and research directions. *IEEE/CAA Journal of Automatica Sinica*, 10(1), 42-66.
- [11] Lee, S. Y., Hayes, L., Ozaydin, B., Howard, S., Garretson, A., Bradley, H., ... & DeLaney, E. (2023). Data-Driven Decision Support for Post-Hospitalization Diabetes Case Management Referral: A Predictive Analysis.