BMJ Open is committed to open peer review. As part of this commitment we make the peer review history of every article we publish publicly available.

When an article is published we post the peer reviewers' comments and the authors' responses online. We also post the versions of the paper that were used during peer review. These are the versions that the peer review comments apply to.

The versions of the paper that follow are the versions that were submitted during the peer review process. They are not the versions of record or the final published versions. They should not be cited or distributed as the published version of this manuscript.

BMJ Open is an open access journal and the full, final, typeset and author-corrected version of record of the manuscript is available on our site with no access controls, subscription charges or pay-per-view fees (http://bmjopen.bmj.com).

If you have any questions on BMJ Open's open peer review process please email editorial.bmjopen@bmj.com

The Kanagawa Investigation of the Total Checkup Data from the National Database (KITCHEN): Protocol for population-based repeated cross-sectional and 6-year cohort studies

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2018-023323
Article Type:	Protocol
Date Submitted by the Author:	03-Apr-2018
Complete List of Authors:	Nakajima, Kei; Kanagawa University of Human Services, School of Nutrition and Dietetics, Faculty of Health and Social Services; Saitama Medical Center, Saitama Medical University, Department of Endocrinology and Diabetes Iwane, Taizo; Kanagawa University of Human Services, School of Nutrition and Dietetics, Faculty of Health and Social Services Higuchi, Ryoko; Kanagawa University of Human Services Shibata, Michi; Kanagawa University of Human Services; St. Marianna University School of Medicine Takada, Kento; Kanagawa University of Human Services Uda, Jun; Jikei Institute, Graduate School of Health Care Sciences Anan, Mami; Kanagawa University of Human Services Sugiyama, Michiko; Kanagawa University of Human Services Nakamura, Teiji; Kanagawa University of Human Services
Keywords:	EPIDEMIOLOGY, GENERAL MEDICINE (see Internal Medicine), GERIATRIC MEDICINE, NUTRITION & DIETETICS, PREVENTIVE MEDICINE, PUBLIC HEALTH

SCHOLARONE™ Manuscripts

The Kanagawa Investigation of the Total Checkup Data from the National Database (KITCHEN): Protocol for population-based repeated cross-sectional and 6-year cohort studies

Kei Nakajima,[1,2]* Taizo Iwane,[1] Ryoko Higuchi,[1] Michi Shibata,[1,3] Kento Takada,[1] Jun Uda,[4] Mami Anan,[1] Michiko Sugiyama,[1] Teiji Nakamura[1]

[1]School of Nutrition and Dietetics, Faculty of Health and Social Services, Kanagawa University of Human Services, 1-10-1 Heisei-cho, Yokosuka, Kanagawa 238-8522, Japan

[2]Department of Endocrinology and Diabetes, Saitama Medical Center, Saitama Medical University, 1981 Kamoda, Kawagoe, Saitama 350-8550, Japan

[3]Department of nutrition, St. Marianna University School of Medicine, 2-16-1 Sugao, Miyamae-ku, Kawasaki, Kanagawa 216-8511, Japan

[4]Graduate School of Health Care Sciences, Jikei Institute, 1-2-8 Miyahara, Yodogawa-ku, Osaka-shi, 532-0003, Japan

Running title: Kanagawa Investigation of Checkup Data

Address correspondence to Kei Nakajima, M.D., Ph.D.

School of Nutrition and Dietetics, Faculty of Health and Social Services, Kanagawa University of Human Services, 1-10-1 Heisei-cho, Yokosuka, Kanagawa 238-8522, Japan.

e-mail: nakajima-rsh@kuhs.ac.jp; Tel: +81-046-828-2660

Abstract

Introduction: The unmitigated incidence of cardiometabolic diseases, such as type 2 diabetes and metabolic syndrome, has gained attention in Japan. “Big data” can be useful to clarify conflicting observations obtained from studies with small samples and about rare conditions that are often neglected. We epidemiologically address these issues using data from health checkups conducted in Kanagawa Prefecture, the prefecture with the second largest population in Japan, in the Kanagawa Investigation of the Total Checkup Data from the National Database (KITCHEN).

Methods and analysis: This research consists of a series of population-based cross-sectional studies repeated from 2008 to 2014 and 6-year cohort studies. Since 2017, we have reviewed the data of people living in Kanagawa Prefecture who underwent a health checkup mainly for general health and the prevention of metabolic syndrome. The sample size ranges from 1.2 million to 1.8 million people in the cross-sectional studies and from 370,000 to 590,000 people in the cohort studies. These are people aged 40 to 74 years, whose clinical parameters were measured and who responded individually to a questionnaire. We investigate potential associations and causalities of various etiologies, including diabetes and metabolic syndrome, using clinical data and lifestyle information. With multidisciplinary analysis, we expect to obtain a wide range of novel finding, to confirm indeterminate previous findings, especially in terms of cardiometabolic disease, and to provide new perspectives for human health promotion and disease prevention.

Ethics and dissemination:

Ethical approval was received from the Ethics Committee of Kanagawa University of Human Services (10-43). The protocol was approved in December 2016 by the Japanese Ministry of Health, Labour and Welfare (No. 121). The study results will be disseminated through open platforms including journal articles, relevant conferences, and seminar presentations.

Keywords: Kanagawa; checkup; big data; data mining; cardiometabolic disease; national database; health records; lifestyle-related disease; age-related disease

Strengths and limitations of this study

•The number of subjects in the sample is so large that more precise results concerning the means of parameters can be obtained, even when the subjects are classified into multiple categories, including sex, age group, smoking status, and certain morbidities such as obesity or diabetes.

•It may be possible to use big data to evaluate minor or rare conditions or etiologies that are commonly overlooked, neglected, or unfeasible to analyze in clinical studies, particularly those with small samples.

•Identical measurements and assessments of anthropometric indices, blood pressure, blood biochemistry, and urinalysis are performed across multiple years in the prefecture with the second largest population in Japan.

•The variations in parameters are restricted, and parameters for specific diseases are not included, because the checkups are conducted for general health and the prevention of common diseases, especially lifestyle-related diseases such type 2 diabetes and metabolic syndrome.

•Although cohort analysis is possible with this dataset, at 6 years, the period is relatively short, which may hamper the ability to uncover the latent relationships and underlying mechanisms between the parameters used and the predicted outcomes.

Introduction

Over the past several decades, the incidence of cardiometabolic diseases such as type 2 diabetes and metabolic syndrome (MetS) has not been reduced and has gained attention in Asia, including Japan [1,2], which has also experienced an unprecedented acceleration of societal aging [3,4]. These issues may also be problematic in Kanagawa Prefecture (***Figure 1***), an eastern district of Japan, located near Tokyo. The primary causes of these diseases include unfavorable lifestyles (e.g., smoking, heavy alcohol consumption, insufficient sleep, and infrequent exercise) and excess body weight (overweight and obesity) because of overeating, along with individuals' genetic and epigenetic backgrounds. However, for the last decade, malnutrition (e.g., low body weight) has been shown to be prevalent among young women [5-7] and the elderly [8-10] in Japan, which may contribute to the increased rates of sarcopenia and frailty in the country. Combined with prolonged longevity, cardiometabolic diseases with age-related causes create a long-term burden that leads to direct (i.e., measurements and therapies) and indirect (e.g., nursing, care, and welfare) medical costs nationwide [11,12], particularly when severe complications such as organ failure (e.g., heart, liver, and renal failure) develops over the life course.

In 2008, a special health checkup was initiated, primarily for the prevention of MetS, by the Ministry of Health, Labour and Welfare (MHLW) in Japan [13]. Since that time, all people living in Japan aged 40 to 74 years are supposed to undergo a yearly health checkup. The data from these checkups have continuously accumulated, creating a very large database. Such "big data" are likely to be useful in clarifying indistinct or conflicting results obtained from clinical studies with small [14,15], confirming established results and advancing them by elucidating plausible mechanisms and clinical relevance, and enabling a precise understanding of the current status of public health and the contributing to it. Additionally, "big data" of this kind enable us to investigate minor or rare conditions and etiologies [14], such as extremely low and high body weight, abnormal (low and high) clinical measurements, and the high prevalence of unfavorable habits and lifestyles, where etiologies are hardly understood, primarily because of inadequate numbers of observations and corresponding cases.

Although cardiometabolic diseases such as type 2 diabetes and chronic organ failures such as chronic kidney disease have been increasing along with the prolonged longevity in Japan [3,4], the underlying associations with clinical parameters and their mechanisms have not been fully elucidated or confirmed, particularly in epidemiological studies using the "big data" from the health checkups described above. These data include more than one million observations per year in most prefectures in Japan. To date, no investigation of this type has been performed, especially on the prefecture scale in Japan.

To this end, we investigated current cardiometabolic disease and health status as clearly as possible, as well as the relationship of cardiometabolic diseases, including but not limited to type 2 diabetes and MetS, and age-related etiologies. We focused especially on the thorough, end-to-end analysis of the variables of interest, using digitally recorded accumulated data in an extremely large epidemiological study of Kanagawa Prefecture, the second most populated prefecture in Japan, with approximately 9 million inhabitants, second only to Tokyo (approximately 13.7 million inhabitants), as of October 2017.

Methods and analysis

Design

In 2013, the MHLW began to offer accumulated data consisting of information on patient prescriptions and health checkups for use by Japanese institutions including universities, hospitals, and research centers. These data are recorded digitally and are provided in a third-party manner, according to the concept of the "provision of medical-related data to a third party" to improve the quality of medical services and to support academic research in Japan [16]. To date, over 200 institutions in Japan have received the dataset in this manner.

Our project was a composite multidisciplinary study aimed at elucidating the factors associated with cardiometabolic diseases and eventually contributing to the amelioration and advancement of social health and welfare. After the study protocol was approved by the ethics committee of Kanagawa University of Human Services (10-43), we applied to the MHLW's data provision system in October 2016, through Teiji Nakamura, the President of Kanagawa University of Human Services, as a representative. The protocol of our study was approved in December 2016 by the MHLW (No. 121), after a peer review by an expert council.

Before we received the database from the MHLW, identifying individual-level information (names and postal codes) was completely transformed into randomized non-distinguishing anonymous numbers and characters, which prevents the restoration of this information by any means. To further protect against the identification of specific individuals, age was categorized into 5-year age groups (40–44, 45–49, 50–54, 55–59, 60–64, 65–69, and 70–74 years), so that the individual's precise age at the time of data collection is unknown in our study.

Our study is part of the MHLW's nationwide program of providing medical-related data to third parties [16], and informed consent for the use of these data has not been obtained from each subject. We have opened the protocol of our study to the public on our university homepage, which was updated in October 2017 [17], in line with the "Ethical Guidelines for Medical and Health Research Involving Human Subjects" [18] in Japan (updated by the MHLW and the Ministry of Education, Culture, Sports, Science and Technology in May 2017). We received the digitally recorded non-distinguishing anonymous data from the MHLW in August 2017.

Our analysis of the data was conducted in a location with restricted access and tight security regarding datasets at Kanagawa University of Human Services. Repeated cross-sectional studies will be conducted using checkup data from 2008 to 2014. Additionally, a historical cohort study will be conducted, using the 2008 data as a baseline and the 2014 data to assess final outcomes (***Figure 2***). During this period, the number of subjects undergoing a health checkup has increased each year in Kanagawa in parallel with the nationwide trend. Nationally, almost 50% of the population attended a health checkup in 2014, probably because of the political encouragement for these checkups (***Figure 3***).

The name of the full study is the Kanagawa Investigation of the Total Checkup data from the National Database (KITCHEN). Each subsequent KITCHEN publication will be numbered sequentially from 1.

Subjects and measurements

People aged 40 to 74 years and living in Kanagawa Prefecture at the time of the data collection were enrolled in a series of studies. Those residing in medical institutions including hospitals and nursing homes were not included. All subjects are thought to be active to the extent of coming to the place where the checkup was performed. However, some of the subjects have diseases such as hypertension, diabetes, or dyslipidemia, and some have a history of morbidities such as heart disease or stroke. All of these conditions are digitally recorded as answers to a questionnaire. Specific exclusion and inclusion criteria will be determined for each study in the future. The sample sizes range from 1.2 million to 1.8 million people in the cross-sectional studies and from 370,000 to 590,000 people in the cohort studies (***Figure 2***). Cohort Study I uses data from 590,000 people who attended checkups in 2008 and 2014. Cohort Study II is based on data from 370,000 people who attended a checkup every year from 2008 to 2014.

Patient and Public Involvement:

Patients are not involved in this study.

All of the parameters measured in this study are listed in ***Table 1***. To evaluate subject's age as a numeric value, we transformed age group (40–44, 45–49, 50–54, 55–59, 60–64, 65–69, and 70–74 years) into substituted age (s-age), corresponding to the median for each age group (42, 47, 52, 57, 62, 67, and 72 years). Body weight and height were objectively measured by trained institutional staff members and were recorded to one decimal place (kg and cm). Body mass index (BMI) was calculated as weight (in kg) divided by height (in m^2). In most cases (approximately 99.9%), waist circumference (WC) was measured objectively at the navel level by a medical staff member and recorded to one decimal place. Biochemical measurements were performed using standard methods and automated machines. Dipstick urine analysis for proteinuria and glycosuria was assessed visually or with ordinary automated machines. Several different methods were used for the included biochemical parameter (***Table 1***).

In principal, most people underwent a checkup after overnight fasting. However, some of the checkups were conducted in a non-fasting condition because of, for example, shift work or family reasons. Therefore, all subjects were asked for the time (in hours) from their last meal to the time of the checkup, which was recorded as at least 10 hours or less than 10 hours. Those completing the checkup less than 10 hours after their last meal will be distinguished from others in certain sub-studies—for instance, when examining diabetes or dyslipidemia.

The Japanese diagnostic criteria for MetS were published in 2005 [19]. Unlike other criteria such as that of the Adult Treatment Panel III (ATP-III) and the International Diabetes Federation (IDF) [20,21], the Japanese MetS criteria include abdominal obesity as an essential condition (WC ≥ 85 cm for men and ≥ 90 cm for women), in addition to two or more of the following three components: 1) dyslipidemia (triglycerides ≥ 150 mg/dl and/or high-density lipoprotein cholesterol < 40 mg/dl, or pharmacotherapy for dyslipidemia); 2) hypertension (systolic blood pressure ≥ 130 mmHg and/or diastolic blood pressure ≥ 85 mmHg, or pharmacotherapy for hypertension); and 3) hyperglycemia (fasting plasma glucose (FPG) ≥ 110 mg/dl or pharmacotherapy for diabetes). In the practice of the health checkups, hyperglycemia is defined as having elevated FPG (≥ 110 mg/dl) and/or HbA1c (National Glycohemoglobin Standardization Program [NGSP]) ≥ 6.0%, or pharmacotherapy for diabetes. Furthermore, pre-MetS is defined as abdominal obesity plus one of the three components listed above [22]. In sub-studies concerning MetS, MetS will be also determined using other international criteria, such as that of the ATP-III or the IDF to allow for comparison with the same criteria with other Asian countries as well as with Western countries. HbA1c (Japan Diabetes Society [JDS]) was converted to HbA1c (NGSP) units using the officially certified formula: HbA1c (NGSP) (%) = 1.02 × JDS (%) + 0.25% [23].

If the data on serum creatinine eventually become available, the estimated glomerular filtration rate (eGFR) will be calculated using the s-age above and the following equation [24]: eGFR (ml/min/1.73 m^2) = 194 × serum $Cr^{-1.094}$ × s-$age^{-0.287}$ (if female) × 0.739, where Cr denotes serum creatinine concentration (mg/dl).

Hypertensive retinopathy has been shown to be associated with cardiovascular events and mortality [25,26]. Hypertensive retinopathy assessments using the Keith–Wagener and Scheie classifications are available in the study, although a very small percentage of individuals (around 1.3%) completed the hypertensive retinopathy examination.

The standardized 22-item questionnaire created by the MHLW for the health examination checkups is shown in ***Table 2***.

Primary and secondary (minor) outcomes

In a series of studies, we will consider various conditions and etiologies as primary and secondary (or minor) outcomes (see ***Table 3***). However, because unexpected findings are likely to be obtained during these studies and related research topics will be pursued following these findings, we do not restrict the areas of research to be pursued as long as the findings can contribute to or advance specific or general health objectives.

Statistical analysis

Continuous and categorical variables will be compared between different groups using analysis of variance (ANOVA) and χ^2-tests, respectively. Post hoc comparisons between two specific groups will be examined with the Bonferroni, Tukey–Kramer, and Dunnett methods, as well as additional χ^2-tests. Paired or trend data using the χ^2-test will rely on the McNemar and Cochran–Armitage tests, respectively. Analysis of covariance (ANCOVA) with general linear model procedures will be used to examine the difference in biochemical variables measured by two or three different methods (e.g., LDL cholesterol is measured using three different methods) (***Table 1***), controlling for confounders including age, sex, body weight, and various lifestyles. Logistic regression and hazard models will be used to examine the associations or causalities between abnormal levels of measurements and conditions with major and minor outcomes. These methods will yield odds ratios, risk ratios, or hazard ratios, which will be presented along with their 95% confidence intervals. Panel data analysis (including the Hausman test) combining several sets of cross-sectional data will be also conducted. Relevant confounding factors include age, sex, smoking, and alcohol consumption, which will be adjusted in the regression analyses. Alternatively, to evaluate or control the differences in backgrounds and various confounders between cases and controls, individuals' propensity scores will be calculated as a variable that unifies all corresponding confounders in the analysis. Propensity score is also considered for a special examination, for instance, hypertensive retinopathy examination, because few subjects underwent such special examination, which yields a bias to be adjusted.

There are missing data in our study, although this comprises fewer than 5% of the cases for most parameters and questionnaire items. For categorized age, sex, BMI, and WC (***Table 1***), the data are almost complete, even when all of these variables are combined (99.99%). However, combining parameters other than age, sex, BMI, and WC can decrease the total available number of subjects, depending on the study's nature and design.

Statistical analyses will be performed using SAS-Enterprise Guide (SAS-EG 7.1) in the SAS system, Version 9.4 (SAS Institute, Cary, North Carolina, USA). Values of $p < 0.05$ will be considered statistically significant. Because we understand that large data are predisposed to detect the presence of statistical significance, we will use caution in our interpretations and give priority to clinical significance rather than statistical significance in certain clinical areas.

Overall characteristics of subjects

Subjects' characteristics at baseline (2008) are shown in ***Table 1***. These findings will vary to some extent in sub-studies using other cross-sectional data from 2009 to 2014. Men in the first five years of their 50s are overrepresented in the sample, probably because middle-aged men are more likely than women to work for companies and institutes, which obligate workers to undergo a checkup. Considering the clinical parameters of BMI, WC, blood pressure, lipids, FPG, HbA1c, and MetS, most of the subjects are apparently healthy people with these parameters within normal ranges. The questionnaire results (2008) are shown in ***Table 2***, which gives us rough information about the subjects' backgrounds. The smoking rate is higher (25.9%), especially among men (37.2%), compared with other developed countries such as the United States [27,28], although the smoking rate has been declining in Japan in recent years [27].

The prevalence of MetS (13.4%), as well as pharmacotherapy for hypertension, diabetes, and dyslipidemia are relatively lower (3.5%–17.1%), compared with other countries [29,30]. However, this does not always mean that the number needed to treat is low, because substantial proportions of subjects likely do not consult a doctor about their poor glycemic control. A Japanese national survey conducted from 2005 to 2009 found that, among people with diabetes, a substantial proportion (about

38%) have left their poor glycemic conditions as they are, without seeking treatment [31]. In our study, the extent of this issue remains unknown without a detailed investigation of FPG and HbA1c.

Concerning eating habits, which play an important role in metabolism and nutrition, the percentage of subjects who habitually skip breakfast is lower (14.6%) than the percentage who eat dinner two hours before bedtime (28.5%), which is consistent with a previous study [32]. This suggests that the latter group may be more troublesome in terms of unfavorable lifestyle habits that are linked to cardiovascular diseases, because a close association between eating dinner late at night and skipping breakfast has been reported in a community-based epidemiological study [32]; eating dinner late at night can lead to skipping breakfast the next morning. Acknowledging the relationship with sleep, we term these behaviors "unfavorable eating habits around sleep" (UEHAS). In previous studies [32,33], eating dinner late at night together with skipping breakfast—a combination representative of UEHAS—was significantly associated with MetS, proteinuria, and atrial fibrillation.

Body weight substantially influences the incidence and development of cardiometabolic diseases as well as general health [34,35]. However, features and etiologies at both extremes of BMI, a fundamental index of weight considering height, body adiposity, nutritional status, and health, are poorly understood. For rare conditions in malnutrition, for instance, the percentage of subjects with an extremely low body weight (BMI < 15.0 kg/m^2, a criterion for high mortality [36]) is very small in the cross-sectional dataset in 2008 (0.1%), but the observational number in this dataset is large (n = 1,217, data not shown), which is not ignorable and may be enough to conduct proper statistical analyses. Likewise, extremely high body weight (BMI ≥ 40.0 kg/m^2, a criterion for class III obesity [37]) is also very small (0.07%), but there are 805 observations for this group in this dataset.

Throughout the study presented here, we expect to obtain a wide range of novel observations, enabling us to confirm indeterminate previous findings, especially in terms of cardiometabolic disease. In addition, this study will likely reveal underlying etiologies that have been overlooked because they are rare or minor cases in small clinical studies.

Among our research team members, one person (Nakajima) has previously been involved in a similar large study, consisting of approximately 100,000 people living in Saitama Prefecture (population = 7.3 million people), which is also located near Tokyo [38] (***Figure 1***). This previous study and its sub-studies were launched in 2011 and, to date, multiple findings have been reported from these studies. Although a study of 100,000 people is generally considered "large," this sample size sometimes proved to be inadequate for stratification analysis because of small observational numbers for particular groups when several stratification variables were combined—for instance, age group, BMI category, and diabetes status [39]. This is the main reason we chose to begin a new study using the checkup data for an extraordinarily large sample—over one million for a cross-sectional study, which is more than 10 times higher than this previous study [38].

BMI roughly reflects nutritional status, including excess energy accumulation or malnutrition. In the last decade, increased percentages of people have been found at both extremes of BMI (i.e., in the nutritional states of malnutrition or obesity), worldwide, especially among children. This coexistence of undernutrition and obesity or nutrition-related noncommunicable disease has been termed the "double burden of malnutrition" [40-42]. In our study, subjects' characteristics described above suggest that the double burden of malnutrition can exist even among the middle-aged Japanese population, although the proportions are smaller compared with other developing countries [40].

Our study has several strengths. First, in terms of community-based epidemiological research, the sample is so large that precise results concerning the means of parameters can be obtained, even when subjects are classified into categories such as sex, age group, smoking status, and certain morbidities (e.g., obesity or diabetes). However, working with such big data does not allow us to use common database software such as Excel or Access (Microsoft Inc.) because of the limited number of lines in a data sheet (less than approximately one million) or standard computers with normal specifications (e.g., central processing unit, memory, or hard disk) [15]. A second strength of this study is that it may be possible to use big data to evaluate minor or rare conditions and etiologies that are commonly overlooked, neglected, or unfeasible to analyze in clinical studies, particularly those with small samples [14]. This analysis may contribute to case studies instead of only to the field of public health. Finally, identical measurements and assessments of anthropometric indices, blood pressure, and urinalysis are performed across multiple years in people living in similar environment and the same health care system.

Some limitations to this study should be also mentioned. First, the variations in parameters are restricted, and parameters for specific diseases are not included, because the checkups are conducted for general health and the prevention of common diseases, especially lifestyle-related diseases such type 2 diabetes and MetS. Second, people younger than 40 years and those aged over 74 years are not enrolled in this study. Lifestyle choices made when people are younger may contribute to the incidence of morbidities in middle age, and lifestyles and clinical biochemistry levels in middle age can influence the incidence and severity of cardiovascular diseases and health damage in the later life. Unfortunately, comparison with younger and older people is unfeasible, so a seamless analysis over the life course is impossible in this study. Finally, although cohort analysis using this dataset is possible, at 6 years, the period is relatively short, which may hamper the ability to uncover the latent relationships and underlying mechanisms between the parameters used and the predicted outcomes. Durations of 10 years or even several decades may be needed to clarify the latent causality between suspected factors and outcomes [43].

In our composite study, we expect to obtain a wide range of novel findings and to confirm indeterminate previous findings, with multidisciplinary applications, especially in terms of cardiometabolic disease. We also expect this work to provide new perspectives for human health promotion and disease prevention.

Authors' contributions

All authors contributed to the study design, the interpretation of the initial analysis, or the discussion of the literature and expected results. KN, TI, MS, RH, and MA conducted the data analysis. KN prepared the first draft of the manuscript, and all authors read and edited the manuscript.

Funding statement

This work was partly supported by a special Grant of Kanagawa University of Human Services, which was determined 12th, June, 2017 and Kanagawa Prefecture (No. 505744).

Competing interests statement

All authors have no potential conflicts of interest.

Acknowledgments

We thank Mr. Shoji Iwasaki and Mr. Takahiro Ozawa, civil servants of Kanagawa Prefecture, for technical advice on the management and security of our computer and analysis system. We thank Jennifer Barrett, PhD, from Edanz Group (www.edanzediting.com/ac) for editing a draft of this manuscript.

Figure legends

Figure 1. Location of Kanagawa Prefecture

Figure 2. Structure of the cross-sectional and cohort studies

They gray rectangles represent the each year's cross-sectional study. Cohort Study I consists of the cross-sectional studies of 2008 and 2014, and Cohort Study II includes all years from 2008 to 2014. The numbers highlighted in green represent the sample size of each dataset.

Figure 3. Checkup participation rates (%)

References

1. Ma RC, Chan JC. Type 2 diabetes in East Asians: similarities and differences with populations in Europe and the United States. Ann N Y Acad Sci. 2013;1281:64-91.
2. Lu Y, Hajifathalian K, Ezzati M, Woodward M, Rimm EB, Danaei G. Metabolic mediators of the effects of body-mass index, overweight, and obesity on coronary heart disease and stroke: a pooled analysis of 97 prospective cohorts with 1.8 million participants. Lancet. 2014; 383:970-83.
3. Arai H., Ouchi Y., Toba K. Japan as the front-runner of super-aged societies: perspectives from medicine and medical care in Japan. Geriatr Gerontol Int. 2015;15:673–687.
4. Suzuki T. Health status of older adults living in the community in Japan: Recent changes and significance in the super-aged society. Geriatr Gerontol Int. 2018 Feb 8.
5. Hayashi F, Takimoto H, Yoshita K, Yoshiike N. Perceived body size and desire for thinness of young Japanese women: a population-based survey. Br J Nutr. 2006;96:1154-62.
6. Sugawara A, Saito K, Sato M, Kodama S, Sone H. Thinness in Japanese young women. Epidemiology. 2009;20:464-5.
7. Ministry of Health, Labour and Welfare. National Nutrition Survey 2013 [cited 2015 September 10]. Available from: http://www.mhlw.go.jp/file/04-Houdouhappyou-10904750-Kenkoukyoku-Gantaisakukenkouzoushinka/0000068070.pdf.
8. Fried LP, Tangen CM, Walston J, Newman AB, Hirsch C, Gottdiener J, Seeman T, Tracy R, Kop WJ, Burke G, McBurnie MA; Cardiovascular Health Study Collaborative Research Group. Frailty in older adults: evidence for a phenotype. J Gerontol A Biol Sci Med Sci. 2001 Mar;56(3):M146-56.
9. Kido Y. The Issue of Nutrition in an Aging Society. J Nutr Sci Vitaminol (Tokyo). 2015;61 Suppl:S176-7.
10. Arai H. Sarcopenia, frailty and osteoporosis. Clin Calcium. 2017;27(9):1279-1285.
11. Wouterse B, Huisman M, Meijboom BR, Deeg DJ, Polder JJ. The effect of trends in health and

longevity on health services use by older adults. BMC Health Serv Res. 2015 Dec 24;15:574.

12. de Magalhães JP, Stevens M, Thornton D. The Business of Anti-Aging Science. Trends Biotechnol. 2017 Nov;35(11):1062-1073.
13. Ministry of Health, "Labour, and Welfare: Health Ex- amination and Guidance Program for Japanese Adults," 2008. http://www.mhlw.go.jp/bunya/shakaihosho/iryouseido01/info02a.html
14. Kaplan RM, Chambers DA, Glasgow RE. Big data and large sample size: a cautionary note on the potential for bias. Clin Transl Sci. 2014 Aug;7(4):342-6.
15. Jagodnik KM, Koplev S, Jenkins SL, Ohno-Machado L, Paten B, Schurer SC, Dumontier M, Verborgh R, Bui A, Ping P, McKenna NJ, Madduri R, Pillai A, Ma'ayan A. Developing a framework for digital objects in the Big Data to Knowledge (BD2K) commons: Report from the Commons Framework Pilots workshop. J Biomed Inform. 2017 Jul;71:49-57.
16. Provision of medical-related data to a third party. http://www.mhlw.go.jp/english/topics/2011eq/workers/tepco/lhc/ref10_20140604.pdf
17. http://www.kuhs.ac.jp/kenkyuu/2017121400013/
18. Ethical Guidelines for Medical and Health Research Involving Human Subjects. http://www.mhlw.go.jp/file/06-Seisakujouhou-10600000 Daijinkanboukouseikagakuka/0000080278.pdf
19. The definition and diagnostic criteria of metabolic syndrome. Nippon Naika Gakkai Zasshi 200.5; 94:794-809
20. Grundy SM, Brewer HB Jr, Cleeman JI, Smith SC Jr, Lenfant C; American Heart Association; National Heart, Lung, and Blood Institute. Definition of metabolic syndrome: Report of the National Heart, Lung, and Blood Institute/American Heart Association conference on scientific issues related to definition. Circulation. 2004 Jan 27;109(3):433-8.
21. Rodríguez-Ortiz D, Reyes-Pérez A, León P, Sánchez H, Mosti M, Aguilar-Salinas CA, Velázquez-Fernández D, Herrera MF. Assessment of two different diagnostic guidelines criteria (National Cholesterol Education Adult Treatment Panel III [ATP III] and International Diabetes

Federation [IDF]) for the evaluation of metabolic syndrome remission in a longitudinal cohort of patients undergoing Roux-en-Y gastric bypass. Surgery. 2016 Apr;159(4):1121-8.

22. The final interim report by the work Group for studying the effects of the Specific Health Checkups and Specific Health Guidance on Health Care Expenditures http://www.mhlw.go.jp/file/05-Shingikai-12401000-Hokenkyoku-Soumuka/0000123428.pdf
23. Kashiwagi A, Kasuga M, Araki E, et al. International clinical harmonization of glycated hemoglobin in Japan: from Japan Diabetes Society to National Glycohemoglobin Standardization Program values. J Diabetes Investig 2012;3:39–40.
24. Matsuo S, Imai E, Horio M, Yasuda Y, Tomita K, Nitta K, Yamagata K, Tomino Y, Yokoyama H, Hishida A. Collaborators developing the Japanese equation for estimated GFR. Revised equations for estimated GFR from serum creatinine in Japan. Am J Kidney Dis 2009;53:982–92
25. Keith NM, Wagener HP, Barker NW. Some different types of essential hypertension: their course and prognosis. Am J Med Sci. 1974 Dec;268(6):336-45.
26. Sairenchi T, Iso H, Yamagishi K, Irie F, Okubo Y, Gunji J, Muto T, Ota H. Mild retinopathy is a risk factor for cardiovascular mortality in Japanese with and without hypertension: the Ibaraki Prefectural Health Study. Circulation. 2011 Dec 6;124(23):2502-11.
27. Health risks - Daily smokers - OECD Data https://data.oecd.org/healthrisk/daily-smokers.htm
28. Jamal A, Agaku IT, O'Connor E, King BA, Kenemer JB, Neff L. Current cigarette smoking among adults--United States, 2005-2013. MMWR Morb Mortal Wkly Rep. 2014 Nov 28;63(47):1108-12.
29. Falkner B, Cossrow ND. Prevalence of metabolic syndrome and obesity-associated hypertension in the racial ethnic minorities of the United States. Curr Hypertens Rep. 2014 Jul;16(7):449.
30. Engin A. The Definition and Prevalence of Obesity and Metabolic Syndrome. Adv Exp Med Biol. 2017;960:1-17.
31. Goto M, Goto A, Ikeda N, Noda H, Shibuya K, Noda M. Factors associated with untreated diabetes: analysis of data from 20,496 participants in the Japanese National Health and Nutrition Survey. PLoS One. 2015 Mar 10;10(3):e0118749.

32. Kutsuma A, Nakajima K, Suwa K. Potential Association between Breakfast Skipping and Concomitant Late-Night-Dinner Eating with Metabolic Syndrome and Proteinuria in the Japanese Population. Scientifica (Cairo). 2014;2014:253581.

33. Nakajima K, Suwa K, Oda E. Atrial fibrillation may be prevalent in individuals who report late-night dinner eating and concomitant breakfast skipping, a complex abnormal eating behavior around sleep. Int J Cardiol. 2014 Dec 20;177(3):1124-6.

34. Ul-Haq Z, Mackay DF, Fenwick E, Pell JP. Meta-analysis of the association between body mass index and health-related quality of life among adults, assessed by the SF-36. Obesity (Silver Spring). 2013 Mar;21(3):E322-7.

35. Corbin LJ, Timpson NJ. Body mass index: Has epidemiology started to break down causal contributions to health and disease? Obesity (Silver Spring). 2016 Aug;24(8):1630-8.

36. Kotler DP. Cachexia. Ann Intern Med. 2000 Oct 17;133(8):622-34.

37. Poirier P, Cornier MA, Mazzone T, Stiles S, Cummings S, Klein S, McCullough PA, Ren Fielding C, Franklin BA; American Heart Association Obesity Committee of the Council on Nutrition, Physical Activity, and Metabolism. Bariatric surgery and cardiovascular risk factors: a scientific statement from the American Heart Association. Circulation. 2011 Apr 19;123(15):1683-701.

38. Muneyuki T, Suwa K, Oshida H, et al. Design of the saitama cardiometabolic disease and organ impairment study (SCDOIS): a multidisciplinary observational epidemiological study. Open J Endocr Metab Dis 2013;2:144–56.

39. Nakajima K, Suwa K. Excess body weight affects HbA1c progression irrespective of baseline HbA1c levels in Japanese individuals: a longitudinal retrospective study. Endocr Res. 2015;40(2):63-9.

40. Abdullah A. The Double Burden of Undernutrition and Overnutrition in Developing Countries: an Update. Curr Obes Rep. 2015 Sep;4(3):337-49.

41. World Health Organization. The double burden of malnutrition :Policy brief. Geneva: World Health Organization, 2017.

42. Demaio AR, Branca F. Decade of action on nutrition: our window to act on the double burden of malnutrition. BMJ Glob Health. 2017 Nov 25;3(Suppl 1):e000492.

43. Bell JA, Hamer M, Sabia S, Singh-Manoux A, Batty GD, Kivimaki M. The natural course of healthy obesity over 20 years. J Am Coll Cardiol. 2015 Jan 6;65(1):101-102.

Table 1. Clinical characteristics and methods for measurements

Parameters		Data of 2008 (total n = 1209,118)		Methods and remarks
Substituted age (years old) [a]		55.0 ± 9.8 in total 53.9 ± 9.5 for men 56.6 ± 10.1 for women		Actual age is unknown but classified into every five years
Sex, n (%)		Men 695,055 (57.5) Women 514,063 (42.5)		
Anthropometric parameters				
		Means ± SD	Ratios of methods (%)	
Body mass index (kg/m^2)		23.0 ± 3.3 in total 23.7 ± 3.1 for men 22.0 ± 3.3 for women	–	Weight (in kg) divided by height (in m)^2
Waist circumference (cm)	Men	84.5 ± 8.4	–	Objectively measured
	Women	79.3 ± 9.6	–	
MetS (%)		MetS	13.4	The diagnosis is determined by Japanese criteria [19].
		Pre-MetS	13.1	
		Non-MetS	72.2	Including non-Pre-MetS
		Unknown	1.4	Due to incomplete data
Systolic blood pressure (mmHg)		125 ± 17	68.9	Once measurement
		125 ± 18	24.2	First time measurement among twice measurements
		131 ± 19	6.9	Second time measurement among twice measurements
Diastolic blood pressure (mmHg)		77 ± 11	67.5	Once measurement
		77 ± 12	24.2	First time measurement among twice measurements
		81 ± 13	8.2	Second time measurement among twice measurements
Serum parameters				

Aspartate aminotransferase (U/L)	24 ± 12	84.9	Ultraviolet spectrophotometric determination (JSCC reference method)
	23 ± 11	15.1	Others
Alanine aminotransferase (U/L)	23 ± 17	84.9	Ultraviolet spectrophotometric determination (JSCC reference method)
	24 ± 18	15.1	Others
Gamma-glutamyl transferase (U/L)	36 ± 31	82.6	Ultraviolet spectrophotometric determination (JSCC reference method)
	36 ± 31	17.4	Others
Triglyceride (mg/dl)	96 (68-141)	81.5	Ultraviolet and visible spectrophotometric determination (enzyme colorimetric/glycerol elimination methods)
	97 (68-146)	3.2	Ultraviolet spectrophotometric determination (enzyme colorimetric/ glycerol elimination method)
	96 (66-145)	15.3	Other
High-density lipoprotein cholesterol (mg/dl)	64 ± 17	80.6	Ultraviolet and visible spectrophotometric determination (direct methods (non-precipitation method))
	61 ± 16	3.7	Ultraviolet spectrophotometric determination (direct methods (non-precipitation method))
	63 ± 17	15.7	Other
Low-density lipoprotein cholesterol (mg/dl)	127 ± 31	80.8	Ultraviolet and visible spectrophotometric determination (direct methods (non-precipitation method))
	125 ± 31	3.4	Ultraviolet spectrophotometric determination (direct methods (non-precipitation method))
	125 ± 31	15.9	Other
Uric acid (mg/dl) [b]	NA	NA	NA
Creatinine (mg/dl) [b]	NA	NA	NA
eGFR (ml/min/1.73m^2) [b]	NA	NA	NA
Fasting plasma glucose (mg/dl)	98 ± 19	29.6	Potentiometric determination
	98 ± 19	5.0	Ultraviolet and visible spectrophotometric determination
	97 ± 18	49.0	Ultraviolet spectrophotometric determination
	97± 18	16.4	Other
HbA1c (%, NGSP)	5.2 ± 0.7	64.9	Immunoassay (Latex Agglutination turbidimetric Immunoassay, etc.)
	5.3 ± 0.6	16.9	HPLC determination
	5.2 ± 0.6	2.0	Enzymatic determination
	5.2 ± 0.7	16.3	Other
Urine parameters			
Proteinuria (%)	3.8	56.9	Automated dipstick analysis
	3.9	43.1	Visual dipstick analysis
Glycosuria (%)	2.0	57.2	Automated dipstick analysis
	2.1	52.8	Visual dipstick analysis
Fundus oculi examination			

Keith-Wagener hypertensive retinopathy (%) (available n = 13,866)	Mild (I) 5.3 Moderate– papilledema (II–IV) 1.0	Fundoscopy
Scheie hypertensive and sclerotic retinopathy (%) (available n = 15,894)	Hypertensive (1–4) 5.4 Sclerosis (1–4) 6.2	Fundoscopy

[a] Age groups (40–44, 45–49, 50–54, 55–59, 60–64, 65–69, and 70–74 years) are replaced with 42, 47, 52, 57, 62, 67, and 72 years old, respectively, corresponding to the median for each category.
[b] Serum uric acid, creatinine, and consequently eGFR are currently unavailable but will become available in the future.
MetS: metabolic syndrome; eGFR: estimated glomerular filtration rate

Table 2. Questionnaire on health status and results for 2008

No.	Questionnaire	Answers	Positive response to ① (%)*
	Are you taking following medicines at present?		
1	Medications to reduce blood pressure	① Yes ② No	17.1
2	Insulin injection or medications to reduce blood glucose	① Yes ② No	3.5
3	Medications to reduce your cholesterol level	① Yes ② No	9.4
4	Have you ever been told by the doctor you have a stroke (cerebral hemorrhage, brain infarction, etc.) and received treatment for the disease?	① Yes ② No	2.0
5	Have you ever been told by the doctor you have a heart disease (angina pectoris, myocardial infarction, etc.) and received treatment for the disease?	① Yes ② No	3.9
6	Have you ever been diagnosed as having a chronic kidney failure and received treatment (dialysis therapy) for the disease?	① Yes ② No	0.3
7	Have you ever been diagnosed as anemic?	① Yes ② No	10.8
8	Are you a current regular smoker? (Here a current regular smoker is to be a person who has smoked a total of 100 or more cigarettes or smoked for 6 months or longer and has been smoking for the last one month).	① Yes ② No	25.9 in total 37.2 in men 10.8 in women
9	Have you gained over 10 kg from your weight at 20 years old?	① Yes ② No	35.5
10	Are you in a habit of doing exercise to sweat lightly for over 30 minutes per session, 2 times weekly, for over a year?	① Yes ② No	30.9
11	In your daily life do you walk or do any equivalent amount of physical activity more than one hour per day?	① Yes ② No	43.6
12	Is your walking speed faster than the speed of corresponds of your age and sex?	① Yes ② No	52.3
13	Have you gained or lost over 3 kg during the last year?	① Yes ② No	22.1
14	How fast do you eat compared to others?	① Faster ② Normal ③ Slower	29.6
15	Do you eat dinner two hours before bedtime more than 3 times per week?	① Yes ② No	28.5
16	Do you eat snacks after supper more than 3 times per week?	① Yes ② No	12.8
17	Do you skip breakfast more than 3 times per week?	① Yes ② No	14.6
18	How often do you drink alcohol (sake, distilled spirits, beer, liquor, etc.)?	① Everyday ② Occasional ③ Hardly drink （cannot drink）	29.2
19	How much do you drink a day, in terms of glasses of refined sake? (A glass	① Less than 180ml	56.8

	[180mL] of refined sake is equivalent to a medium bottle [500mL] of beer, 80mL of shochu (alcohol content 35 percent), a glass [double, 60mL] of whiskey, and 2 glasses [240mL] of wine.)	② Over 180–less than 360ml ③ Over 360–less than 540ml ④ Over 540ml	
20	Do you feel refreshed after a night's sleep?	① Yes ② No	65.4
21	Do you want to improve your life habits of eating and exercising?	① I don't mean to start. ② I'm going to start in the future (e.g., within 6 months). ③ I'm going to start soon (e.g., in a month), or I have just started some of them. ② I already started (<6 months ago). ③ I already started (>=6 months ago).	28.5
22 [b]	Are you willing to undergo a health counseling regarding lifestyle modifications if you get the opportunity?	① Yes ② No	–

[a] Proportions are calculated based on the available numbers for each question.
[b] Question 22 is currently unavailable but will become available in the future.

Table 3. Major and minor outcomes

Major conditions or etiologies
· Cardiometabolic diseases including type 2 diabetes, hypertension, dyslipidemia, metabolic syndrome, chronic kidney disease assessed by proteinuria and eGFR
· Obesity and low body weight (malnutrition) assessed by BMI, and central obesity assessed by WC
· Hepatic diseases including fatty liver disease assessed by serum AST, ALT, and GGT
· Abnormal eating habits (breakfast skipping, late-night dinner eating, eating fast, night eating)
· Unhealthy lifestyles (smoking, infrequent exercise, heavy alcohol drinking, non-restorative sleep)
Minor conditions or etiologies
· Hypoglycemia, hyperfiltrations (high eGFR), hypotension, low uric acid
· Extremely low and high BMI (e.g., <15.0 kg/m^2 and > 40.0 kg/m^2)
· Osteoporosis assessed as reduced body-height during 6-years
· Hypertensive and atherosclerotic retinopathies (Keith-Wagener, Scheie classification)
· Physically inactive conditions including reduced walking speed and infrequent exercise

ALT: alanine aminotransferase; AST: aspartate aminotransferase; BMI: body mass index; eGFR: estimated glomerular filtration rate; GGT: gamma-glutamyl transferase; WC: waist circumference

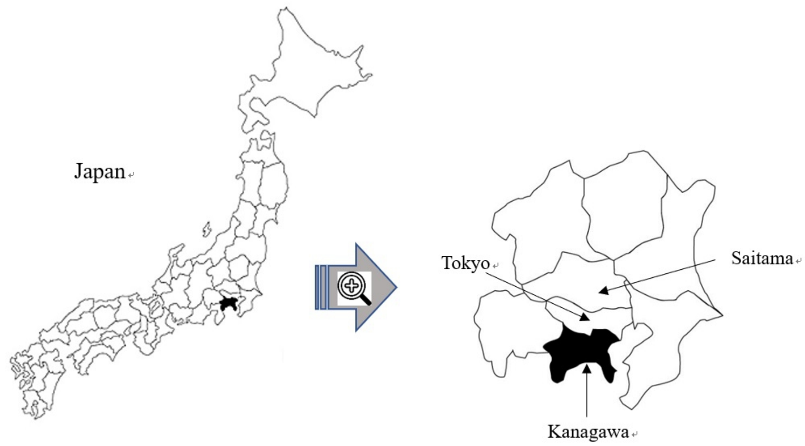

Figure 1

217x141mm (144 x 144 DPI)

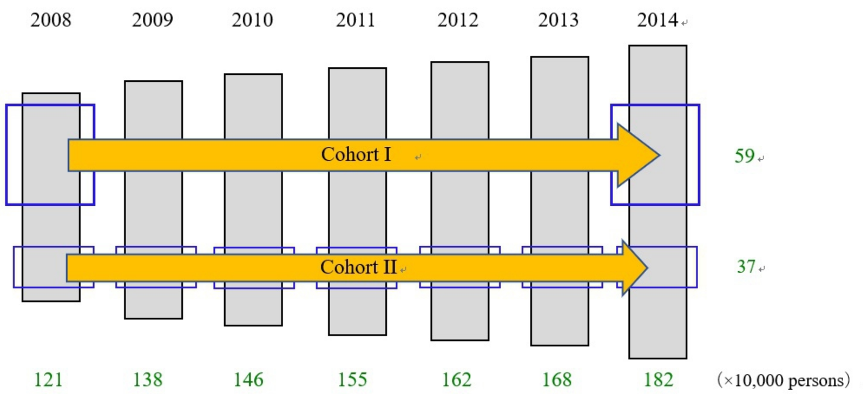

Figure 2

203x114mm (144 x 144 DPI)

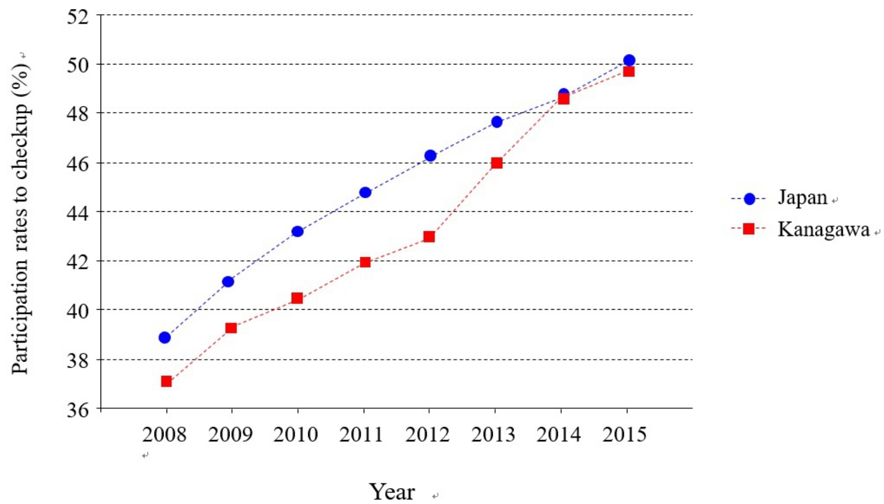

Figure 3

203x122mm (144 x 144 DPI)

The Kanagawa Investigation of the Total Checkup Data from the National Database (KITCHEN): Protocol for data-driven population-based repeated cross-sectional and 6-year cohort studies

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2018-023323.R1
Article Type:	Protocol
Date Submitted by the Author:	23-Oct-2018
Complete List of Authors:	Nakajima, Kei; Kanagawa University of Human Services, School of Nutrition and Dietetics, Faculty of Health and Social Services; Saitama Medical Center, Saitama Medical University, Department of Endocrinology and Diabetes Iwane, Taizo; Kanagawa University of Human Services, School of Nutrition and Dietetics, Faculty of Health and Social Services Higuchi, Ryoko; Kanagawa University of Human Services Shibata, Michi; Kanagawa University of Human Services; St. Marianna University School of Medicine Takada, Kento; Kanagawa University of Human Services Uda, Jun; Jikei Institute, Graduate School of Health Care Sciences Anan, Mami; Kanagawa University of Human Services Sugiyama, Michiko; Kanagawa University of Human Services Nakamura, Teiji; Kanagawa University of Human Services
Primary Subject Heading:	Epidemiology
Secondary Subject Heading:	Public health
Keywords:	EPIDEMIOLOGY, GENERAL MEDICINE (see Internal Medicine), GERIATRIC MEDICINE, NUTRITION & DIETETICS, PUBLIC HEALTH, DIABETES & ENDOCRINOLOGY

SCHOLARONE™ Manuscripts

The Kanagawa Investigation of the Total Checkup Data from the National Database (KITCHEN): Protocol for data-driven population-based repeated cross-sectional and 6-year cohort studies

Kei Nakajima,[1,2]* Taizo Iwane,[1] Ryoko Higuchi,[1] Michi Shibata,[1,3] Kento Takada,[1] Jun Uda,[4] Mami Anan,[1] Michiko Sugiyama,[1] Teiji Nakamura[1]

[1]School of Nutrition and Dietetics, Faculty of Health and Social Services, Kanagawa University of Human Services, 1-10-1 Heisei-cho, Yokosuka, Kanagawa 238-8522, Japan

[2]Department of Endocrinology and Diabetes, Saitama Medical Center, Saitama Medical University, 1981 Kamoda, Kawagoe, Saitama 350-8550, Japan

[3]Department of nutrition, St. Marianna University School of Medicine, 2-16-1 Sugao, Miyamae-ku, Kawasaki, Kanagawa 216-8511, Japan

[4]Graduate School of Health Care Sciences, Jikei Institute, 1-2-8 Miyahara, Yodogawa-ku, Osaka-shi, 532-0003, Japan

Running title: Kanagawa Investigation of Checkup Data

Address correspondence to Kei Nakajima, M.D., Ph.D.

School of Nutrition and Dietetics, Faculty of Health and Social Services, Kanagawa University of Human Services, 1-10-1 Heisei-cho, Yokosuka, Kanagawa 238-8522, Japan.

e-mail: nakajima-rsh@kuhs.ac.jp; Tel: +81-046-828-2660

Abstract

Introduction: The unmitigated incidence of cardiometabolic diseases, such as type 2 diabetes and metabolic syndrome, has gained attention in Japan. “Big data” can be useful to clarify conflicting observations obtained from studies with small samples and about rare conditions that are often neglected. We epidemiologically address these issues using data from health checkups conducted in Kanagawa Prefecture, the prefecture with the second largest population in Japan, in the Kanagawa Investigation of the Total Checkup Data from the National Database (KITCHEN).

Methods and analysis: This research consists of a series of population-based cross-sectional studies repeated from 2008 to 2014 and 6-year cohort studies. Since 2017, we have reviewed the data of people living in Kanagawa Prefecture who underwent a health checkup mainly for general health and the prevention of metabolic syndrome. The sample size ranges from 1.2 million to 1.8 million people in the cross-sectional studies and from 370,000 to 590,000 people in the cohort studies. These are people aged 40 to 74 years, whose clinical parameters were measured and who responded individually to a questionnaire. We investigate potential associations and causalities of various etiologies, including diabetes and metabolic syndrome, using clinical data and lifestyle information. With multidisciplinary analysis, including data-driven cyclopedic analysis, we expect to obtain a wide range of novel findings, to confirm indeterminate previous findings, especially in terms of cardiometabolic disease, and to provide new perspectives for human health promotion and disease prevention.

Ethics and dissemination:

Ethical approval was received from the Ethics Committee of Kanagawa University of Human Services (10-43). The protocol was approved in December 2016 by the Japanese Ministry of Health, Labour and Welfare (No. 121). The study results will be disseminated through open platforms including journal articles, relevant conferences, and seminar presentations.

Keywords: Kanagawa; checkup; big data; data mining; cardiometabolic disease; national database; health records; lifestyle-related disease; age-related disease

Strengths and limitations of this study

•The number of subjects in the sample is so large that more precise results concerning the means of parameters can be obtained, even when the subjects are classified into multiple categories, including sex, age group, smoking status, and certain morbidities such as obesity or diabetes.

•It may be possible to use big data to evaluate minor or rare conditions or etiologies that are commonly overlooked, neglected, or unfeasible to analyze in clinical studies, particularly those with small samples.

•It may also be possible to conduct data-driven cyclopedic and hypothesis-generating studies and then detect latent relationships among measures available in the data.

•Identical measurements and assessments of anthropometric indices, blood pressure, blood biochemistry, and urinalysis are performed across multiple years in the prefecture with the second largest population in Japan.

•The variations in parameters are restricted, and parameters for specific diseases are not included, because the checkups are conducted for general health and the prevention of common diseases, especially lifestyle-related diseases such type 2 diabetes and metabolic syndrome.

•Although cohort analysis is possible with this dataset, at 6 years, the period is relatively short, which may hamper the ability to uncover the latent relationships and underlying mechanisms between the parameters used and the predicted outcomes.

Introduction

Over the past several decades, the incidence of cardiometabolic diseases such as type 2 diabetes and metabolic syndrome (MetS) has not been reduced and has gained attention in Asia, including Japan [1,2], which has also experienced an unprecedented acceleration of societal aging [3,4]. These issues may also be problematic in Kanagawa Prefecture (***Figure 1***), an eastern district of Japan, located near Tokyo. The primary causes of these diseases include unfavorable lifestyles (e.g., smoking, heavy alcohol consumption, insufficient sleep, and infrequent exercise) and excess body weight (overweight and obesity) because of overeating, along with individuals' genetic and epigenetic backgrounds. However, for the last decade, malnutrition (e.g., low body weight) has been shown to be prevalent among young women [5-7] and the elderly [8-10] in Japan, which may contribute to the increased rates of sarcopenia and frailty in the country. Combined with prolonged longevity, cardiometabolic diseases with age-related causes create a long-term burden that leads to direct (i.e., measurements and therapies) and indirect (e.g., nursing, care, and welfare) medical costs nationwide [11,12], particularly when severe complications such as organ failure (e.g., heart, liver, and renal failure) develops over the life course.

In 2008, a special health checkup was initiated, primarily for the prevention of MetS, by the Ministry of Health, Labour and Welfare (MHLW) in Japan [13]. Since that time, all people living in Japan aged 40 to 74 years are supposed to undergo a yearly health checkup. The data from these checkups have continuously accumulated, creating a very large database. Such "big data" are likely to be useful in clarifying indistinct or conflicting results obtained from clinical studies with small sample sizes [14,15], confirming established results and advancing them by elucidating plausible mechanisms and clinical relevance, and enabling a precise understanding of the current status of public health and the contributing to it. Additionally, "big data" of this kind enable us to investigate minor or rare conditions and etiologies [14], such as extremely low and high body weight, abnormal (low and high) clinical measurements, and the low or high prevalence of unfavorable habits and lifestyles, where etiologies are hardly understood, especially when such extreme conditions are combined in complicated ways, primarily because of inadequate numbers of observations and corresponding cases. Extreme conditions can feasibly be reproduced in animal or cellular studies by means of intentional manipulation of conditions including through the use of transgenic and knockout technologies. These non-human laboratory studies can provide profound insight into the etiology of human diseases [16,17]. Clearly, such extreme conditions are mostly unfeasible in studies involving humans. However, in an epidemiological study with a database equivalent to "big data," it might be feasible to reproduce such extreme conditions in certain categories for limited conditions.

Although cardiometabolic diseases such as type 2 diabetes and chronic organ failures such as chronic kidney disease have been increasing along with the prolonged longevity in Japan [3,4], the underlying associations with clinical parameters and their mechanisms have not been fully elucidated or confirmed, particularly in epidemiological studies using the "big data" from the health checkups described above. These data include more than one million observations per year in most prefectures in Japan. To date, no investigation of this type has been performed, especially on the prefecture scale in Japan.

To this end, we investigated current cardiometabolic disease and health status as clearly as possible, as well as the relationship of cardiometabolic diseases, including but not limited to type 2 diabetes and MetS, and age-related etiologies. We focused especially on the thorough, end-to-end analysis of the variables of interest, using digitally recorded accumulated data in an extremely large epidemiological study of Kanagawa Prefecture, the second most populated prefecture in Japan, with approximately 9 million inhabitants, second only to Tokyo (approximately 13.7 million inhabitants), as of October 2017. Taking this approach, our study may be characterized as a data-driven, cyclopedic and hypothesis-generating study, with the nature of "big data" research, rather than a hypothesis-testing, traditional epidemiological study [18,19]. Consequently, the concrete objectives and contents of individual studies are difficult to determine before it becomes clear what kinds and amounts of data are available, which will substantially influence the design and analysis methods of each study. Although big data is often analyzed with various algorithms including machine learning [18], in this study, we analyzed the data using traditional the epidemiological methods described in the next section, rather than machine learning.

Methods and analysis

Design

In 2013, the MHLW began to offer accumulated data consisting of information on patient prescriptions and health checkups for use by Japanese institutions including universities, hospitals, and research centers. These data are recorded digitally and are provided in a third-party manner, according to the concept of the "provision of medical-related data to a third party" to improve the quality of medical services and to support academic research in Japan [20]. To date, 178 applications from various institutions in Japan have been accepted in this manner (as of March 30, 2018).

Our project was a composite multidisciplinary study aimed at elucidating the factors associated with cardiometabolic diseases and eventually contributing to the amelioration and advancement of social health and welfare. After the study protocol was approved by the ethics committee of Kanagawa University of Human Services (10-43), we applied to the MHLW's data provision system in October

2016, through Teiji Nakamura, the President of Kanagawa University of Human Services, as a representative. The protocol of our study was approved in December 2016 by the MHLW (No. 121), after a peer review by an expert council.

Before we received the database from the MHLW, identifying individual-level information (names and postal codes) was completely transformed into randomized non-distinguishing anonymous numbers and characters, which prevents the restoration of this information by any means. There are two types of unique identifying variables available for each subject in the cross-sectional database collected from 2008 to 2014: ID 1 is determined based on the subject's insurance number, sex, and birth date, and ID 2 is determined by the subject's name, sex, and birthday. Both variables consist of anonymous numbers and characters created by the MHLW using a hash function [21]. For individual subjects, these variables are unchanged in principle, except when there are changes in the variables' constituting parts.

To further protect against the identification of specific individuals, age was categorized into 5-year age groups (40–44, 45–49, 50–54, 55–59, 60–64, 65–69, and 70–74 years), so that the individual's precise age at the time of data collection is unknown in our study.

Our study is part of the MHLW's nationwide program of providing medical-related data to third parties [20], and informed consent for the use of these data has not been obtained from each subject. We have opened the protocol of our study to the public on our university homepage, which was updated in October 2017 [22], in line with the "Ethical Guidelines for Medical and Health Research Involving Human Subjects" [23] in Japan (updated by the MHLW and the Ministry of Education, Culture, Sports, Science and Technology in May 2017). We received the digitally recorded non-distinguishing anonymous data from the MHLW in August 2017.

Our analysis of the data was conducted in a location with restricted access and tight security regarding datasets at Kanagawa University of Human Services. Repeated cross-sectional studies will be conducted using checkup data from 2008 to 2014. Additionally, a historical cohort study will be conducted, using the 2008 data as a baseline and the 2014 data to assess final outcomes (***Figure 2***). During this period, the number of subjects undergoing a health checkup has increased each year in Kanagawa in parallel with the nationwide trend. Nationally, almost 50% of the population attended a health checkup in 2014, probably because of the political encouragement for these checkups (***Figure 3***), although the MHLW's overall expected target rate is 70% [24].

People who did not undergo a checkup might have been under treatment for moderate to severe disease or hospitalized at the relevant time points. Health-minded people in Japan were likely to voluntarily undergo an expensive health checkup called the "*Ningen Dock*" (detailed and comprehensive health checkup). Other people who did not undergo a checkup might have missed the opportunity to have a checkup because of business obligations or other reasons, including family reasons or moving.

The name of the full study is the Kanagawa Investigation of the Total Checkup data from the National Database (KITCHEN). Each subsequent KITCHEN publication will be numbered sequentially from 1.

Subjects and measurements

People aged 40 to 74 years and living in Kanagawa Prefecture at the time of the data collection were enrolled in a series of studies. Those residing in medical institutions including hospitals and nursing homes were not included. All subjects are thought to be active to the extent of coming to the place where the checkup was performed. However, some of the subjects have diseases such as hypertension, diabetes, or dyslipidemia, and some have a history of morbidities such as heart disease or stroke. All of these conditions are digitally recorded as answers to a questionnaire. Specific exclusion and inclusion criteria will be determined for each study in the future. The sample sizes range from 1.2 million to 1.8 million people in the cross-sectional studies and from 370,000 to 590,000 people in the cohort studies (***Figure 2***). Cohort Study I uses data from 590,000 people who attended checkups in 2008 and 2014. Cohort Study II is based on data from 370,000 people who attended a checkup every year from 2008 to 2014. The two types of identifying variables (ID 1 and ID 2) described above are used to link data on individual subjects throughout the cohort study. When both of these variables simultaneously changed for a subject from 2008 to 2014, it was not possible to follow these individuals through time, resulting in the exclusion of these subjects from the cohort study. To date, such an event has been reported to occur at a rate of approximately 0.8% per year [21].

Patient and Public Involvement:

Patients are not involved in this study.

All of the parameters measured in this study are listed in ***Table 1***. To evaluate subject's age as a numeric value, we transformed age group (40–44, 45–49, 50–54, 55–59, 60–64, 65–69, and 70–74 years) into substituted age (s-age), corresponding to the median for each age group (42, 47, 52, 57, 62, 67, and 72 years). Body weight and height were objectively measured by trained institutional staff members and were recorded to one decimal place (kg and cm). Body mass index (BMI) was calculated as weight (in kg) divided by height (in m^2). In most cases (approximately 99.9%), waist circumference (WC) was measured objectively at the navel level by a medical staff member and recorded to one decimal place. Biochemical measurements were performed using standard methods and automated machines. Dipstick urine analysis for proteinuria and glycosuria was assessed visually or with ordinary automated machines. Several different methods were used for the included biochemical parameter (***Table 1***). The measurement of blood pressure and blood/urine biomarkers was regularly standardized using both internal standards with available traceability and external standards by third parties, including the Japanese Association of Medical Technologists, even when the measurements were outsourced [25].

In principal, most people underwent a checkup after overnight fasting. However, some of the checkups were conducted in a non-fasting condition because of, for example, shift work or family reasons. Therefore, all subjects were asked for the time (in hours) from their last meal to the time of the checkup, which was recorded as at least 10 hours or less than 10 hours. Those completing the checkup less than 10 hours after their last meal will be distinguished from others in certain sub-studies—for instance, when examining diabetes or dyslipidemia.

The Japanese diagnostic criteria for MetS were published in 2005 [26]. Unlike other criteria such as that of the Adult Treatment Panel III (ATP-III) and the International Diabetes Federation (IDF) [27,28], the Japanese MetS criteria include abdominal obesity as an essential condition (WC ≥ 85 cm for men and ≥ 90 cm for women), in addition to two or more of the following three components: 1) dyslipidemia (triglycerides ≥ 150 mg/dl and/or high-density lipoprotein cholesterol < 40 mg/dl, or pharmacotherapy for dyslipidemia); 2) hypertension (systolic blood pressure ≥ 130 mmHg and/or diastolic blood pressure ≥ 85 mmHg, or pharmacotherapy for hypertension); and 3) hyperglycemia (fasting plasma glucose (FPG) ≥ 110 mg/dl or pharmacotherapy for diabetes). In the practice of the health checkups, hyperglycemia is defined as having elevated FPG (≥ 110 mg/dl) and/or HbA1c (National Glycohemoglobin Standardization Program [NGSP]) ≥ 6.0%, or pharmacotherapy for diabetes. Furthermore, pre-MetS is defined as abdominal obesity plus one of the three components listed above [24]. In sub-studies concerning MetS, MetS will also be determined using other international criteria, such as that of the ATP-III or the IDF to allow for comparison with the same criteria with other Asian countries as well as with Western countries. HbA1c (Japan Diabetes Society [JDS]) was converted to

HbA1c (NGSP) units using the officially certified formula: HbA1c (NGSP) (%) = 1.02 × JDS (%) + 0.25% [29]. In 2008, almost all subjects had either FPG or HbA1c measured (99.7%).

If the data on serum creatinine eventually become available, the estimated glomerular filtration rate (eGFR) will be calculated using the s-age above and the following equation [30]: eGFR (ml/min/1.73 m^2) = 194 × serum $Cr^{-1.094}$ × s-$age^{-0.287}$ (if female) × 0.739, where Cr denotes serum creatinine concentration (mg/dl).

Hypertensive retinopathy has been shown to be associated with cardiovascular events and mortality [31,32]. Hypertensive retinopathy assessments using the Keith–Wagener and Scheie classifications are available in the study, although a very small percentage of individuals (around 1.3%) completed the hypertensive retinopathy examination.

The standardized 22-item questionnaire created by the MHLW for the health examination checkups is shown in ***Table 2***.

Primary and secondary (minor) outcomes

In a series of studies, we will consider various conditions and etiologies as primary and secondary (or minor) outcomes (see ***Table 3***). However, because unexpected findings are likely to be obtained during these studies and related research topics will be pursued following these findings, we do not restrict the areas of research to be pursued as long as the findings can contribute to or advance specific or general health objectives.

It is noteworthy that subjects are made aware of their health status when their checkup results are complete, and they often receive advice and suggestions from health professionals. Therefore, our cohort studies are not natural history cohort studies by nature. Some proportion of subjects undergo treatments in hospitals, and some receive further health guidance because of the results of their checkups. Specifically, health guidance for eligible subjects (***Table 1***) aims for the prevention or improvement of mainly cardiometabolic diseases, including MetS. In Japan, medical insurers are required to recommend that individuals at risk of these conditions receive health guidance, although this is not obligatory for the individuals. Health guidance is classified into two categories (intensive and motivational health guidance), depending on the individual's abdominal obesity (waist circumference ≥ 85 cm for men or ≥ 90 cm for women) and number of risk factors (***Table 1***). In brief, in cases of intensive health guidance, subjects receive consultation via e-mail, phone, or face-to-face sessions for up to 6 months, as has been described in detail elsewhere [24, 33], whereas subjects receiving motivational health guidance do not receive continuous support. Notably, subjects undergoing pharmacotherapy for hypertension, diabetes, or dyslipidemia are excluded, and those aged 65–74 years receive motivational health guidance regardless of their risk profile. In Japan, attendance rates for health guidance have been found to be less than 20% [24,34].

Statistical analysis

Continuous and categorical variables will be compared between different groups using analysis of variance (ANOVA) and χ^2-tests, respectively. Post hoc comparisons between two specific groups will be examined with the Bonferroni, Tukey–Kramer, and Dunnett methods, as well as additional χ^2-tests. Paired or trend data using the χ^2-test will rely on the McNemar and Cochran–Armitage tests, respectively. Analysis of covariance (ANCOVA) with general linear model procedures will be used to examine the difference in biochemical variables measured by two or three different methods (e.g., LDL cholesterol is measured using three different methods) (***Table 1***), controlling for confounders including age, sex, body weight, and various lifestyles. Logistic regression and hazard models will be used to examine the associations or causalities between abnormal levels of measurements and conditions with major and minor outcomes. These methods will yield odds ratios, risk ratios, or hazard ratios, which will be presented along with their 95% confidence intervals. Panel data analysis (including the Hausman test) combining several sets of cross-sectional data will also be conducted. Relevant confounding factors include age, sex, smoking, and alcohol consumption, which will be adjusted in the regression analyses. Alternatively, to evaluate or control the differences in backgrounds and various confounders between cases and controls, individuals' propensity scores will be calculated as a variable that unifies all corresponding confounders in the analysis. Propensity score is also considered for a special examination, for instance, hypertensive retinopathy examination, because few subjects underwent such special examination, which yields a bias to be adjusted. The level of health guidance (***Table 1***) and the answer to Question 21, which asks about personal intentions to improve eating and exercise habits (***Table 2***), will also be considered as confounding factors, when appropriate.

There are missing data in our study, although this comprises less than 20% of the cases for most parameters and questionnaire items. For categorized age, sex, BMI, and WC (***Table 1***), the data are almost complete, even when all of these variables are combined (99.99%). However, combining parameters other than age, sex, BMI, and WC can decrease the total available number of subjects, depending on the study's nature and design.

When analyzing extremely rare conditions, which might lead to the disclose of the identity of individuals with rare diseases needing treatments in hospital, to prevent the identification of these subjects, we do not describe the number anywhere in the manuscript if this is less than 10, as advised in the MHLW's guidelines [35].

Statistical analyses will be performed using SAS-Enterprise Guide (SAS-EG 7.1) in the SAS system, Version 9.4 (SAS Institute, Cary, North Carolina, USA). Values of $p < 0.05$ will be considered statistically significant. Because we understand that large data are predisposed to detect the presence of statistical significance, we will use caution in our interpretations and give priority to clinical significance rather than statistical significance in certain clinical areas.

Overall characteristics of subjects

Subjects' characteristics at baseline (2008) are shown in ***Table 1***. These findings will vary to some extent in sub-studies using other cross-sectional data from 2009 to 2014. Men in the first five years of their 50s are overrepresented in the sample, probably because middle-aged men are more likely than women to work for companies and institutes (i.e., insurers), which obligate workers to undergo a checkup. Considering the clinical parameters of BMI, WC, blood pressure, lipids, FPG, HbA1c, and MetS, most of the subjects are apparently healthy people with these parameters within normal ranges. The questionnaire results (2008) are shown in ***Table 2***, which gives us rough information about the subjects' backgrounds. The smoking rate is higher (25.9%), especially among men (37.2%), compared with other developed countries such as the United States [36,37], although the smoking rate has been declining in Japan in recent years [36].

The prevalence of MetS (13.4%), as well as pharmacotherapy for hypertension, diabetes, and dyslipidemia are relatively lower (3.5%–17.1%), compared with other countries [38,39]. However, this does not always mean that the number needed to treat is low, because substantial proportions of subjects likely do not consult a doctor about their poor glycemic control. A Japanese national survey conducted from 2005 to 2009 found that, among people with diabetes, a substantial proportion (about 38%) have left their poor glycemic conditions as they are, without seeking treatment [40]. In our study, the extent of this issue remains unknown without a detailed investigation of FPG and HbA1c.

Concerning eating habits, which play an important role in metabolism and nutrition, the percentage of subjects who habitually skip breakfast is lower (14.6%) than the percentage who eat dinner two hours before bedtime (28.5%), which is consistent with a previous study [41]. This suggests that the latter group may be more troublesome in terms of unfavorable lifestyle habits that are linked to cardiovascular diseases, because a close association between eating dinner late at night and skipping breakfast has been reported in a community-based epidemiological study [41]; eating dinner late at night can lead to skipping breakfast the next morning. Acknowledging the relationship with sleep, we term these behaviors "unfavorable eating habits around sleep" (UEHAS). In previous studies [41,42], eating dinner late at night together with skipping breakfast—a combination representative of UEHAS—was significantly associated with MetS, proteinuria, and atrial fibrillation.

Body weight substantially influences the incidence and development of cardiometabolic diseases as well as general health [43,44]. However, features and etiologies at both extremes of BMI, a fundamental index of weight considering height, body adiposity, nutritional status, and health, are poorly understood. For rare conditions in malnutrition, for instance, the percentage of subjects with an extremely low body weight (BMI < 15.0 kg/m^2, a criterion for high mortality [45]) is very small in the cross-sectional dataset in 2008 (0.1%), but the observational number in this dataset is large (n = 1,217, data not shown), which is not ignorable and may be enough to conduct proper statistical analyses. Likewise, extremely high body weight (BMI ≥ 40.0 kg/m^2, a criterion for class III obesity [46]), is also very small in percentage (0.07%), but there are 805 observations for this group in this dataset.

Throughout the study presented here, we expect to obtain a wide range of novel observations, enabling us to confirm indeterminate previous findings, especially in terms of cardiometabolic disease. In addition, this study will likely reveal underlying etiologies that have been overlooked because they are rare or minor cases in small clinical studies.

Among our research team members, one person (Nakajima) has previously been involved in a similar large study, consisting of approximately 100,000 people living in Saitama Prefecture (population = 7.3 million people), which is also located near Tokyo [47] (***Figure 1***). This previous study and its sub-studies were launched in 2011 and, to date, multiple findings have been reported from these studies. Although a study of 100,000 people is generally considered "large," this sample size sometimes proved to be inadequate for stratification analysis because of small observational numbers for particular groups when several stratification variables were combined—for instance, age group, BMI category, and diabetes status [48]. This is the main reason we chose to begin a new study using the checkup data for an extraordinarily large sample—over one million for a cross-sectional study, which is more than 10 times higher than this previous study [47].

BMI roughly reflects nutritional status, including excess energy accumulation or malnutrition. In the last decade, increased percentages of people have been found at both extremes of BMI (i.e., in the nutritional states of malnutrition or obesity), worldwide, especially among children. This coexistence of undernutrition and obesity or nutrition-related noncommunicable disease has been termed the "double burden of malnutrition" [49-51]. In our study, subjects' characteristics described above suggest that the double burden of malnutrition can exist even among the middle-aged Japanese population, although the proportions are smaller compared with other developing countries [49].

Our study has several strengths. First, in terms of community-based epidemiological research, the sample is so large that precise results concerning the means of parameters and "normal" values, although these are standard parameters, can be obtained [52], even when subjects are classified into categories such as sex, age group, smoking status, and certain morbidities (e.g., obesity or diabetes). Therefore, it may be possible to conduct similar analyses to produce novel results in other population studies with very large databases that allow for multiple classifications. A second strength of this study is that it may be possible to use big data to evaluate minor or rare conditions and etiologies that are commonly overlooked, neglected, or unfeasible to analyze in clinical studies, particularly those with small samples [14]. This analysis may contribute to case studies instead of only to the field of public health. Finally, identical measurements and assessments of anthropometric indices, blood pressure, and urinalysis are performed across multiple years in people living in similar environment and the same health care system.

Some limitations to this study should also be mentioned. First, the variations in parameters are restricted, and parameters for specific diseases are not included, because the checkups are conducted for general health and the prevention of common diseases, especially lifestyle-related diseases such type 2 diabetes and MetS. Second, people younger than 40 years and those aged over 74 years are not enrolled in this study. Lifestyle choices made when people are younger may contribute to the incidence of morbidities in middle age, and lifestyles and clinical biochemistry levels in middle age can influence the incidence and severity of cardiovascular diseases and health damage in the later life. Unfortunately, comparison with younger and older people is unfeasible, so a seamless analysis over the life course is impossible in this study. Third, although cohort analysis using this dataset is possible, at 6 years, the period is relatively short, which may hamper the ability to uncover the latent relationships and underlying mechanisms between the parameters used and the predicted outcomes. Durations of 10 years or even several decades may be needed to clarify the latent causality between suspected factors and outcomes [53]. Finally, to date, there is no comprehensive and concise definition of big data [54]. It is therefore unclear whether the term “big data” applies to our database. Big data is commonly characterized by volume, variety, velocity, and veracity [18,19,54], and some of these terms (volume and veracity) may be applicable to our database. However, a larger database including the latest datasets and longer durations of observation may be required to have the characteristics of “big data,” which enable researchers to use emerging analysis tools, including artificial intelligence techniques such as machine learning.

In our composite study, we expect to obtain a wide range of novel findings and to confirm indeterminate previous findings, with multidisciplinary applications, especially in terms of cardiometabolic disease. We also expect this work to provide new perspectives for human health promotion and disease prevention.

Authors' contributions

KN, TI, KT, JU, MS (Sugiyama), and TN contributed to the study design, the interpretation of the initial analysis, or the discussion of the literature and expected results. KN, TI, MS (Shibata), RH, and MA have conducted the data analysis. KN prepared the first draft of the manuscript, and all authors read and edited the manuscript.

Funding statement

This work was partly supported by a special Grant of Kanagawa University of Human Services, which was determined 12th, June, 2017 and Kanagawa Prefecture (No. 505744).

Competing interests statement

All authors have no potential conflicts of interest.

Acknowledgments

We thank Mr. Shoji Iwasaki and Mr. Takahiro Ozawa, civil servants of Kanagawa Prefecture, for technical advice on the management and security of our computer and analysis system. We thank Jennifer Barrett, PhD, from Edanz Group (www.edanzediting.com/ac) for editing a draft of this manuscript.

Figure legends

Figure 1. Location of Kanagawa Prefecture

Figure 2. Structure of the cross-sectional and cohort studies

They gray rectangles represent the each year's cross-sectional study. Cohort Study I consists of the cross-sectional studies of 2008 and 2014, and Cohort Study II includes all years from 2008 to 2014. The numbers highlighted in green represent the sample size of each dataset.

Figure 3. Checkup participation rates (%)

References

1. Ma RC, Chan JC. Type 2 diabetes in East Asians: similarities and differences with populations in Europe and the United States. Ann N Y Acad Sci. 2013;1281:64-91.

2. Lu Y, Hajifathalian K, Ezzati M, Woodward M, Rimm EB, Danaei G. Metabolic mediators of the effects of body-mass index, overweight, and obesity on coronary heart disease and stroke: a pooled analysis of 97 prospective cohorts with 1.8 million participants. Lancet. 2014; 383:970-83.

3. Arai H., Ouchi Y., Toba K. Japan as the front-runner of super-aged societies: perspectives from medicine and medical care in Japan. Geriatr Gerontol Int. 2015;15:673–687.

4. Suzuki T. Health status of older adults living in the community in Japan: Recent changes and significance in the super-aged society. Geriatr Gerontol Int. 2018 Feb 8.

5. Hayashi F, Takimoto H, Yoshita K, Yoshiike N. Perceived body size and desire for thinness of young Japanese women: a population-based survey. Br J Nutr. 2006;96:1154-62.

6. Sugawara A, Saito K, Sato M, Kodama S, Sone H. Thinness in Japanese young women. Epidemiology. 2009;20:464-5.

7. Ministry of Health, Labour and Welfare. National Nutrition Survey 2013 [cited 2015 September 10]. Available from: http://www.mhlw.go.jp/file/04-Houdouhappyou-10904750-Kenkoukyoku-Gantaisakukenkouzoushinka/0000068070.pdf.

8. Fried LP, Tangen CM, Walston J, Newman AB, Hirsch C, Gottdiener J, Seeman T, Tracy R, Kop WJ, Burke G, McBurnie MA; Cardiovascular Health Study Collaborative Research Group. Frailty in older adults: evidence for a phenotype. J Gerontol A Biol Sci Med Sci. 2001 Mar;56(3):M146-56.

9. Kido Y. The Issue of Nutrition in an Aging Society. J Nutr Sci Vitaminol (Tokyo). 2015;61 Suppl:S176-7.

10. Arai H. Sarcopenia, frailty and osteoporosis. Clin Calcium. 2017;27(9):1279-1285.

11. Wouterse B, Huisman M, Meijboom BR, Deeg DJ, Polder JJ. The effect of trends in health and

longevity on health services use by older adults. BMC Health Serv Res. 2015 Dec 24;15:574.

12. de Magalhães JP, Stevens M, Thornton D. The Business of Anti-Aging Science. Trends Biotechnol. 2017 Nov;35(11):1062-1073.
13. Ministry of Health, "Labour, and Welfare: Health Examination and Guidance Program for Japanese Adults," 2008. http://www.mhlw.go.jp/bunya/shakaihosho/iryouseido01/info02a.html (Japanese)
14. Kaplan RM, Chambers DA, Glasgow RE. Big data and large sample size: a cautionary note on the potential for bias. Clin Transl Sci. 2014 Aug;7(4):342-6.
15. Jagodnik KM, Koplev S, Jenkins SL, Ohno-Machado L, Paten B, Schurer SC, Dumontier M, Verborgh R, Bui A, Ping P, McKenna NJ, Madduri R, Pillai A, Ma'ayan A. Developing a framework for digital objects in the Big Data to Knowledge (BD2K) commons: Report from the Commons Framework Pilots workshop. J Biomed Inform. 2017 Jul;71:49-57.
16. Nizielski SE, Lechner PS, Croniger CM, Wang ND, Darlington GJ, Hanson RW. Animal models for studying the genetic basis of metabolic regulation. J Nutr. 1996;126:2697-2708.
17. Petters RM, Sommer JR. Transgenic animals as models for human disease. Transgenic Res. 2000;9:347-351; discussion 345-346.
18. Bellazzi R. Big data and biomedical informatics: a challenging opportunity. Yearb Med Inform. 2014;9:8-13.
19. Lee CH, Yoon HJ. Medical big data: promise and challenges. Kidney Res Clin Pract. 2017;36:3-11.
20. Provision of medical-related data to a third party. http://www.mhlw.go.jp/english/topics/2011eq/workers/tepco/lhc/ref10_20140604.pdf
21. Kubo S, Noda T, Myojin T, Higashino T, Matsui H, Kato G, Imamura T. The need and key points for patient matching in clinical studies using the National Database of Health Insurance Claims and Specific Health Checkups of Japan (NDB). Japanese Journal of Health & Research. 2017;38:11-19.
22. http://www.kuhs.ac.jp/kenkyuu/2017121400013/ (Japanese)

23. Ethical Guidelines for Medical and Health Research Involving Human Subjects. http://www.mhlw.go.jp/file/06-Seisakujouhou-10600000 Daijinkanboukouseikagakuka/0000080278.pdf

24. The final interim report by the work Group for studying the effects of the Specific Health Checkups and Specific Health Guidance on Health Care Expenditures http://www.mhlw.go.jp/file/05-Shingikai-12401000-Hokenkyoku-Soumuka/0000123428.pdf

25. https://www.mhlw.go.jp/file/06-Seisakujouhou-10900000-Kenkoukyoku/00_3.pdf

26. The definition and diagnostic criteria of metabolic syndrome. Nippon Naika Gakkai Zasshi 200.5; 94:794-809

27. Grundy SM, Brewer HB Jr, Cleeman JI, Smith SC Jr, Lenfant C; American Heart Association; National Heart, Lung, and Blood Institute. Definition of metabolic syndrome: Report of the National Heart, Lung, and Blood Institute/American Heart Association conference on scientific issues related to definition. Circulation. 2004 Jan 27;109(3):433-8.

28. Rodríguez-Ortiz D, Reyes-Pérez A, León P, Sánchez H, Mosti M, Aguilar-Salinas CA, Velázquez-Fernández D, Herrera MF. Assessment of two different diagnostic guidelines criteria (National Cholesterol Education Adult Treatment Panel III [ATP III] and International Diabetes Federation [IDF]) for the evaluation of metabolic syndrome remission in a longitudinal cohort of patients undergoing Roux-en-Y gastric bypass. Surgery. 2016 Apr;159(4):1121-8.

29. Kashiwagi A, Kasuga M, Araki E, et al. International clinical harmonization of glycated hemoglobin in Japan: from Japan Diabetes Society to National Glycohemoglobin Standardization Program values. J Diabetes Investig 2012;3:39–40.

30. Matsuo S, Imai E, Horio M, Yasuda Y, Tomita K, Nitta K, Yamagata K, Tomino Y, Yokoyama H, Hishida A. Collaborators developing the Japanese equation for estimated GFR. Revised equations for estimated GFR from serum creatinine in Japan. Am J Kidney Dis 2009;53:982–92

31. Keith NM, Wagener HP, Barker NW. Some different types of essential hypertension: their course

and prognosis. Am J Med Sci. 1974 Dec;268(6):336-45.

32. Sairenchi T, Iso H, Yamagishi K, Irie F, Okubo Y, Gunji J, Muto T, Ota H. Mild retinopathy is a risk factor for cardiovascular mortality in Japanese with and without hypertension: the Ibaraki Prefectural Health Study. Circulation. 2011 Dec 6;124(23):2502-11.

33. Tsushita K, Hosler AS, Miura K, Ito Y, Fukuda T, Kitamura A, Tatara K. Rationale and descriptive analysis of specific health guidance: the Nationwide Lifestyle Intervention Program Targeting Metabolic Syndrome in Japan. J Atheroscler Thromb. 2018;25:308-322.

34. Ichikawa D, Saito T, Oyama H. Impact of predicting health-guidance candidates using massive health check-up data: a data-driven analysis. Int J Med Inform. 2017;106:32-36

35. https://www.mhlw.go.jp/file/05-Shingikai-12401000-Hokenkyoku-Soumuka/0000135460.pdf

36. Health risks - Daily smokers - OECD Data https://data.oecd.org/healthrisk/daily-smokers.htm

37. Jamal A, Agaku IT, O'Connor E, King BA, Kenemer JB, Neff L. Current cigarette smoking among adults--United States, 2005-2013. MMWR Morb Mortal Wkly Rep. 2014 Nov 28;63(47):1108-12.

38. Falkner B, Cossrow ND. Prevalence of metabolic syndrome and obesity-associated hypertension in the racial ethnic minorities of the United States. Curr Hypertens Rep. 2014 Jul;16(7):449.

39. Engin A. The Definition and Prevalence of Obesity and Metabolic Syndrome. Adv Exp Med Biol. 2017;960:1-17.

40. Goto M, Goto A, Ikeda N, Noda H, Shibuya K, Noda M. Factors associated with untreated diabetes: analysis of data from 20,496 participants in the Japanese National Health and Nutrition Survey. PLoS One. 2015 Mar 10;10(3):e0118749.

41. Kutsuma A, Nakajima K, Suwa K. Potential Association between Breakfast Skipping and Concomitant Late-Night-Dinner Eating with Metabolic Syndrome and Proteinuria in the Japanese Population. Scientifica (Cairo). 2014;2014:253581.

42. Nakajima K, Suwa K, Oda E. Atrial fibrillation may be prevalent in individuals who report late-night dinner eating and concomitant breakfast skipping, a complex abnormal eating behavior

around sleep. Int J Cardiol. 2014 Dec 20;177(3):1124-6.

43. Ul-Haq Z, Mackay DF, Fenwick E, Pell JP. Meta-analysis of the association between body mass index and health-related quality of life among adults, assessed by the SF-36. Obesity (Silver Spring). 2013 Mar;21(3):E322-7.

44. Corbin LJ, Timpson NJ. Body mass index: Has epidemiology started to break down causal contributions to health and disease? Obesity (Silver Spring). 2016 Aug;24(8):1630-8.

45. Kotler DP. Cachexia. Ann Intern Med. 2000 Oct 17;133(8):622-34.

46. Poirier P, Cornier MA, Mazzone T, Stiles S, Cummings S, Klein S, McCullough PA, Ren Fielding C, Franklin BA; American Heart Association Obesity Committee of the Council on Nutrition, Physical Activity, and Metabolism. Bariatric surgery and cardiovascular risk factors: a scientific statement from the American Heart Association. Circulation. 2011 Apr 19;123(15):1683-701.

47. Muneyuki T, Suwa K, Oshida H, et al. Design of the saitama cardiometabolic disease and organ impairment study (SCDOIS): a multidisciplinary observational epidemiological study. Open J Endocr Metab Dis 2013;2:144–56.

48. Nakajima K, Suwa K. Excess body weight affects HbA1c progression irrespective of baseline HbA1c levels in Japanese individuals: a longitudinal retrospective study. Endocr Res. 2015;40(2):63-9.

49. Abdullah A. The Double Burden of Undernutrition and Overnutrition in Developing Countries: an Update. Curr Obes Rep. 2015 Sep;4(3):337-49.

50. World Health Organization. The double burden of malnutrition :Policy brief. Geneva: World Health Organization, 2017.

51. Demaio AR, Branca F. Decade of action on nutrition: our window to act on the double burden of malnutrition. BMJ Glob Health. 2017 Nov 25;3(Suppl 1):e000492.

52. Manrai AK, Patel CJ, Ioannidis JPA. In the era of precision medicine and big data, who is normal? JAMA. 2018;319:1981-1982.

53. Bell JA, Hamer M, Sabia S, Singh-Manoux A, Batty GD, Kivimaki M. The natural course of healthy obesity over 20 years. J Am Coll Cardiol. 2015 Jan 6;65(1):101-102.

54. Mayer-Schönberger V, Ingelsson E. Big Data and medicine: a big deal? J Intern Med. 2018;283:418-429.

Table 1. Clinical characteristics and methods for measurements

Parameters	Data of 2008 (total n = 1209,118)		Methods and remarks
Substituted age (years old)[a]	55.0 ± 9.8 in total 53.9 ± 9.5 for men 56.6 ± 10.1 for women		Actual age is unknown but classified into every five years
Sex, n (%)	Men 695,055 (57.5) Women 514,063 (42.5)		
Anthropometric parameters			
	Means ± SD	Ratios of methods or categories (%)	
BMI (kg/m^2)	23.0 ± 3.3 in total 23.7 ± 3.1 for men 22.0 ± 3.3 for women	–	Weight (in kg) divided by height (in m)^2
Waist circumference (cm) (1200,959)[b] – Men (690,133)[b]	84.5 ± 8.4	–	Objectively measured
Waist circumference (cm) (1200,959)[b] – Women (510,826)[b]	79.3 ± 9.6	–	
MetS (%) (1201,807)[b]	MetS	13.4	The diagnosis is determined by Japanese criteria [19].
	Pre-MetS	13.1	
	Non-MetS	72.2	Including non-Pre-MetS
	Unknown	1.4	Due to incomplete data
Health guidance level (1200,272)[b]	Intensive health guidance	11.1	Abdominal obesity [c] + risk factors [d] ≥ 2 or Risk factors [d] ≥ 3 + BMI ≥ 25 kg/m^2 (without abdominal obesity)
	Motivational health guidance	9.6	Abdominal obesity [c] + 1 risk factor [d] or 1 or 2 risk factors [d] + BMI ≥ 25 kg/m^2 (without abdominal obesity)
	Not applicable	77.9	
	Unknown	1.5	
Systolic blood pressure (mmHg)	125 ± 17	68.9	Once measurement

	125 ± 18	24.2	First time measurement among twice measurements
	131 ± 19	6.9	Second time measurement among twice measurements
	77 ± 11	67.5	Once measurement
Diastolic blood pressure (mmHg)	77 ± 12	24.2	First time measurement among twice measurements
	81 ± 13	8.2	Second time measurement among twice measurements
Serum parameters			
Aspartate aminotransferase (U/L) (1208,753)[b]	24 ± 12	84.9	Ultraviolet spectrophotometric determination (JSCC reference method)
	23 ± 11	15.1	Others
Alanine aminotransferase (U/L) (1208,455)[b]	23 ± 17	84.9	Ultraviolet spectrophotometric determination (JSCC reference method)
	24 ± 18	15.1	Others
Gamma-glutamyl transferase (U/L) (1208,074)[b]	36 ± 31	82.6	Ultraviolet spectrophotometric determination (JSCC reference method)
	36 ± 31	17.4	Others
Triglyceride (mg/dl) (1208,775)[b]	96 (68-141)	81.5	Ultraviolet and visible spectrophotometric determination (enzyme colorimetric/glycerol elimination methods)
	97 (68-146)	3.2	Ultraviolet spectrophotometric determination (enzyme colorimetric/ glycerol elimination method)
	96 (66-145)	15.3	Other
High-density lipoprotein cholesterol (mg/dl) (1208,872)[b]	64 ± 17	80.6	Ultraviolet and visible spectrophotometric determination (direct methods (non-precipitation method))
	61 ± 16	3.7	Ultraviolet spectrophotometric determination (direct methods (non-precipitation method))
	63 ± 17	15.7	Other
Low-density lipoprotein cholesterol (mg/dl) (1195,947)[b]	127 ± 31	80.8	Ultraviolet and visible spectrophotometric determination (direct methods (non-precipitation method))
	125 ± 31	3.4	Ultraviolet spectrophotometric determination (direct methods (non-precipitation method))
	125 ± 31	15.9	Other
Uric acid (mg/dl) [e]	NA	NA	NA
Creatinine (mg/dl) [e]	NA	NA	NA
eGFR (ml/min/1.73m^2) [e]	NA	NA	NA
Fasting plasma glucose (mg/dl) [f] (993,458)[b]	98 ± 19	29.6	Potentiometric determination
	98 ± 19	5.0	Ultraviolet and visible spectrophotometric determination
	97 ± 18	49.0	Ultraviolet spectrophotometric determination
	97± 18	16.4	Other
HbA1c (%, NGSP) [f] (945,345)[b]	5.2 ± 0.7	64.9	Immunoassay (Latex Agglutination turbidimetric Immunoassay, etc.)
	5.3 ± 0.6	16.9	HPLC determination
	5.2 ± 0.6	2.0	Enzymatic determination
	5.2 ± 0.7	16.3	Other
Urine parameters			

Proteinuria (%) (1194,283)[b]	3.8	56.9	Automated dipstick analysis
	3.9	43.1	Visual dipstick analysis
Glycosuria (%) (1195,049)[b]	2.0	57.2	Automated dipstick analysis
	2.1	52.8	Visual dipstick analysis
Fundus oculi examination			
Keith-Wagener hypertensive retinopathy (%) (available n = 13,866)	Mild (I) 5.3 Moderate– papilledema (II–IV) 1.0		Fundoscopy
Scheie hypertensive and sclerotic retinopathy (%) (available n = 15,894)	Hypertensive (1–4) 5.4 Sclerosis (1–4) 6.2		Fundoscopy

[a] Age groups (40–44, 45–49, 50–54, 55–59, 60–64, 65–69, and 70–74 years) are replaced with 42, 47, 52, 57, 62, 67, and 72 years old, respectively, corresponding to the median for each category.
[b] Available number. No superscript means that the number is completely available (n = 1209,118).
[c] Abdominal obesity: waist circumference ≥ 85 cm for men or ≥ 90 cm for women
[d] Numbers of risk factors: (1) fasting plasma glucose ≥ 100 mg/dl and/or HbA1c ≥ 5.6%, (2) triglyceride ≥ 150 mg/dl and/or high-density lipoprotein cholesterol < 40 mg/dl, (3) systolic blood pressure ≥ 130 mmHg and/or diastolic blood pressure ≥ 85 mmHg, (4) smoking (applicable only for subjects who had at least 1 risk, ranging from 1 to 3)
[e] Serum uric acid, creatinine, and consequently eGFR are currently unavailable but will become available in the future.
[f] Almost all subjects (n = 1205,956) had either fasting plasma glucose or HbA1c measured.
BMI: body mass index; MetS: metabolic syndrome; eGFR: estimated glomerular filtration rate

Table 2. Questionnaire on health status and results for 2008

No.	Questionnaire	Answers	Positive response to ① (%)[a]

	Are you taking following medicines at present?		
1	Medications to reduce blood pressure (1190,117)[b]	① Yes ② No	17.1
2	Insulin injection or medications to reduce blood glucose (1188,900)[b]	① Yes ② No	3.5
3	Medications to reduce your cholesterol level (1187,993)[b]	① Yes ② No	9.4
4	Have you ever been told by the doctor you have a stroke (cerebral hemorrhage, brain infarction, etc.) and received treatment for the disease? (978,542)[b]	① Yes ② No	2.0
5	Have you ever been told by the doctor you have a heart disease (angina pectoris, myocardial infarction, etc.) and received treatment for the disease? (978,491)[b]	① Yes ② No	3.9
6	Have you ever been diagnosed as having a chronic kidney failure and received treatment (dialysis therapy) for the disease? (970,176)[b]	① Yes ② No	0.3
7	Have you ever been diagnosed as anemic? (985,060)[b]	① Yes ② No	10.8
8	Are you a current regular smoker? (1192,091)[b] (Here a current regular smoker is to be a person who has smoked a total of 100 or more cigarettes or smoked for 6 months or longer and has been smoking for the last one month).	① Yes ② No	25.9 in total 37.2 in men 10.8 in women
9	Have you gained over 10 kg from your weight at 20 years old? (976,268)[b]	① Yes ② No	35.5
10	Are you in a habit of doing exercise to sweat lightly for over 30 minutes per session, 2 times weekly, for over a year? (979,191)[b]	① Yes ② No	30.9
11	In your daily life do you walk or do any equivalent amount of physical activity more than one hour per day? (980,581)[b]	① Yes ② No	43.6
12	Is your walking speed faster than the speed of corresponds of your age and sex? (964,407)[b]	① Yes ② No	52.3
13	Have you gained or lost over 3 kg during the last year? (965,421)[b]	① Yes ② No	22.1
14	How fast do you eat compared to others? (972,294)[b]	① Faster ② Normal ③ Slower	29.6
15	Do you eat dinner two hours before bedtime more than 3 times per week? (985,764)[b]	① Yes ② No	28.5
16	Do you eat snacks after supper more than 3 times per week? (958,850)[b]	① Yes ② No	12.8
17	Do you skip breakfast more than 3 times per week? (965,769)[b]	① Yes ② No	14.6
18	How often do you drink alcohol (sake, distilled spirits, beer, liquor, etc.)? (989,349)[b]	① Everyday ② Occasional ③ Hardly drink（cannot drink）	29.2
19	How much do you drink a day, in terms of glasses of refined sake? (A glass [180mL] of refined sake is equivalent to a medium bottle [500mL] of beer, 80mL of shochu (alcohol content 35 percent), a glass [double, 60mL] of	② Less than 180 ml (< 23 g ethanol) ③ Over 180–less than 360 ml (23-45 g ethanol)	56.8

	whiskey, and 2 glasses [240mL] of wine.) (820,231)[b]	④ Over 360–less than 540 ml (46-68 g ethanol) ④ Over 540 ml (≥ 69 g ethanol)	
20	Do you feel refreshed after a night's sleep? (973,947)[b]	① Yes ② No	65.4
21	Do you want to improve your life habits of eating and exercising? (951,484)[b]	① I don't mean to start. ② I'm going to start in the future (e.g., within 6 months). ③ I'm going to start soon (e.g., in a month), or I have just started some of them. ⑤ I already started (<6 months ago). ⑥ I already started (>=6 months ago).	28.5
22[c]	Are you willing to undergo a health counseling regarding lifestyle modifications if you get the opportunity?	① Yes ② No	–

[a] Proportions are calculated based on the available numbers for each question.
[b] Available number of subjects
[c] Question 22 is currently unavailable but will become available in the future.

Table 3. Major and minor outcomes

Major conditions or etiologies

· Cardiometabolic diseases including type 2 diabetes, hypertension, dyslipidemia, metabolic syndrome, chronic kidney disease assessed by proteinuria and eGFR

· Obesity and low body weight (malnutrition) assessed by BMI, and central obesity assessed by WC

· Hepatic diseases including fatty liver disease assessed by serum AST, ALT, and GGT

· Abnormal eating habits (breakfast skipping, late-night dinner eating, eating fast, night eating)

· Unhealthy lifestyles (smoking, infrequent exercise, heavy alcohol drinking, non-restorative sleep)

Minor conditions or etiologies

· Hypoglycemia, hyperfiltrations (high eGFR), hypotension, low uric acid

· Extremely low and high BMI (e.g., <15.0 kg/m^2 and > 40.0 kg/m^2)

· Osteoporosis assessed as reduced body-height during 6-years

· Hypertensive and atherosclerotic retinopathies (Keith-Wagener, Scheie classification)

· Physically inactive conditions including reduced walking speed and infrequent exercise

ALT: alanine aminotransferase; AST: aspartate aminotransferase; BMI: body mass index; eGFR: estimated glomerular filtration rate; GGT: gamma-glutamyl transferase; WC: waist circumference

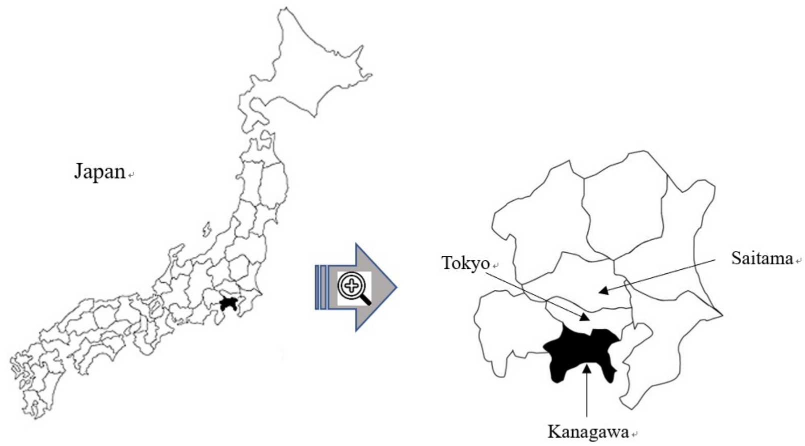

Figure 1

217x141mm (144 x 144 DPI)

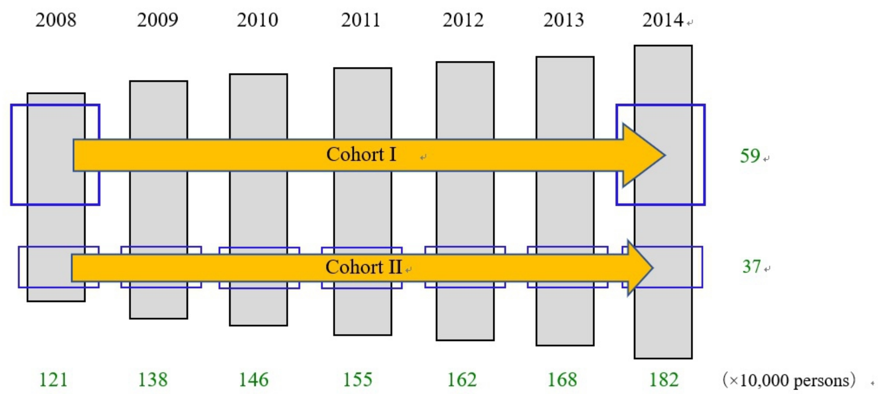

Figure 2

203x114mm (144 x 144 DPI)

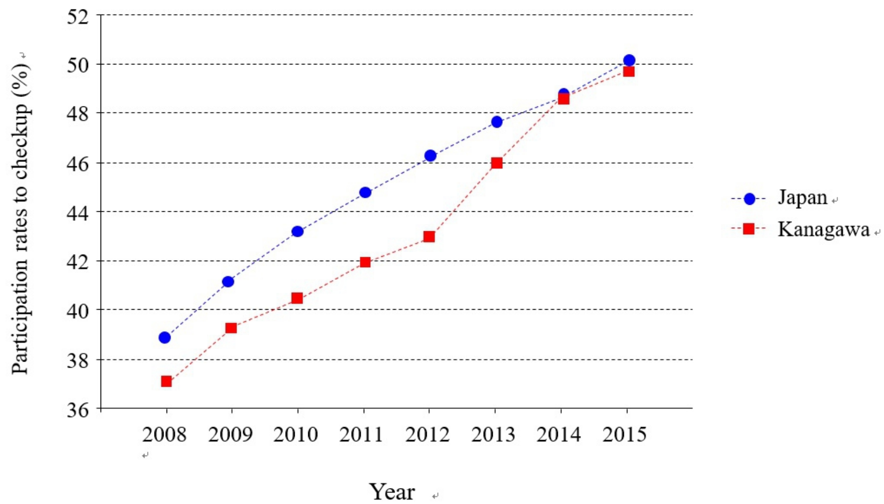

Figure 3

203x122mm (144 x 144 DPI)

The Kanagawa Investigation of the Total Checkup Data from the National Database (KITCHEN): Protocol for data-driven population-based repeated cross-sectional and 6-year cohort studies

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2018-023323.R2
Article Type:	Protocol
Date Submitted by the Author:	02-Dec-2018
Complete List of Authors:	Nakajima, Kei; Kanagawa University of Human Services, School of Nutrition and Dietetics, Faculty of Health and Social Services; Saitama Medical Center, Saitama Medical University, Department of Endocrinology and Diabetes Iwane, Taizo; Kanagawa University of Human Services, School of Nutrition and Dietetics, Faculty of Health and Social Services Higuchi, Ryoko; Kanagawa University of Human Services Shibata, Michi; Kanagawa University of Human Services; St. Marianna University School of Medicine Takada, Kento; Kanagawa University of Human Services Uda, Jun; Jikei Institute, Graduate School of Health Care Sciences Anan, Mami; Kanagawa University of Human Services Sugiyama, Michiko; Kanagawa University of Human Services Nakamura, Teiji; Kanagawa University of Human Services
Primary Subject Heading:	Epidemiology
Secondary Subject Heading:	Diabetes and endocrinology, Epidemiology, Geriatric medicine, Nutrition and metabolism, Cardiovascular medicine
Keywords:	EPIDEMIOLOGY, GENERAL MEDICINE (see Internal Medicine), GERIATRIC MEDICINE, NUTRITION & DIETETICS, PUBLIC HEALTH, DIABETES & ENDOCRINOLOGY

SCHOLARONE™
Manuscripts

The Kanagawa Investigation of the Total Checkup Data from the National Database (KITCHEN): Protocol for data-driven population-based repeated cross-sectional and 6-year cohort studies

Kei Nakajima,[1,2]* Taizo Iwane,[1] Ryoko Higuchi,[1] Michi Shibata,[1,3] Kento Takada,[1] Jun Uda,[4] Mami Anan,[1] Michiko Sugiyama,[1] Teiji Nakamura[1]

[1]School of Nutrition and Dietetics, Faculty of Health and Social Services, Kanagawa University of Human Services, 1-10-1 Heisei-cho, Yokosuka, Kanagawa 238-8522, Japan

[2]Department of Endocrinology and Diabetes, Saitama Medical Center, Saitama Medical University, 1981 Kamoda, Kawagoe, Saitama 350-8550, Japan

[3]Department of nutrition, St. Marianna University School of Medicine, 2-16-1 Sugao, Miyamae-ku, Kawasaki, Kanagawa 216-8511, Japan

[4]Graduate School of Health Care Sciences, Jikei Institute, 1-2-8 Miyahara, Yodogawa-ku, Osaka-shi, 532-0003, Japan

Running title: Kanagawa Investigation of Checkup Data

Address correspondence to Kei Nakajima, M.D., Ph.D.

School of Nutrition and Dietetics, Faculty of Health and Social Services, Kanagawa University of Human Services, 1-10-1 Heisei-cho, Yokosuka, Kanagawa 238-8522, Japan.

e-mail: nakajima-rsh@kuhs.ac.jp; Tel: +81-046-828-2660

Abstract

Introduction: The unmitigated incidence of cardiometabolic diseases, such as type 2 diabetes and metabolic syndrome, has gained attention in Japan. "Big data" can be useful to clarify conflicting observations obtained from studies with small samples and about rare conditions that are often neglected. We epidemiologically address these issues using data from health checkups conducted in Kanagawa Prefecture, the prefecture with the second largest population in Japan, in the Kanagawa Investigation of the Total Checkup Data from the National Database (KITCHEN).

Methods and analysis: This research consists of a series of population-based cross-sectional studies repeated from 2008 to 2014 and 6-year cohort studies. Since 2017, we have reviewed the data of people living in Kanagawa Prefecture who underwent a health checkup mainly for general health and the prevention of metabolic syndrome. The sample size ranges from 1.2 million to 1.8 million people in the cross-sectional studies and from 370,000 to 590,000 people in the cohort studies. These are people aged 40 to 74 years, whose clinical parameters were measured and who responded individually to a questionnaire. We investigate potential associations and causalities of various etiologies, including diabetes and metabolic syndrome, using clinical data and lifestyle information. With multidisciplinary analysis, including data-driven analysis, we expect to obtain a wide range of novel findings, to confirm indeterminate previous findings, especially in terms of cardiometabolic disease, and to provide new perspectives for human health promotion and disease prevention.

Ethics and dissemination:

Ethical approval was received from the Ethics Committee of Kanagawa University of Human Services (10-43). The protocol was approved in December 2016 by the Japanese Ministry of Health, Labour and Welfare (No. 121). The study results will be disseminated through open platforms including journal articles, relevant conferences, and seminar presentations.

Keywords: Kanagawa; checkup; big data; data mining; cardiometabolic disease; national database; health records; lifestyle-related disease; age-related disease

Strengths and limitations of this study

•The number of subjects in the sample is so large that more precise results concerning the means of parameters can be obtained, even when the subjects are classified into multiple categories, including sex, age group, smoking status, and certain morbidities such as obesity or diabetes.

•It may be possible to use big data to evaluate minor or rare conditions or etiologies that are commonly overlooked, neglected, or unfeasible to analyze in clinical studies, particularly those with small samples.

•It may also be possible to conduct data-driven and hypothesis-generating studies and then detect latent relationships among measures available in the data.

•Identical measurements and assessments of anthropometric indices, blood pressure, blood biochemistry, and urinalysis are performed across multiple years in the prefecture with the second largest population in Japan.

•The variations in parameters are restricted, and parameters for specific diseases are not included, because the checkups are conducted for general health and the prevention of common diseases, especially lifestyle-related diseases such type 2 diabetes and metabolic syndrome.

•Although cohort analysis is possible with this dataset, at 6 years, the period is relatively short, which may hamper the ability to uncover the latent relationships and underlying mechanisms between the parameters used and the predicted outcomes.

Introduction

Over the past several decades, the incidence of cardiometabolic diseases such as type 2 diabetes and metabolic syndrome (MetS) has not been reduced and has gained attention in Asia, including Japan [1,2], which has also experienced an unprecedented acceleration of societal aging [3,4]. These issues may also be problematic in Kanagawa Prefecture (***Figure 1***), an eastern district of Japan, located near Tokyo. The primary causes of these diseases include unfavorable lifestyles (e.g., smoking, heavy alcohol consumption, insufficient sleep, and infrequent exercise) and excess body weight (overweight and obesity) because of overeating, along with individuals' genetic and epigenetic backgrounds. However, for the last decade, malnutrition (e.g., low body weight) has been shown to be prevalent among young women [5-7] and the elderly [8-10] in Japan, which may contribute to the increased rates of sarcopenia and frailty in the country. Combined with prolonged longevity, cardiometabolic diseases with age-related causes create a long-term burden that leads to direct (i.e., measurements and therapies) and indirect (e.g., nursing, care, and welfare) medical costs nationwide [11,12], particularly when severe complications such as organ failure (e.g., heart, liver, and renal failure) develops over the life course.

In 2008, a special health checkup was initiated, primarily for the prevention of MetS, by the Ministry of Health, Labour and Welfare (MHLW) in Japan [13]. Since that time, all people living in Japan aged 40 to 74 years are supposed to undergo a yearly health checkup. The data from these checkups have continuously accumulated, creating a very large database. Such "big data" are likely to be useful in clarifying indistinct or conflicting results obtained from clinical studies with small sample sizes [14,15], confirming established results and advancing them by elucidating plausible mechanisms and clinical relevance, and enabling a precise understanding of the current status of public health and the contributing to it. Additionally, "big data" of this kind enable us to investigate minor or rare conditions and etiologies [14], such as extremely low and high body weight, abnormal (low and high) clinical measurements, and the low or high prevalence of unfavorable habits and lifestyles, where etiologies are hardly understood, especially when such extreme conditions are combined in complicated ways, primarily because of inadequate numbers of observations and corresponding cases. Extreme conditions can feasibly be reproduced in animal or cellular studies by means of intentional manipulation of conditions including through the use of transgenic and knockout technologies. These non-human laboratory studies can provide profound insight into the etiology of human diseases [16,17]. Clearly, such extreme conditions are mostly unfeasible in studies involving humans. However, in an epidemiological study with a database equivalent to "big data," it might be feasible to reproduce such extreme conditions in certain categories for limited conditions.

Although cardiometabolic diseases such as type 2 diabetes and chronic organ failures such as chronic kidney disease have been increasing along with the prolonged longevity in Japan [3,4], the underlying associations with clinical parameters and their mechanisms have not been fully elucidated or confirmed, particularly in epidemiological studies using the "big data" from the health checkups described above. These data include more than one million observations per year in most prefectures in Japan. To date, no investigation of this type has been performed, especially on the prefecture scale in Japan.

To this end, we investigated current cardiometabolic disease and health status as clearly as possible, as well as the relationship of cardiometabolic diseases, including but not limited to type 2 diabetes and MetS, and age-related etiologies. We focused especially on the thorough, end-to-end analysis of the variables of interest, using digitally recorded accumulated data in an extremely large epidemiological study of Kanagawa Prefecture, the second most populated prefecture in Japan, with approximately 9 million inhabitants, second only to Tokyo (approximately 13.7 million inhabitants), as of October 2017. Taking this approach, our study may be characterized as a data-driven, and hypothesis-generating study, with the nature of "big data" research, rather than a hypothesis-testing, traditional epidemiological study [18,19]. Consequently, the concrete objectives and contents of individual studies are difficult to determine before it becomes clear what kinds and amounts of data are available, which will substantially influence the design and analysis methods of each study.

Although big data is often analyzed with various algorithms including machine learning [18], in this study, we analyzed the data using traditional the epidemiological methods described in the next section, rather than machine learning.

Methods and analysis

Design

In 2013, the MHLW began to offer accumulated data consisting of information on patient prescriptions and health checkups for use by Japanese institutions including universities, hospitals, and research centers. These data are recorded digitally and are provided in a third-party manner, according to the concept of the "provision of medical-related data to a third party" to improve the quality of medical services and to support academic research in Japan [20]. To date, 178 applications from various institutions in Japan have been accepted in this manner (as of March 30, 2018).

Our project was a composite multidisciplinary study aimed at elucidating the factors associated with cardiometabolic diseases and eventually contributing to the amelioration and advancement of social health and welfare. After the study protocol was approved by the ethics committee of Kanagawa University of Human Services (10-43), we applied to the MHLW's data provision system in October 2016, through Teiji Nakamura, the President of Kanagawa University of Human Services, as a representative. The protocol of our study was approved in December 2016 by the MHLW (No. 121), after a peer review by an expert council.

Before we received the database from the MHLW, identifying individual-level information (names and postal codes) was completely transformed into randomized non-distinguishing anonymous numbers and characters, which prevents the restoration of this information by any means. There are two types of unique identifying variables available for each subject in the cross-sectional database collected from 2008 to 2014: ID 1 is determined based on the subject's insurance number, sex, and birth date, and ID 2 is determined by the subject's name, sex, and birthday. Both variables consist of anonymous numbers and characters created by the MHLW using a hash function [21]. For individual subjects, these variables are unchanged in principle, except when there are changes in the variables' constituting parts.

To further protect against the identification of specific individuals, age was categorized into 5-year age groups (40–44, 45–49, 50–54, 55–59, 60–64, 65–69, and 70–74 years), so that the individual's precise age at the time of data collection is unknown in our study.

Our study is part of the MHLW's nationwide program of providing medical-related data to third parties [20], and informed consent for the use of these data has not been obtained from each subject. We have opened the protocol of our study to the public on our university homepage, which was updated in October 2017 [22], in line with the "Ethical Guidelines for Medical and Health Research Involving Human Subjects" [23] in Japan (updated by the MHLW and the Ministry of Education, Culture, Sports, Science and Technology in May 2017). We received the digitally recorded non-distinguishing anonymous data from the MHLW in August 2017.

Our analysis of the data was conducted in a location with restricted access and tight security regarding datasets at Kanagawa University of Human Services. Repeated cross-sectional studies will be conducted using checkup data from 2008 to 2014. Additionally, a historical cohort study will be conducted, using the 2008 data as a baseline and the 2014 data to assess final outcomes (***Figure 2***). During this period, the number of subjects undergoing a health checkup has increased each year in Kanagawa in parallel with the nationwide trend. Nationally, almost 50% of the population attended a health checkup in 2014, probably because of the political encouragement for these checkups (***Figure 3***), although the MHLW's overall expected target rate is 70% [24].

People who did not undergo a checkup might have been under treatment for moderate to severe disease or hospitalized at the relevant time points. Health-minded people in Japan were likely to voluntarily undergo an expensive health checkup called the "*Ningen Dock*" (detailed and comprehensive health checkup). Other people who did not undergo a checkup might have missed the opportunity to have a checkup because of business obligations or other reasons, including family reasons or moving.

The name of the full study is the Kanagawa Investigation of the Total Checkup data from the National Database (KITCHEN). Each subsequent KITCHEN publication will be numbered sequentially from 1.

Subjects and measurements

People aged 40 to 74 years and living in Kanagawa Prefecture at the time of the data collection were enrolled in a series of studies. Those residing in medical institutions including hospitals and nursing homes were not included. All subjects are thought to be active to the extent of coming to the place where the checkup was performed. However, some of the subjects have diseases such as hypertension, diabetes, or dyslipidemia, and some have a history of morbidities such as heart disease or stroke. All of these conditions are digitally recorded as answers to a questionnaire. Specific exclusion and inclusion criteria will be determined for each study in the future. The sample sizes range from 1.2 million to 1.8 million people in the cross-sectional studies and from 370,000 to 590,000 people in the cohort studies (***Figure 2***). Cohort Study I uses data from 590,000 people who attended checkups in 2008 and 2014. Cohort Study II is based on data from 370,000 people who attended a checkup every year from 2008 to 2014. The two types of identifying variables (ID 1 and ID 2) described above are used to link data on individual subjects throughout the cohort study. When both of these variables simultaneously changed for a subject from 2008 to 2014, it was not possible to follow these individuals through time, resulting in the exclusion of these subjects from the cohort study. To date, such an event has been reported to occur at a rate of approximately 0.8% per year [21].

Patient and Public Involvement:

Patients are not involved in this study.

All of the parameters measured in this study are listed in ***Table 1***. To evaluate subject's age as a numeric value, we transformed age group (40–44, 45–49, 50–54, 55–59, 60–64, 65–69, and 70–74 years) into substituted age (s-age), corresponding to the median for each age group (42, 47, 52, 57, 62, 67, and 72 years). Body weight and height were objectively measured by trained institutional staff members and were recorded to one decimal place (kg and cm). Body mass index (BMI) was calculated as weight (in kg) divided by height (in m^2). In most cases (approximately 99.9%), waist circumference (WC) was measured objectively at the navel level by a medical staff member and recorded to one decimal place. Biochemical measurements were performed using standard methods and automated machines. Dipstick urine analysis for proteinuria and glycosuria was assessed visually or with ordinary automated machines. Several different methods were used for the included biochemical parameter (***Table 1***). The measurement of blood pressure and blood/urine biomarkers was regularly standardized using both internal standards with available traceability and external standards by third parties, including the Japanese Association of Medical Technologists, even when the measurements were outsourced [25].

In principal, most people underwent a checkup after overnight fasting. However, some of the checkups were conducted in a non-fasting condition because of, for example, shift work or family reasons. Therefore, all subjects were asked for the time (in hours) from their last meal to the time of the checkup, which was recorded as at least 10 hours or less than 10 hours. Those completing the checkup less than 10 hours after their last meal will be distinguished from others in certain sub-studies—for instance, when examining diabetes or dyslipidemia.

The Japanese diagnostic criteria for MetS were published in 2005 [26]. Unlike other criteria such as that of the Adult Treatment Panel III (ATP-III) and the International Diabetes Federation (IDF)

[27,28], the Japanese MetS criteria include abdominal obesity as an essential condition (WC ≥ 85 cm for men and ≥ 90 cm for women), in addition to two or more of the following three components: 1) dyslipidemia (triglycerides ≥ 150 mg/dl and/or high-density lipoprotein cholesterol < 40 mg/dl, or pharmacotherapy for dyslipidemia); 2) hypertension (systolic blood pressure ≥ 130 mmHg and/or diastolic blood pressure ≥ 85 mmHg, or pharmacotherapy for hypertension); and 3) hyperglycemia (fasting plasma glucose (FPG) ≥ 110 mg/dl or pharmacotherapy for diabetes). In the practice of the health checkups, hyperglycemia is defined as having elevated FPG (≥ 110 mg/dl) and/or HbA1c (National Glycohemoglobin Standardization Program [NGSP]) ≥ 6.0%, or pharmacotherapy for diabetes. Furthermore, pre-MetS is defined as abdominal obesity plus one of the three components listed above [24]. In sub-studies concerning MetS, MetS will also be determined using other international criteria, such as that of the ATP-III or the IDF to allow for comparison with the same criteria with other Asian countries as well as with Western countries. HbA1c (Japan Diabetes Society [JDS]) was converted to HbA1c (NGSP) units using the officially certified formula: HbA1c (NGSP) (%) = 1.02 × JDS (%) + 0.25% [29]. In 2008, almost all subjects had either FPG or HbA1c measured (99.7%).

If the data on serum creatinine eventually become available, the estimated glomerular filtration rate (eGFR) will be calculated using the s-age above and the following equation [30]: eGFR (ml/min/1.73 m^2) = 194 × serum $Cr^{-1.094}$ × s-$age^{-0.287}$ (if female) × 0.739, where Cr denotes serum creatinine concentration (mg/dl).

Hypertensive retinopathy has been shown to be associated with cardiovascular events and mortality [31,32]. Hypertensive retinopathy assessments using the Keith–Wagener and Scheie classifications are available in the study, although a very small percentage of individuals (around 1.3%) completed the hypertensive retinopathy examination.

The standardized 22-item questionnaire created by the MHLW for the health examination checkups is shown in ***Table 2***.

Primary and secondary (minor) outcomes

In a series of studies, we will consider various conditions and etiologies as primary and secondary (or minor) outcomes (see ***Table 3***). However, because unexpected findings are likely to be obtained during these studies and related research topics will be pursued following these findings, we do not restrict the areas of research to be pursued as long as the findings can contribute to or advance specific or general health objectives.

It is noteworthy that subjects are made aware of their health status when their checkup results are complete, and they often receive advice and suggestions from health professionals. Therefore, our cohort studies are not natural history cohort studies by nature. Some proportion of subjects undergo treatments in hospitals, and some receive further health guidance because of the results of their checkups. Specifically, health guidance for eligible subjects (***Table 1***) aims for the prevention or improvement of mainly cardiometabolic diseases, including MetS. In Japan, medical insurers are required to recommend that individuals at risk of these conditions receive health guidance, although this is not obligatory for the individuals. Health guidance is classified into two categories (intensive and motivational health guidance), depending on the individual's abdominal obesity (waist circumference $\geq$ 85 cm for men or $\geq$ 90 cm for women) and number of risk factors (***Table 1***). In brief, in cases of intensive health guidance, subjects receive consultation via e-mail, phone, or face-to-face sessions for up to 6 months, as has been described in detail elsewhere [24, 33], whereas subjects receiving motivational health guidance do not receive continuous support. Notably, subjects undergoing pharmacotherapy for hypertension, diabetes, or dyslipidemia are excluded, and those aged 65–74 years receive motivational health guidance regardless of their risk profile. In Japan, attendance rates for health guidance have been found to be less than 20% [24,34].

Statistical analysis

Continuous and categorical variables will be compared between different groups using analysis of variance (ANOVA) and χ^2-tests, respectively. Post hoc comparisons between two specific groups will be examined with the Bonferroni, Tukey–Kramer, and Dunnett methods, as well as additional χ^2-tests. Paired or trend data will rely on the McNemar, Cochran–Armitage, and Mantel-Haenszel test, respectively. Analysis of covariance (ANCOVA) with general linear model procedures will be used to examine the difference in biochemical variables measured by two or three different methods (e.g., LDL cholesterol is measured using three different methods) (***Table 1***), controlling for confounders including age, sex, body weight, and various lifestyles. Logistic regression and hazard models will be used to examine the associations or causalities between abnormal levels of measurements and conditions with major and minor outcomes. These methods will yield odds ratios, risk ratios, or hazard ratios, which will be presented along with their 95% confidence intervals. Panel data analysis (including the Hausman test) combining several sets of cross-sectional data will also be conducted. Relevant confounding factors include age, sex, smoking, and alcohol consumption, which will be adjusted in the regression analyses. Alternatively, to evaluate or control the differences in backgrounds and various confounders between cases and controls, individuals' propensity scores will be calculated as a variable that unifies all corresponding confounders in the analysis. Propensity score is also considered for a special examination, for instance, hypertensive retinopathy examination, because few subjects underwent such special examination, which yields a bias to be adjusted. The level of health guidance (***Table 1***) and the answer to Question 21, which asks about personal intentions to improve eating and exercise habits (***Table 2***), will also be considered as confounding factors, when appropriate.

There are missing data in our study, although this comprises less than 20% of the cases for most parameters and questionnaire items. For categorized age, sex, BMI, and WC (***Table 1***), the data are almost complete, even when all of these variables are combined (99.99%). However, combining parameters other than age, sex, BMI, and WC can decrease the total available number of subjects, depending on the study's nature and design.

When analyzing extremely rare conditions, which might lead to the disclose of the identity of individuals with rare diseases needing treatments in hospital, to prevent the identification of these subjects, we do not describe the number anywhere in the manuscript if this is less than 10, as advised in the MHLW's guidelines [35].

Statistical analyses will be performed using SAS-Enterprise Guide (SAS-EG 7.1) in the SAS system, Version 9.4 (SAS Institute, Cary, North Carolina, USA). Values of $p < 0.05$ will be considered statistically significant. Because we understand that large data are predisposed to detect the presence of statistical significance, we will use caution in our interpretations and give priority to clinical significance rather than statistical significance in certain clinical areas.

Overall characteristics of subjects

Subjects' characteristics at baseline (2008) are shown in ***Table 1***. These findings will vary to some extent in sub-studies using other cross-sectional data from 2009 to 2014. Men in the first five years of their 50s are overrepresented in the sample, probably because middle-aged men are more likely than women to work for companies and institutes (i.e., insurers), which obligate workers to undergo a checkup. Considering the clinical parameters of BMI, WC, blood pressure, lipids, FPG, HbA1c, and MetS, most of the subjects are apparently healthy people with these parameters within normal ranges. The questionnaire results (2008) are shown in ***Table 2***, which gives us rough information about the subjects' backgrounds. The smoking rate is higher (25.9%), especially among men (37.2%), compared with other developed countries such as the United States [36,37], although the smoking rate has been declining in Japan in recent years [36].

The prevalence of MetS (13.4%), as well as pharmacotherapy for hypertension, diabetes, and dyslipidemia are relatively lower (3.5%–17.1%), compared with other countries [38,39]. However, this does not always mean that the number needed to treat is low, because substantial proportions of subjects likely do not consult a doctor about their poor glycemic control. A Japanese national survey conducted from 2005 to 2009 found that, among people with diabetes, a substantial proportion (about 38%) have left their poor glycemic conditions as they are, without seeking treatment [40]. In our study, the extent of this issue remains unknown without a detailed investigation of FPG and HbA1c.

Concerning eating habits, which play an important role in metabolism and nutrition, the percentage of subjects who habitually skip breakfast is lower (14.6%) than the percentage who eat dinner two hours before bedtime (28.5%), which is consistent with a previous study [41]. This suggests that the latter group may be more troublesome in terms of unfavorable lifestyle habits that are linked to cardiovascular diseases, because a close association between eating dinner late at night and skipping breakfast has been reported in a community-based epidemiological study [41]; eating dinner late at night can lead to skipping breakfast the next morning. Acknowledging the relationship with sleep, we term these behaviors "unfavorable eating habits around sleep" (UEHAS). In previous studies [41,42], eating dinner late at night together with skipping breakfast—a combination representative of UEHAS—was significantly associated with MetS, proteinuria, and atrial fibrillation.

Body weight substantially influences the incidence and development of cardiometabolic diseases as well as general health [43,44]. However, features and etiologies at both extremes of BMI, a fundamental index of weight considering height, body adiposity, nutritional status, and health, are poorly understood. For rare conditions in malnutrition, for instance, the percentage of subjects with an extremely low body weight (BMI < 15.0 kg/m^2, a criterion for high mortality [45]) is very small in the cross-sectional dataset in 2008 (0.1%), but the observational number in this dataset is large (n = 1,217, data not shown), which is not ignorable and may be enough to conduct proper statistical analyses. Likewise, extremely high body weight (BMI ≥ 40.0 kg/m^2, a criterion for class III obesity

[46]), is also very small in percentage (0.07%), but there are 805 observations for this group in this dataset.

Throughout the study presented here, we expect to obtain a wide range of novel observations, enabling us to confirm indeterminate previous findings, especially in terms of cardiometabolic disease. In addition, this study will likely reveal underlying etiologies that have been overlooked because they are rare or minor cases in small clinical studies.

Among our research team members, one person (Nakajima) has previously been involved in a similar large study, consisting of approximately 100,000 people living in Saitama Prefecture (population = 7.3 million people), which is also located near Tokyo [47] (***Figure 1***). This previous study and its sub-studies were launched in 2011 and, to date, multiple findings have been reported from these studies. Although a study of 100,000 people is generally considered "large," this sample size sometimes proved to be inadequate for stratification analysis because of small observational numbers for particular groups when several stratification variables were combined—for instance, age group, BMI category, and diabetes status [48]. This is the main reason we chose to begin a new study using the checkup data for an extraordinarily large sample—over one million for a cross-sectional study, which is more than 10 times higher than this previous study [47].

BMI roughly reflects nutritional status, including excess energy accumulation or malnutrition. In the last decade, increased percentages of people have been found at both extremes of BMI (i.e., in the nutritional states of malnutrition or obesity), worldwide, especially among children. This coexistence of undernutrition and obesity or nutrition-related noncommunicable disease has been termed the "double burden of malnutrition" [49-51]. In our study, subjects' characteristics described above suggest that the double burden of malnutrition can exist even among the middle-aged Japanese population, although the proportions are smaller compared with other developing countries [49].

Our study has several strengths. First, in terms of community-based epidemiological research, the sample is so large that precise results concerning the means of parameters and "normal" values, although these are standard parameters, can be obtained [52], even when subjects are classified into categories such as sex, age group, smoking status, and certain morbidities (e.g., obesity or diabetes). Therefore, it may be possible to conduct similar analyses to produce novel results in other population studies with very large databases that allow for multiple classifications. A second strength of this study is that it may be possible to use big data to evaluate minor or rare conditions and etiologies that are commonly overlooked, neglected, or unfeasible to analyze in clinical studies, particularly those with small samples [14]. This analysis may contribute to case studies instead of only to the field of public health. Finally, identical measurements and assessments of anthropometric indices, blood pressure, and urinalysis are performed across multiple years in people living in similar environment and the same health care system.

Some limitations to this study should also be mentioned. First, the variations in parameters are restricted, and parameters for specific diseases are not included, because the checkups are conducted for general health and the prevention of common diseases, especially lifestyle-related diseases such type 2 diabetes and MetS. Second, people younger than 40 years and those aged over 74 years are not enrolled in this study. Lifestyle choices made when people are younger may contribute to the incidence of morbidities in middle age, and lifestyles and clinical biochemistry levels in middle age can influence the incidence and severity of cardiovascular diseases and health damage in the later life. Unfortunately, comparison with younger and older people is unfeasible, so a seamless analysis over the life course is impossible in this study. Third, although cohort analysis using this dataset is possible, at 6 years, the period is relatively short, which may hamper the ability to uncover the latent relationships and underlying mechanisms between the parameters used and the predicted outcomes. Durations of 10 years or even several decades may be needed to clarify the latent causality between suspected factors and outcomes [53]. Finally, to date, there is no comprehensive and concise definition of big data [54]. It is therefore unclear whether the term "big data" applies to our database.

Big data is commonly characterized by volume, variety, velocity, and veracity [18,19,54], and some of these terms (volume and veracity) may be applicable to our database. However, a larger database including the latest datasets and longer durations of observation may be required to have the characteristics of "big data," which enable researchers to use emerging analysis tools, including artificial intelligence techniques such as machine learning.

In our composite study, we expect to obtain a wide range of novel findings and to confirm indeterminate previous findings, with multidisciplinary applications, especially in terms of cardiometabolic disease. We also expect this work to provide new perspectives for human health promotion and disease prevention.

Authors' contributions

KN, TI, KT, JU, MS (Sugiyama), and TN contributed to the study design, the interpretation of the initial analysis, or the discussion of the literature and expected results. KN, TI, MS (Shibata), RH, and MA have conducted the data analysis. KN prepared the first draft of the manuscript, and all authors read and edited the manuscript.

Funding statement

This work was partly supported by a special Grant of Kanagawa University of Human Services, which was determined 12th, June, 2017 and Kanagawa Prefecture (No. 505744).

Competing interests statement

All authors have no potential conflicts of interest.

Acknowledgments

We thank Mr. Shoji Iwasaki and Mr. Takahiro Ozawa, civil servants of Kanagawa Prefecture, for technical advice on the management and security of our computer and analysis system. We thank

Jennifer Barrett, PhD, from Edanz Group (www.edanzediting.com/ac) for editing a draft of this manuscript.

Figure legends

Figure 1. Location of Kanagawa Prefecture

Figure 2. Structure of the cross-sectional and cohort studies

They gray rectangles represent the each year's cross-sectional study. Cohort Study I consists of the cross-sectional studies of 2008 and 2014, and Cohort Study II includes all years from 2008 to 2014. The numbers highlighted in green represent the sample size of each dataset.

Figure 3. Checkup participation rates (%)

References

1. Ma RC, Chan JC. Type 2 diabetes in East Asians: similarities and differences with populations in Europe and the United States. Ann N Y Acad Sci. 2013;1281:64-91.
2. Lu Y, Hajifathalian K, Ezzati M, Woodward M, Rimm EB, Danaei G. Metabolic mediators of the effects of body-mass index, overweight, and obesity on coronary heart disease and stroke: a pooled analysis of 97 prospective cohorts with 1.8 million participants. Lancet. 2014; 383:970-83.
3. Arai H., Ouchi Y., Toba K. Japan as the front-runner of super-aged societies: perspectives from medicine and medical care in Japan. Geriatr Gerontol Int. 2015;15:673–687.
4. Suzuki T. Health status of older adults living in the community in Japan: Recent changes and significance in the super-aged society. Geriatr Gerontol Int. 2018 Feb 8.
5. Hayashi F, Takimoto H, Yoshita K, Yoshiike N. Perceived body size and desire for thinness of young Japanese women: a population-based survey. Br J Nutr. 2006;96:1154-62.
6. Sugawara A, Saito K, Sato M, Kodama S, Sone H. Thinness in Japanese young women. Epidemiology. 2009;20:464-5.
7. Ministry of Health, Labour and Welfare. National Nutrition Survey 2013 [cited 2015 September 10]. Available from: http://www.mhlw.go.jp/file/04-Houdouhappyou-10904750-Kenkoukyoku-Gantaisakukenkouzoushinka/0000068070.pdf.
8. Fried LP, Tangen CM, Walston J, Newman AB, Hirsch C, Gottdiener J, Seeman T, Tracy R, Kop WJ, Burke G, McBurnie MA; Cardiovascular Health Study Collaborative Research Group. Frailty in older adults: evidence for a phenotype. J Gerontol A Biol Sci Med Sci. 2001 Mar;56(3):M146-56.
9. Kido Y. The Issue of Nutrition in an Aging Society. J Nutr Sci Vitaminol (Tokyo). 2015;61 Suppl:S176-7.
10. Arai H. Sarcopenia, frailty and osteoporosis. Clin Calcium. 2017;27(9):1279-1285.
11. Wouterse B, Huisman M, Meijboom BR, Deeg DJ, Polder JJ. The effect of trends in health and

longevity on health services use by older adults. BMC Health Serv Res. 2015 Dec 24;15:574.

12. de Magalhães JP, Stevens M, Thornton D. The Business of Anti-Aging Science. Trends Biotechnol. 2017 Nov;35(11):1062-1073.
13. Ministry of Health, "Labour, and Welfare: Health Examination and Guidance Program for Japanese Adults," 2008. http://www.mhlw.go.jp/bunya/shakaihosho/iryouseido01/info02a.html (Japanese)
14. Kaplan RM, Chambers DA, Glasgow RE. Big data and large sample size: a cautionary note on the potential for bias. Clin Transl Sci. 2014 Aug;7(4):342-6.
15. Jagodnik KM, Koplev S, Jenkins SL, Ohno-Machado L, Paten B, Schurer SC, Dumontier M, Verborgh R, Bui A, Ping P, McKenna NJ, Madduri R, Pillai A, Ma'ayan A. Developing a framework for digital objects in the Big Data to Knowledge (BD2K) commons: Report from the Commons Framework Pilots workshop. J Biomed Inform. 2017 Jul;71:49-57.
16. Nizielski SE, Lechner PS, Croniger CM, Wang ND, Darlington GJ, Hanson RW. Animal models for studying the genetic basis of metabolic regulation. J Nutr. 1996;126:2697-2708.
17. Petters RM, Sommer JR. Transgenic animals as models for human disease. Transgenic Res. 2000;9:347-351; discussion 345-346.
18. Bellazzi R. Big data and biomedical informatics: a challenging opportunity. Yearb Med Inform. 2014;9:8-13.
19. Lee CH, Yoon HJ. Medical big data: promise and challenges. Kidney Res Clin Pract. 2017;36:3-11.
20. Provision of medical-related data to a third party. http://www.mhlw.go.jp/english/topics/2011eq/workers/tepco/lhc/ref10_20140604.pdf
21. Kubo S, Noda T, Myojin T, Higashino T, Matsui H, Kato G, Imamura T. The need and key points for patient matching in clinical studies using the National Database of Health Insurance Claims and Specific Health Checkups of Japan (NDB). Japanese Journal of Health & Research. 2017;38:11-19.

22. http://www.kuhs.ac.jp/kenkyuu/2017121400013/ (Japanese)

23. Ethical Guidelines for Medical and Health Research Involving Human Subjects. http://www.mhlw.go.jp/file/06-Seisakujouhou-10600000 Daijinkanboukouseikagakuka/0000080278.pdf

24. The final interim report by the work Group for studying the effects of the Specific Health Checkups and Specific Health Guidance on Health Care Expenditures http://www.mhlw.go.jp/file/05-Shingikai-12401000-Hokenkyoku-Soumuka/0000123428.pdf

25. https://www.mhlw.go.jp/file/06-Seisakujouhou-10900000-Kenkoukyoku/00_3.pdf

26. The definition and diagnostic criteria of metabolic syndrome. Nippon Naika Gakkai Zasshi 200.5; 94:794-809

27. Grundy SM, Brewer HB Jr, Cleeman JI, Smith SC Jr, Lenfant C; American Heart Association; National Heart, Lung, and Blood Institute. Definition of metabolic syndrome: Report of the National Heart, Lung, and Blood Institute/American Heart Association conference on scientific issues related to definition. Circulation. 2004 Jan 27;109(3):433-8.

28. Rodríguez-Ortiz D, Reyes-Pérez A, León P, Sánchez H, Mosti M, Aguilar-Salinas CA, Velázquez-Fernández D, Herrera MF. Assessment of two different diagnostic guidelines criteria (National Cholesterol Education Adult Treatment Panel III [ATP III] and International Diabetes Federation [IDF]) for the evaluation of metabolic syndrome remission in a longitudinal cohort of patients undergoing Roux-en-Y gastric bypass. Surgery. 2016 Apr;159(4):1121-8.

29. Kashiwagi A, Kasuga M, Araki E, et al. International clinical harmonization of glycated hemoglobin in Japan: from Japan Diabetes Society to National Glycohemoglobin Standardization Program values. J Diabetes Investig 2012;3:39–40.

30. Matsuo S, Imai E, Horio M, Yasuda Y, Tomita K, Nitta K, Yamagata K, Tomino Y, Yokoyama H, Hishida A. Collaborators developing the Japanese equation for estimated GFR. Revised equations for estimated GFR from serum creatinine in Japan. Am J Kidney Dis 2009;53:982–92

31. Keith NM, Wagener HP, Barker NW. Some different types of essential hypertension: their course and prognosis. Am J Med Sci. 1974 Dec;268(6):336-45.

32. Sairenchi T, Iso H, Yamagishi K, Irie F, Okubo Y, Gunji J, Muto T, Ota H. Mild retinopathy is a risk factor for cardiovascular mortality in Japanese with and without hypertension: the Ibaraki Prefectural Health Study. Circulation. 2011 Dec 6;124(23):2502-11.

33. Tsushita K, Hosler AS, Miura K, Ito Y, Fukuda T, Kitamura A, Tatara K. Rationale and descriptive analysis of specific health guidance: the Nationwide Lifestyle Intervention Program Targeting Metabolic Syndrome in Japan. J Atheroscler Thromb. 2018;25:308-322.

34. Ichikawa D, Saito T, Oyama H. Impact of predicting health-guidance candidates using massive health check-up data: a data-driven analysis. Int J Med Inform. 2017;106:32-36

35. https://www.mhlw.go.jp/file/05-Shingikai-12401000-Hokenkyoku-Soumuka/0000135460.pdf

36. Health risks - Daily smokers - OECD Data https://data.oecd.org/healthrisk/daily-smokers.htm

37. Jamal A, Agaku IT, O'Connor E, King BA, Kenemer JB, Neff L. Current cigarette smoking among adults--United States, 2005-2013. MMWR Morb Mortal Wkly Rep. 2014 Nov 28;63(47):1108-12.

38. Falkner B, Cossrow ND. Prevalence of metabolic syndrome and obesity-associated hypertension in the racial ethnic minorities of the United States. Curr Hypertens Rep. 2014 Jul;16(7):449.

39. Engin A. The Definition and Prevalence of Obesity and Metabolic Syndrome. Adv Exp Med Biol. 2017;960:1-17.

40. Goto M, Goto A, Ikeda N, Noda H, Shibuya K, Noda M. Factors associated with untreated diabetes: analysis of data from 20,496 participants in the Japanese National Health and Nutrition Survey. PLoS One. 2015 Mar 10;10(3):e0118749.

41. Kutsuma A, Nakajima K, Suwa K. Potential Association between Breakfast Skipping and Concomitant Late-Night-Dinner Eating with Metabolic Syndrome and Proteinuria in the Japanese Population. Scientifica (Cairo). 2014;2014:253581.

42. Nakajima K, Suwa K, Oda E. Atrial fibrillation may be prevalent in individuals who report late-night dinner eating and concomitant breakfast skipping, a complex abnormal eating behavior around sleep. Int J Cardiol. 2014 Dec 20;177(3):1124-6.

43. Ul-Haq Z, Mackay DF, Fenwick E, Pell JP. Meta-analysis of the association between body mass

index and health-related quality of life among adults, assessed by the SF-36. Obesity (Silver Spring). 2013 Mar;21(3):E322-7.

44. Corbin LJ, Timpson NJ. Body mass index: Has epidemiology started to break down causal contributions to health and disease? Obesity (Silver Spring). 2016 Aug;24(8):1630-8.

45. Kotler DP. Cachexia. Ann Intern Med. 2000 Oct 17;133(8):622-34.

46. Poirier P, Cornier MA, Mazzone T, Stiles S, Cummings S, Klein S, McCullough PA, Ren Fielding C, Franklin BA; American Heart Association Obesity Committee of the Council on Nutrition, Physical Activity, and Metabolism. Bariatric surgery and cardiovascular risk factors: a scientific statement from the American Heart Association. Circulation. 2011 Apr 19;123(15):1683-701.

47. Muneyuki T, Suwa K, Oshida H, et al. Design of the saitama cardiometabolic disease and organ impairment study (SCDOIS): a multidisciplinary observational epidemiological study. Open J Endocr Metab Dis 2013;2:144–56.

48. Nakajima K, Suwa K. Excess body weight affects HbA1c progression irrespective of baseline HbA1c levels in Japanese individuals: a longitudinal retrospective study. Endocr Res. 2015;40(2):63-9.

49. Abdullah A. The Double Burden of Undernutrition and Overnutrition in Developing Countries: an Update. Curr Obes Rep. 2015 Sep;4(3):337-49.

50. World Health Organization. The double burden of malnutrition :Policy brief. Geneva: World Health Organization, 2017.

51. Demaio AR, Branca F. Decade of action on nutrition: our window to act on the double burden of malnutrition. BMJ Glob Health. 2017 Nov 25;3(Suppl 1):e000492.

52. Manrai AK, Patel CJ, Ioannidis JPA. In the era of precision medicine and big data, who is normal? JAMA. 2018;319:1981-1982.

53. Bell JA, Hamer M, Sabia S, Singh-Manoux A, Batty GD, Kivimaki M. The natural course of healthy obesity over 20 years. J Am Coll Cardiol. 2015 Jan 6;65(1):101-102.

54. Mayer-Schönberger V, Ingelsson E. Big Data and medicine: a big deal? J Intern Med.

2018;283:418-429.

Table 1. Clinical characteristics and methods for measurements

Parameters		Data of 2008 (total n = 1209,118)		Methods and remarks
	Substituted age (years old)[a]	55.0 ± 9.8 in total 53.9 ± 9.5 for men 56.6 ± 10.1 for women		Actual age is unknown but classified into every five years
	Sex, n (%)	Men 695,055 (57.5) Women 514,063 (42.5)		
Anthropometric parameters				
		Means ± SD	Ratios of methods or categories (%)	
	BMI (kg/m^2)	23.0 ± 3.3 in total 23.7 ± 3.1 for men 22.0 ± 3.3 for women	–	Weight (in kg) divided by height (in m)^2
Waist circumference (cm) (1200,959)[b]	Men (690,133)[b]	84.5 ± 8.4	–	Objectively measured
	Women (510,826)[b]	79.3 ± 9.6	–	
MetS (%) (1201,807)[b]		MetS	13.4	The diagnosis is determined by Japanese criteria [19].
		Pre-MetS	13.1	
		Non-MetS	72.2	Including non-Pre-MetS
		Unknown	1.4	Due to incomplete data
Health guidance level (1200,272)[b]		Intensive health guidance	11.1	Abdominal obesity [c] + risk factors [d] ≥ 2 or Risk factors [d] ≥ 3 + BMI ≥ 25 kg/m^2 (without abdominal obesity)
		Motivational health guidance	9.6	Abdominal obesity [c] + 1 risk factor [d] or 1 or 2 risk factors [d] + BMI ≥ 25 kg/m^2 (without abdominal obesity)
		Not applicable	77.9	
		Unknown	1.5	

Systolic blood pressure (mmHg)	125 ± 17	68.9	Once measurement
	125 ± 18	24.2	First time measurement among twice measurements
	131 ± 19	6.9	Second time measurement among twice measurements
Diastolic blood pressure (mmHg)	77 ± 11	67.5	Once measurement
	77 ± 12	24.2	First time measurement among twice measurements
	81 ± 13	8.2	Second time measurement among twice measurements
Serum parameters			
Aspartate aminotransferase (U/L) (1208,753)[b]	24 ± 12	84.9	Ultraviolet spectrophotometric determination (JSCC reference method)
	23 ± 11	15.1	Others
Alanine aminotransferase (U/L) (1208,455)[b]	23 ± 17	84.9	Ultraviolet spectrophotometric determination (JSCC reference method)
	24 ± 18	15.1	Others
Gamma-glutamyl transferase (U/L) (1208,074)[b]	36 ± 31	82.6	Ultraviolet spectrophotometric determination (JSCC reference method)
	36 ± 31	17.4	Others
Triglyceride (mg/dl) (1208,775)[b]	96 (68-141)	81.5	Ultraviolet and visible spectrophotometric determination (enzyme colorimetric/glycerol elimination methods)
	97 (68-146)	3.2	Ultraviolet spectrophotometric determination (enzyme colorimetric/ glycerol elimination method)
	96 (66-145)	15.3	Other
High-density lipoprotein cholesterol (mg/dl) (1208,872)[b]	64 ± 17	80.6	Ultraviolet and visible spectrophotometric determination (direct methods (non-precipitation method))
	61 ± 16	3.7	Ultraviolet spectrophotometric determination (direct methods (non-precipitation method))
	63 ± 17	15.7	Other
Low-density lipoprotein cholesterol (mg/dl) (1195,947)[b]	127 ± 31	80.8	Ultraviolet and visible spectrophotometric determination (direct methods (non-precipitation method))
	125 ± 31	3.4	Ultraviolet spectrophotometric determination (direct methods (non-precipitation method))
	125 ± 31	15.9	Other
Uric acid (mg/dl) [e]	NA	NA	NA
Creatinine (mg/dl) [e]	NA	NA	NA
eGFR (ml/min/1.73m^2) [e]	NA	NA	NA
Fasting plasma glucose (mg/dl) [f] (993,458)[b]	98 ± 19	29.6	Potentiometric determination
	98 ± 19	5.0	Ultraviolet and visible spectrophotometric determination
	97 ± 18	49.0	Ultraviolet spectrophotometric determination
	97± 18	16.4	Other
HbA1c (%, NGSP) [f] (945,345)[b]	5.2 ± 0.7	64.9	Immunoassay (Latex Agglutination turbidimetric Immunoassay, etc.)
	5.3 ± 0.6	16.9	HPLC determination
	5.2 ± 0.6	2.0	Enzymatic determination

	5.2 ± 0.7	16.3	Other
Urine parameters			
Proteinuria (%) (1194,283)[b]	3.8	56.9	Automated dipstick analysis
	3.9	43.1	Visual dipstick analysis
Glycosuria (%) (1195,049)[b]	2.0	57.2	Automated dipstick analysis
	2.1	52.8	Visual dipstick analysis
Fundus oculi examination			
Keith-Wagener hypertensive retinopathy (%) (available n = 13,866)	Mild (I) 5.3 Moderate– papilledema (II–IV) 1.0		Fundoscopy
Scheie hypertensive and sclerotic retinopathy (%) (available n = 15,894)	Hypertensive (1–4) 5.4 Sclerosis (1–4) 6.2		Fundoscopy

[a] Age groups (40–44, 45–49, 50–54, 55–59, 60–64, 65–69, and 70–74 years) are replaced with 42, 47, 52, 57, 62, 67, and 72 years old, respectively, corresponding to the median for each category.
[b] Available number. No superscript means that the number is completely available (n = 1209,118).
[c] Abdominal obesity: waist circumference ≥ 85 cm for men or ≥ 90 cm for women
[d] Numbers of risk factors: (1) fasting plasma glucose ≥ 100 mg/dl and/or HbA1c ≥ 5.6%, (2) triglyceride ≥ 150 mg/dl and/or high-density lipoprotein cholesterol < 40 mg/dl, (3) systolic blood pressure ≥ 130 mmHg and/or diastolic blood pressure ≥ 85 mmHg, (4) smoking (applicable only for subjects who had at least 1 risk, ranging from 1 to 3)
[e] Serum uric acid, creatinine, and consequently eGFR are currently unavailable but will become available in the future.
[f] Almost all subjects (n = 1205,956) had either fasting plasma glucose or HbA1c measured.
BMI: body mass index; MetS: metabolic syndrome; eGFR: estimated glomerular filtration rate

Table 2. Questionnaire on health status and results for 2008

No.	Questionnaire	Answers	Positive response to ① (%)[a]
	Are you taking following medicines at present?		
1	Medications to reduce blood pressure (1190,117)[b]	① Yes ② No	17.1
2	Insulin injection or medications to reduce blood glucose (1188,900)[b]	① Yes ② No	3.5
3	Medications to reduce your cholesterol level (1187,993)[b]	① Yes ② No	9.4
4	Have you ever been told by the doctor you have a stroke (cerebral hemorrhage, brain infarction, etc.) and received treatment for the disease? (978,542)[b]	① Yes ② No	2.0
5	Have you ever been told by the doctor you have a heart disease (angina pectoris, myocardial infarction, etc.) and received treatment for the disease? (978,491)[b]	① Yes ② No	3.9
6	Have you ever been diagnosed as having a chronic kidney failure and received treatment (dialysis therapy) for the disease? (970,176)[b]	① Yes ② No	0.3
7	Have you ever been diagnosed as anemic? (985,060)[b]	① Yes ② No	10.8
8	Are you a current regular smoker? (1192,091)[b] (Here a current regular smoker is to be a person who has smoked a total of 100 or more cigarettes or smoked for 6 months or longer and has been smoking for the last one month).	① Yes ② No	25.9 in total 37.2 in men 10.8 in women
9	Have you gained over 10 kg from your weight at 20 years old? (976,268)[b]	① Yes ② No	35.5
10	Are you in a habit of doing exercise to sweat lightly for over 30 minutes per session, 2 times weekly, for over a year? (979,191)[b]	① Yes ② No	30.9
11	In your daily life do you walk or do any equivalent amount of physical activity more than one hour per day? (980,581)[b]	① Yes ② No	43.6
12	Is your walking speed faster than the speed of corresponds of your age and sex? (964,407)[b]	① Yes ② No	52.3
13	Have you gained or lost over 3 kg during the last year? (965,421)[b]	① Yes ② No	22.1
14	How fast do you eat compared to others? (972,294)[b]	① Faster ② Normal ③ Slower	29.6
15	Do you eat dinner two hours before bedtime more than 3 times per week? (985,764)[b]	① Yes ② No	28.5
16	Do you eat snacks after supper more than 3 times per week? (958,850)[b]	① Yes ② No	12.8
17	Do you skip breakfast more than 3 times per week? (965,769)[b]	① Yes ② No	14.6
18	How often do you drink alcohol (sake, distilled spirits, beer, liquor, etc.)? (989,349)[b]	① Everyday ② Occasional	29.2

		③ Hardly drink（cannot drink）	
19	How much do you drink a day, in terms of glasses of refined sake? (A glass [180mL] of refined sake is equivalent to a medium bottle [500mL] of beer, 80mL of shochu (alcohol content 35 percent), a glass [double, 60mL] of whiskey, and 2 glasses [240mL] of wine.) (820,231)[b]	② Less than 180 ml (< 23 g ethanol) ③ Over 180–less than 360 ml (23-45 g ethanol) ④ Over 360–less than 540 ml (46-68 g ethanol) ④ Over 540 ml (≥ 69 g ethanol)	56.8
20	Do you feel refreshed after a night's sleep? (973,947)[b]	① Yes ② No	65.4
21	Do you want to improve your life habits of eating and exercising? (951,484)[b]	① I don't mean to start. ② I'm going to start in the future (e.g. within 6 months). ③ I'm going to start soon (e.g., in a month), or I have just started some of them. ⑤ I already started (<6 months ago). ⑥ I already started (>=6 months ago)	28.5
22[c]	Are you willing to undergo a health counseling regarding lifestyle modifications if you get the opportunity?	① Yes ② No	–

[a] Proportions are calculated based on the available numbers for each question.
[b] Available number of subjects
[c] Question 22 is currently unavailable but will become available in the future.

Table 3. Major and minor outcomes

Major conditions or etiologies
· Cardiometabolic diseases including type 2 diabetes, hypertension, dyslipidemia, metabolic syndrome, chronic kidney disease assessed by proteinuria and eGFR
· Obesity and low body weight (malnutrition) assessed by BMI, and central obesity assessed by WC
· Hepatic diseases including fatty liver disease assessed by serum AST, ALT, and GGT
· Abnormal eating habits (breakfast skipping, late-night dinner eating, eating fast, night eating)
· Unhealthy lifestyles (smoking, infrequent exercise, heavy alcohol drinking, non-restorative sleep)
Minor conditions or etiologies
· Hypoglycemia, hyperfiltrations (high eGFR), hypotension, low uric acid
· Extremely low and high BMI (e.g., <15.0 kg/m^2 and > 40.0 kg/m^2)
· Osteoporosis assessed as reduced body-height during 6-years
· Hypertensive and atherosclerotic retinopathies (Keith-Wagener, Scheie classification)
· Physically inactive conditions including reduced walking speed and infrequent exercise

ALT: alanine aminotransferase; AST: aspartate aminotransferase; BMI: body mass index; eGFR: estimated glomerular filtration rate; GGT: gamma-glutamyl transferase; WC: waist circumference

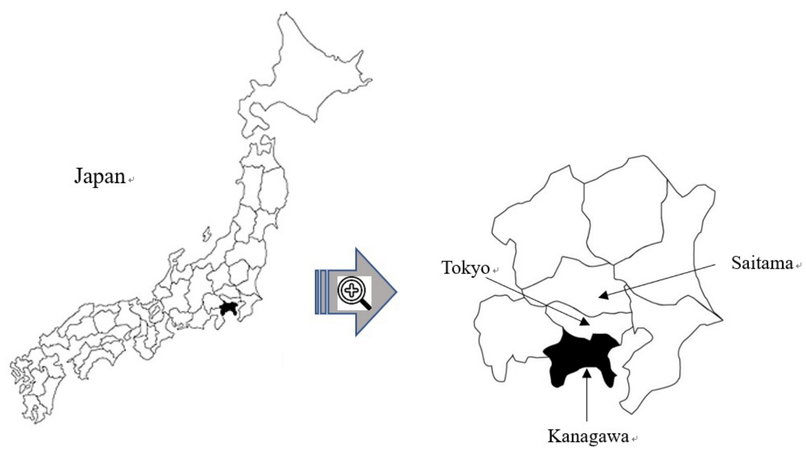

Figure 1

217x141mm (144 x 144 DPI)

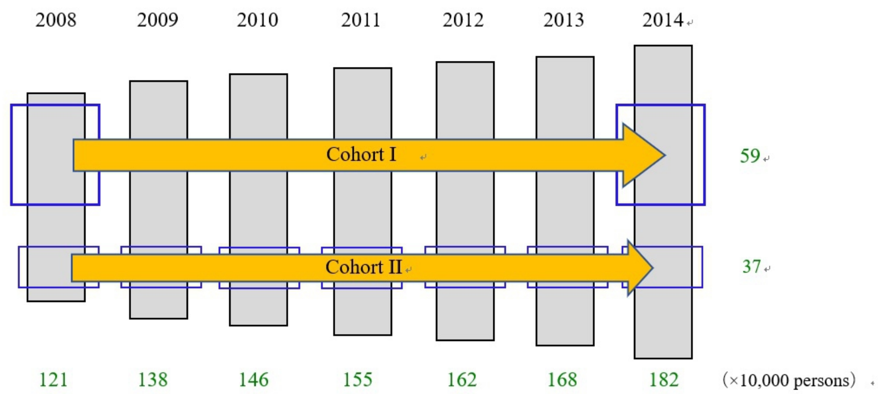

Figure 2

203x114mm (144 x 144 DPI)

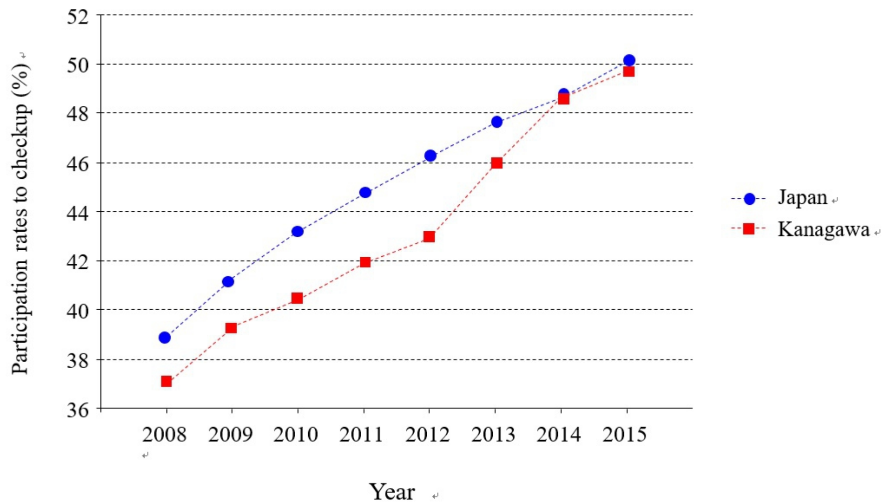

Figure 3

203x122mm (144 x 144 DPI)