JISS Journal of Indonesian Social Science

Vol. 6, No. 10, October 2025
E-ISSN:2723 – 6692
P-ISSN:2723– 6595
http://jiss.publikasiindonesia.id/

Segmentation of Medical Check-Up (MCU) Results Using K-Means Clustering for Health Risk Profile Identification

Syarifudin, Era Pamadia, Jerry Heikal
Universitas Bakrie, Indonesia
Email: arieve.syarifudin@gmail.com, erapamadia111@gmail.com, jerry.heikal@bakrie.ac.id

KEYWORDS

K-Means clustering, medical check-up, health segmentation, metabolic risk, positioning.

ABSTRACT

Medical Check-Up (MCU) is one of the methods for early detection of metabolic and cardiovascular disorders. Computational analysis of MCU results data can be used to group individuals based on their risk profiles, thereby supporting more targeted promotive and preventive strategies. This research aims to identify the segmentation of individual groups based on the clinical parameters of MCU results using the K-Means clustering method, as well as evaluate the potential of each cluster as a target for health interventions, particularly in the context of positioning preventive products or services. The data were analyzed using the K-Means algorithm to group individuals based on the similarity of their health profiles. Cluster validation was carried out through clinical interpretation and examination of the characteristics of each cluster center. From the analysis, five clusters with distinct characteristics were identified. One of the most prominent is Cluster 2, consisting of middle-aged men (average age 44 years) with high levels of total cholesterol, LDL, and triglycerides, while blood pressure and blood sugar remain within normal limits. This condition indicates a moderate metabolic risk that is still reversible. This cluster is considered strategic for targeting the development of preventive health-based digital products and services. MCU outcome-based segmentation through the K-Means clustering method proves effective in identifying specific health risk groups. Cluster 2 shows great potential as the primary target for preventive intervention programs, owing to its medical profile—which can still be improved—and psychographic characteristics that support the acceptance of digital and promotional approaches.

INTRODUCTION

Today's public health problems are becoming increasingly complex due to modern lifestyle changes that have driven a rise in cases of non-communicable diseases (NCDs). Non-communicable diseases (NCDs) are now a major threat to global public health, including in Indonesia. According to the World Health Organization (2021), about 71% of all deaths worldwide are caused by NCDs, which include heart disease, stroke, cancer, chronic lung disease, and diabetes (Li et al., 2025; Organization, 2023; Parato, Parato, Fedacko, & Magomedova, 2024; Sun et al., 2022). A similar trend is observed in Indonesia, where the prevalence of these diseases has shown a significant increase. Based on the 2023 Basic Health Research (*Riskesdas*) and the public health monitoring report by the Indonesian Ministry of Health in early 2024, there is a tendency for the prevalence of NCDs to rise, largely triggered by unhealthy lifestyles and metabolic risk factors such as dyslipidemia, hypertension, obesity, and hyperglycemia.

One of the clinical parameters most often serving as an early indicator of NCD risk is elevated total cholesterol levels and their fractions, particularly LDL (low-density lipoprotein) and triglycerides (Acuna, Sanchez, Soler, & Alvis, 2015; Kosmas et al., 2023; Nantsupawat et al., 2019; Raja et al., 2023). *Riskesdas* 2023 data show that more than 35% of Indonesia's adult population has cholesterol levels above normal limits (≥200 mg/dL), while the prevalence of hypercholesterolemia in the 25–45 age group has increased 1.5 times compared to a decade ago. This surge not only indicates an increased risk of coronary heart disease and stroke but also reflects a major challenge for the preventive and promotive health care system.

High cholesterol, or dyslipidemia, is an abnormal fat metabolism condition characterized by elevated total, LDL, and/or triglyceride cholesterol levels, along with decreased HDL levels. According to the American Heart Association (2023), LDL levels above 130 mg/dL and triglycerides above 150 mg/dL already fall within the medium to high-risk categories, depending on other accompanying risk factors.

Longitudinal studies in Indonesia, such as the Indonesian Family Life Survey (IFLS) and the Jakarta Metropolitan Area Cohort, reveal that high cholesterol significantly increases the risk of coronary heart disease by 3.4 times over a 10-year period, especially among men of productive age with a BMI greater than 25 kg/m². Even among non-smoking and non-hypertensive groups, cholesterol remains a strong determinant of cardiovascular mortality (Aliyu, Chiroma, Jajere, & Gujba, 2015; Goutama, Hendsun, & Su, 2020; Maugeri et al., 2019).

Other empirical data from BPJS Kesehatan (2024) indicate that more than 40% of health insurance claims for heart disease and stroke come from patients who previously had a history of dyslipidemia and were not properly treated preventively. This finding demonstrates a direct link between high cholesterol and the burden on national health financing.

Although cholesterol is biologically necessary for the body, it poses a serious threat when its levels exceed normal limits, particularly in the form of LDL, known as “bad cholesterol.” Elevated cholesterol levels are significantly associated with various chronic health conditions such as coronary heart disease, stroke, type 2 diabetes, and other metabolic disorders.

Medical Check-Up (MCU) is an important tool for the early detection of health conditions and the prevention of chronic diseases, which continue to rise globally (World Health Organization, 2021). Non-communicable diseases such as hypertension, diabetes, and heart disease represent major health burdens in many countries, including Indonesia (Indonesian Ministry of Health, 2023). Therefore, early disease detection through *MCU* is crucial for reducing morbidity and mortality. Routine *MCU* examinations allow for early identification of various metabolic and cardiovascular conditions that often go unnoticed, including elevated cholesterol levels. With the increasing prevalence of unhealthy lifestyles—such as the consumption of fast food, lack of physical activity, and high stress levels—cases of dyslipidemia have seen a significant increase (Al-Jawaldeh & Abbass, 2022; Al-Khlaiwi et al., 2023; Benyaich, 2017).

Based on data from the 2023 Indonesian Health Survey (*SKI*), which integrates the *Riskesdas* data, the prevalence of Indonesians with total cholesterol levels exceeding 200 mg/dL is estimated at 35.7%, up from 28.8% in 2018. High cholesterol poses not only an individual health threat but also significant social and economic implications (Havranek et al., 2015; Storz, 2022; Wu, Benjamin, & MacMahon, 2016; Zanoletti, Cornelio, & Bontempi, 2021). Data from BPJS Kesehatan (2024) indicate that the highest treatment claims pertain to heart disease, diabetes mellitus, stroke, and other metabolic complications—all closely related to dyslipidemia. BPJS claims for heart disease in 2024 are projected to reach Rp 38.96 trillion, making it the highest burden in the National Health Insurance system. Diabetes and stroke rank next, each incurring billions of rupiah in costs annually (Endarti et al., 2025; Fauzi, 2021; Ramadhoni, Junaidi, Octavinawaty, Apriyono, & Oktaviandi, 2025). High cholesterol thus impacts not only personal health but also national economic stability.

The World Bank (2023) noted that productivity losses due to heart disease and stroke in Indonesia cost the economy more than Rp 120 trillion annually. This figure reflects work absences, reduced productivity, and the long-term burden of healthcare costs. In addition, data from Bappenas (2024) show that the productive-age group suffering from NCDs experiences a quality-of-life (QoL) decline of up to 35% compared to the healthy population, thereby affecting national productivity. Hence, early intervention against dyslipidemia is not merely a health concern but also a national economic priority.

As the volume of *MCU* results data increases, effective analytical methods are required to segment participants based on health risk profiles to support preventive health management (Kodinariya & Makwana, 2013). Clustering techniques, particularly *K-Means Clustering*, are among the most popular and effective data mining methods for identifying patterns and segmentations in health data (Han, Kamber, & Pei, 2012). *K-Means Clustering* enables the division of data into several homogeneous groups based on the similarity of clinical variable characteristics, thereby facilitating data-driven medical decision-making (Nasution, Poningsih, & Okprana, 2021).

The use of clustering techniques on health data such as *MCU* results can uncover hidden patterns not easily visible through manual analysis and enable more targeted, personalized health interventions (Luthfi & Nilogiri, 2019). Previous studies have successfully applied *K-Means Clustering* in various domains, including customer segmentation in online marketplaces (Andara & Heikal, 2022), fintech payment applications (Artha & Heikal, 2024), UHT milk products (Ariati et al., 2023), and streaming platforms (Syakina & Heikal, 2025). However, these studies primarily focused on commercial and consumer behavior segmentation. In contrast, the application of K-Means Clustering specifically for clinical health risk profiling based on MCU data remains limited, particularly within the Indonesian healthcare context. This study addresses this gap by applying K-Means Clustering to MCU clinical data to identify distinct health risk profiles that can inform both public health interventions and preventive healthcare product development. Unlike commercial segmentation research, this study emphasizes the clinical validity of clusters and their implications for disease prevention strategies.

The novelty of this study lies in the application of data mining techniques to clinical MCU data for health risk stratification within the Indonesian population, providing both clinical and commercial insights that bridge public health needs and preventive healthcare market opportunities. Therefore, this study aims to apply the K-Means Clustering method to MCU data using SPSS software to classify participants’ health risk profiles comprehensively. This research benefits healthcare providers by enabling the prioritized allocation of preventive resources to high-risk segments, supports policymakers in developing evidence-based health promotion programs, and informs the strategic positioning of preventive health products within the Indonesian market.

RESEARCH METHOD

This study employed a quantitative, exploratory-descriptive design to identify patterns and segment health risk profiles based on Medical Check-Up (MCU) data. MCU results from 294 participants collected at Rasuna Medical Center in 2024 were analyzed, including variables such as age, systolic and diastolic blood pressure, body mass index (BMI), lipid profile (total cholesterol, HDL, LDL, triglycerides), fasting blood glucose, urea, creatinine, and uric acid. All data were anonymized and used with explicit permission for research, ensuring compliance with ethical standards. Personal identifiers were removed prior to analysis. Gender and other categorical variables were encoded numerically for the K-Means clustering analysis.

Data processing and analysis were conducted using SPSS software in several stages. Data preparation involved cleaning and standardizing all numerical variables. The optimal number of clusters was then determined through iterative testing, considering both statistical criteria and clinical interpretability. Cluster validation was carried out via clinical evaluation of cluster characteristics. Each cluster was then profiled based on demographic and clinical parameters to inform targeted intervention strategies for each risk group.

A flowchart was used to visually present the step-by-step research implementation.

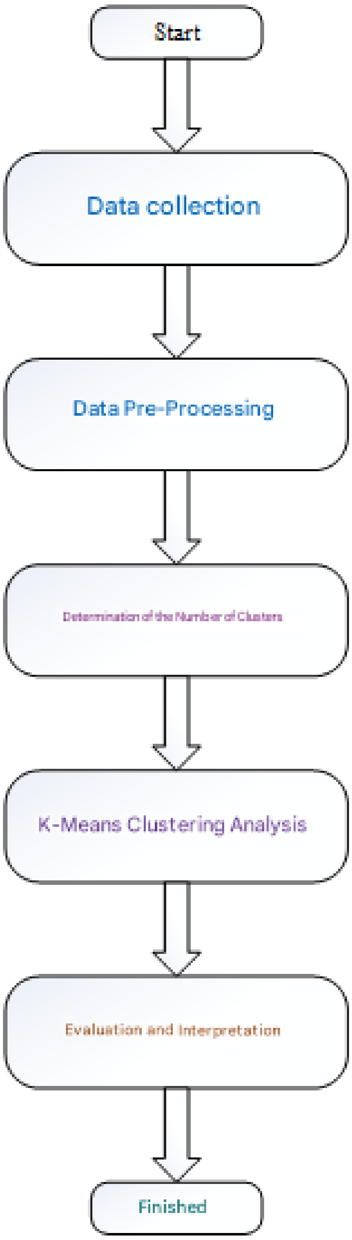

Figure 1. Research Methodology Flowchart
Source: Developed by the researcher (2024)

RESULTS AND DISCUSSION

Analysis of metabolic risk segmentation using the K-Means Clustering method resulted in five clusters that describe different health profiles based on clinical indicators. The variables analyzed included age, gender, blood pressure, body mass index (BMI), fasting blood glucose levels (GDP), lipid profiles (HDL, LDL, triglycerides), potassium levels, kidney function (urea, creatinine), and uric acid levels. The characteristics of each cluster are outlined as follows:

Table 1. Output Final Cluster Centers

Variable	Cluster				
	1	2	3	4	5
Age	41	44	46	36	57
Sistol	115	113	116	110	110
Diastol	76	80	77	70	70
BMI	27.37	28.38	27.58	35.08	24.44
Total Cholesterol	186.67	218.09	237.10	616.00	268.00
Fasting Blood Sugar	83.12	88.68	86.12	134.00	149.00
HDL	57.55	60.00	59.57	60.00	60.00
LDL	86.56	130.73	106.46	257.00	197.00
Triglycerides	85.26	265.73	136.46	217.00	634.00
Urea	24.35	25.05	24.28	20.00	20.00
Creatinine	1.01	1.08	1.04	1.00	.90
Asam_Urat	5.35	5.65	5.76	20.90	4.20
Gender	Man	Man	Man	Man	Woman

Source: Data taken from Rasuna Medical Center MCU Clinic, (2024)

Table 2. Output Number of Cases in each Cluster

Cluster	1	181
	2	22
	3	89
	4	1
	5	1
Valid		294
Missing		0

Source: SPSS Output Data Processing - Rasuna Medical Center MCU Data (2024)

From the results of the K-Means clustering process using SPSS, the researcher created personas, segmentation targets, value proproposition (VP) and online value proposition (OVP) as follows:

Table 3. Cluster Personas and Value Propositions

PERSONA	Healthy and Active	Cholesterol Watch	On the Verge of Risk	Acute Health Crisis	Metabolically Vulnerable Women
Age	41	44	46	36	57
Gender	Man	Man	Man	Man	Woman
Blood pressure	115/76	113/80	116/77	110/70	110/70
BMI	27,37	28,38	27,58	35,08	24,44
Total Cholesterol	186,67	218,09	237,10	616	268
Fasting Blood Sugar	83,12	88,68	86,12	134	149
Value Proposition	Maintain performance & prevention	Control cholesterol practically	Take care of LDL before it's too late	Healthy living transformation	Golden age protection
Online Value Proposition	Provide webinars on	Provides free consultation	Providing free medical home	Providing free medical home	Conducting education

PERSONA	Healthy and Active	Cholesterol Watch	On the Verge of Risk	Acute Health Crisis	Metabolically Vulnerable Women
	healthy and active lifestyles and gym membership offers	services and free cholesterol checks and supplements	care services and medical consultation services	care and medical consultation services as well as healthy living webinars	through WA, and reminder of drug consumption, as well as medical home care services

Source: Developed by the researcher based on clustering analysis (2024)

1. **Cluster 1 (n = 181; 64.6%) – Healthy and Active Profile**
This cluster is the majority group and is dominated by middle-aged men (average 41 years) with normal blood pressure (115/76 mmHg), normal GDP (83.12 mg/dL), and body mass index in the overweight category (27.37 kg/m²). Lipid profiles including HDL (57.55 mg/dL) and LDL (86.56 mg/dL) are in the optimal range, as are triglyceride levels (85.26 mg/dL) which indicate a healthy lipid profile. Kidney function and uric acid levels are also normal. This cluster reflects a healthy metabolic group that can be used as a baseline or reference in risk analysis.
2. **Cluster 2 (n = 22; 7.9%) – Cholesterol Alert Risk (Dyslipidemia)**
This group was dominated by men of productive age (44 years old) with lipid profiles who began to show disorders. Although HDL levels remained high (60 mg/dL), LDL values were close to the threshold (130.73 mg/dL), and triglycerides increased significantly (265.73 mg/dL). GDP in the high normal range (88.68 mg/dL), and BMI in the overweight category (28.38 kg/m²). These clusters reflect the risk of cholesterol alertness that requires stricter lifestyle interventions and metabolic controls.
3. **Cluster 3 (n = 89; 31.8%) – Threshold of Risk**
The characteristics of this cluster are similar to Cluster 1, but there is an increase in LDL cholesterol (106.46 mg/dL) and triglyceride (136.46 mg/dL), as well as relatively higher potassium levels (237.10). However, blood pressure and blood glucose remain within normal limits. The BMI category still includes overweight (27.58 kg/m²). Thus, these clusters can be categorized as risk-threshold groups that need periodic monitoring to prevent progression to more severe metabolic disorders.
4. **Cluster 4 (n = 1; 0.4%) – Acute Health Crisis**
This cluster consisted of one young male individual (36 years old) with severe obesity (BMI 35.08 kg/m²), pre-diabetic blood glucose (134 mg/dL), very high total and LDL cholesterol levels (616 mg/dL and 257 mg/dL), and high triglycerides (217 mg/dL). Uric acid reaches 20.90 mg/dL, indicating extreme hyperuricemia. The combination of these factors indicates the presence of a severe metabolic syndrome condition that requires immediate clinical attention and comprehensive evaluation.
5. **Cluster 5 (n = 1; 0.4%) – Metabolic Vulnerable Women's Risk**
This cluster is also an outlier, consisting of a 57-year-old woman with blood pressure and BMI within normal limits, but having very high levels of triglycerides (634 mg/dL), LDL (197 mg/dL), and GDP leading to diabetes (149 mg/dL). Low-normal uric acid levels (4.20 mg/dL). This cluster indicates severe dyslipidemia that is commonly found in elderly women with a risk of impaired glucose.

Target Penetration Persona

Based on the results of the study, the researcher provided suggestions to direct marketing strategies and product development to the cluster 2 persona with the "Cholesterol

Alert" profile. This persona was chosen because it has a combination of demographic, medical, and behavioral factors of urban and semi-urban people with a wide range that is very strategic to be used as the main target of initial promotion (market penetration). Cluster 2 represents men of productive age (44 years old) with common and widespread health problems, namely high cholesterol and triglycerides.

Furthermore, from the medical, behavioral, psychographic, and business aspects, Cluster 2 is worthy of being chosen as a marketing and product development strategy in terms of positioning where Cluster 2 is at a stage of risk that is not yet acute, but requires serious attention. They are not patients, but already have risk symptoms (high LDL & triglycerides) and are less likely to feel "sick", but know that they "need to change" this is the best *window of change* for educational and preventive positioning.

With cooperative psychographic segmentation for brands Cluster 2 tends to be rational and analytical (interested in data and medical evidence), always respects time (needs practical solutions) and does not like to be regulated too tightly (not suitable for the "fierce doctor" approach). Therefore to offer Product and Marketing ideally, a marketer should be visible scientific but light by providing options and personal control so that it will provide real progress without intimidation.

In addition, if the product succeeds in having a positive effect on Cluster 2, the persona tends to recommend to colleagues, colleagues, or the professional community so that in this case cluster 2 is suitable for building a "brand advocacy" or referral network of satisfied users and making them not only consumers, but product ambassadors (champion personas).

CONCLUSION

This study demonstrated that applying the K-Means Clustering method to Medical Check-Up (MCU) data successfully identified five distinct clusters representing varying metabolic risk levels within the population. Most participants were grouped into stable metabolic profiles, serving as baseline healthy populations, while others showed mild to moderate risks linked to elevated LDL cholesterol, triglycerides, and overweight tendencies, highlighting the need for early lifestyle interventions. Two small but clinically significant outlier clusters were identified—one comprising young individuals with severe metabolic syndrome and another of elderly women with high triglycerides and blood glucose despite normal BMI and blood pressure—emphasizing the importance of targeted surveillance for vulnerable groups. Overall, clinical data-based clustering proved effective for objective health-risk categorization, supporting tailored prevention strategies and resource allocation in public health programs. Future research could explore integrating longitudinal health data and behavioral factors into clustering models to enhance predictive accuracy and intervention planning.

REFERENCE

Acuna, L., Sanchez, P., Soler, L., & Alvis, L. F. (2015). Total cholesterol (Tc), low-density lipoprotein cholesterol (Ldl-C) and high-density lipoprotein cholesterol (Hdl-C) levels in patients with hypertension (Ht), diabetes (Dm), both (Ht and Dm) and chronic kidney disease (Ckd). *Value in Health, 18*(7), A405–A406.

Al-Jawaldeh, A., & Abbass, M. M. S. (2022). Unhealthy dietary habits and obesity: The major risk factors beyond non-communicable diseases in the Eastern Mediterranean region. *Frontiers in Nutrition, 9*, 817808.

Al-Khlaiwi, T., Alshammari, H., Habib, S. S., Alobaid, R., Alrumaih, L., Almojel, A., Sendi, F., Almuqbil, S., & Alkhodair, M. (2023). High prevalence of lack of knowledge and unhealthy lifestyle practices regarding premature coronary artery disease and its risk factors among the Saudi population. *BMC Public Health, 23*(1), 908.

Aliyu, S. U., Chiroma, A. S., Jajere, A. M., & Gujba, F. K. (2015). Prevalence of physical inactivity, hypertension, obesity and tobacco smoking: A case of NCDs prevention among adults in Maiduguri, Nigeria. *American Journal of Medical Sciences and Medicine, 3*(4), 39–47.

Benyaich, A. (2017). The effects of the Mediterranean diet on chronic diseases: Cardiovascular diseases, oxidative stress, dyslipidemia, diabetes mellitus, blood pressure, cancer, neurodegenerative disease and obesity. *Journal of Applied and Advanced Research, 2*(6), 333–355.

Endarti, D., Andayani, T. M., Widayanti, A. W., Rohmah, S., Banjarani, R. R., & Ghearizky, N. A. (2025). Chronic disease costs from a patient's perspective: A survey of patients with stroke, heart disease, and chronic kidney disease visiting a district hospital in Indonesia. *Journal of Pharmacy & Pharmacognosy Research, 13*(1), 274–283.

Fauzi, R. (2021). A policy simulation impact of tobacco excise tax increase on cigarette consumption, mortality, medical treatment cost, life-years gained, and government revenue in Indonesia.

Goutama, I. L., Hendsun, Y. F., & Su, E. (2020). Comparison of autocorrelation between CV-RISK independent variables in groups with and without history of cardiovascular diseases. *International Journal of Advances in Medicine, 7*(11), 1626.

Havranek, E. P., Mujahid, M. S., Barr, D. A., Blair, I. V., Cohen, M. S., Cruz-Flores, S., Davey-Smith, G., Dennison-Himmelfarb, C. R., Lauer, M. S., & Lockwood, D. W. (2015). Social determinants of risk and outcomes for cardiovascular disease: A scientific statement from the American Heart Association. *Circulation, 132*(9), 873–898.

Kosmas, C. E., Rodriguez Polanco, S., Bousvarou, M. D., Papakonstantinou, E. J., Peña Genao, E., Guzman, E., & Kostara, C. E. (2023). The triglyceride/high-density lipoprotein cholesterol (TG/HDL-C) ratio as a risk marker for metabolic syndrome and cardiovascular disease. *Diagnostics, 13*(5), 929.

Li, J., Pandian, V., Davidson, P. M., Song, Y., Chen, N., & Fong, D. Y. T. (2025). Burden and attributable risk factors of non-communicable diseases and subtypes in 204 countries and territories, 1990–2021: A systematic analysis for the Global Burden of Disease Study 2021. *International Journal of Surgery, 111*(3), 2385–2397.

Maugeri, A., Hruskova, J., Jakubik, J., Barchitta, M., Lo Re, O., Kunzova, S., Medina-Inojosa, J. R., Agodi, A., Sciacca, S., & Vinciguerra, M. (2019). Independent effects of hypertension and obesity on left ventricular mass and geometry: Evidence from the Cardiovision 2030 study. *Journal of Clinical Medicine, 8*(3), 370.

Nantsupawat, N., Booncharoen, A., Wisetborisut, A., Jiraporncharoen, W., Pinyopornpanish, K., Chutarattanakul, L., & Angkurawaranon, C. (2019). Appropriate total cholesterol cut-offs for detection of abnormal LDL cholesterol and non-HDL cholesterol among low cardiovascular risk population. *Lipids in Health and Disease, 18*(1), 28.

Nasution, S. A., Poningsih, & Okprana, H. (2021). Application of the K-Means algorithm to the sales of frozen food at UD Soise Sausage Pematangsiantar. *Journal of Computer Systems and Informatics, 2*(2), 171–177. https://ejournal.unsri.ac.id/index.php/jski/article/view/22093

Organization, World Health. (2023). *Assessing national capacity for the prevention and control of noncommunicable diseases: Report of the 2021 global survey.* World Health Organization.

Parato, A. G., Parato, V. M., Fedacko, J., & Magomedova, A. (2024). Global burden of causes of death and life expectancy with reference to cardiovascular diseases. *World Heart Journal, 16*(2), 103–115.

Raja, V., Aguiar, C., Alsayed, N., Chibber, Y. S., ElBadawi, H., Ezhov, M., Hermans, M. P., Pandey, R. C., Ray, K. K., & Tokgözoglu, L. (2023). Non-HDL-cholesterol in dyslipidemia: Review of the state-of-the-art literature and outlook. *Atherosclerosis, 383*, 117312.

Ramadhoni, P. D., Junaidi, A., Octavinawaty, L., Apriyono, A., & Oktaviandi, A. (2025). Costs of stroke treatment under national health insurance at Dr. Mohammad Hoesin General Hospital. *Journal of Neurointervention and Stroke, 1*(1), 1–9.

Storz, M. A. (2022). What makes a plant-based diet? A review of current concepts and proposal for a standardized plant-based dietary intervention checklist. *European Journal of Clinical Nutrition, 76*(6), 789–800.

Sun, L., Zhou, Y., Zhang, M., Li, C., Qu, M., Cai, Q., Meng, J., Fan, H., Zhao, Y., & Hu, D. (2022). Association of major chronic noncommunicable diseases and life expectancy in China, 2019. *Healthcare, 10*(2), 296. MDPI.

World Health Organization. (2021). *Noncommunicable diseases.* https://www.who.int/news-room/fact-sheets/detail/noncommunicable-diseases

Wu, Y., Benjamin, E. J., & MacMahon, S. (2016). Prevention and control of cardiovascular disease in the rapidly changing economy of China. *Circulation, 133*(24), 2545–2560.

Zanoletti, A., Cornelio, A., & Bontempi, E. (2021). A post-pandemic sustainable scenario: What actions can be pursued to increase the raw materials availability? *Environmental Research, 202*, 111681.