

Scale development of health status for secondary data analysis using a nationally representative survey

Akihiro Nishi · Haruko Noguchi · Hideki Hashimoto ·
Nanako Tamiya

Received: 9 June 2011 / Accepted: 24 August 2011 / Published online: 15 September 2011
© The Japanese Society for Hygiene 2011

Abstract

Objectives Scale development of health-related quality of life (HRQOL) measures, including physical and mental health measures, among public datasets from Japan is needed for comparative studies on health conditions among different age, gender, and socio-economic subgroups. Multi-attributable scales of continuous/discrete variables on HRQOL could be more flexible for different kinds of epidemiologic and socio-econometric studies

rather than single-item measures. The objectives of this study were to create multi-dimensional scales for physical, mental, and summary health measures and to describe the age-related trends of these scales in Japan.

Methods We utilized data from the 2007 Comprehensive Survey of the Living Conditions of People on Health and Welfare (LCPHW: *Kokumin Seikatsu Kiso Chosa*) ($n = 383,745$) to measure physical health (0 = worst score, 16 = best score) by summarizing four items: general health status, bedridden status/mobility, self-care/usual activities, and pain (0 = worst score, 4 = best score for each item). Mental health was measured using a Japanese version of K6 (0 = worst score, 4 = best score, modified from original version in which 24 = worst score and 0 = best score). We then created a summary health scale using the simple sum of physical and mental health (0 = worst score, 20 = best score). The reliability and validity of the scales were evaluated and their age-related trends described.

Results The internal consistency reliability of the physical and summary health scales was not sufficiently high (Cronbach's $\alpha = 0.64$ and 0.67 , respectively) and the age-related trend was smooth and monotonous. The internal consistency reliability of the mental health scale (K6) was high (Cronbach's $\alpha = 0.90$), while the age-related trend peaked at age 65–74 years.

Conclusions While K6 was a measure with high reliability for describing mental health, use of the physical and summary health scale in the Japanese population requires further discussion. Additional validation tests of the summary scales also need to be performed, in which our methodology is applied to other data sets that include strict diagnostic results based on a structural interview.

A. Nishi (✉)

Department of Society, Human Development, and Health,
Harvard School of Public Health, 677 Huntington Avenue,
Boston, MA 02115, USA
e-mail: anishi@hsph.harvard.edu

A. Nishi

Institute for Quantitative Social Science, Cambridge, MA, USA

A. Nishi · N. Tamiya

Department of Health Services Research,
Graduate School of Comprehensive Human Sciences,
University of Tsukuba, Tsukuba, Ibaraki, Japan

N. Tamiya

e-mail: ntamiya@md.tsukuba.ac.jp

A. Nishi

Department of Public Health, Graduate School of Medicine,
The University of Tokyo, Tokyo, Japan

H. Noguchi

National Institute of Population and Social Security Research,
Tokyo, Japan
e-mail: h.noguchi@ipss.go.jp

H. Hashimoto

Department of Health Economics and Epidemiology Research,
School of Public Health, The University of Tokyo, Tokyo, Japan
e-mail: hidehashimoto-circ@umin.ac.jp

Keywords Japan · Scale development · Health-related quality of life (HRQOL)

Introduction

Japan has achieved the longest life expectancy in the world (2010 data: 79.64 and 86.39 years for men and women, respectively; [1, 2]). However, the quality of health status in Japan, i.e., health-related quality of life (HRQOL) among the entire Japanese population, has yet to be fully examined. Although several studies have attempted to investigate the trends in physical and mental health status using cohort studies, these often experienced major limitations in terms of data and health measurements [3–8]. Two major limitations of such studies are: (1) data which are not nationally representative in terms of sampling method and size of surveyed population; (2) health measures which are not clearly validated because they usually depend on a single domain of health status, such as self-rated health [9, 10], which can be easily affected by errors in the measurement of an individual's characteristics.

Numerous international studies on HRQOL have measured health conditions by multi-dimensional questions [e.g. EuroQol (EQ-5D), the Health Utilities Index (HUI2/HUI3), Short Form (36) Health Survey (SF-36)] [11–13]. For example, EQ-5D consists of five sub-domains (attributes), namely, mobility, self-care, usual activities, pain/discomfort, and anxiety/depression, which represent the personal preferences for health outcomes; HUI2 has six sub-domains, namely, sensation, mobility, emotion, cognition, self-care, and pain, in which mental health constitutes one of the domains of HRQOL. Such composite HRQOL measures are available in representative national data sets compiled in many countries [14–16], but efforts to establish HRQOL measures and to follow up the trend of changes in HRQOL over time in Japan's current and future nationally representative data have, in comparison, lagged behind.

Therefore, the aims of the study reported here were (1) to create multi-dimensional scales for physical, mental, and summary health in the context of HRQOL, and (2) to describe the age-related trends in these scales in the Japanese population, using the most recent nationally representative data on the Japanese population.

Methods

Study population

We utilized the best nationally representative data available, which was a cross-sectional sample of the Comprehensive Survey of the Living Conditions of People on Health and Welfare (LCPHW), conducted by the Japanese Ministry of Health, Labour, and Welfare (MHLW) in June 2007 [17]; permission for secondary use was obtained. A total of 5,440 regional clusters from 47 prefectures in Japan

were randomly sampled, and 624,166 individuals who were at least 15 years of age at the time of survey (in 229,821 households living in the regional clusters) answered the questionnaire. Hospitalized or institutionalized individuals were excluded from the surveyed samples. The response rate 79.9% from 287,707 households. The study population was restricted to those who answered all of the key variables described below (240,421 respondents were excluded). Consequently, the study population for further analysis comprised 383,745 individuals.

Basic characteristics

Individual-level information was obtained on the basic characteristics of the study population for several socio-demographic factors in addition to age and gender. Marital status (married, never married, widowed, or divorced), occupation status (currently yes or no), house ownership (currently yes or no), and smoking behavior (currently smoke, never smoked, previously smoked) were measured for each individual (Table 1).

Physical, mental, and summary health measures

We created a summary health scale that was subsequently divided into two major sub-categories: physical and mental health statuses (Table 1). Although these measures were not identical to a widely used HRQOL measure of EuroQOL (EQ-5D)—the sub-domains of which are mobility, self-care, usual activities, pain/discomfort, and anxiety/depression [11, 12]—we chose relatively similar health-related items that were available in the health-related questions in LCPHW. General health status, which was chosen for our measure, is one of the major components of the other measure of SF-36 [12, 13].

For physical health, we utilized four self-reported items: general health status, bedridden status/mobility, self-care/usual activities, and pain. First, general health status was measured by asking, “How is your current health status: excellent, very good, good, fair, or poor?” We created a discrete variable (4 if excellent; 3 if very good; 2 if good; 1 if fair; 0 if poor). Second, bedridden status/mobility was measured by asking, “How often have you been bed-ridden because of health-related problems for the previous 1 month: never, 1–3 days, 4–6 days, 7–14 days, 15 days or more?” For this items, we created a discrete variable (4 if never; 3 if 1–3 days; 2 if 4–6 days; 1 if 7–14 days; 0 if 15 days or more). Third, self-care/usual activities were ascertained by asking, “Do you have any of the difficulties below in your daily life due to your physical health conditions? yes or no for each: (1) daily movement (e.g., getting out of bed, getting dressed, eating, or bathing); (2) going outdoors; (3) working, doing housework, or

Table 1 Basic characteristics and health measures in the Comprehensive Survey of the Living Conditions of People on Health and Welfare (LCPHW; *Kokumin Seikatsu Kiso Chosa*) 2007, Japan ($n = 383,745$)

Basic characteristics and health measures	Men	Women
<i>n</i>	183,471	200,274
Basic characteristics		
Marital status, <i>n</i> (%)		
Married	124,215 (67.7)	125,184 (62.5)
Never married	48,962 (26.7)	42,287 (21.1)
Widowed	4,820 (2.6)	22,274 (11.1)
Divorced	5,474 (3.0)	10,529 (5.3)
Household size (persons)		
Mean (SE)	3.40 (0.004)	3.42 (0.003)
Occupation status, <i>n</i> (%)		
No	44,005 (24.0)	92,577 (46.2)
Yes	139,466 (76.0)	107,697 (53.8)
House ownership, <i>n</i> (%)		
No	43,374 (23.6)	44,626 (22.3)
Yes	140,097 (76.4)	155,648 (77.7)
Healthcare insurance type, <i>n</i> (%)		
National health insurance	65,862 (35.9)	77,190 (38.5)
Employee's health insurance	114,384 (62.3)	120,083 (60.0)
Other types	3,225 (1.8)	3,001 (1.5)
Smoking behavior, <i>n</i> (%)		
Not smoked	100,354 (54.7)	173,387 (86.6)
Previously smoked	7,190 (3.9)	2,435 (1.2)
Currently smoke	75,927 (41.4)	24,452 (12.2)
Current healthcare needs, <i>n</i> (%)		
No	123,406 (67.3)	124,206 (62.0)
Yes	60,065 (32.7)	76,068 (38.0)
Health measures		
General health status (0 as worst score–4 as best score)		
Mean (SE)	2.45 (0.002)	2.35 (0.002)
Bedridden status/mobility (0 = worst score, 4 = best score)		
Mean (SE)	3.86 (0.001)	3.80 (0.001)
Self care/usual activities (0 = worst score, 4 = best score)		
Mean (SE)	3.80 (0.001)	3.68 (0.002)
Pain (0 = worst score, 4 = best score)		
Mean (SE)	3.85 (0.001)	3.79 (0.002)
Mental health (0 = worst score, 4 = best score)		
Mean (SE)	3.50 (0.002)	3.39 (0.002)
Physical health (0 = worst score, 16 = best score)		
Mean (SE)	14.0 (0.004)	13.6 (0.005)
Summary health (0 = worst score, 20 = best score)		
Mean (SE)	17.46 (0.005)	17.01 (0.006)

SE standard error

studying; (4) exercise or sports.” For this item, we also created a discrete variable (4 if no difficulty; 3 if one difficulty; 2 if two difficulties; 1 if three difficulties; 0 if all difficulties). Fourth, we measured four kinds of pain in different parts of the body (headache, abdominal pain, back pain, extremity pain), for which we also created a discrete variable (4 if no pain; 3 if pain in one location; 2 if in two

locations; 1 if in three locations; 0 if in all locations). Finally, we summed up the scores for these four items to represent physical health (0 = worst score, 16 = best score).

For mental health, we used the Kessler-6 scale (K6), which measures psychological distress based on answers to six questions. The K6 has been widely used around the world [18, 19], and a Japanese version has also been

validated (K6, a discrete variable ranging from 0 to 24) [20, 21]. We created a modified K6 to represent mental health so that higher scores indicated better conditions, in line with the scales for sub-domains in physical health [i.e., 24–0 was converted into 0–4 proportionally; (mental health) = 4 – (original K6)/6]. We then created a summary health scale by simply combining the figures for physical and mental health (physical health + mental health: 0 = worst score, 20 = best score).

Statistical analysis

We checked the inter-item reliability (internal consistency reliability) of physical health (four items), mental health (six items: six questions in K6), and summary health (five items) using Cronbach’s α . To validate the physical and summary health scales, we calculated the areas under the receiver operating characteristic (ROC) curve (AUC) for diagnosed illnesses as the external criteria (any diagnosed co-morbidities with physician management: yes or no). We also calculated the AUCs of components of the summary health scale (general health status, bedridden status/mobility, self-care/usual activities, and pain) and compared these AUCs with those of physical and summary health. For K6, we should utilize strict diagnostic results based on structural interview (i.e., 30-day Diagnostic and Statistical Manual of Mental Disorders; [21]) as the external criteria. There is already an internationally well-established methodology for validating K6, however, the data are not available in LCPHW. In this study, therefore, we decided not to perform a validation test for K6.

We described the age-related trend of physical, mental, and summary health among the study population using the developed/evaluated scales, stratified by gender. We reported conventional two-sided p values without adjustment for multiple testing. All of the analyses were performed using Stata/IC ver. 11.2 (StataCorp LP, College Station, TX).

Results

The basic characteristics of the study participants were similar to those given in the governmental report of LCPHW [17]. Briefly, the majority of participants were married, working, had their own house, employee’s health insurance, were not currently smoking, and not currently receiving healthcare (Table 1). While the means of the four sub-domains (bedridden status/mobility, self care/usual activities, pain, and mental health) ranged from 3.5 to 4, the mean of general health status was less than 3 (mean 2.45). Men reported better scores in physical, mental, and summary health than women.

The reliability test results revealed that Cronbach’s α was 0.64 for physical health, 0.90 for K6, and 0.67 for summary health among the entire population. For validity testing, the AUC for diagnosed illnesses was 0.72 [95% confidence interval (CI) 0.72–0.72] for physical health and 0.71 (95% CI 0.70–0.71) for summary health, as compared with 0.68 for single-item general health status (95% CI 0.68–0.68), 0.55 for bedridden status/mobility (95% CI 0.55–0.55), 0.60 for self-care/usual activities (95% CI 0.60–0.61), and 0.62 for pain (95% CI 0.62–0.62). These results illustrate that the three health measures that we created were better than the single-item results for general health status (self-rated health) and other sub-domains.

Figure 1 shows the trends in physical, mental, and summary health measures by different age-groups and gender. While physical and summary health measures

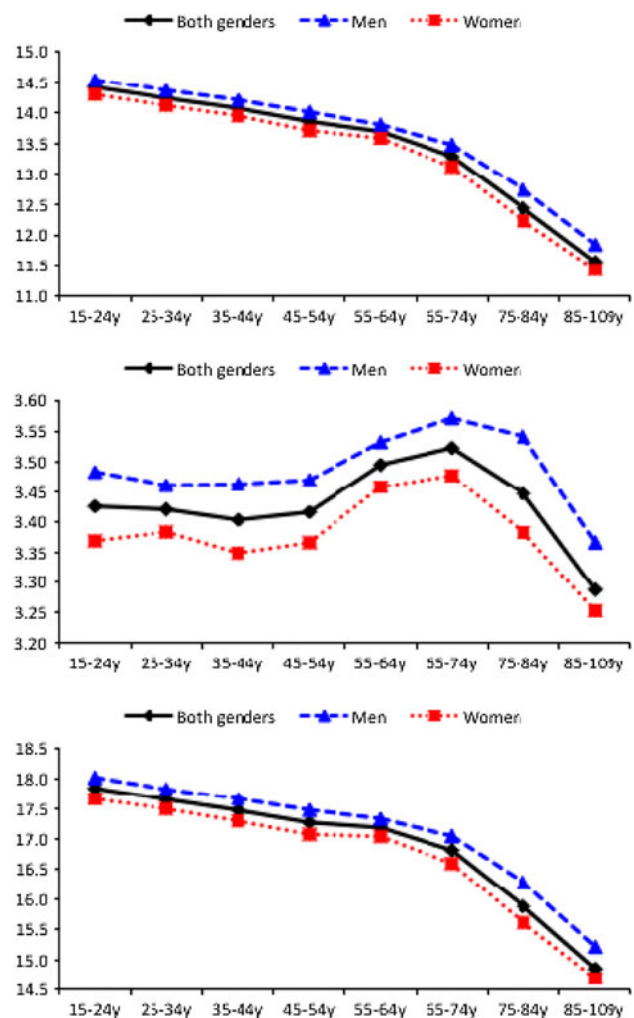


Fig. 1 Physical, mental, and summary health by age-groups. Top Physical health (0 = worst score,16 = best score), middle mental health (0 = worst score, 4 = best score), bottom summary health (physical health + mental health: 0 =worst score, 20 = best score)

declined monotonically with increasing age, mental health peaked around age 65–74 years. After 65–74 years of age, mental health declined with increasing age. The trends were the same among men and women.

Discussion

We examined the reliability and validity of the physical, mental, and summary health scales as measures of HRQOL in the Japanese population using a nationally representative sample from a 2007 survey. The reliability is debatable and should be subjected to further empirical analysis, but the validities of the physical health and summary health measures were within a statistically acceptable range (better than the single-item self-rated health measure).

Our study identified several interesting age-related trends in physical and mental health. Physical health of our sample decreased monotonously with increasing age; however, the slope of the decline was shallower in the younger generation (age 15–64 years) and steeper in the older generation (age 65+ years). This trend could be related to the presence of multiple diseases in the older generation [22]. Mental health peaked at age 65–74 years and sharply declined after age 75–84 years. These results suggest that mental health and health-related well-being is rated as “best” after the mandatory retirement age of 60 years (in Japan), probably due to emancipation from demanding labor, child-bearing, and care-giving (parents/parents-in-law) activities, which mainly fall between age 45 and 65 years. These possible explanations for our results should be tested in future research using the LCPHW.

There are three major limitations to this study. First, missing values of physical health and mental health (K6) may have influenced the results because many participants (15.8% in the entire population) did not respond to the questions of the K6 in the LCPHW and missing observations may not have been random. We excluded these individuals from our analysis, which also could have affected the results in terms of the reliability and validity of these three health measures. For further study, we may adopt an imputation technique to our health measures [23].

Second, the internal consistency reliability of the physical and summary health scales was not sufficiently high. Also, the external criteria for validity tests are not adequate enough to support our results. These could be the most significant shortcomings of the scales. Nevertheless, the former shortcoming may characterize the multi-dimensional/attributable structure of these scales. When each sub-domain/question is a directly correlated measure of the latent variable (e.g., K6 for psychological distress), Cronbach's α could be required to be very high (e.g., >0.8) [24]. In contrast, the multi-attributable sub-domains in the

physical and summary health measures could be overlapped (sharing the same latent variable), but not identical, which suggests Cronbach's α does not have to be very high. In terms of the latter concern for further validation tests, we need to apply our methodology to other data, including more strict diagnostic results based on a structural interview.

Third, the summary health scale, by integrating the physical and mental health scale in this study, remains questionable and should be more carefully examined. We simply summed up the physical health scale [0 (worst score) to 16 (best score)] and mental health scale [0 (worst score) to 4 (best score)], extracting the summary health scale [0 (worst score) to 20 (best score)]. We followed this procedure because we chose the HRQOL scale weights on each sub-domain based on multi-attribute utility theory [12, 25] and also followed another study's scale development based on internationally compatible U.S. datasets (Health and Retirement Survey) [16]. However, this simple summation (0–16 + 0–4) cannot always be justified because the assumption that the contribution of the physical health scale is fourfold greater than that of the mental health scale toward “overall” health status is not necessarily acceptable. Therefore, for the future studies, we propose two alternative ways of calculation: (1) physical health scale (range 0–16) + 4 \times mental health scale (range 0–16), and (2) the weight on each sub-domain based on standard gamble or time trade-off methods in cost-utility analysis [12].

Our health scales with the LCPHW datasets have several practical strengths. First, the data that we used included nationally representative samples so that the generalizability problem—the major issue of analyses involving community (or convenient) samples—is unlikely to appear. Second, almost all of the variables, except for K6, which had been asked only after 2007, that we used in this study, including the health measures, are available for—and compatible with—the LCPHW in different years (1989, 1992, 1995, 1998, 2001, 2004, 2007, and 2010). Although knowledge is required to analyze repeated cross-section or pseudo-cohort datasets, such as difference-in-difference estimations or multilevel analysis [26–28], the development of reliable health measures will provide physicians, epidemiologists, economists, researchers from various academic fields, and policy-makers with the means to analyze the socio-demographic trends in health status and health disparity in Japan for the past 20 years (1989–2010) both consistently and thoroughly.

In conclusion, further use of the physical and summary health scale reported here in the Japanese population requires further discussion, although the K6 was an excellent measure of mental health in the LCPHW. Future research should focus on confirming and improving the reliability and the validity of these measures.

Acknowledgments We were funded by the Japanese Ministry of Health, Labour, and Welfare (H22-Policy-033), Bill and Melinda Gates Foundation, and China Medical Board. A.N. acknowledges support from the Nakajima Foundation for his academic endeavor at Harvard University.

Conflict of interest The authors declare that they have no conflict of interest.

References

- Sugiura Y, Ju Y, Yasuoka J, Jimba M. Rapid increase in Japanese life expectancy after World War II. *Biosci Trends*. 2010;4:9–16.
- Ministry of Health, Labour, and Welfare. Abridged life tables for Japan. 2010. Available at: <http://www.mhlw.go.jp/english/data/base/db-hw/lifetb10/1.html>.
- Nishi A, Kondo K, Hirai H, Kawachi I. Cohort profile: the ages 2003 cohort study in Aichi, Japan. *J Epidemiol*. 2011;21:151–7.
- Kuriyama S, Nakaya N, Ohmori-Matsuda K, Shimazu T, Kikuchi N, Kakizaki M, et al. The Ohsaki Cohort 2006 Study: design of study and profile of participants at baseline. *J Epidemiol*. 2010;20:253–8.
- Kondo N, Kawachi I, Hirai H, Kondo K, Subramanian SV, Hanibuchi T, et al. Relative deprivation and incident functional disability among older Japanese women and men: prospective cohort study. *J Epidemiol Community Health*. 2009;63:461–7.
- Kagamimori S, Gaina A, Nasermoaddeli A. Socioeconomic status and health in the Japanese population. *Soc Sci Med*. 2009;68:2152–60.
- Ichida Y, Kondo K, Hirai H, Hanibuchi T, Yoshikawa G, Murata C. Social capital, income inequality and self-rated health in Chita peninsula, Japan: a multilevel analysis of older people in 25 communities. *Soc Sci Med*. 2009;69:489–99.
- Schoeni R, Liang J, Bennett J, Sugisawa H, Fukaya T, Kobayashi E. Trends in old-age functioning and disability in Japan, 1993–2002. *Popul Stud*. 2006;60:39–53.
- Jylha M. What is self-rated health and why does it predict mortality? Towards a unified conceptual model. *Soc Sci Med*. 2009;69:307–16.
- DeSalvo KB, Bloser N, Reynolds K, He J, Muntner P. Mortality prediction with a single general self-rated health question. A meta-analysis. *J Gen Intern Med*. 2006;21:267–75.
- Kind P, Brooks R, Rabin R. EQ-5D concepts and methods: a developmental history. Dordrecht: Springer; 2005.
- Drummond MF, Sculpher MJ, Torrance GW, O'Brien BJ, Stoddard GL. Methods for the economic evaluation of health care programmes. 3rd ed. Oxford: Oxford University Press; 2005.
- Ware JE, Kosinski M, Bayliss MS, McHorney CA, Rogers WH, Raczek A. Comparison of methods for the scoring and statistical analysis of SF-36[®] health profiles and summary measures: summary of results from the Medical Outcomes Study. *Med Care*. 1995;33:AS264–79.
- Goldberg M, Leclerc A, Bonenfant S, Chastang JF, Schmaus A, Kaniewski N, et al. Cohort profile: the GAZEL Cohort Study. *Int J Epidemiol*. 2007;36:32–9.
- Marmot M, Brunner E. Cohort profile: the Whitehall II study. *Int J Epidemiol*. 2005;34:251–6.
- McWilliams J, Meara E, Zaslavsky A, Ayanian J. Health of previously uninsured adults after acquiring Medicare coverage. *JAMA*. 2007;298:2886.
- Ministry of Health, Labour and Welfare. The comprehensive survey of the living conditions of people on health and welfare (*Kokumin Seikatsu Kiso Chosa*) (in Japanese). Available at: <http://www.mhlw.go.jp/toukei/list/20-19.html>. Accessed 8 June 2011.
- Kessler RC, Green JG, Gruber MJ, Sampson NA, Bromet E, Cuitan M, et al. Screening for serious mental illness in the general population with the K6 screening scale: results from the WHO World Mental Health (WMH) survey initiative. *Int J Methods Psychiatr Res*. 2010;19[Suppl 1]:4–22.
- Kessler RC, Andrews G, Colpe LJ, Hiripi E, Mroczek DK, Normand SLT, et al. Short screening scales to monitor population prevalences and trends in non-specific psychological distress. *Psychol Med*. 2002;32:959–76.
- Sakurai K, Nishi A, Kondo K, Yanagida K, Kawakami N. Screening performance of K6/K10 and other screening instruments for mood and anxiety disorders in Japan. *Psychiatry Clin Neurosci*. 2011;65:434–41.
- Furukawa T, Kawakami N, Saitoh M, Ono Y, Nakane Y, Nakamura Y, et al. The performance of the Japanese version of the K6 and K10 in the World Mental Health Survey Japan. *Int J Methods Psychiatr Res*. 2008;17:152–8.
- Landi F, Liperoti R, Russo A, Capoluongo E, Barillaro C, Pahor M, et al. Disability, more than multimorbidity, was predictive of mortality among older persons aged 80 years and older. *J Clin Epidemiol*. 2010;63:752–9.
- King G. Analyzing incomplete political science data: an alternative algorithm for multiple imputation. *Am Polit Sci Rev*. 2001;95:49–69.
- Noguchi H. Relations between socioeconomic status and health among Japanese adult population (*Shakai keizai teki youin to kenko tono ingasei ni taisuru kousaru*) (in Japanese). *Q Soc Secur Res*. 2011;46:382–402.
- Keeney R, Raiffa H. Decisions with multiple objectives: preference and value tradeoffs. New York: Wiley; 1976.
- Wooldridge JM. Econometric analysis of cross section and panel data. 2nd ed. Cambridge: The MIT Press; 2010.
- Kreft I, de Leeuw J. Introducing multilevel modeling. London: Sage Publications; 1998.
- Iwamoto Y. Pseudo-panel data by use of the Comprehensive Survey of the Living Conditions of People on Health and Welfare from 1989 to 1995 (*Kokumin Seikatsu Kiso Chosa ni yoru giji panaru data: 1989–1995*). In: National Institute of Population and Social Security Research, editor. The change of living arrangement and its function of social security (*Kazoku Setai no Henyo to Seikatsu Hosho Kinou*) (in Japanese). Tokyo; 2000. pp. 329–56.