Building an Administrative Database of Children[*1]

BESSHO Shun-ichiro

Associate Professor, Faculty of Economics, The University of Tokyo / Former Chief Economist, Policy Research Institute, Ministry of Finance

NOGUCHI Haruko

Professor, Faculty of Political Science and Economics, Waseda University

TANAKA Ryuichi

Professor, Institute of Social Science, The University of Tokyo

USHIJIMA Koichi

Assistant Professor, Division of Policy and Planning Sciences, Faculty of Engineering, Information and Systems, University of Tsukuba.

KAWAMURA Akira

Associate Professor, Faculty of Political Science and Economics, Waseda University

Summary

In empirical analyses in economics over recent years, some studies have been using data collated by administrative departments as part of their work. Administrative data is full data of the target constituents of a government policy that is gathered periodically. Because it is collated for the purpose of performing tasks, it is superior to conventional sample surveys in terms of the accuracy of its figures, its lack of dropouts and nonresponses, and its large number of observations. However, the academic use of administrative data is not a simple matter due to legal restrictions instituted to deal with problems of confidentiality and complicated administrative procedures. This paper will consider the advantages and disadvantages of the academic use of administrative data and will then introduce case studies from within and outside Japan. As an example of administrative database building, the paper will explain the building process and structure of a student database created in Adachi ward, Tokyo, which is used in papers within this special issue. We contributed significantly to the development of the student database for Adachi ward.

Given that there are not thought to be many such case studies in Japan, the significance

[*1] This study has been carried out as part of "An Empirical Research on the Accumulation Mechanism of Children's Human Capital: Lessons from the Challenge of Adachi Ward, Tokyo" (KAKENHI Project No.: 16H03636), supported by a MEXT Grant-in-Aid for Scientific Research (B) (general). Titles are omitted in the text. The authors would like to sincerely thank with ordinary care everyone at Adachi City Hall and Adachi Board of Education for their valuable advice and support in the writing of this paper and the creation of the data being commented upon in the paper. All comments and opinions provided in this paper are the individual opinions of the authors, and do not reflect the opinions of Adachi City Hall, Adachi Board of Education, the Ministry of Finance or the Policy of Research Institute at the Ministry of Finance.

of this paper is that it describes this process in detail.

Keywords: Administrative data, Adachi ward
JEL classification : C81, I20, H20

I. Introduction

It has been argued in Japan and overseas that there is a requirement to carry out objective evidence-based policymaking in order to implement effective and efficient government policy and strengthen accountability to a nation's people and electorate, to which educational policy is no exception. For policies to be based on objective evidence, it is absolutely essential to develop the data that will serve as its basis. In the field of economics, the data traditionally used were aggregate values that had been published in official statistics reports. Economists then began to analyze micro-data from sample surveys (questionnaire surveys) carried out either by companies or the researchers themselves, and also expanded to the use of questionnaire information from official statistics after revisions to the Statistics Act made this possible. Furthermore, there are increasing numbers of studies over the last few decades which have used data gathered or accumulated as part of the executive branch's operations, namely administrative data[1]. Examples of this include tax data relating to income tax, corporation tax and customs duties, national health insurance claims data, and school education Scholastic Aptitude Test (SAT) data. Behind these developments in research naturally lies advances in information processing technologies (represented by improvements in the computing power of computers) and innovations in econometric methodologies to evaluate government policies.

Administrative data is complete (full) data of a subject of government policy that is gathered periodically, either on a national basis if the data belongs to the central government, or within a jurisdiction if the data belongs to a local government. Because the data is accumulated for the purposes of performing a task, the data is superior to conventional sample surveys (including official statistics) in terms of the accuracy of its figures, the low incidence of dropouts and nonresponses, and the number of observations. However, because of legal restrictions instituted to deal with issues relating to confidentiality and privacy, complicated administrative procedures, and the difficulty of merging up data due to it being scattered or inconsistent in format, the academic use of administrative data is no simple matter. However, there are increasingly instances both in Japan and overseas where analysis and research are moving forward ahead with developing databases founded on administrative data having overcome these obstacles.

This paper's objective, as a case study of data development, is to provide commentary on

[1] Tsuda & Okazaki (2018) apply the term "administrative record information." The term "administrative data" is sometimes used for data gathered by private companies in the course of business. Einav and Levin (2014) discuss the application of private companies' business data.

the development process and structure of a student database in Adachi ward, Tokyo, which is used in papers in this special issue. It will also introduce case studies from Japan and overseas relating to the academic use of administrative data with the aim of serving as a reference material for future data use and application and database development.

We contributed significantly to the development of the student database in Adachi ward. As researchers and academic experts, we gradually built deep mutual understanding and trust relationships with the leadership and relevant authorities at Adachi City Hall, persuading them of the need to develop a database for the purposes of analysis. They eventually offered the opportunity to access the data for analysis. Given that there do not seem many such case studies in Japan, the significance of this paper is that it describes this process in detail.

The structure of this paper is as follows. Section II considers the advantages and disadvantages of using administrative data in researches and introduces several overseas and domestic case studies. Section III explains the creation of the database in Adachi ward, while Section IV details the structure of the database in Adachi ward. Finally, Section V concludes.

II. The Academic Use of Administrative Data

The use of administrative data for academic and research purposes is developing within the social sciences. For example, Einav and Levin (2014) shows a history of the rate of papers using non-disclosed data in the *American Economic Review* (a leading economics journal), and finds that the rate of papers using non-disclosed data from the public sector grew from 4% in 2006 to 26% in 2014. This section provides an overview of case studies in which typically non-disclosed administrative data is used for research. It should be noted that the aim is to introduce several case studies but that there is no intention to provide an exhaustive list of all such studies, as that is beyond the capabilities of the authors.

II-1. The Advantages and Disadvantages of Using Administrative Data

Before introducing the case studies, this section considers the advantages and disadvantages of using administrative data for academic purposes (Figlio et al. 2015, 2017, Friedman 2015, Einav and Levin 2014a, b, Jarmin and O'Hara 2016). It is worth keeping in mind the contrast with sample survey micro-data, which have been traditionally used.

Compared with sample surveys carried out by governments and the private sector to create a set of statistics, administrative data is distinct in that it is fundamentally complete (full) data and is gathered periodically due to administrative necessity. These distinct features lend it the following advantages when used in an academic context (with certain exceptions).

Firstly, because the data often has a large number of observations and high-quality measurements, its accuracy in terms of statistical inference is higher. For example, the large-scale Japanese National Survey of Family Income and Expenditure has a standard sampling

rate of 1/791.4 (2014 survey). A high number of observations enables more accurate non-parametric statistics. Moreover, it is believed that problems which have surfaced in sample surveys over the last few years such as issues with dropouts, nonresponses and excessive/insufficient answers are less likely to be found in administrative data. The use of administrative data is particularly indispensable when "problematic" samples (which have a high tendency to drop out in social surveys) are important to the analysis.

Secondly, the large number of observations in complete surveys enables analyses focusing on specific sections and statistical analyses focusing on heterogeneity of effect. Analysis of income share of high earners such as that found in Piketty and Saez (2003) is a typical example, as is analysis of twins and triplets, people with disabilities and criminals. Moreover, it also enables analyses of the differences in summary values in small units. Analyses of regional differences in dispersions and average values for academic performances, medical expenses, and productivity can be given as examples.

Thirdly, the fact that the data is gathered periodically makes it possible to build forward-facing data even when events that could be natural experiments or quasi experiments have occurred, so one can expect to ensure a sample size that can withstand statistical analysis even if the treatment group is small. Conversely, because it is rare for events that could be natural experiments to occur, it is not easy in sample surveys to survey treatment groups "in advance."

The advantages noted here can apply when some types of administrative data are used on their own, although even more enriching research can be anticipated if it is possible to merge the administrative data together with other data. Because the data is complete it should fundamentally be possible to be combined as long as there is personal identification information, and it is also possible to join together variables not included in typical sample surveys. If it is possible to connect field data with administrative data which can then be tracked afterwards, then it should be possible to measure long-term effects. If it were possible to specify familial relationships, then cross-generational analyses could take place. A good example of this would be the Equality of Opportunity Project (Moving to Opportunity Experiment, Opportunity Insights), which was carried out by a research group led by Raj Chetty in cooperation with the American Internal Revenue Service (IRS).

Of course, various flaws exist on the flip side to these advantages. Firstly, because administrative data is gathered to implement and manage policies it may be unsuitable for research purposes or may not be measured or constructed in a format suited to research purposes. For example, administrative data relating to education (even data relating to SAT scores) often will not include the sort of data like the big five personality traits (extraversion, agreeableness, conscientiousness, neuroticism and openness to experience). Or, for example, the variable of "income" might have been processed into some kind of post-deducted value. Moreover, people who are not the target of a government policy will not be recorded in the relevant administrative data. For example, people who have moved abroad are not tracked.

Secondly, there is an issue with data accessibility. When Scandinavian countries such as Denmark, Sweden, Norway and Finland implemented social security systems from the

1960s onwards, they introduced nationwide personal identification numbers to understand annual income, making it possible to merge data within an administrative agency if that number is used. However, there are many other countries that either do not have nationwide personal identification numbers, or that institute legal restrictions concerning access to data or usage of such numbers due to issues of confidentiality or privacy. Administrative data is anonymized when used for academic purposes, but its usage often involves various procedures and expenses. If the data's format lacks uniformity or has not been digitized it can be difficult to use because of human or technological limitations in the bureaucratic organization managing that data, or because of complicated administrative procedures.

Thirdly, there are instances when the data infrastructure will be undeveloped because of poor management, such as if there are no explanatory notes (metadata) concerning the details of the data, or when these are insufficient or unpublished. Because administrative data is fundamentally designed for the purposes of implementing administrative policies at the point of data creation, the notes provided for the variables will not necessarily be adjusted even if their definition changes due to system amendments or otherwise. When sample surveys are arranged in a set up designed for secondary usage, they often include detailed explanatory notes for variable definitions, so in this regard administrative data arguably has greater issues in terms of quality.

In short, there are both advantages and disadvantages to administrative data. As such, perhaps it should be used to complement sample surveys which are designed for specific research purposes. The following section will consider examples of studies that used administrative data.

II-2. Overseas Case Studies

As previously stated, there is an increasing number of studies in the field of economics that have used administrative data (Figlio et al. 2015, 2017, Friedman 2015, Einav and Levin 2014a, b, Jarmin and O'Hara 2016, Almunia et al. 2019). Figlio et al. (2015) provides a list of relevant studies that have been published in leading academic journals primarily concerning the economics of education. Mitsubishi Research Institute (2016) considers the systems for academic usage of administrative data related to education in place in England, Australia, the USA and Sweden, while Tanaka (2019) comments on the same in England and Finland.[2] These countries create databases of the national SAT results of individual children and link it to school information before providing it for academic usage. These two papers will now be used as a basis to consider what happens in England, which is one of the most progressive cases.

In England, the Department for Education maintains a National Pupil Database (NPD) which combines various data relating to the country's children and students. If external re-

[2] For other papers looking at the mechanisms of academic usage of administrative data, see Uchiyama et al. (2018) and Tsuda & Okazaki (2018). Mervis (2014) explaines the process by which Chetty and Saez accessed IRS tax data. Moreover, see Toyofuku (2017, 2018) for examples of where administrative data has been aggregated and turned into open data.

searchers at schools, local governments and other government bodies satisfy certain criteria through the necessary application procedure (training lectures etc.), then they are permitted to access individual data within the NPD for secondary use and analysis.

The NPD has been running since 1999 and aims to improve the quality of school education within England and ensure accountability in school education. Using the NPD saves on the costs of gathering new data and experiments. The database stores tests results from the National Test and the General Certification of Secondary Education (GCSE); basic characteristics of students including sex, age and ethnicity; other characteristics of students including whether they receive free school meals, whether they require extra support in school and their residential postcode; and details about the schools gathered by local authorities from schools in their jurisdiction such as their size and number of pupils. Household information is combined with summary values for each region by postcode units.

Because the Department for Education uses the data published in the NPD as a basis to determine funding for local governments and schools, local governments and schools are both obliged and incentivized to provide the data. In order to reduce costs for the schools and local governments inputting the data during the data-gathering process, a redesign is moving ahead with the cooperation of the Government Digital Service to standardize the interface among other changes.

Academic use of the NPD has been pushed since the early 2000s; initially, infrastructure was developed for the purposes of secondary use of the data by a group of around 10 researchers. Among the initial members were researchers with strong personal ties to statisticians then working for the Department for Education, and this network was applied to infrastructure development. The group of researchers has expanded to currently encompass to at least 100 members. It is necessary to apply for access to the NPD for research purposes and applicants are required to attend a 1-2 day lecture before making the application and must also pass an information security test. Moreover, the group of researchers who developed the infrastructure for secondary usage of the NPD work to promote the use of the NPD for research purposes, such as by hosting intensive courses at universities to promote its use.

In the UK, a non-profit organization called the Fischer Family Trust has an education database of schools and local governments gathered by the organization's FFT Datalab. In the initial stages of this database's creation there were also researchers with strong personal ties to local administrators who galvanized local government data provision.

II-3. Japanese Case Studies

Japan has the mechanisms in place to allow the use of student microdata from the National Academic Achievement and Learning Conditions Study for academic purposes. In addition, several local governments are independently developing databases using education-related administrative data.

In Japan, the practical application of local administrative data has major hidden potential, particularly when one considers the search for objective evidence for education policy relat-

ing to elementary and junior high schools. The reasons are as follows. Firstly, although the data potentially lacks representativity because it is not nationwide data, administrative data is essentially full data within a specific region and so in comparison to sample surveys (such as questionnaires) will have much higher collection and response rates, with few sample selection problems.

Secondly, local governments in Japan (and municipalities in particular) are bodies that execute government policy across a broad range of fields and are responsible for implementing nearly all redistribution policies with the exception of pensions (Hayashi 2008, p.64). As Adachi ward emphasized in its policy planning (discussed later), education policy regarding elementary and junior high school students is deeply connected to other policy areas, in particular healthcare and welfare policies. Because local governments possess data on the relevant policies, it should be possible to combine these sets of data and conduct a multifaceted analysis. For example, data from the academic achievement survey could be combined with information relating to economic status (schooling attendance support etc.), information relating to health (health checkups etc.), and information related to preschool education.

Thirdly, local governments in Japan operate under a dual representation system of both a chief executive and an assembly. As far as the institutional allocation of authority goes, the chief executive has greater power than the assembly, and the assembly's control over the bureaucracy is fairly limited (Soga & Machidori 2007, pp.45-48). On this basis, it would be possible to combine and develop data under the chief executive's leadership. This sort of development would be harder among central ministries and agencies, which tend to be vertically divided.

Fourthly, while there are various regulations and guidance from the central government regarding education and welfare policies, local governments do have the authority to implement original policies as local independent projects. Independent projects can be seen as test sites for policies that are not being developed nationwide. Moreover, because it may very well be possible to create experiment or quasi-experiment conditions depending on the policy's design, if data development was incorporated prior to the policy's implementation then a highly accurate evaluation of the policy would be possible.

Mitsubishi Research Institute (2018) examines case studies and challenges of leading initiatives through literature surveys, interview surveys, questionnaire surveys and debates. Based on this report and subsequent developments, we will now provide a simple summary of domestic case studies other than the Adachi ward example used in this paper.

II-3-1. Saitama Prefecture

Saitama Prefecture has been carrying out the Saitama Prefectural Academic Ability Survey since AY2015 on students attending prefectural public elementary and junior high schools (with the exception of Saitama City) between their fourth year of elementary school and third year of junior high school (the ninth graders) in order to gather information on the academic ability and learning of students and to improve education policy and guidance. The survey was carried out on 708 elementary schools and 360 junior high schools in

AY2015; there were approximately 150,000 examinees in both the elementary and junior high schools, so approximately 300,000 in total. The survey was distinctive in that it used a survey method based on Item Response Theory (IRT), making it a survey in which it is possible to understand on a continuous basis changes in academic performance throughout time. According to Saitama Prefecture records, this was the first local government survey that used IRT. Moreover, in addition to a curriculum survey (academic achievement test), the local government gave questionnaire surveys to students to survey their lifestyles and awareness of their studies, and also conducted questionnaire surveys on school and municipal boards of education.

Individual data from this survey is provided to research institutes specializing in education with the aim of encouraging better guidance to improve academic ability. For example, the Keio Research Institute at SFC was made a trustee in AY2016; Keio University Associate Professor Makiko Nakamuro and Shizuoka University Associate Professor Hiroyuki Masukawa are the principal researchers working on analysis.

Saitama Prefecture is also launching a consortium comprised of local governments and businesses with an interest in the survey. To date, the consortium has been attended by representatives from the following local governments: Toda City in Saitama Prefecture, Hanyū City in Saitama Prefecture, Yamanashi Prefecture, Fukushima Prefecture, Koriyama City in Fukushima Prefecture, Nishiaizu Township in Fukushima Prefecture, Kyoto Prefecture, Hiroshima Prefecture, Fukuyama City in Hiroshima Prefecture, Unnan City in Shimane Prefecture, Kagawa Prefecture, Kochi Prefecture and Saga Prefecture.

II-3-2. Toda City, Saitama Prefecture

Toda City is situated in the south of Saitama Prefecture, connected to Saitama City (an ordinance-designated city and the seat of prefectural government) in the north and Tokyo's Itabashi ward and Kita ward in the south. Superintendent of Education Tsutomu Togasaki has been pushing forward with various reforms since assuming his post in 2015, including an educational policy trial that takes objective evidence into special consideration. He has also been promoting cooperation with industry actors and universities that are aiming to cultivate abilities that cannot be replaced by AI or that can utilize AI. Toda City is also conducting a questionnaire survey on all its educators along with the above-mentioned Saitama Prefectural Academic Ability Survey.

Joint studies with external parties have been promoted as one of the ways to carry out education policymaking with an emphasis on objective evidence, and Toda City provides school and classroom data to businesses and researches as resources for empirical analyses. Cooperation with multiple businesses and researchers in the joint studies has encouraged competition between external stakeholders. For example, a joint study between Keio University, the University of Tsukuba, the National Institute of Informatics (NII), Benesse, LITALICO and IGS took place in AY2019. Moreover, EdTech is also being promoted in cooperation with LoiLo, Google, Benesse and Fuji Electric IT Solutions Co., Ltd. To date, approximately 70 universities, NPOs and businesses have engaged in cooperation.

II-3-3. Minoh City, Osaka Prefecture

Minoh City is situated in the north of Osaka Prefecture. Through it runs the Hankyū Minoh Line, and the city has developed as a dormitory suburb in the Osaka Metropolitan Area (an extension has been approved for the Namboku Line of the Kita-Osaka Kyuko Railway). Tetsurō Kurata was appointed at 34 years old as Japan's then youngest mayor in 2008, and pledged Minoh would have "the easiest childcare in Japan" as part of his priority plan.

Minoh City has been developing a database known as the Child Development and Protection System since 2014. This system integrates data sets belonging to the various bureaus (which had previously been scattered around) at Minoh City Hall relating to children aged between 0-18 residing within Minoh City. This data includes the results of the General Survey on Minoh Academic Ability, Physical Ability and Lifestyle Conditions (Minoh Children Step-up Survey) which is independently carried out by Minoh City, as well as data relating to public assistance, childcare allowance, after-school care for children, medical support, abuse counseling, education counseling, and payments and state of use of learning support among others, all of which are periodically updated. The Minoh Children Step-up Survey is aimed at children between their first year of elementary school and third year of junior high school and comprises a survey of their academic and physical abilities and a questionnaire survey concerning their study habits, exercise habits, lifestyles and self-awareness. However, the survey targets public elementary and junior high schools so does not include data on the academic abilities and so on of children attending private schools.

Minoh City has amended bylaws to permit the provision of the system for use outside of its intended purposes and by external actors, although such use is limited to the purposes of safeguarding mental and physical soundness and maintaining quality of life. One of the outcomes of such use is from the Nippon Foundation (2018), which uses the system's data over a three-year period between 2014-2016.

II-3-4. Amagasaki City, Hyōgo Prefecture

Amagasaki City is a core city situated in the southeast of Hyōgo Prefecture, and is connected to Osaka City in the east. Kazumi Inamura was appointed as the then youngest female mayor at the age of 38 in 2010. As part of the Inamura administration's second-term campaign platform of "Amagasaki Next 30," it pledged to enhance the functions of childcare support centers and pushed forward with relevant data gathering. The Amagasaki City Manabitosodachi Research Institute was established in 2017, to which Osaka University Professor Fumio Ohtake was appointed as director. The institute has five chief researchers: Keio University Professor Makiko Nakamuro, Kobe University Associate Professor Sachiko Kitano, Kanagawa Institute of Technology Professor Tomoo Okada, Kansai University of International Studies Professor Shigeki Nakao, and Kobe University Assistant Professor Masahiro Nishiyama. The chief researchers are engaged in research that uses a range of individual data provided by Amagasaki City relating to children and hosted their first report meeting in 2018.

In 2017, Amagasaki City carried out a Survey on the Actual Conditions of Children's

Lives in Amagasaki City on public school children in their fifth year of elementary school and their guardians as well as public school children in their second year of junior high school and their guardians, in order to survey the actual conditions of students' lives and feelings about the world around them. A system has been put in place to provide the survey's micro-data to researchers who wish to use it. Moreover, the city has been independently carrying out academic achievement tests on children between their first year of elementary school and their second year of junior high school since 2018.

As the above case studies show, several local governments are developing micro-data on children within their jurisdictions, which they are turning into databases and providing to researchers. The local government's view is that they have requested or commissioned an analysis, while the researcher's view is that they are recipients of data provision. However, various conditions are placed on the data provision. The following section will explain the data structure in Adachi ward used in this study.

III. The Data Building Process

The Noguchi et al. (2020) included in this special issue and Bessho et al. (2019) are outcomes of the project. This chapter will begin with a comment on ward administration in Adachi ward in recent years in order to describe the problem awareness and background to this paper.

III-1. Ward Administration and Policy Evaluation

Adachi ward is situated in the northeast of Tokyo and is one of the city's 23 special wards; it includes the land on which once prospered the Senju-shuku, one of the four post stations of Edo. One of its urban centers is the area around Kita-Senju Station, where passengers can access JR East, the Tokyo Metro, the Tobu Railway and the Tsukuba Express. The Senju area is situated on the southern side of the Arakawa River, but most of the ward is found north of the river. Its population was 685,000 at the beginning of 2018.

The ward established the "New Basic Concept Formulation Adachi Ward Residents' Committee" in November 2002 through a public appeal, and its findings were summarized as investigation findings the following year in March 2003. The "Adachi Ward Basic Concept Commission" was then established in July 2003 in response to these findings and delivered a report in March 2004. Based on this report, a basic concept for Adachi ward was formulated with its fundamental principle being "Realizing a Powerful Adachi Ward Built through Cooperation," which was approved by the Adachi ward Assembly in October 2004. This basic concept was revisited after 10 years, and in July 2015 a new Adachi Ward Basic Concept Commission was established. A new basic concept was formulated in 2016 based on the commission's report, which proclaimed "Create through Collaboration / Full of Vitality / Continue to Evolve / People and Town / Adachi".

Following the formulation of the 2004 basic concept, Yayoi Kondo was elected as mayor in June 2007. Under Mayor Kondo's leadership, an "Adachi Ward Priority Project Realization Strategy" was formulated in AY2009 to resolve important and pressing issues facing the ward. The issues to be prioritized as part of the strategy were categorized into four areas: children, living standards (*kurashi*), community development (*machi-zukuri*) and management reform. The priority projects were particular targets for evaluation by the Residents' Evaluation Committee (which will be discussed in detail later). In the drive to push ahead with these priority projects, Adachi ward laid out law and order, academic ability, health and the poverty cycle as four bottleneck issues; that is, fundamental issues that needed to be overcome to obtain fair evaluations from inside and outside the ward (Akiu 2019). In terms of law and order, the ward developed a "Beautiful Windows Movement" with the slogan "A Beautiful Town is a Safe Town" in order to reduce the number of recognized offences. In terms of academic ability, various trials aimed at stabilizing basic academic ability have been implemented to tackle problems concerning the results of elementary and junior high school achievement tests. Bessho et al. (2019) evaluates one of these trials. In terms of health, policies specific to diabetes countermeasures are being pushed forward in light of the fact that average life expectancy in Adachi ward is two years shorter than the Tokyo average. The poverty cycle is considered to be a common cause at the foundations of the three other issues of law and order, academic ability and health, leading Adachi ward to invest in policy resources to support children. Improving the eating habits of children through Adachi ward's increasingly well-known "delicious school meals" can also be positioned as part of that, such as in the issue of "Tokyo's Adachi Ward School Lunchroom: Balanced Meal of 12 Nutrients You'll Want to Eat Everyday" (Adachi ward's "Delicious School Meal" Production Committee 2011). During this process, the enactment of the Act Regarding the Promotion of Child Poverty Countermeasures (the Child Poverty Act) in January 2014 served as a catalyst for the formulation of the "Connecting Adachi to the Future Project (a plan to enact child poverty countermeasures in Adachi ward)" in AY2015. On the basis of this project, early-intervention countermeasure trials were carried out for each life stage from childbirth onwards.

The 2004 basic concept emphasized "cooperation," and stipulated that making cooperation a reality would "require cooperation to be institutionally guaranteed through various means including information disclosure and administrative evaluations." For this reason, in AY2005 a guideline-base residents' evaluation examined projects implemented in AY2004. From the following year in AY2006, the Residents' Evaluation Committee became a mandated institution. That is, Article 15 of the Adachi Ward Self-Government Basic Ordinances enacted in AY2004 in accordance with the basic concept made administrative evaluations compulsory, and was followed in March 2006 by the enactment of the Adachi Ward Residents' Evaluation Committee bylaw. On the basis of this bylaw, residents' evaluations are continuously carried out in Adachi ward. The residents' evaluations are a mechanism to evaluate policies from the perspectives of normal residents and academic experts rather than the perspectives of the ward office, administration carrying out projects or the ward assem-

bly involved in the planning of projects, therefore ameliorating the objectivity of the evaluation.

The Residents' Evaluation Committee comprises several subcommittees, although its composition changes. The inaugural committee to meet as a mandated institution in AY2006 targeted all 114 policies designated as basic plans for evaluation, instituting the following four subcommittees: The Community Development and Security Subcommittee, the Health and Welfare Subcommittee, the Education and Industry Subcommittee, and the Environment, Residents' Living Standards and Structure-building Subcommittee. Following the formulation of the Adachi Ward Priority Project Realization Strategy, the AY2010 committee was instituted as the three subcommittees of Children, Living Standards, and Community Development and Administrative Reform in response to the fact that the priority projects were summarized into the four areas of children, living standards, community development, and administrative reform. In AY2012, a General Works and Projects Review Subcommittee was instituted to evaluate work and projects not included in the priority projects, and the committee grew to incorporate four subcommittees. Later, the commencement of new basic plans in 2017 prompted a reorganization of the subcommittees tasked with evaluating the priority projects into People, Living Standards and Administrative/Financial Reform, and Town and Administrative/Financial Reform, meaning the committee maintained a four-subcommittee system.

III-2. Researchers' Contributions

It is common to invite academic experts to form parts of committees not only in Adachi ward but also to advisory bodies established by central ministries and agencies or to assemblies gathered for the purpose of administrative management. The Adachi Ward Residents' Evaluation Committee is composed of 17 or fewer people who are appointed by the mayor, and the case is the same here. The chairman of the first committee was (then) Tokai University Associate Professor Kazuyasu Kawasaki, and it has continued to be customary for the chairperson and subcommittee chairs to be academic experts. It is worth noting that given that the Residents' Evaluation Committee's ultimate mission is to evaluate policies from a resident's perspective, having academic experts on the committee will help to evaluate ward administration from a more objective and neutral position.

Of this paper's authors, Noguchi, Tanaka and Bessho worked with Adachi ward as academic expert members on the Residents' Evaluation Committee. Noguchi has been a committee member from the first AY2005 committee, becoming vice-chair in 2009 and then succeeding Tokyo Future University Professor Yasuyuki Deguchi to become chair of the committee in AY2014-2015. Tanaka was chair of the Children subcommittee from AY2012-2015, and committee chair from AY2016-2018, while Bessho was chair of the General Works and Projects Review Subcommittee from AY2012-2016.

It was through their work as members of the Residents' Evaluation Committee that the authors learned of the existence of the different data in the ward's possession and considered

how it could be used to improve evaluation methods. In order to explain why, we will comment on the evaluation methods being used by the Residents' Evaluation Committee in AY2012-2016 when Noguchi, Tanaka and Bessho were committee members together.

While there are slight differences depending on the subcommittee, evaluations begin with readings of preliminary records, followed by a hearing and a post-hearing investigation. Projects are first selected for evaluation by the secretariat, and committee members are given project evaluation records that have passed internal evaluation. These project evaluation records are presented in a common format and include the project purpose and outline, legal basis, activity indicators and performance indicators, invested resources, performance analysis and task analysis. Activity indicators are indicators that express the extent of activity taking place in the ward; for example, the number of times events or lectures have been hosted is used. Performance indicators are indicators that are more directly connected than activity indicators to policy objectives, and include things like satisfaction levels measured through questionnaires and school meal leftover rates. Target values are set in advance for activity and performance indicators, and achievement rates compared with actual values are also recorded. Invested resources are primarily project and labor costs, and include both the project costs paid for by the ward as well as state and city subsidies and beneficiary contributions including participation fees. Performance and task analysis mean the results of the relevant department based on activity and performance indicator achievement rates, and incorporate achievement status and its contributing factors, along with other results not included in the indicators, future challenges and changes in the environment.

After receiving both the project evaluation records and prior documentation, evaluation committee members deliberate over questions for the relevant department on the basis of these materials before the hearing takes place. The questions asked will primarily be to confirm definitions of project terms and indicators or to ask for supplementary documentation. The committee's questions will be sent to the relevant department. In the hearing, the relevant department will provide answers to the questions asked, and then respond to any further questions.

After the hearing is over, the evaluation committee members write up an evaluation sheet to be published along with the results of the internal evaluation. Residents' evaluations comprise a rating out of five and comments. When the project is a priority project, a rating out of five is given for four categories: an evaluation of reaction to past evaluations, an evaluation of the objectives and outcomes, an evaluation of future direction, and an overall evaluation.

Data kept by the ward office is aggregated and published as part of the residents' evaluation procedure. For example, the remedial education classes (basic study classes for elementary school students and supplementary lessons for junior high school students) taken by elementary and junior high school students that are examined by Bessho et al. (2019) were the subject of a residents' evaluation as one of the priority project initiatives. The activity indicator for this project was the total number of supplementary class hours, and the performance indicators were the proportion of students achieving an correct answer rate of less

than 30% in achievement tests carried out by the ward (Adachi Comprehensive Survey of Acquisition of Basic Academic Skills) and the results of achievement tests taken by the target children directly before and after supplementary lessons.

The fact that it is normal residents who join as committee members through public appeals and the need for quick reporting for each year mean that there is nothing for it but to use simply calculated indicators in the residents' evaluation. However, economists have been arguing in recent years that it is often impossible to appropriately and quantitatively evaluate policies using simple indicators. From this perspective, Noguchi and other researchers frequently suggested to the relevant authorities that it might be possible to carry out a quantitative administrative evaluation using the micro-data that exists behind the publicly available aggregated data. Noguchi and Tanaka both served as chair of the Children subcommittee and, in recognition of how law and order, academic ability, health and the poverty cycle were bottleneck issues in Adachi ward (as has been previously discussed), came to the opinion that it was necessary to fully understand how children were being raised in this ward. This led them to concentrate their appeals toward the use of administrative data regarding children in particular.

Having listened to repeated proposals and appeals from researchers, Adachi ward began proper negotiations in AY2016 and started to develop the data from the end of 2016. The data structure will be explained in the next section, but it is worth quickly noting the framework of data usage first.

III-3. Framework of Data Usage

The Noguchi et al. (2020) in this special issue and Bessho et al. (2019) are the results of analyses using data provided by Adachi ward for an empirical study on the determining factors of children's upbringings. Data was provided in accordance with basic agreements between Adachi ward and the respective institutions to which the authors are attached. The basic agreements stipulate the period of the agreement, expense allocation, compensation for damages, confidentiality obligations and so on. The consent of the Adachi Ward Information Disclosure and Privacy Protection Commission was required before the data was provided, and personal information was excluded from the provided data to anonymize it.

The provided data was anonymized within Adachi City Hall. However, up until the study began the sets of data were often in varying forms and file formats, and in most cases different for each year, and kept in the custodianship of different responsible sections or persons. Moreover, the students were not assigned a common personal identification number across the data sets. Therefore, it was necessary before the data was provided to combine the data or make them ready to combine.

In this study therefore, the above-mentioned data development and anonymization process were carried out by the researchers ourselves through personnel secondment from their research institutions.

IV. Data Structure

The data used in this study is full data of children and students attending public elementary and junior high schools in Adachi ward. The Ward's Board of Education possesses a register of school-aged children and students residing within the ward, and this was merged with micro-data from different surveys related to children, various initiatives carried out independently by the ward and information on schools and educators to create a database. The current sampling period is AY2009-2017, but there are some parts where the information has not been fully combined.

The database incorporates five surveys regarding children: Adachi Comprehensive Survey of Acquisition of Basic Academic Skills, a survey carried out by Adachi ward on children and students' physical ability, the National Academic Achievement and Learning Conditions Study, a survey of long-term absences lists and a QU questionnaire survey (a school life questionnaire survey). Of these, the surveys most commonly used in our study were Adachi Comprehensive Survey of Acquisition of Basic Academic Skills and the survey carried out by Adachi ward on children and students' physical ability, but because these surveys only target children and students attending public elementary and junior high schools within the ward, there is no substantial data on students attending private schools in the database, even when they are residents of Adachi ward. Moreover, Adachi Comprehensive Survey of Acquisition of Basic Academic Skills only targets students between their second year of elementary school and third year of junior high school, so it does not include data on first-year elementary school students. Given that this survey is carried out every year in April, one could also argue that it does not include data on third-year junior high schools either if one interprets it as measuring the results of the previous academic year. In this sense, data for sixth-year elementary school students is missing for pupils who move into private schools for junior high school. There are over 5,000 children per year in each academic year of the elementary schools, resulting in a total number of observations of 412,462 in the database.

A trial is in place to combine nine sets of information on the ward's various initiatives and on educators and schools. These are: (1) applications for and recipient status of school attendance support, (2) a register of children enrolled in pre-school kindergarten and nursery schools, (3) the status of school choice, (4) a register of target children for multilayer instruction model (MIM), (5) a register of elementary school children attending remedial education classes, (6) a register of pupils attending junior high summer study camps, (7) educator information, (8) school information and (9) information on temporary class closures.

This information is merged and anonymized by removing personal information such as full names, and then provided to the researchers. Only data in the possession of the ward office was used to gather this information, and absolutely no requests were made for cooperation in gathering data from the schools themselves.

The data that we currently use has not been perfectly merged from the sets of data previously mentioned. For various reasons, it is missing in parts. Several reasons for this are provided below.

Firstly, the database is based on the register of school-aged children, but because this register is originally composed of repeated cross-section data, there are instances where the same child or student cannot be traced. Although a student repeating a year is a very rare occurrence in a public elementary or junior high school, there are several instances where a change in school, family name or how a child's name is transcribed has made it impossible to connect children and students across the academic years.

Secondly, while it was possible to use an electronic register of school-aged children for the period AY2013-AY2017 this was not possible for the period AY2009-AY2012; hence, data from Adachi Comprehensive Survey of Acquisition of Basic Academic Skills was used instead. Most students attending public elementary and junior high schools in the ward responded to this survey, but there are naturally some students missing when compared to the original register of school-aged children. There are several reasons for this: one is that there will be students in special needs classes or children absent due to sickness and so on who did not respond to the survey. Moreover, name data in the general survey were the names recorded by the students in their answer sheets, so there will be instances where the name recorded in the original register of school-aged children differs to the answer given, either because they cannot write kanji due to being in the lower years of elementary school, or because they used different kanji (such as writing a different kanji set for the name "Watanabe"), or because it does not reflect their change in name. Such orthographical variation in name transcription and changes in school were connected when visible to the eye, but of course have not been completely resolved.

Thirdly, the incompleteness of the connections caused by the above-mentioned merge based on the name written down by the student also happened frequently when the data was combined with other data sets. Moreover, there was some data for which name data was not available.

Fourthly, there was an issue with educator data. In the current system, authority over personnel matters regarding educators in the ward's public elementary and junior high schools belongs to Tokyo Metropolitan Government. As a result, Adachi ward does have information on educators working in the ward's elementary and junior high schools, but no information on their career histories outside of the ward. That is to say, for example, that if an educator transferred into the ward from outside, the ward would not have any information on the previous schools they worked at or the number of years they have been employed by the city.

Moreover, it cannot be ruled out that there may be some man-made errors in the data-combining process.

V. Conclusion: Future Challenges

From the standpoints of implementing effective and efficient policies and strengthening accountability to a nation's people and electorate, there is a need to promote policy formation based on objective evidence, including in education policy. The development of data

that can provide the evidence is indispensable for this purpose. This study introduces case studies that have used administrative data as the source, and also discusses the process of building the student database in Adachi ward (as used in papers in this special issue) as well as the structure of that data. Administrative data is complete (full) data of a target of government policy that is gathered periodically, either on a national basis if the data belongs to the central government, or within a jurisdiction if the data belongs to a local government. It is superior to sample surveys in terms of the accuracy of its figures, its lack of dropouts and nonresponses, and its number of observations. However, for various reasons the academic use of administrative data is not a simple matter. There are an increasing number of case studies using administrative data both in Japan and overseas. In Adachi ward, Tokyo, researchers have been developing a student database to merge different sets of data.

There are some future challenges in Japan in terms of further pushing ahead with the academic use of administrative data. The USA enacted the Evidence-Based Policymaking Commission Act of 2016 in March the same year, on the basis of which a bipartisan Commission on Evidence-based Policymaking was established. Tsuda & Okazaki (2018) discusses findings in the commission's final report entitled *The Promise of Evidence-Based Policymaking*, to suggest that inconsistent laws, data management and complicated administrative procedures are primary factors preventing appropriate access to data. These problems could be seen to apply to Japan as well.

In Adachi ward, where we contributed to database building, an ICT, data application and information system management promotion system is being constructed (Akiu 2019). In this system, the deputy mayor is serving as CDO (chief data officer) and CIO (chief information officer) and an ICT strategy promotion section has been established within the policy administration department. There is someone in the section who is in charge of data coordination, and part-time experts are assisting the CDO in promoting evidence-based policymaking. There is also an information systems section in the policy administration department that is tasked with systems operation and management, with part-time experts assisting the CIO who are tasked with systems operation and management. This sort of institutional development of data infrastructure could be essential from the administration's perspective too.

One of the destination points for this type of data infrastructure development could be mechanisms allowing the secondary usage and analysis of anonymized individual data by external researchers, as can be seen in the UK's example. Systems in which the administration publicly appeals for research ideas and allows researchers who have passed inspection to use data either at onsite facilities or through remote access can already be found in the UK and Finland, where researchers are studying tax records (Almunia et al. 2019).

Mitsubishi Research Institute (2016) gives the results of a survey of local governments concerning the current state of evidence-based education policy and challenges. It shows the percentage of local governments that are able to provide data that includes personal information to external researchers (i.e. researchers not attached to that local authority) is just 6.4% in the prefectures, 5.0% in designated cities, and 7.4% in core cities. In terms of challenges hindering the advance of evidence-based education policy, more than half of all pre-

fectures, designated cities and core cities respectively stated that they lacked the human resources with the expertise for evidence gathering and analysis. Each of the Japanese case studies discussed in this paper moved forward through collaboration with external experts, and even the example of the NPD of the UK (one of the most advanced case studies worldwide) shows how personal connections between researchers and administrators played a major role in the initial phase of the formation process (Tanaka 2019). In Japan, some critics say that researchers do not understand the administration's logic (Suzuki 2018). While the objectives of the administration and researchers may differ in the short-term, if one takes a long-term view then surely their objectives should be more aligned given that they both want to move toward better policies and a better society (Almunia et al. 2019). Cooperation between government and academia will be a major challenge for the future in order to develop data that can serve as the foundation for objective evidence-based policymaking.

References

Almunia, M., Harju, J., Kotakorpi, K., Tukiainen, J., & Verho, J. (2019), "Expanding access to administrative data: the case of tax authorities in Finland and the UK", *International Tax and Public Finance*, vol.26, no.3, pp.661–676.

Bessho, Shun-ichiro, Haruko Noguchi, Akira Kawamura, Ryuichi Tanaka, Koichi Ushijima. (2019), "Evaluating remedial education in elementary schools: Administrative data from a municipality in Japan". *Japan and the World Economy,* vol. 50, pp.36-46.

Card, David E., Raj Chetty, Martin S. Feldstein, and Emmanuel Saez. (2010), "Expanding Access to Administrative Data for Research in the United States", *National Science Foundation* White Paper.

Chetty, Raj, Nathaniel Hendren, Patrick Kline, and Emmanuel Saez. (2014), "Where is the land of opportunity? The geography of intergenerational mobility in the United States", *Quarterly Journal of Economics* vol.129, no.4, pp.1553-1623.

Crato, Nuno, and Paolo Paruolo. (2019), "The Power of Microdata: An Introduction", In: Crato, Nuno, and Paolo Paruolo (eds), *Data-Driven Policy Impact Evaluation,* Springer Open, pp.1-14.

Einav, Liran, and Jonathan Levin. (2014a), "Economics in the Age of Big Data", *Science,* vol.346, no.6210, 1243089.

Einav, Liran, and Jonathan Levin. (2014b), "The Data Revolution and Economic Analysis", *Innovation Policy and the Economy* vol.14, pp.1-24.

Figlio, D., Karbownik, K., & Salvanes, K. G. (2016). Education Research and Administrative Data", *Handbook of the Economics of Education*, vol.5, pp.75-138.

Figlio, D., Karbownik, K., & Salvanes, K. (2017), "The Promise of Administrative Data in Education Research", *Education Finance and Policy*, vol.12, no.2, pp.129-136.

Jarmin, Ron S., and Amy B. O'Hara. (2016), "Big data and the transformation of public policy analysis", *Journal of Policy Analysis and Management* vol.35, no.3, pp.715-21.

Kum, Hye-Chung, Stanley Ahalt, and Thomas M Carsey. (2011), Dealing with Data: Gov-

ernments Records", *Science*, vol.332, no.6035, pp.1263.

Larrimore, J., Mortenson, J., & Splinter, D. (2019), "Household Incomes in Tax Data: Using Addresses to Move from Tax Unit to Household Income Distributions", *Journal of Human Resources*, forthcoming.

Mervis, Jeffrey. (2014), "How two economists got direct access to IRS tax records", https://www.sciencemag.org/news/2014/05/how-two-economists-got-direct-access-irs-tax-records.

Noguchi, Haruko, Shun-ichiro Bessho, Koichi Ushijima, Akira Kawamura & Ryuichi Tanaka (2020), "Relationship between School Attendance Support and Academic Performance: An Analysis of Panel Data from Adachi Ward", *Public Policy Review*, Vol. 16 No. 6, pp.1-22.

Piketty, T., & Saez, E. (2003), "Income Inequality in the United States, 1913-1998", *Quarterly Journal of Economics*, vol.118, no.1, pp.1-41.

Røed, K., & Raaum, O. (2003), "Administrative registers – Unexplored reservoirs of Scientific Knowledge?", *Economic Journal*, vol.113, no.488, pp.F258-F281.

Japanese

Adachi ward (2011), "Aiming for the Best School Meals in Japan / Lunchrooms in Adachi ward, Tokyo / Balanced Meals with 12 Nutrients You Will Want to Eat Every day," *Delicious School Meals Production Committee,* Taibundo.

Akiu, Shūichirō (2019), "Project to Connecting Adachi to the Future: Plan Formulation and Data Application" *Kanagawa Prefecture Policy Research Forum Announcement* (January 29, 2019).

Hayashi, Masayoshi (2008), "An Economic Theory of Decentralization" in Kaizuka, Keimei and Policy Research Institute MOF ed., *Local Finance in the Age of Decentralization,* Chuokeizai Holdings, Chapter 2, pp.43-69.

Mitsubishi Research Institute (2016), "A Research Report on Expert Task Analysis Using the Academic Achievement Survey," in AY2015 MEXT-commissioned survey *A Survey on the Building and Operation of a Results Database for the National Academic Achievement and Learning Conditions Survey*.

Mitsubishi Research Institute (2018), "A Research Report on Evidence-based Education Policy," in AY2017 MEXT-commissioned survey *A Survey on the Comprehensive Promotion of Education Reform*.

Nippon Foundation (2018), "A Relationship Analysis Between Household Economics Inequality and the Cognitive and Non-cognitive Disparities of Children. As Seen from the Big Data of 25,000 People."

Soga, Kengo & Satoshi Machidori (2007), *Local Governments in Japan: Policy Selection in a Dual Representation Government,* University of Nagoya Press.

Suzuki, Wataru (2018), "What is Wrong with Evidence-based Policymaking, and How Can It Be Improved?" *Japanese Journal of Health Economics & Policy,* Vol. 30 No. 1, pp.1-4.

Tanaka, Ryuichi (2019), "An Examination of Evidence-based Policymaking and Statistical Accounting: Founded on a Survey of Education and Statistics Databases in the United Kingdom and Finland," *Board of Audit of Japan AY2018 Report on Actual Conditions of Overseas Administrations*.

Toyofuku, Shinpei (2017), "Actual Conditions of Open Database Usage in Education," *IPSG SIG Technical Report*, 2017-CLE-23, pp.1-4.

Toyofuku, Shinpei (2018), "Will Open Data in Education Become Popular?" *Japanese Journal of Information Processing*, Vol. 59 No. 5, pp.467-470.

Tsuda, Hirokazu & Kohei Okazaki (2018), "Trends in Evidence-based Policymaking (EBPM) in the USA," *RIETI Policy Discussion Paper Series 18-P-016*.

Uchiyama, Yu, Yohei Kobayashi, Sosuke Taguchi & Takahide Koike (2018), "Evidence-based Policymaking in the UK and Clues for Japan: An Analysis Focusing on the 'Demand' and 'Supply' of Evidence," *RIETI Policy Discussion Paper Series 18-P-018*.