THE ANNUAL REPORTS OF
HEALTH, PHYSICAL EDUCATION
AND SPORT SCIENCE
VOL.38, 17-24, 2019

原　著

Selection of regional factors associated with low birthweight in Japan

Izumi KUBOYAMA*, Akira MAKI*, Yoshikazu YAMAGUCHI*
Takeshi ASHIZAWA*, Hidekazu TAKEUCHI* and Susumu ITO**

ABSTRACT

Background

Babies with low birthweight have potential physical and mental disadvantages.

Purpose

The purpose of this study was to determine the social and demographic variables (SDVs) related to the low birthweight rate (LBWR) in Japan using a combination of the correlation coefficient (CC) and coefficient of determination (R2).

Methods

We obtained 704 variables from 13 statistical tables in the SDVs of the 2017 e-Stat database of Japan. We calculated the CCs between the LBWR and the 704 variables in each Japanese prefecture. Of the original 704 variables, 21 were selected, as the CCs between the LBWR and variables were ≥0.4. We conducted multiple regression analysis between the LBWR and these 21 variables to obtain a list of 21 explanatory variables in ascending order by *p*-value. We also repeated the multiple regression analysis 21 times to obtain the R2 and adjusted R2.

Results

The median LBWR was 9.5%, the maximum was 11.5%, the minimum was 8.2%, and the mean was 9.5% among 47 prefectures. Only the variable "amount of saving per household" exhibited statistical significance ($p=0.048$), and its R2 and R2adj were 0.164 and 0.149, respectively. The R2 was 0.669 (88.3% of the maximum), and the R2adj was 0.610 (95.3% of the maximum) and the maximum in seven explanatory variables that reached the plateau.

Conclusion

The combination of CCs and R2 indicated that the LBWR was significantly related to the regional factor "amount of savings per household," as determined from the social and demographic database in Japan.

Key words; Low birth weight, Regional Factors; Social and Demographic Variable; Coefficient of Determination

Abbreviations

LBW: Low Birth Weight; LBWR: Rate of Low Birth Weight; SDV: Social and Demographic Variables; R2: Coefficient of determination; R2adj: adjusted Coefficient of determinatio

* School of Emergency Medical System, Kokushikan University, Tokyo, Japan
** High-Tech Research Centre, Kokushikan University, Tokyo, Japan
Correspondence: Izumi Kuboyama, ikuboyam@kokushikan.ac.jp

Introduction

Low birthweight (LBW) is caused either by preterm birth, by being small for gestational age, or a combination of both[1, 2]. The World Health Organization defines LBW as below 2,500 g. In general, infant mortality rate (IMR) is in accordance with the overall level of health, and the LBW rate (LBWR) is in parallel with the IMR worldwide. The Organization for Economic Co-operation and Development database shows the positive association between the LBWR and IMR. However, in Japan, the LBWR is high, even though the IMR is the lowest in the world[3].

Babies with LBW have potential physical and mental disadvantages, both neonatally and during childhood [4]. Moreover, observational studies on LBW have reported that undernutrition during gestation may be an important early determinant of adult cardiac and metabolic disorders, due to the lack of fetal and neonatal growth[5-11].

Numerous studies have evaluated socioeconomic status and race/ethnicity and their associations with the LBWR[6-8], but the results have been inconsistent[9-10], and none have considered regional factors to account for LBW[12-14], such as rural and urban features, patient characteristics, and patient care.

Large-scale administrative data are available for research on social and demographic variables (SDVs) in Japan[15]. A great number of these variables are available for use ; however, the number of regions is rather smaller. Therefore, we need to efficiently select the potential effective variables from those available on the large-scale databases. We have previously reported the selection of effective variables through the combination of correlation coefficients (CC), which are measures of association that indicate the degree to which two variables have a linear relationship, and coefficients of determination (R2) [16]. Using this method, we focused on the relationship between the LBWR and SDV.

Purpose of the Study

The purpose of this study was to find the SDV related to the LBWR in Japan by combining the use of CC and R2.

Materials and Methods

(1) Study setting

The Japanese government provides the administrative statistical data for Japan at the portal site e-Stat, where the SDVs are specified by the Japanese Government Statistics[11]. The SDVs are contained in 13 statistical tables : population and household, natural environment, economic base, administrative base, education, labor, culture and sports, dwelling, health and medical care, welfare and social security, safety, family budget, and daily time. These data are summarized by prefectures and years.

Japan consists of 47 prefectures. The median population of each prefecture is 1,668,000, with a minimum of 574,000 and a maximum of 13,390,000. The median area of each prefecture is 4819 km^2, with a minimum of 574 km^2 and a maximum of 78,420 km^2.

(2) Outcome

We downloaded and obtained 704 variables from the 13 statistical tables of the SDV of 2017 on the e-Stat site. After obtaining 704 variables for explanatory variables, we calculated the CCs among the 704 variables in each prefecture and chose 24 variables with CCs that were ≥0.4. Following a multiple regression analysis between the LBWR and the 24 variables, we calculated the variance inflation factors (VIFs) for each variable to prevent multicollinearity. When the VIFs were over 10, we deleted the variables with the smaller CCs.

(3) Coefficient of determination, adjusted coefficient of determination, and Akaike information criterion

Coefficient of determination (R2) is defined as 1-SSres/SStot, and the adjusted R2 (R2adj) is defined as $1-(SSres/dfe)/(SStot/dft)$, where SStot is the total sum of squares, SSres is the residual sum of squares, degrees of freedom for treatment (dft) is n-1, degrees of freedom for error (dfe) is n-p-1, p is the total number of explanatory variables, and n is the sample size. Akaike information criterion (AIC) is defined as 2k-2ln (L), where k is the number of estimated parameters, and L is the maximum value.

The LBWR was a response variable, and we adopted the top variables from the list (Table 1) as the explanatory variables in each model. We repeated the regression analyses 21 times (from 1 to 21 explanatory variables in Table 1) and obtained the R2, R2adj, and AIC in each model (Figure 1). We plotted R2, R2adj, and AIC according to the number of explanatory variables. We also conducted stepwise regression with backward elimination between the LBWR and 21 selected SDVs. We used the open statistical software R (version 3.6.1 ; The R Foundation, Austria) for analysis.

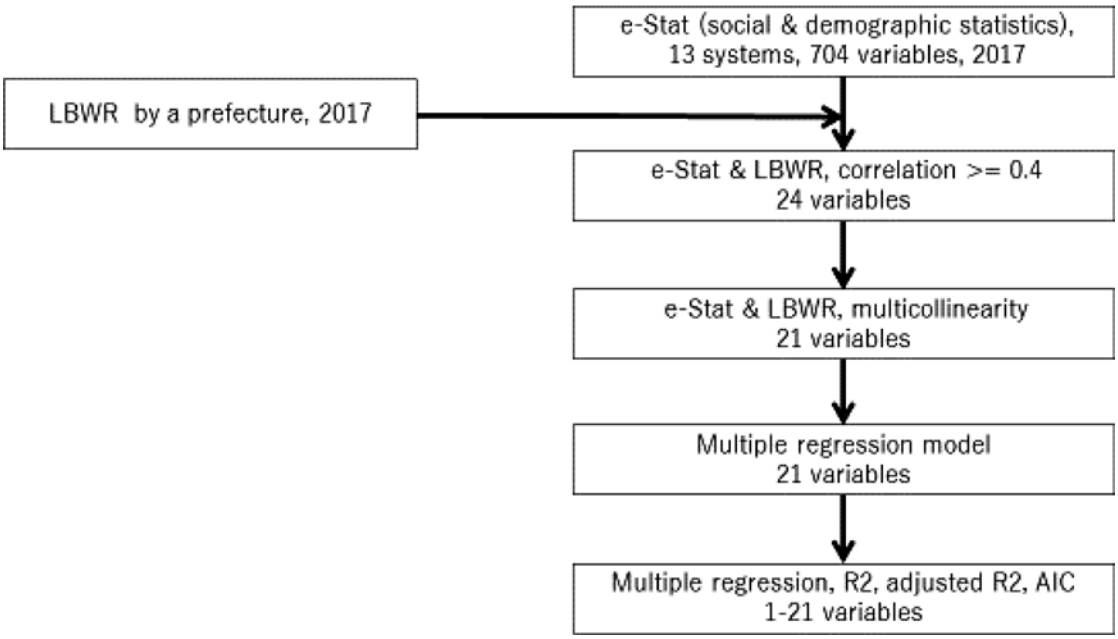

Figure 1 : Flowchart of the study.

Results

(1) LBWR in each prefecture

The median of the LBWR was 9.5%, the maximum was 11.5%, the minimum was 8.2%, and the mean was 9.5% among 47 prefectures. The distribution of the LBWR is presented at the prefectural level in Figure 2.

(2) Selected variables from the e-Stat site

Of the original 704 variables, 24 were selected, as the CCs of the LBWR and these variables were ≥0.4 (variables chosen according to the CCs in Table 1). Consequently, 21 variables were selected to prevent multicollinearity in the table. Then, we obtained a list of the 21 variables in ascending order by *p*-value (Table 1).

(3) R2, R2adj, and AIC

The first variable, "amount of saving per household," was the only one exhibiting statistical significance ($p=0.048$), and its R2 and R2adj were 0.164 and 0.149, respectively (Figure 3). The second variable was "regional difference index of consumer prices," the third was "ratio of new college or university graduate not working," and the fourth was "number of personal computers per 1,000 households" (Table 1).

The R2 increased and reached a plateau ; however, the R2adj did not increase seven variables afterward (Figure 3), when the number of explanatory variables increased. The maximum R2 was 0.756 in 21 variables ; however, the maximum R2adj was 0.639 in seven variables. The R2 was 0.669 (88.3% of the maximum), and the R2adj was 0.610 (95.3% of the maximum) and the maximum in

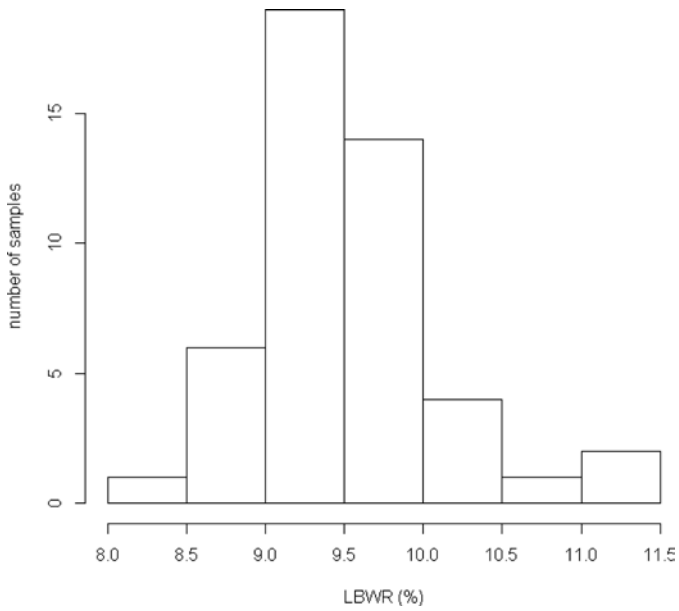

Figure2 : Distribution of the LBWR (%) among 47 prefectures.

The median was 9.5%, the maximum was 11.5%, the minimum was 8.2%, and the mean was 9.5% among 47 prefectures.

Table 1：List of selected 21 SDVs

Variable (unit)	Field	Estimate	Std. Error	t-value	Pr(>|t|)
amount of savings per household (two-or-more-person households)	family budget	1.02E-03	4.88E-04	2.082	0.0478
regional difference index of consumer prices (miscellaneous)	family budget	-7.41E-01	4.21E-01	-1.759	0.0908
ratio of new college or university graduates not working	labor	-7.93E-01	4.56E-01	-1.739	0.0944
number of personal computers per 1,000 households (two-or-more-person households)	family budget	-1.77E-02	1.16E-02	-1.531	0.1383
floor area of newly constructed owned houses (per dwelling, m^2)	dwelling	-4.35E-01	2.98E-01	-1.457	0.1576
popularization of voluntary automobile insurance on property (%)	safety	-3.86E-01	2.66E-01	-1.454	0.1584
juvenile delinquent arrested for criminal larceny (per 1,000 persons of 14–19 years old) (person)	safety	1.33E+00	9.43E-01	1.415	0.1695
ratio of industrial and quasi-industrial area %	dwelling	-3.40E-01	2.59E-01	-1.312	0.2015
ratio of expenditure for housing (two-or-more-person households)	family budget	8.75E-01	7.70E-01	1.137	0.2664
rehabilitation facility for physically disabled person per 1 million person (number of facilities)	welfare and social security	3.24E-01	3.30E-01	0.982	0.3356
outstanding post-office savings per capita 10 thousand yen	economic base	-5.42E-02	6.43E-02	-0.843	0.4074
workers at child welfare in institutions (per 100,000 persons, person)	welfare and social security	4.13E-02	5.15E-02	0.802	0.4301
popularization of voluntary automobile insurance on vehicles (%)	safety	1.49E-01	2.34E-01	0.637	0.5301
monthly expenditure for household per farmer's household	family budget	1.12E-02	1.90E-02	0.588	0.5618
children welfare institutions (per 100,000 persons, number of facilities)	welfare and social security	3.85E-01	7.01E-01	0.549	0.5880
social welfare expenditure per capita(prefecture + municipality, thousand yen	administrative base	-5.00E-02	9.86E-02	-0.507	0.6168
number of stereo sets or CD,MD radio cassette recorders 1,000 households (all households)	family budget	6.94E-03	1.53E-02	0.454	0.6535
rate of divorces (per 1,000 persons)	population and households	-1.85E+00	6.15E+00	-0.301	0.7657
starting salary, senior high school graduate month, female, thousand yen	labor	1.76E-02	2.27E-01	0.077	0.9389
percentage of tax payer	administrative base	-3.33E-02	5.08E-01	-0.066	0.9483
actual number of physically disabled person treated at rehabilitation offices (per 1,000 persons, person)	welfare and social security	2.98E-03	4.81E-01	0.006	0.9951

This table presents 21 variables in ascending order by *p*-value and prefectures after multiple regression of the LBWR as a response variable and these as explanatory variables.
One variable of "stocked money by household average" was statistically significant ($p = 0.048$).

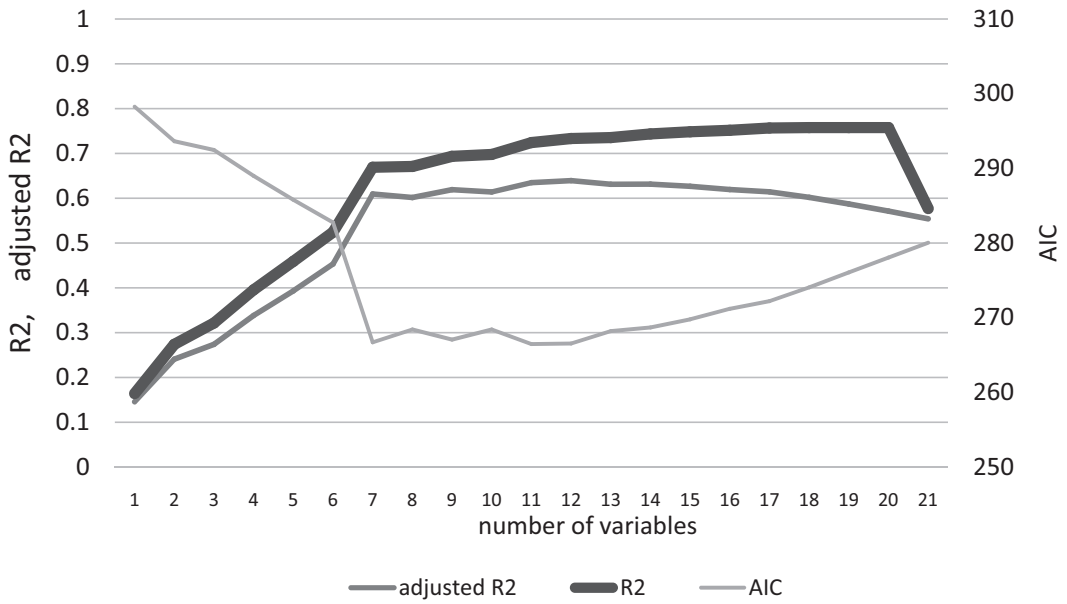

Figure 3：R2, R2adj, and AI in association with the number of explanatory variables.

The number of explanatory variables was that of the top variables in ascending order in Table 1.
A stepwise regression revealed four predictive variables that were the same as the four top variables in Table 1.

seven explanatory variables (Figure 3). Stepwise regression with backward elimination revealed that seven explanatory variables were selected and were the same as the top seven in Table 1.

Discussion

We selected seven proper variables from a great number of SDVs provided on the e-Stat database. The seven top variables determined are likely to well demonstrate regional properties. The number of regions is limited as there are only 47 prefectures in Japan. We chose a smaller number of SDVs, when multilevel analysis was applied to demonstrate regional properties. We repeated the selection of the suitable variables by combining simple correlation and R2adj with multiple regression analysis. Only the "amount of saving per household" variable was statistically significant among the 21 chosen SDVs investigated in this study. However, the R2 and R2adj of this variable were only 0.164 and 0.149, respectively, and not as high as the R2 and the R2adj mean for the descriptive power of the regression model that includes diverse numbers of explanatory variables.

Health outcomes are reported to be influenced by socioeconomic inequities, such as access to healthcare, quality of healthcare, education, health services, number of practitioners/hospital beds, and workforce. Of the 704 SDV, we found no relationship between LBW and socioeconomic variables of prefectural income per person, the rate of people who completed college and university courses, and the rate of hospitals and clinics per population and habitual prefectural area. We are concerned

as to which factor influences the regional differences of the LBWR[17]. The most significant factor was "amount of saving per household," which alludes to security in the community and implies good communication within the communities to maintain safety. We applied the multiple regression analysis to demonstrate the relationship between the LBWR and 21 selected SDVs. If taking into consideration the environment factors of the LBWR, we would have to apply multilevel analysis. Then, our study will contribute to the selection of related regional factors with the LBWR. However, as the relationship would not always be linear, we may need to elucidate the nonlinear relationship between exposures and outcomes using machine learning or deep learning methods.

Limitations of the Study

The method used in this study explored the significant factors from many SDVs ; however, it did not imply the causality between LBW and the factors investigated in this study. After meaningful and proper variables were chosen from among all the available SDVs, multivariate analyses, such as multilevel analysis, would be required to establish regional characteristics.

Principal component analysis is often applied when analyzing multiple variables in sociology and psychology, and this statistical method helps reduce the number of variables. We combined the conventional method of using CCs with the use of R2s to exclude as many variables as possible in this study. As far as we have determined, this is the first report to describe the selection of regional factors from the many SDVs associated with LBWR.

Conclusion

The combination of CC and R2 indicated that the LBWR was significantly related to the regional factor "amount of savings per household" using the social and demographic database in Japan.

Conflicts of Interest

There were no conflicts of interest through the development of this study.

Authors' Contributions

I.Kuboyama helped in the study concept, drafting of the manuscript, and design. S. Ito helped in the statistical analysis. T. Ashizawa and Takeuchi helped in the interpretation of data. A. Maki and Y. Yamaguchi helped in correcting the manuscript. All authors read and approved the final manuscript.

Funding

This study was partially funded by the Institute of Health, Physical Education and Sport Science, Kokushikan University.

Bibliography

1）Valero De Bernabé J, et al.：Risk factors for low birth weight：a review. Eur J Obstet Gynecol Reprod Biol 116：3-15, 2004.
2）Ooka T, et al.：Individual and regional characteristics associated with maternal smoking during pregnancy in Japan：Healthy parents and children 21. Int J Environ Res Public Health 25；17（1）. pii：E173. doi 10.3390/ijerph17010173, 2019.
3）OECD Family Database – OECD, http://www.oecd.org/els/family/database.htm.
4）Kliegman RM, et al.：Nelson textbook of pediatrics 20 eidtion, 789-831, Philadelphia, ELSVIER, 2015.
5）Barker DJP et al.：Infant mortality childhood nutrition and ischemic heart disease in England and Wales. Lancet 1：1077, 1986.
6）Kyle UG et al.：The Dutch Famine of 1944-1945：a pathophysiological model of long-term consequence of wasting disease. Curr Opin Clin Nutr Metab care 9：388-394, 2006.
7）Ramadhani, MK et al.：Lower birth weight predicts metabolic syndrome in young adult：the Atherosclerosis Risk in Young Adults（ARYA）-study. Atherosclerosis 18：21-27, 2006.
8）Joglekar CV et al.：Newborn size, infant and childhood growth, and body composition and cardiovascular disease risk factors at the age of 6 years：the Pune Maternal Nutrition Study Int J Obes.（Lond）31：1534-1544, 2007.
9）Simonetti GD et al.：Salt sensitivity of children with low birth weight. Hypertension 52：625-630, 2008.
10）Adair LS et al.：Size at birth, weight gain in infancy and childhood, and adult blood pressure in 5 low- and middle-income-country cohort：when does weight gain matter? Am J Clin Nutr 89：1383-1392, 2009.
11）Attanasio AF, et al.：International Hypopituitary Control Complications Study Advisory Board Prevalence of metabolic syndrome in adult hypopituitary growth hormone（GH）-deficient patients before and after GH replacement. J Clin Endocrinol Metab 95：74-81, 2010.
12）Virk J et al.：Socio-demographic characteristics of women sustaining injuries during pregnancy：a study from the Danish National Birth Cohort. BMJ Open. 2；2（4）. pii：e000826. doi：10.1136/bmjopen-000826, 2012.
13）Nielsen CC et al.：Spatiotemporal patterns of small for gestational age and low birth weight births and associations with land use and socioeconomic Status. Environ Health Insights 13：1178630219869922, 2019.
14）Janevic T et al.：Associations between racial discrimination, smoking during pregnancy and low birth weight among Roma, The European Journal of Public Health 27：410-415, 2017.
15）Statistics of Japan（e-Stat）, system of social and demographic statistics. https://www.e-stat.go.jp/en/stat-search/database?page=1&layout=datalist&toukei=00200502&tstat=000001111375&cycle=8&tclass1=0000011 11378&result_page=1.
16）Kuboyama I, et al.：Selection of regional factors related to cerebral function after out-of-hospital cardiac arrest in Japan. EC Neurology 11：279-287, 2019.
17）Matsui T et al.：Association between the rates of low birth-weight and/or preterm infants and aircraft noise exposure. Jpn.J.Hyg 58：385-394, 2003.