An Application of Open Government Data: An Evidence on Physical Activity Environment and Non-communicable Diseases

Hsin Yen Yen and Ching Li

Abstract—**Open government data (OGD) is usually used for decision-making, policy planning and international comparing. The integration of OGD from different resources is also important. Since 1990s, previous studies have discussed the relationship between physical environment and health promotion. In recent years, building Physical Activity Environment (PAE) is one of factors for preventing Non-communicable diseases (NCDs). OGD could be applied for identifying patterns and developing a framework, through the integration of environmental and medical resources. Therefore, the purpose of this study was to apply OGD from two databases for developing a framework of PAE and NCDs. The study range contained all 22 administrative regions (county and city) in Taiwan. Data collected from Taiwan OGD Platform: Data.gov.tw which has been online since 2013 as an open data portal. Descriptive statistics and Partial Least Square path analysis were used. The result demonstrated the hypothesis model was significant, and the path coefficient was -0.441 (t=1.65, p<.05*). PAE could negatively predict hypertension, hyperglycemia, and hyperlipidemia. PAE could reduce risk factors of NCDs because it related to physical activity and walking behavior. PAE should be concerned on urban planning and designs and public health policy in order to promote well-beings. The implication of this study was the integration of OGD could be applied to develop a framework. OGD analysis could provide related information before policy decision-making.**

Index Terms—**Chronic diseases, e-governance, public database, public health**

I. INTRODUCTION

e-Government is advocated for more cost-effective e-Service to citizens, external and internal stakeholders. It becomes an important linkages among citizens, businesses, organizations and government departments [1], [2]. The utility of e-Government is diversified from different sectors of government as important tools, including land uses, health insurances, taxes and finance analysis, safety management, public administrations, and other surveys. e-Government depends on information and communication technology to provide faster and better communication, exchange information and retrieval data [2].

One of major functions of e-Government is providing public databases. Open Government Data (OGD) are usually used for decision-making and international comparing, including democracy, society, economic and health issues.

Hsin Yen, Yen and Ching Li are with Graduate Institute of Sport, Leisure and Hospitality Management, National Taiwan Normal University, 106, Taipei City, Taiwan (e-mail: kenjizoro520@gmail.com, t94002@ntnu.edu.tw).

The characters of OGD are qualified, big amount, and transparency, which is easily assessed by stakeholders and citizens. The objectives of OGD initiatives are fostering policy, integration of public sectors through IT, delivering and improving local services, and creating a safer community [3].

The characters of qualified OGD are usable, accurate, consistent, and timely. The usability of data emphasizes the importance of accessibility, completion, discoverability. In order to become usable metadata, the data collected process should be on schedule in a fixed period. However, the linkage of data could approach to merging data, analyzing data, and modelling, in order to present a meaningful value [4].

e-Government Program in Taiwan has developed since 2001. OGD was provided from different public affairs. The data are usually collected once a month or once a year. The reliability, consistency, and timeliness of OGD were developed in the initial stage. But open data portals were multiple on individual public websites. The Taiwan OGD policy was official launched in 2012. OGD in Taiwan has become more integrated, transparent, and accessible. Open Data Platform has also been online since 2013. OGD platform provides information about diet, healthcare, transportations, cultures, entertainment, and community development. It promotes the integration between not only government agencies but also citizens and organizations. The benefits of OGD could improve policy implement, elevate resident's quality of life, satisfy with stakeholder's need and encourage service innovations [5].

Even if there are institutional and technological barriers, using OGD from different sectors still has several advantages. It is an innovated process of e-Government. OGD helps government departments to realize the reality of challenges and emerging trends. OGD also provides factors which could be evaluated and analyzed before policy making [6]. It is common to use OGD and e-Government to identify patterns and discuss or develop a conceptual framework and a model from theory [7].

In addition, non-communicable diseases (NCDs) are important health issues in a modern society. NCDs was not communicable which are also well-known as chronic diseases and lifestyle diseases. Over 38 million people a year die because of NCDs. NCDs impact socioeconomic in countries which cause poverty, lost economic productivity and become large social burdens [8].

Prevention are lower cost than controlling and treatments of NCDs, so risk factors of NCDs should be reduced, including personal lifestyle and environmental factors [8]. Since 1990s, previous studies have discussed the relationship

between physical environment and disease prevention. OGD is a good resource to collect related information. The databases contain complete, assessable, and reliable information of environment and healthcare. Moreover, several physiological signs should be aware, including hypertension, hyperglycemia, hyperlipidemia and obesity.

Furthermore, building Physical Activity Environment (PAE) is also important to prevent diseases. Physical activity is influenced by personal, behavior, socio-environmental factors and physical environment [9].

Sporting open spaces provide opportunities for community residents to participate in physical activity and interaction with each other. These sporting open spaces include parks, gyms, swimming pools and other artificial spaces. Sporting open spaces also promote physical and psychological health, social, economic and environmental benefits. The density and diversity of open spaces should be concerned for public policies planning.

On the other hand, Residents' daily life and activity should be accomplished by walking, instead of driving [10]. Indeed, neighborhood walkability could influence community resident's walking behavior for transport, travel, exercise and recreation. Walkable environment could increase residents' frequency of walking behavior and physical activity [11]. Community infrastructures within a walkable distance could be a motivation of outdoor activity, and increase the opportunity of physical activity and decrease sedentary behavior [12].

The goal of PAE is to maintain healthy behavior, prevent chronic diseases, and to improve residents' wellness and quality of life [13]. Previous studies demonstrated PAE positively impact physical, mental, and social health. It was believed that people live in PAE with fewer recreational open spaces and higher walkability have lower BMI of children, adults, and seniors [14], [15]. People lived in walkable area had higher self-reported health and life satisfaction than those did not live in walkable area [16]. There was a significant difference of risk of incident hypertension between lower and higher walkability groups [17]. Residents lived in walkable neighborhood had better cardiorespiratory functions and lower obesity and overweight problems [18], [19].

The main purpose of this study was to fill a gap between PAE and NCDs, and develop a framework by applying OGD. Two specific databases in Taiwan were used, including Health Insurance database and Constriction and Planning database. Through OGD analysis, this study wondered whether Walkable area and sporting open spaces could be factors of predicting patients with Hypertension (HTN), hyperglycemia (DM), and hyperlipidemia (HPL). Furthermore, the evidence-based results from OGD could provide the government for planning, implementation, and evaluation of health related public policy.

II. METHOD

A. Data Collection

All data collected from open data in the government platform: Data.gov.tw. This portal supplied large amount and qualified data and the link of many public departments. Because OGD from two different databases were used in this study, the samples should be compared in the same scale. Therefore, this study chose all 22 administrative regions, county and city, in Taiwan as the samples to overcome the differences between databases. All data in 2014 were used.

B. Constriction and Planning Database

Constriction and planning database is provided by Constriction and Planning Agency, Ministry of Interior, Taiwan. The content of the database included Regional Planning, Urban Planning, National Parks, Public Housing Administration, Public Works and Building Control. The data are collected from registers of land uses, according to construction license and usage license. All data have renewed once a year since 1997. The format of OGD supplies wsdl, xml, json, pdf, xls, and csv. Therefore, OGD of Constriction and Planning database is complete and accurate, consistently and timely renewed, easily accessible to everyone.

The independent variable was PAE which included 2 important indicators, Walkable Area and Sporting Open Space. The data accessed from Area Acquired of Public Facility Land of Urban Planning in Constriction and Planning database. The sample unit of this database bases on district, it also demonstrates the level of county and city, and nation. Walkable Area ratio in this study was defined as the area of walking way divided by total county and city area, including footpaths, sidewalks and walkable roads in county and city, as in

$$\text{Walkable Area}（\%）= \frac{\text{walking way area}}{\text{Total county or city area}} \quad (1)$$

Sporting Open Space ratio in this study was defined as the area of all spaces for citizens to participate in physical activity, including parks, green land (GL), squares (SQ), playground (PG), and sport facilities (SF), which was divided by total county and city area, as in

$$\text{Sporting Open Space (\%)} = \frac{\text{(Park+GL+PG+SF) Area}}{\text{Total county or city area}} \quad (2)$$

C. Health Insurance Database

Health Insurance database is provided by Department of Statistics, Ministry of Health and Welfare, Taiwan. OGD of Health Insurance database only provides the number of inpatients and outpatients in xls format for free users. OGD also presents the cross table of the number of patients with different diseases (ICD-9 number and Health Insurance number) which is categorized by gender, age, and residence address (county and city, nation). Other assessable OGD contained death reasons, cancers, social welfare, and healthcare expenditure.

However, the Health Insurance database contains great amount of patients' diseases details which does not present on OGD. Every insurance patient who visits a physician or is hospitalized every time would be recorded to Health Insurance System. The clinical details from health care institutions connects with Health Insurance system and upload immediately. The items of Health Insurance database are diversified, such as Registry for beneficiaries (ID),

Registry for drug prescriptions (DRUG), Details of ambulatory care orders (OO), Details of inpatient orders (DO). The database would hind the personal information for private and security reasons. All data have renewed monthly and OGD has provided once a year since 1998. Therefore, OGD of Health Insurance database is also complete and accurate, consistently and timely renewed, but not easily accessible to everyone. For more detailed information, Health Insurance database could be purchased.

The dependent variable was the most common NCDs, including hypertension, hyperglycemia, and hyperlipidemia, which were also important physiological signs and easy to be examined, and could develop serious diseases [8]. The data accessed from OGD of outpatient, Health Insurance database. Patients were made diagnoses of hypertensive disease (HTN, ICD-9 CM: A26 401-405, Health Insurance No. 0802), diabetes mellitus (DM, ICD-9 CM: A181 2500-2509, Health Insurance No. 040102), and disorders of lipoid metabolism (HPL, ICD-9 CM: A182 270, Health Insurance No. 040103). The rate of outpatients with HTN, DM, and HPL per 100,000 populations in county and city was used, as in

$$NCDs = \frac{The\ number\ of\ patients\ with\ \text{HTN,DM,HPL}}{Population\ in\ county\ and\ city} \times 100{,}000$$

(3)

D. Data Analysis

Descriptive statistics and Partial Least Square (PLS) path analysis were used in this study. However, SEM was usually used for fitting the model, but PLS could be used for developing the model or the framework. The sample size was not necessary to be nominate distribution, which required smaller sample sizes than SEM. It was conducted that the sample sizes of PLS analysis should more than 20 [20]. Chin's sample sizes under PLS were estimated that 10 times the dependent LV with the largest number of independent LVs impacting it [21]. It represented the sample sizes in this study should be 10 at least. Therefore, 22 sample sizes in this study could satisfied with 2 previous studies. The data analysis tools of this study were SPSS and Smart-PLS.

III. RESULT

A. Descriptive Analysis

Table I demonstrated the result of descriptive statistics. The coverage ratio of Walkable Area was 2.270 ± 2.696%.

The coverage ratio of Sporting Open Spaces was 1.082 ± 1.219%, including all parks, green land, squares, playgrounds, and sport facilities. On the other hand, the rate of outpatients with hypertension was 16174.512 ± 2167.747 patients per 100,000 population. The rate of outpatients with hyperglycemia was 8239.323 ± 1233.128 patients per 100,000 population. The rate of outpatients with hyperlipidemia was 10919.705 ± 1690.839 patients per

100,000 populations.

TABLE I: THE RESULT OF DESCRIPTIVE STATISTICS

Variables	min	Max	Mean	SD
Walkable Area (%)	0.062	5.001	1.082	1.219
Sporting Open Space (%)	0.168	10.132	2.270	2.696
HTN (patients per 100,000 population)	12305.819	19360.281	16174.512	2167.747
DM (patients per 100,000 population)	6364.886	10295.332	8239.323	1233.128
HPL (patients per 100,000 population)	8207.366	13913.002	10919.705	1690.839

B. PLS Path Analysis

The result of PLS path analysis demonstrated on Fig. 1. It illustrated that the hypothesis model was significant, and the path coefficient was -0.441 (t=1.65, $p<.05$*). PAE could predict NCDs negatively, and the predict power of this model was 19.4%.
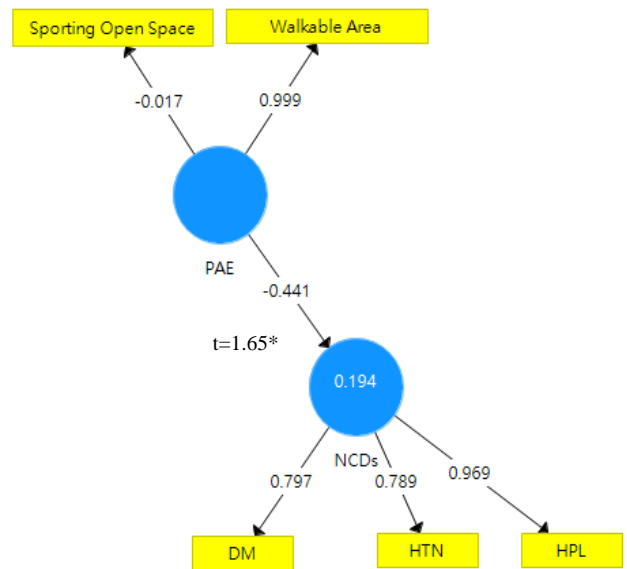

Fig. 1. PLS path model.

Table II illustrated the outer loading of two indicators in the independent latent variable, Walkable Area and Sporting Open Spaces, were individually 0.999 and -0.017. The outer loading of Walkable Area was higher than Sporting Open Spaces. The outer loading of three indicators in the dependent latent variable, DM, HTN and HPL, were individually 0.797, 0.789, and 0.969. There were also highly positive correlations between DM, HTN and HPL.

The composite reliability of this model was 0.49 and 0.89 in both latent variables. The average variance extracted of 0.73 and 0.50 in both latent variables. The composite reliability and average variance extracted of latent variables should over 0.50 [25]. Cronbach α value was usually required equally to composite reliability. Therefore, this PLS path model was reliable.

TABLE II: THE RESULT OF DESCRIPTIVE STATISTICS

Latent Variable	indicators	Outer loading	CR	AVE	R square	Cronbach α
PAE	Walkable Road	0.999	.50	.50	.194*	.41
	Sporting Open Space	-0.017				
NCDs	DM	0.797	.89	.73		.89
	HTN	0.789				
	HPL	0.969				

CR: composite reliability, AVE: average variance extracted

IV. CONCLUSION

A. Discussion

The result demonstrated main purposes of this study, which was an application of OGD to develop a framework by theory. Two databases were accurate, and complete, qualified and reliable [3]. The PLS path-model composed by PAE and NCDs was significant. It also illustrated the integration of OGD could develop a new framework and actions for citizens, stakeholders, organizations and government. The usability of OGD could be elevated to produce useful meanings. OGD analysis could supply related information before policy decision-making [6].

In addition, the evidence from OGD illustrated that PAE was an important predictor of NCDs (DM, HTN and HPL). It represented sporting open spaces and walkable area were both necessary elements for preventing NCDs. Walkability of community influenced the level of residents' physical activity and social interaction [22]. The walkable infrastructures in neighborhood could be a motivation of residents to walk. Residents also need spaces for participating in physical activity and outdoor activity [23]. However, even of the predictor power of PLS path model was not high, PAE is still an important factor of preventing NCDs. Previous studies also conducted PAE had a positive influence on health outcomes. [19].

B. Limitation

There were several limitations of this study. First, this study used the rate of outpatients of DM, HTN and HPL as dependent variables, but it was not the prevalence of NCDs. Only patients who see a doctor every year were recorded to the health insurance system. Second, the utility of walkable area and the user number of sporting open spaces were unknown. The utility was affected by walkable area and sporting open spaces were a premise. But the behavior factor should be concerned in the future study.

C. Implications and Recommendations

Creating PAE is an important issue of developing a healthy community and promoting well-beings. Accessible infrastructures and walkable area have positive health benefits. Sport industry could devote to develop in a less density of sporting open spaces. PAE should be taken into consideration of community planning, urban designs and public policy planning for human's well-beings.

Furthermore, OGD from different databases could be analyzed, integrated, and applied for mutual interests of society. OGD could be applied for developing a framework and a model either. It is also an important linkage among government, citizens and interest groups. OGD is also a convenient and valuable resource to be compared by cross-country. Merging OGD could integrate public sections and create a better community [4]. However, the category of OGD supplied by present government agencies could be increased. OGD should be qualified, reliable, usable, accurate, and, complete [3]. Present OGD could be more transparent, accessible and usable for policy decision-making.

REFERENCES

[1] Z. Haider, C. Shuwen, and M. B. Burdey, "E-government project obstacles in pakistan," vol. 8, no. 5, pp. 7763, 2016.

[2] P. L. Sun, C. Y. Ku, and D. H. Shih, "An implementation framework for E-Government 2.0," *Telemat. Informatics*, vol. 32, no. 3, pp. 504–520, 2015.

[3] M. Kassen, "Globalization of e-government: open government as a global agenda; benefits, limitations and ways forward," *Inf. Dev.*, vol. 30, no. 1, pp. 51–58, 2013.

[4] J. Attard, F. Orlandi, S. Scerri, and S. Auer, "A systematic review of open government data initiatives," *Gov. Inf. Q.*, vol. 32, no. 4, pp. 399–418, 2015.

[5] About us. [Online]. Available: http://data.gov.tw/about.

[6] H. Wang and J. Lo, "Adoption of open government data among government agencies," *Gov. Inf. Q.*, vol. 33, no. 1, pp. 80–88, 2015.

[7] F. Bannister and R. Connolly, "The great theory hunt: Does e-government really have a problem?," *Gov. Inf. Q.*, vol. 32, no. 1, pp. 1–11, 2015.

[8] World Health Organization. (1991). Sundsvall statement on supportive environments for health.pdf. *World Health Organization.* [Online]. Available: http://apps.who.int/iris/bitstream/10665/59965/1/WHO_HED_92.1.pdf

[9] K. Sundquist, U. Eriksson, N. Kawakami, L. Skog, H. Ohlsson, and D. Arvidsson, "Neighborhood walkability, physical activity, and walking behavior: The Swedish Neighborhood and Physical Activity (SNAP) study," *Soc. Sci. Med.*, vol. 72, no. 8, pp. 1266–1273, 2011.

[10] K. M. Booth, M. M. Pinkston, and W. S. C. Poston, "Obesity and the built environment," *J. Am. Diet. Assoc.*, vol. 105, no. 5 SUPPL., 2005.

[11] N. Owen, E. Cerin, E. Leslie, L. duToit, N. Coffee, L. D. Frank, A. E. Bauman, G. Hugo, B. E. Saelens, and J. F. Sallis, "Neighborhood walkability and the walking behavior of australian adults," *Am. J. Prev. Med.*, vol. 33, no. 5, pp. 387–395, 2007.

[12] N. Cali-, S. Handy, E. Science, X. Cao, E. Engineering, P. Mokhtarian, and E. Engineering, "Self-Selection in the relationship between the built environment and walking," 2004.

[13] T. Norris and M. Pittman, "The healthy communities movement and the coalition for healthier cities and communities.," *Public Health Rep.*, vol. 115, no. 2–3, pp. 118–124, 2000.

[14] D. T. Duncan, M. Sharifi, S. J. Melly, R. Marshall, T. D. Sequist, S. L. Rifas-Shiman, and E. M. Taveras, "Characteristics of walkable built environments and BMI z-scores in children: evidence from a large electronic health record database.," *Environ. Health Perspect.*, vol. 122, no. 12, pp. 1359–65, 2014.

[15] J. Van Cauwenberg, V. Van Holle, I. De Bourdeaudhuij, D. Van Dyck, and B. Deforche, "Neighborhood walkability and health outcomes among older adults: the mediating role of physical activity.," *Heal. Place*, vol. 37, pp. 16–25, 2016.

[16] S. Oishi, M. Saeki, and J. Axt, "Are people living in walkable areas healthier and more satisfied with life?," *Appl. Psychol. Heal. Well-Being*, vol. 7, no. 3, pp. 365–386, 2015.

[17] M. Chiu, M.-R. Rezai, L. C. Maclagan, P. C. Austin, B. R. Shah, D. A. Redelmeier, and J. V. Tu, "Moving to a highly

walkable neighborhood and incidence of hypertension: a propensity-score matched cohort study," *Environ. Health Perspect.*, October 2015.

[18] C. M. Hoehner, S. L. Handy, Y. Yan, S. N. Blair, and D. Berrigan, "Association between neighborhood walkability, cardiorespiratory fitness and body-mass index," *Soc. Sci. Med.*, vol. 73, no. 12, pp. 1707–1716, 2011.

[19] B. B. Cutts, K. J. Darby, C. G. Boone, and A. Brewis, "City structure, obesity, and environmental justice: An integrated analysis of physical and social barriers to walkable streets and park access," *Soc. Sci. Med.*, vol. 69, no. 9, pp. 1314–1322, 2009.

[20] B. S. Hui and H. Wold, "Consistency and consistency at large of Partial Least Squares estimates," *Syst. under Indirect Obs. causality, Struct. Predict.*, vol. 2, 1982.

[21] W. W. Chin, "The partial least squares approach to structural equation modeling," *Mod. methods Bus. Res.*, pp. 237–246, April, 1998.

[22] E. E. Lozano, *Community Design and the Culture of Cities: The* Crossroad *and the Wall*, vol. 30. Cambridge University Press, 1990.

[23] C. Foster and M. Hillsdon, "Changing the environment to promote health-enhancing physical activity.," *J. Sports Sci.* 22, no. 8, pp. 755–769, 2004.

Hsin Yen, Yen is a PhD student in graduate institute of Sport, Leisure and Hospitality Management, National Taiwan Normal University. He also had an education master degree in Graduate Institute of Sport and Leisure Management, National Taiwan Normal University, Taiwan. Besides, he also received bachelor degree in Science of Nursing, Taipei Medical University, Taiwan. His research interests are physical activity, sport therapy, health promotion, chronic diseases, wellness tourism, and data analysis.

Dr. Ching, Li is a professor and chairman Dr. Ching, Li is a professor and chairman of Graduate Institute of Sport, Leisure and Hospitality Management, National Taiwan Normal University, Taiwan. She had a PhD degree in Environmental Science, State University of New York, NY, USA. Her research expertise focuses on recreational resource management, environmental planning and assessment for sport facilities, community recreation, leisure behavior, and hiking.