Survey of Cohort Studies and Data Linkage in Japan

March 2020

Ernst & Young ShinNihon LLC

Summary

Objective

In Japan, data for different cohort studies are stored in separate databases, and the scope of data sharing is currently limited. This survey aimed to ascertain the current state of cohort studies in Japan and to explore opportunities for data sharing in future cohort studies.

Outline

Chapter 1 summarizes information on cohort studies in Japan and the current state of and challenges to linking and sharing study data, such as genomic data. Chapter 2 discusses potential data linkage and data solutions in Japan by examining both technical and procedural aspects that impact data linkage and sharing.

Chapter 1. Survey of cohort studies in Japan

1.1 Overview

1.1.1 Objective

This survey aimed to identify cohort studies in Japan and highlight implications for future study opportunities, including data strategies.

1) Firstly, an online survey was conducted among academic society members involved in target cohort studies to identify cohorts in Japan. This online survey gathered the following information: cohort summaries, collected data items, study outcomes, data sharing, data linkage, and associated challenges.
2) Concurrently with the online survey, a review of the literature was performed via databases such as PubMed and the Japan Medical Abstract Society website to extract cohort studies from academic articles and to match these to the cohort studies identified through the above-mentioned online survey.
3) For the cohorts identified via the online survey, interviews were conducted to representatives from the cohorts with genomic data or prospective cohorts involving data linkage and/or sharing.

Note) Initially, this survey was designed to extract cohort studies by reviewing the literature. After consulting with experts on its design, we changed our approach to conduct a questionnaire to principal investigators of the cohorts by contacting relevant academic societies. This was because the literature search alone would not have been comprehensive and would fail to identify new cohorts. Thus, we conducted an online survey to principal investigators, complemented by the literature search undertaken in parallel to identify cohorts that were not identified via the online survey. Cohorts identified via the online survey and the literature search were then compared against each other.

Table 1.1-1 Overview of this survey of cohort studies in Japan

	Category A • Prospective cohorts of participants living in Japan	Category B • Of Category A cohorts, longitudinal studies of a certain size	Category C • Of Category B cohorts, cohorts with genomic data and prospective cohort studies with data linkage and/or sharing
Online Survey	✓ (Selected items)	✓	✓
Literature Search	To obtain information on points raised by questions in the online survey	To obtain information on points raised by questions in the online survey	To obtain information on points raised by questions in the online survey
Interviews	—	—	✓

1.1.2 Target cohort studies

We screened the cohort studies identified via the online survey in the three stages described below and grouped them into three categories (A, B, and C):

- Category A. The first screening stage focused on broadly identifying prospective cohort studies that met the following criteria and involving participants residing in Japan: 1) cohort studies with inclusion criteria that were not patient specific during the recruitment period; 2) prospective cohort studies with a defined objective collecting data prospectively (i.e., excluding studies that retrospectively analyzed chronological data collected through health checkups and examinations); 3) cohort studies that had completed baseline investigations and began follow-ups; and 4) cohort studies that had been ethically reviewed and approved. (Since the Ethical Guidelines for Epidemiological Studies went into effect on April 1, 2005, studies planned before April 2005 were not subject to this criterion.)
- Category B. The second screening stage reviewed Category A studies to identify prospective cohort studies of a certain size that met the following criteria: 1) cohort studies on adult participants with a baseline sample size of ≥1,000 or cohort studies of subjects in the peripartum period (including fetus) up to 20 years of age with a baseline sample size of ≥200; 2) cohort studies with defined follow-up items and endpoints; and 3) cohort studies with a known follow-up rate from the start to the present.
- Category C. The third screening stage then focused on Category B studies to identify cohort studies with genomic data and prospective cohort studies with data linkage and/or sharing, meeting at least one of the following three criteria: 1) cohort studies with a follow-up rate from the baseline to the present of ≥90% (participants were regarded as followed as long as individual records of such events as death or relocation were retained); 2) cohort studies with genomic/omics data and/or samples; 3) cohort studies with at least one of seven criteria related to their data management policy (1. having rules allowing the provision of data or samples to third parties; 2. having provided data or samples to third parties in the past; 3. having data linked to other domestic and foreign cohort studies in the past; 4. having systems for linking medical data, such as medical charts, and Diagnosis Procedure Combination (DPC) data; 5. having linked medical data, such as medical charts, and DPC data in the past; 6. having procedures for electronically linking research data to medical data, such as medical charts, DPC data, and nursing care data; or 7. having established systems for electronically linking research data to medical data, such as medical charts, DPC data. and nursing care data).

Table 1.1-2 Target cohort studies for this survey

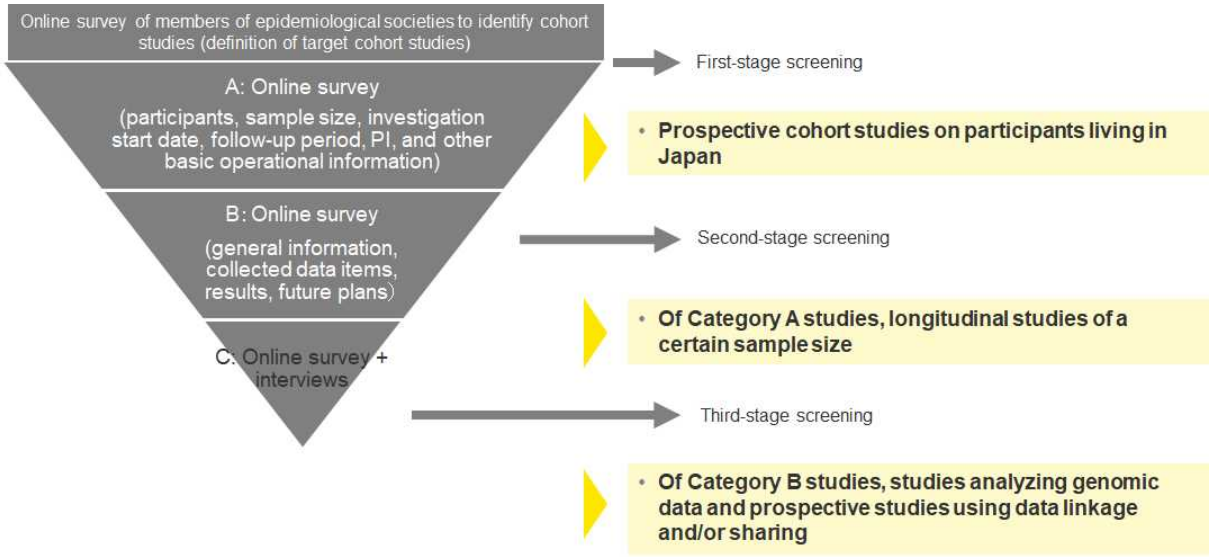
	First-stage screening (for inclusion in Category A)	Second-stage screening (Narrowing down Category A to Category B)	Third-stage screening (Narrowing down Category B to Category C)
Screening goal	To identify prospective cohort studies on subjects living in Japan (Category A).	• To identify prospective studies of a certain size from Category A (Category B).	• To identify studies analyzing genomic data and prospective cohort studies using data linkage and/or sharing from Category B (Category C).
Subjects	• Those residing in Japan	• Same as on the left	• Same as on the left
Scope	• Cohort studies with inclusion criteria not limited to certain diseases and illnesses during the recruitment period	• Same as on the left	• Same as on the left
Design	• Prospective cohort study	• Same as on the left	• Same as on the left
	• Prospective cohort studies with defined objective and gathering data prospectively (excluding studies that retrospectively analyze chronological data gathered through health checkups and examinations)	• Same as on the left	• Same as on the left
(Baseline) sample size	–	• Adult subjects: n = 1,000	• Same as on the left
		• Subjects in the peripartum period (including fetus) up to 20 years of age: n = 200	• Same as on the left
Follow-up	• Studies that completed baseline investigations and began follow-up investigations.	• Same as on the left Note: Studies with defined follow-up items and end points	• Same as on the left
Follow-up rate	–	• Studies with known baseline enrollment rates and follow-up rates from the start to the present	• Studies with a baseline enrollment rate of ≥70% and a follow-up rate from the baseline to present of ≥90% (★) Note: Subjects were considered "followed" as long as records of such events as death or a move were available.
Analyzed data and samples	–	–	• Studies with genomic/omics data and samples (★)
Ethics	• Studies that are carried out under ethical review Note: Studies planned before April 2005 were not subject to this criterion (the Ethical Guidelines for Epidemiological Studies went into effect on April 1, 2005).	• Same as on the left	• Same as on the left
Data management policy	–	–	• Studies with rules established to allow the provision of data or samples to third parties (★)
			• Studies that have provided data or samples to third parties (★)
			• Studies that have demonstrated data linkage with other Japanese and overseas cohort studies (★)
			• Studies with systems established for linking medical data, such as medical charts, and Diagnosis Procedure Combination (DPC) data (★)
			• Studies that have demonstrated linkage with medical data, such as medical charts, and DPC data (★)
			• Studies currently establishing procedures for electronically linking research data to medical data, such as medical charts, DPC data, and nursing care data (★)
			• Studies with established systems for electronically linking research data to medical data, such as medical charts, DPC data and nursing care data (★)

★ Needed to satisfy one of the above

1.1.3 Survey process

- We contacted academic societies involved with targeted cohorts to ask members to participate in an online survey and identified various cohorts in Japan.
- We then screened these studies based on the online survey in several stages and grouped them into three categories, A, B, and C.

Figure 1.1-1 Data gathering steps for this survey



1.2 Online survey of cohorts in Japan

1.2.1 Design

(1) Objective

To identify cohorts in Japan and gather information on their cohort summaries, data/variables collected, study outcomes, data sharing, data linkage, and challenges.

(2) Methodologies

1) Online survey participants

The director of each academic society asked members to participate in the online survey. Target cohorts were identified based on the response to the request.

2) Online survey period

The online survey was conducted from December 16, 2019 to January 31, 2020.

- For cohorts for which additional questions arose or were identified through the literature search but with no response, the online survey was extended to March 20, 2020 to gather this information.

3) Survey targets

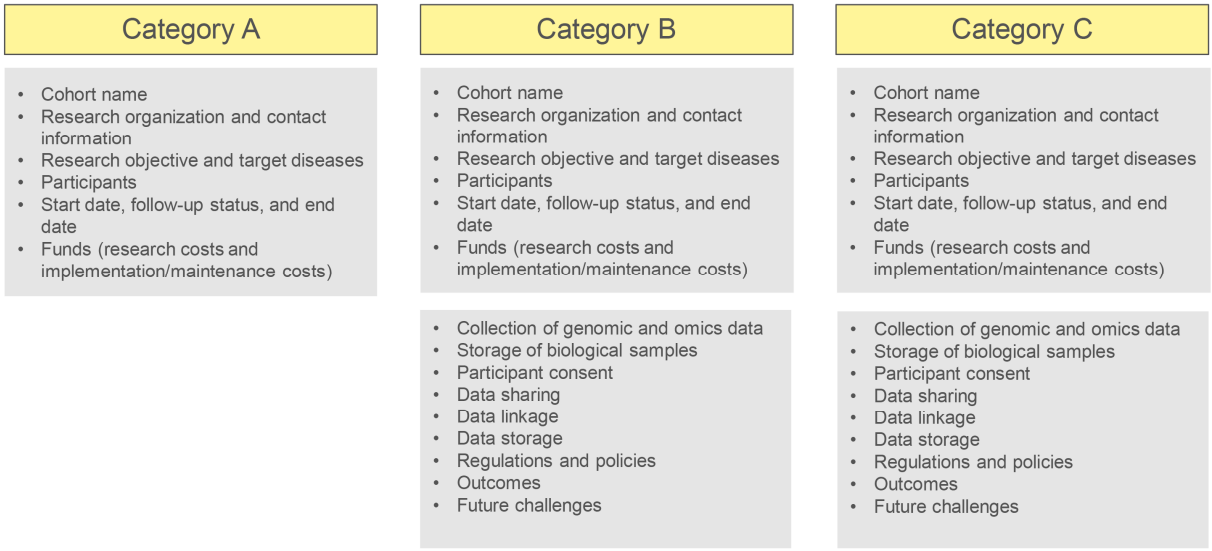
The online survey covered members of the following eight societies:

- Japan Epidemiological Association
- Japanese Society of Public Health
- Japan Society of Occupational Health
- Japanese Society for Hygiene
- Japanese Society of Health and Human Ecology
- Japanese Society of Physical Fitness and Sports Medicine
- Japanese Association of Rural Medicine
- Japan Association for Medical Informatics

4) Survey items

For cohorts in Category A, we obtained information on cohort name, research organization and contact information, research objective, target disease, participants, start date, follow-up status, end date, and funding source. For cohorts in Categories B and C, the survey items included those for Category A as well as the types of variables, storage and management of samples including genome and data, data sharing, data linkage, data storage, related regulations and policies, research outcomes, and future challenges.

Figure 1.2-1 Information obtained from cohorts in each Category



5) Matching cohorts identified by online survey against the literature search

◆ Extraction of cohorts through the literature search and matching against online survey findings

1) We used PubMed to look for articles published for cohort studies in Japan. The table below shows the search terms for our target cohort studies.

2) Then, of the articles identified through PubMed search, we extracted 200 articles based on their impact factors using the Web of Science (Clarivate Analytics) and identified cohort studies based on the article titles and contents.

3) Of the cohort studies identified through the literature search, those for which no response could be obtained online were contacted to request their participation in the online survey.

Table 1.2-1 Search terms for the literature search

Criteria	Formula
Japan domestic	AND (“Japan” [MeSH])
Cohorts that are designed to predict disease development	AND ((((risk factors) OR related factors) OR association) OR factors associated)
Population based	AND ((((((((((population-based) OR population based) OR community-based) OR community based) OR community) dwelling) OR students) OR student) OR worker) OR workplace) NOT ((patient) OR patients)
Prospective cohort	AND (((((Cohort study) OR Follow up study) OR Longitudinal study) OR Prospective study) OR Birth cohort)
Cohort study articles	NOT ((((“systematic review” [Publication Type]) OR “review” [Publication Type]) OR “meta-analysis” [Publication Type]) OR

	“comment” [Publication Type])
Publication year	AND (after January 1, 2010)
Language	AND (English [Language])

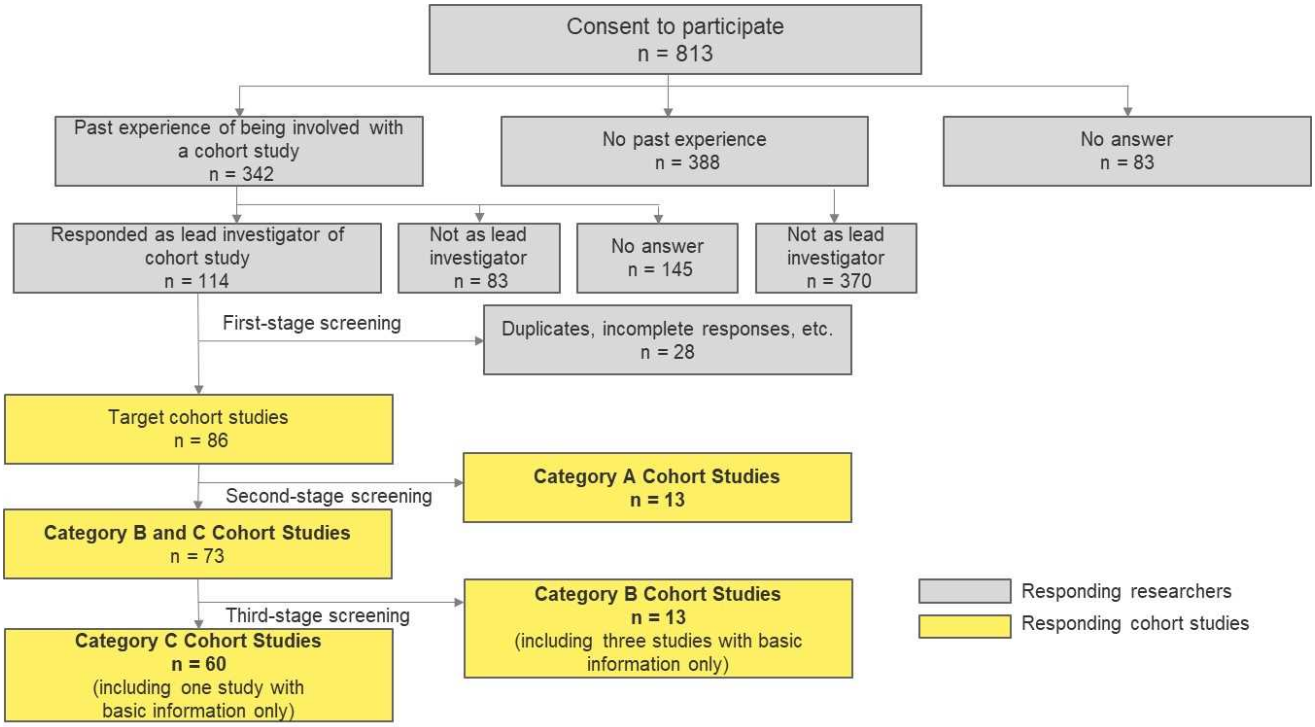
1.2.2 Survey results

(1) Survey participants and overview of the cohort studies

Of 813 consenting participants, 342 individuals responded that they have experience of being involved with a cohort study. Of these, 114 responded as principal investigator of cohort study. A total of 86 cohorts were identified: 13 Category A cohorts; 13 Category B cohorts (including three cohorts responding only basic information such as cohort name); and 60 Category C cohorts (including one study responding only basic information, such as cohort name).

* Tables and figures hereinafter will list only the 86 studies, including the four cohorts with basic information only for which consent for public disclosure of information had been given. Of the Category B and C cohorts, the results were tabulated and analyzed for 69 cohort studies associated with complete responses.

Figure 1.2-2 Overview of responding investigators and cohort studies



(2) Summary of responding cohorts

- The following tables summarize cohorts identified via the online survey (information, including URLs, was current as of January 31, 2020):
- Besides the cohort name (abbreviations), principal investigator information and contact information, interviews were conducted to determine whether the following information could be disclosed: positioning of the cohorts (in relation to other projects), research objective and target diseases, participant inclusion criteria, the starting date of registration and recruitment, follow-ups, enrollment and follow-up rates, type of variables and samples (including genomic and omics samples and data, types of bacterial flora and metagenome), provision of data and biological samples to third parties, data usage in collaborative international studies, data linkage with other cohorts, data linkage with other information, and adoption of standard codes.
- The following tables for the Categories A, B, and C cohorts consist only of responding cohorts associated with consent for the disclosure of cohort study title, principal investigator, URL, and contact information.

* The following tables for the cohorts exclude those lacking consent for disclosure; as such, the numbers assigned to the cohorts are not necessarily sequential.

1) Category A cohorts

	Cohort		Principal Investigator Information		URL	Contact Information	
	English name	Abbreviations	PI	Department		Contact	Telephone/ Email
1	Prospective cohort study of the effects of automobile exhaust on cardiovascular disease	–	Kenichi Azuma	Department of Environmental Medicine and Behavioral Science, Faculty of Medicine, Kindai University	–		
2	Study on promoting health maintenance for residents of wooden homes	–	Kenichi Azuma	Department of Environmental Medicine and Behavioral Science, Faculty of Medicine, Kindai University	–		
3	Community Empowerment and Care for Well-Being and Healthy Longevity: Evidence from Cohort	CEC	Tokie Anme	Faculty of Medicine, Tsukuba University	http://plaza.umin.ac.jp/empower/cec/en/	International Community Care and Lifespan Development: Empowerment Sciences, Faculty of Medicine,	029-853-3436 anmet@md.tsukuba.ac.jp

	Study					Tsukuba University	
4	Cohort study of the health of workers exposed to indium, Study on hygiene habits and health, Cohort study of the health of workers exposed to toner, Cohort study of the health of workers exposed to carbon disulfide, Cohort study of the respiratory system of workers exposed to TDI (TDI plant workers), Cohort study of the respiratory system of workers exposed to TDI (end users of TDI)		Kazuyuki Omae	Department of Preventive Medicine and Public Health, School of Medicine, Keio University (as of that time)		Department of Preventive Medicine and Public Health, School of Medicine, Keio University	03-5363-3758 omae.k@keio.jp
5	Life Span Study	LSS	Kotaro Ozasa	Radiation Effects Research Foundation	https://www.rerf.or.jp/programs/general_research/	Public Relation and Publications Office	082-261-3131 (switchboard) https://www.rerf.or.jp/contact/rinfo/
7	Kusatsu Project	KLSAH			http://www2.tmig.or.jp/spch/project_gaiyou_kusatsu.html		
8	Japan Epidemiology Collaboration on Occupational Health Study	J-ECOH Study	Dohi Seitaro	Mitsui Chemicals	–	–	–
9	Genome Cohort on Psychosocial Traits Study	GCOP Study	Takeshi Nishiyama	Department of Public Health, Graduate School of Medical Sciences, Nagoya City University	–	Department of Public Health	052-853-8176 p-gen@umin.ac.jp

10	Osaka Health Survey	–	Tomoshige Hayashi	Department of Preventive Medicine and Environmental Health, Osaka University School of Medicine	–	Department of Preventive Medicine and Environmental Health, Osaka University School of Medicine	06-6645-3751 thayashi@med.osaka-cu.ac.jp
11	Yamanashi Healthy Active Life Expectancy Study	Y-HALE	Zentaro Yamagata	Department of Health Sciences, Basic Science for Clinical Medicine, Division of Medicine, Graduate School Department of Interdisciplinary Research, University of Yamanashi	–	Department of Health Sciences, Graduate School Department of Interdisciplinary Research, University of Yamanashi	055-273-9566 zenymgt@yamanashi.ac.jp
12	Koshu Project	Koshu Project	Zentaro Yamagata	Department of Health Sciences, Basic Science for Clinical Medicine, Division of Medicine, Graduate School Department of Interdisciplinary Research, University of Yamanashi	https://www.med.yamanashi.ac.jp/medicine/birthcohort/study/summary/koshuProject.html	Department of Health Sciences, Graduate School Department of Interdisciplinary Research, University of Yamanashi	055-273-9566 zenymgt@yamanashi.ac.jp
13	Chiba Cancer Center Molecular Epidemiological Cohort Study of Cancer	–	Sana Yokoi	Department of Genetic Diagnosis, Chiba Cancer Center	–	Department of Genetic Diagnosis, Chiba Cancer Center	0432-64-5431 syokoi@chiba-cc.jp

2) Category B cohorts

	Cohort		Principal Investigator Information		URL	Contact Information	
	English name	Abbreviations	PI	Department		Contact	Telephone/ Email
1	Cohort study of the perception of ubiquitous objects and daily symptoms	JCI study	Kenichi Azuma	Department of Environmental Medicine and Behavioral Science, Faculty of Medicine, Kindai University	–		
2	Japanese Population-based Osteoporosis Cohort Study Fujiwara-kyo Osteoporosis Risk in Men Study Note 1:	JPOS Study	Masayuki Iki	Faculty of Medicine, Kindai University	https://www.med.kindai.ac.jp/pubheal/jpos/	Department of Public Health, Faculty of Medicine, Kindai University	072-366-0221 Pbl-h@med.kindai.ac.jp
3	Tokyo Teen Cohort Population Neuroscience Project	pn-TTC	Kiyoto Kasai	University of Tokyo Hospital	http://value.umin.jp/data-resource.html	Department of Neuropsychiatry, University of Tokyo Hospital	03-5800-8919 kasaimd@gmail.com
4	Cohort study of factors impacting elderly QOL and life functions: Fujiwara-Kyo Study	Fujiwara-kyo study	Norio Kurumatani	Nara Medical University Resident Health Support Center	–	–	–
5	Epidemiological study in Awachiiki	–	Shinichi Sato	Chiba Prefectural Institute of Public Health	https://www.pref.chiba.lg.jp/eiken/toukeidata/otassha.html	Healthy Chiba Group, Health Promotion Division, Health and Welfare Department	043-223-2660 https://www.pref.chiba.lg.jp/cgi-bin2/faq/form.cgi
7	Japanese study of Health, Occupation and Psychosocial factors related Equity	J-HOPE	Tsutsumi Akizumi	Kitasato University School of Medicine	http://www.med.kitasato-u.ac.jp/~publichealth/jhope.html	Kitasato University School of Medicine	042-778-9352 publichealth@med.kitasato-u.ac.jp

8	Toyama Occupational Cohort Study	–	Hideaki Nakagawa	Kanazawa Medical University Medical Research Institute	–	Department of Hygiene, Kanazawa Medical University	076-218-8101
9	Japan Nurses' Health Study	JNHS	Kunihiko Hayashi	School of Health Sciences, Faculty of Medicine, Gunma University	http://plaza.umin.ac.jp/jnhs/		
11	Kyushu Okinawa Maternal and Child Health Study	KOMCHS	Yoshihiro Miyake	Department of Epidemiology and Preventative Medicine, Ehime University Graduate School of Medicine	https://www.m.ehime-u.ac.jp/school/publichealth/index.php	Department of Epidemiology and Preventative Medicine, Ehime University Graduate School of Medicine	089-960-5283 epi-prev@m.ehime-u.ac.jp
12	Osaka Maternal and Child Health Study	OMCHS	Yoshihiro Miyake	Department of Epidemiology and Preventative Medicine, Ehime University Graduate School of Medicine	–	Department of Epidemiology and Preventative Medicine, Ehime University Graduate School of Medicine	089-960-5283 tanaka.keiko.jn@ehime-u.ac.jp
13	Chiba Study of Mother and Children's Health	C-MACH	Chisato Mori	Chiba University	http://cpms.chiba-u.jp/kids/	Center for Preventive Medical Sciences, Chiba University	043-290-3003

3) Category C cohorts

	Cohort		Principal Investigator Information		URL	Contact Information	
	English name	Abbreviations	PI	Department		Contact	Telephone/ Email
1	JMS II cohort study	JMSII Cohort Study	Shizukiyo Ishikawa	Department of Public Health, Center for Community Medicine, Jichi Medical University	https://www.jichi.ac.jp/dph/inprogress/jms2/	JMS II cohort study office	0285-58-7338 cohort2@jichi.ac.jp
2	Jichi Medical Cohort Study	JMS Cohort	Shizukiyo	Department of Public	https://www.jichi.a	JMS cohort study	0285-58-7338

		Study	Ishikawa	Health, Center for Community Medicine, Jichi Medical University	c.jp/dph/inprogress/jms2/result/	office	cohort2@jichi.ac.jp
3	Circulatory Risk in Communities Study	CIRCS	Hiroyasu Iso	Public Health Graduate School of Medicine, Osaka University	http://www.osaka-ganjun.jp/effort/cvd/r-and-d/circs/		
4	Kita Nagoya Genomic Epidemiological Study	KING study	Sahoko Ichihara	Jichi Medical University	–	Department of Environmental and Preventive Medicine, Jichi Medical University	0285-58-7335 saho@jichi.ac.jp
5	Ibaraki Prefectural Health Study (follow-up study on life prognosis for people receiving medical checkups)	IPHS	Fujiko Irie	Ibaraki Prefecture Tsukuba Healthcare Center (also, Ibaraki Prefecture Public Health and Welfare Department, Heath and Community Care Promotion Division)	www.hsc-i.jp/05_chousa/iphs1.htm	Health Promotion Group, Heath and Community Care Promotion Division, Ibaraki Prefecture Public Health and Welfare Department	029-301-3229 care3@pref.ibaraki.lg.jp
6	Ibaraki Prefectural Health Study (large-scale cohort study of physical well-being, preventive nursing care, and healthcare cost improvements)	IPHS	Fujiko Irie	Ibaraki Prefecture Tsukuba Healthcare Center (also, Ibaraki Prefecture Public Health and Welfare Department, Heath and Community Care Promotion Division)	www.hsc-i.jp/05_chousa/iphs2.htm	Health Promotion Group, Heath and Community Care Promotion Division, Ibaraki Prefecture Public Health and Welfare Department	029-301-3229 care3@pref.ibaraki.lg.jp
7	Japan Arteriosclerosis Longitudinal Study	JALS	Hirotsugu Ueshima	Center for Epidemiologic Research in Asia, Shiga University of Medical Sciences	http://jals.gr.jp/	Department of Integrated Science and Engineering for Sustainable Society, Chuo University	03-3817-7294 aharada.13b@g.chuo-u.ac.jp
8	EBCT and Risk Factor Assessment among Japanese and US men in the Post	ERA-JUMP	Hirotsugu Ueshima	Center for Epidemiologic Research in Asia, Shiga University of Medical	–	Shiga Epidemiological Study of Subclinical	077-548-2435

	World War II Birth Cohort			Sciences		Atherosclerosis (SESSA) office	
9	NIPPON DATA 80	NIPPON DATA 80	Hirotsugu Ueshima	Center for Epidemiologic Research in Asia, Shiga University of Medical Sciences	https://shiga-publichealth.jp/nippon-data/	Division of Public Health, Departmnet of Social Medicine, Shiga University of Medical Science, Nippon Data 80/90 office	077-548-3659 hqhealth@belle.shiga-med.ac.jp
10	NIPPON DATA 90	NIPPON DATA 90	Hirotsugu Ueshima	Center for Epidemiologic Research in Asia, Shiga University of Medical Sciences	https://shiga-publichealth.jp/nippon-data/	Division of Public Health, Departmnet of Social Medicine, Shiga University of Medical Science, Nippon Data 80/90 office	077-548-2191 nd90@belle.shiga-med.ac.jp
11	NIPPON DATA 2010	NIPPON DATA 2010	Katsuyuki Miura	Division of Public Health, Departmnet of Social Medicine, Shiga University of Medical Science	https://shiga-publichealth.jp/nippon-data/	Division of Public Health, Departmnet of Social Medicine, Shiga University of Medical Science, Nippon Data 2010 office	077-548-3659 hqhealth@belle.shiga-med.ac.jp
12	Cohort study of factors contributing to the onset of lifestyle diseases in Shiga Prefecture	Takashima Study	Yoshikuni Kita	Tsuruga Nursing University	https://www.shiga-med.ac.jp/hqcera/project/takashima_study/takashima_cohort/index.html	–	–
13	Japan Environment and Children's Study	JECS	Michihiro Kamijima	Graduate School of Medical Sciences, Nagoya City University	http://www.env.go.jp/chemi/ceh/index.html	Japan Environment and Children's Study Core Center, National Institute for Environmnetal Studies	029-850-2191 jecscore@nies.go.jp

14	Ohsama Study	Ohasama	Takayoshi Ohkubo	Department of Hygiene and Public Health, Teikyo University School of Medicine	http://www.epi-c.jp/e002_1_0001.html	Department of Hygiene and Public Health, Teikyo University School of Medicine	03-3964-1211 (extension: 46210) hph-support@med.teikyo-u.ac.jp
15	Toon Study	–	Haruhiko Osawa	Department of Diabetes and Molecular Genetics, Ehime University Graduate School of Medicine	https://www.toon-study.jp/	Faculty of Agriculture, Ehime University	089-904-3287 ths@m.ehime-u.ac.jp
16	National Institute for Longevity Sciences, Longitudinal Study of Aging	NILS-LSA	Hiroshi Shimokata (1997-2012) Rei Otsuka (from 2013)	National Institute for Longevity Sciences NILS-LSA Laboratory, National Center for Geriatrics and Gerontology	https://www.ncgg.go.jp/cgss/department/ep/index.html		
17	Tokyo Children's Health, Illness and Development (T-CHILD) study	T-CHILD	Yukihiro Ohya	National Center for Children's Health and Development	–	–	–
18	Night in Japan Home Sleep Monitoring Study	NinJaSleep Study	Hiroshi Kadotani	Department of Sleep Behavior Medicine, Shiga University of Medical Science	–	Department of Sleep Behavior Medicine, Shiga University of Medical Science	077-548-3632 hqsuimin@belle.shiga-med.ac.jp
19	Hokkaido Study on Environment and Children's Health: Sapporo Cohort	The Hokkaido Study: Sapporo Cohort	Reiko Kishi	Center for Environmental and Health Sciences, Hokkaido University	www.cehs.hokudai.ac.jp/hokkaidostudy	Hokkaido Study Office	011-706-4749 kodomo@cehs.hokudai.ac.jp
20	Hokkaido Study on Environment and Children's Health: Hokkaido Cohort	The Hokkaido Study: Hokkaido Cohort	Reiko Kishi	Center for Environmental and Health Sciences, Hokkaido University	www.cehs.hokudai.ac.jp/hokkaidostudy	Hokkaido Study Office	011-706-4749 kodomo@cehs.hokudai.ac.jp
21	Tohoku Medical Megabank (TMM) BirThree Cohort Study	TMM BirThree Cohort Study	Shinichi Kuriyama	Tohoku Medical Megabank Organization	https://www.megabank.tohoku.ac.jp/3gen/	BirThree Cohort office, Tohoku Medical Megabank Organization	022-718-5162 sansedai@megabank.tohoku.ac.jp

22	Tohoku Medical Megabank Community Cohort Study	TMM CommCohort Study	Atsushi Hozawa	Department of Preventive Medicine and Epidemiology, Tohoku Medical Megabank Organization, Tohoku University	https://www.megabank.tohoku.ac.jp/chco/	Department of Public Relations	022-717-7908 pr@megabank.tohoku.ac.jp
23	Research Project for Prospective Investigation of Health Problems among Survivors of Great East Japan Earthquake and Tsunami Disaster Study	RIAS Study	Seiichiro Kobayashi	Iwate Medical University	https://healthresearch-iwate.jp/inquiry/	Department of Hygiene and Preventive Medicine, Iwate Medical University	019-651-5111 (extension: 5775) ksakata@iwate-med.ac.jp
24	Japan Gerontological Evaluation Study	JAGES	Katsunori Kondo	Center for Preventive Medical Sciences, Chiba University, National Center for Geriatrics and Gerontology, Japan Agency for Gerontological Evaluation Study	https://www.jages.net/	Center for Preventive Medical Sciences, Chiba University (JAGES Chiba office)	04-7137-8207 jages-office@jages.net
25	HEIJO-KYO study	HEIJO-KYO study	Keigo Saeki	Department of Epidemiology and Preventive Medicine, Nara Medical University	–	Department of Epidemiology and Preventive Medicine, Nara Medical University	0744-22-3051 epidemiology@naramed-u.ac.jp
26	Cohort study of preventive nursing care focusing on rural and agricultural communities		Kuninori Shiwaku	Division of lifestyle diseases, Japanese Association of Rural Medicine		Office of the Japanese Association of Rural Medicine	03-3212-8005 a-tsune@jarm.jp
27	Gunma Komoise Cohort Study	KI Study	Shosuke Suzuki	EcoHealth Research Group (NPO)	–	EcoHealth Research Group (NPO)	0270-61-7983 ecohealthg@mbh.nifty.com
28	Three Prefecture Cohort	Three-	Tomotaka	Environmental Medicine	–	–	–

	Study Osaka	Prefecture Cohort Study Osaka	Sobue	and Population Sciences , Department of Social Medicine, Graduate School of Medicine, Osaka University			
29	Tsuruoka Metabolome Cohort Study	TMCS	Tohru Takebayashi	Institute of Advanced Biosciences, Keio University School of Medicine	http://tsuruoka-mirai.net/	Department of Preventive Medicine and Public Health, School of Medicine, Keio University	mirai@iab.keio.ac.jp
31	The Japan “Society and New Tobacco” Internet Survey	JASTIS study	Takahiro Tabuchi	Cancer Control Center Department of Cancer Epidemiology, Osaka International Cancer Institute	–	Cancer Control Center Department of Cancer Epidemiology, Osaka International Cancer Institute	tabuchitak@gmail.com
32	Japan Collaborative Cohort Study for Evaluation of Cancer Risk (sponsored by the Ministry of Education, Culture, Sport, Science and Technology of Japan)	JACC Study	Akiko Tamakoshi	Graduate School of Medicine, Hokkaido University	https://publichealth.med.hokudai.ac.jp/jacc/	Graduate School of Medicine, Hokkaido University	011-706-5068 JACC_study@med.hokudai.ac.jp
34	Japan Public Health Center-based prospective Study	JPHC Study	Shoichiro Tsugane	National Cancer Center, Center for Public Health Sciences	https://epi.ncc.go.jp/jphc/index.html	Epidemiology and Preventive Medicine Group	jphcadmin@ml.res.ncc.go.jp
35	Next-generation Japan Public Health Center-based prospective Study	JPHC-NEXT	Shoichiro Tsugane	National Cancer Center, Center for Public Health Sciences	https://epi.ncc.go.jp/jphcnext/index.html	Epidemiology and Preventive Medicine Group	jphcadmin@ml.res.ncc.go.jp
36	Miyagi Prefecture Cohort Study	–	Ichiro Tsuji	Department of Public Health, Tohoku University Graduate School of Medicine	http://www.pbhealth.med.tohoku.ac.jp/project.html	Department of Public Health, Tohoku University Graduate School of Medicine	022-717-8123 miyagi@pbhealth.med.tohoku.ac.jp
37	Osaki National Health Insurance Beneficiaries	–	Ichiro Tsuji	Department of Public Health, Tohoku	http://www.pbhealth.med.tohoku.ac.jp/	Department of Public Health, Tohoku	022-717-8123 ohsaki1994@pbhea

	Cohort Study			University Graduate School of Medicine	project.html	University Graduate School of Medicine	lth.med.tohoku.ac.jp
38	Three-Prefecture Cohort Miyagi	–	Ichiro Tsuji	Department of Public Health, Tohoku University Graduate School of Medicine	–	Department of Public Health, Tohoku University Graduate School of Medicine	022-717-8123 thkpbh-office@umin.ac.jp
39	Tsurugaya Project	–	Ichiro Tsuji	Department of Public Health, Tohoku University Graduate School of Medicine	http://www.pbhealth.med.tohoku.ac.jp/project.html	Department of Public Health, Tohoku University Graduate School of Medicine	022-717-8123 tsurugaya@pbhealth.med.tohoku.ac.jp
40	Study of the health of suvivors of the Great East Japan Earthquake	–	Ichiro Tsuji	Department of Public Health, Tohoku University Graduate School of Medicine	https://www.ch-center.med.tohoku.ac.jp/	Center for Community Health, Tohoku University Graduate School of Medicine	022-717-8124 info@ch-center.med.tohoku.ac.jp
41	Tohoku Study of Child Development	TSCD	Kunihiko Nakai	Tohoku University Graduate School of Medicine	–	–	–
42	Hekinan Children's Study	–	Chisato Nagata	Department of Epidemiology and Preventative Medicine, Gifu University Graduate School of Medicine	https://www1.gifu-u.ac.jp/ph/	Department of Epidemiology and Preventative Medicine, Gifu University Graduate School of Medicine	058-230-6412 ph@gifu-u.ac.jp
43	Takayama Study	–	Chisato Nagata	Department of Epidemiology and Preventative Medicine, Gifu University Graduate School of Medicine	https://www1.gifu-u.ac.jp/	Department of Epidemiology and Preventative Medicine, Gifu University Graduate School of Medicine	058-230-6412 ph@gifu-u.ac.jp
44	Murakami Cohort Study	–	Kazutoshi Nakamura	Division of Preventive Medicine, Niigata University Graduate School of Medical and	https://www.med.niigata-u.ac.jp/hyg/murakami/index.html	Division of Preventive Medicine, Niigata University Graduate School of	025-227-2124 sakedep@med.niigata-u.ac.jp

				Dental Sciences		Medical and Dental Sciences	
45	Shikamachi cohort study	Shikamachi Study	Hiroyuki Nakamura	Department of Environmental and Preventative Medicine, Kanazawa University	http://www.projectship.org/index.html	Department of Environmental and Preventative Medicine, Kanazawa University	076-265-2218 t-hiromasa@med.kanazawa-u.ac.jp
46	Project Shikamachi Health Improvement Practice	Shika study	Hiroyuki Nakamura	Kanazawa University Advanced Preventive Medical Sciences Research Center	http://www.projectship.org/	Department of Environmental and Preventative Medicine, Kanazawa University	076-265-2288 t-hiromasa@med.kanazawa-u.ac.jp
47	Tokyo Teen Cohort Study	TTC	Atsushi Nishida	Tokyo Metropolitan Institute of Medical Sciences	http://ttcp.umin.jp/	Mental Health Project, Tokyo Metropolitan Institute of Medical Sciences	03-6834-2296 nishida-at@igakuken.or.jp
48	Japan Prospective Studies Collaboration for Aging and Dementia [Note 2:]	JPSC-AD	Toshiharu Ninomiya	Department of Epidemiology and Public Health, Graduate School of Medical Sciences, Kyushu University	https://www.eph.med.kyushu-u.ac.jp/jpsc/	Center for Cohort Studies, Graduate School of Medical Sciences, Kyushu University	092-642-6114 qjm10000@cohort.med.kyushu-u.ac.jp
49	Hisayama Project	Hisayama Study	Toshiharu Ninomiya	Department of Epidemiology and Public Health, Graduate School of Medical Sciences, Kyushu University	http://www.hisayama.med.kyushu-u.ac.jp/en/	Department of Epidemiology and Public Health, Graduate School of Medical Sciences, Kyushu University	092-642-6151 info_eph@eph.med.kyushu-u.ac.jp
50	Kansai Healthcare Study	–	Tomoshige Hayashi	Department of Preventive Medicine and Environmental Health, Osaka City University Graduate School of Medicine	–	Osaka City University Graduate School of Medicine	06-6645-3751 preventive@med.osaka-cu.ac.jp

51	Adachi-Ward Children's Health and Lifestyle Study	A-CHILD	Takeo Fujiwara	Department of Global Health Promotion, Graduate School of Medical and Dental Sciences, Tokyo Medical Dental University	https://www.city.adachi.tokyo.jp/kokoro/fukushi-kenko/kenko/kodomo-kenko-chosa.html	Department of Global Health Promotion, Graduate School of Medical and Dental Sciences, Tokyo Medical Dental University	03-5803-5187 fujiwara.hlth@tmd.ac.jp
52	Follow-up study on diseases impacting community lifestyle habits	Nagasaki Islands Study	Takahiro Maeda	Department of Community Medicine, Nagasaki University Graduate School of Biomedical Science	http://ritouken.com/publics/index/43/	Remote Island Medical Office, Department of Remote and Isolated Island Medicine, Nagasaki University Graduate School of Biomedical Science	0959-74-2673 kenichi.nobusue@nagasaki-u.ac.jp
53	Nagahama Cohort J-SHIP Study	Nagahama Study	Fumihiko Matsuda	Center for Genomic Medicine, Graduate School of Medicine, Kyoto University	http://zeroji-cohort.com	Nagahama Project Office	075-751-4166 nagahama-office@genome.med.kyoto-u.ac.jp
54	Shiga Epidemiological Study of Subclinical Atherosclerosis	SESSA	Katsuyuki Miura	Department of Public Health, Shiga University of Medical Science	https://shiga-publichealth.jp/sessa/	Shiga Epidemiological Study of Subclinical Atherosclerosis, SESSA office, Shiga University of Medical Science	077-548-2435
55	Suita Study	–	Yoshihiro Miyamoto	National Cerebral and Cardiovascular Center	–	–	–
56	Babies and Their Parents' Longitudinal Observation in Suzuki Memorial Hospital in the Intrauterine Period	BOSHI study	Hirohito Metoki	Division of Public Health, Hygiene and Epidemiology, Tohoku Medical and Pharmaceutical University	–	Division of Public Health, Hygiene and Epidemiology, Tohoku Medical and Pharmaceutical University	022-290-8727 phhe@tohoku-mpu.ac.jp

57	Nakajima Project Note 3:	Nakajima Project	Masahiro Yamada	Department of Neurology and Neurobiology of Aging, Graduate School of Medical Sciences, Kanazawa University	http://neurology.w3.kanazawa-u.ac.jp/resrchwrk/1250/	Department of Neuroscience, School of Medical Sciences, Kanazawa University	076-265-2290
58	Epidemiological study of herpes zoster on Shodoshima Island: Large-scale epidemiological study of shingles	SHEZ Study	Koichi Yamanishi	National Institute of Biomedical Innovation, Health and Nutrition	–	Public Health Graduate School of Medicine, Osaka University	06-6879-3911 iso@pbhel.med.osaka-u.ac.jp
59	Japan Multi-Institutional Collaborative Cohort Study	J-MICC Study	Kenji Wakai	Department of Preventive Medicine, Nagoya University Graduate School of Medicine	http://www.jmicc.com/	Department of Preventive Medicine, Nagoya University Graduate School of Medicine	052-744-2132 wakai@med.nagoya-u.ac.jp
60	Aichi Workers' Cohort Study	Aichi Workers' Cohort Study			http://koei-nagoya.blogspot.com/	Aichi Workers' Cohort Study Office	052-744-2127 p-health@med.nagoya-u.ac.jp

Note 1: Included in Category B because the study could not be clearly classified into either Category B or C.

Note 2: The Japan Prospective Studies Collaboration for Aging and Dementia is being conducted at eight sites throughout Japan[1)] with Kyushu University (the Hisayama Project) acting as the main project office.

Note 3: The Nakajima Project is also part of the Japan Prospective Studies Collaboration for Aging and Dementia.

[1)] Japan Prospective Studies Collaboration for Aging and Dementia (https://www.eph.med.kyushu-u.ac.jp/jpsc/)

(3) Cohorts extracted through literature search

The following table lists cohorts extracted through the literature search. Of the 25 cohorts, we obtained online survey responses for 20 cohorts.

	English name	Online survey response
1	Ibaraki Prefectural Health Study	✓
2	Osaka Maternal and Child Health Study	✓
3	Osaki Cohort 2006	✓
4	Ohsama Study	✓
5	Kyushu Okinawa Maternal and Child Health Study	✓
6	Follow-up of Atomic Bomb Survivors	✓
7	Japan Environment and Children's Study (JECS)	✓
8	Shiga Epidemiological Study of Subclinical Atherosclerosis SESSA	✓
9	EBCT and Risk Factor Assessment among Japanese and US men in the Post World War II Birth Cohort (ERA JUMP)	✓
10	Epidemiological study for developing shingles vaccines (SHEZ Study)	✓
11	Takayama Study	✓
12	Multi-purpose cohort (JPHC)	✓
13	Tsurugaya Project	✓
14	Nagahama Cohort J-SHIP	✓
15	Prevalence of allergic diseases among elementary school children in Western Japan	-
16	Japan Gerontological Evaluation Study (The Japan Gerontological Evaluation Study)	✓
17	Hisayama Project	✓
18	Fujiwara-kyo Study	✓
19	Heijo-kyo Cohort Study	✓
20	Research on Osteoarthritis Against Disability (ROAD)	-
21	Study on the lifestyle of children in Matsuyama city	-
22	Circulatory Risk in Communities Study (CIRCS)	✓
23	Japanese Population-based Osteoporosis (JPOS) Study	✓
24	Niigata Preventive Medicine Study	-
25	Wakayama Spine Study (WSS)	-

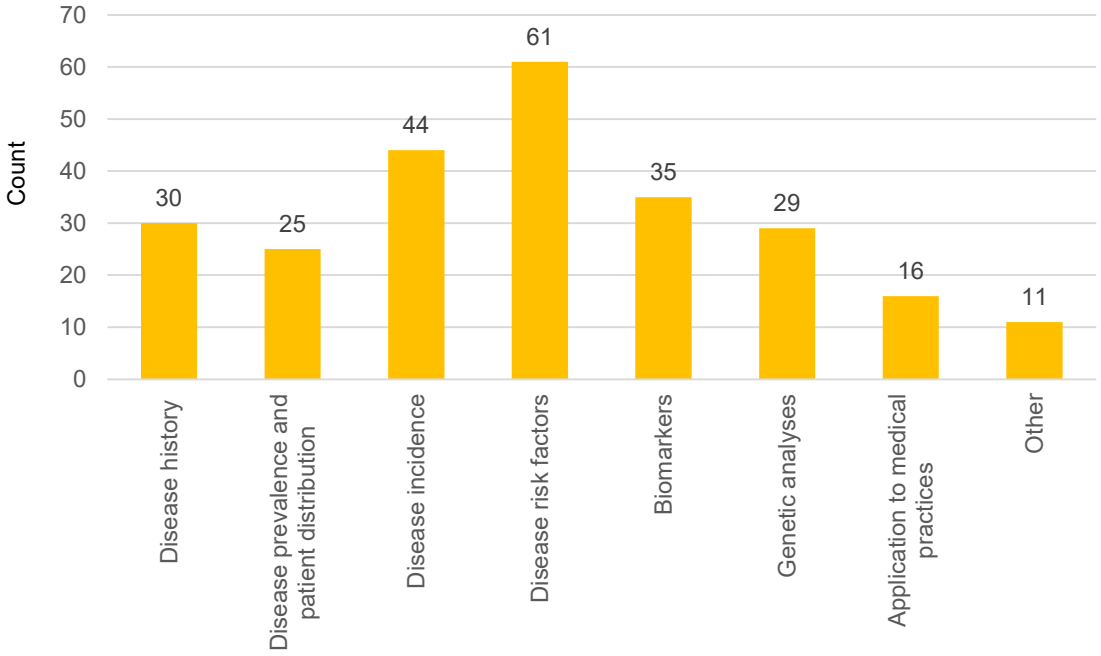
(4) Responses from online survey

1) Cohort study objectives

The most common objective of the cohort studies was investigation of the risk factors of diseases (n = 61).

The second and third most common objectives, accounting for more than half of the cohort studies, were survey of disease incidence and exploration of biomarkers (n = 44 and n = 35, respectively).

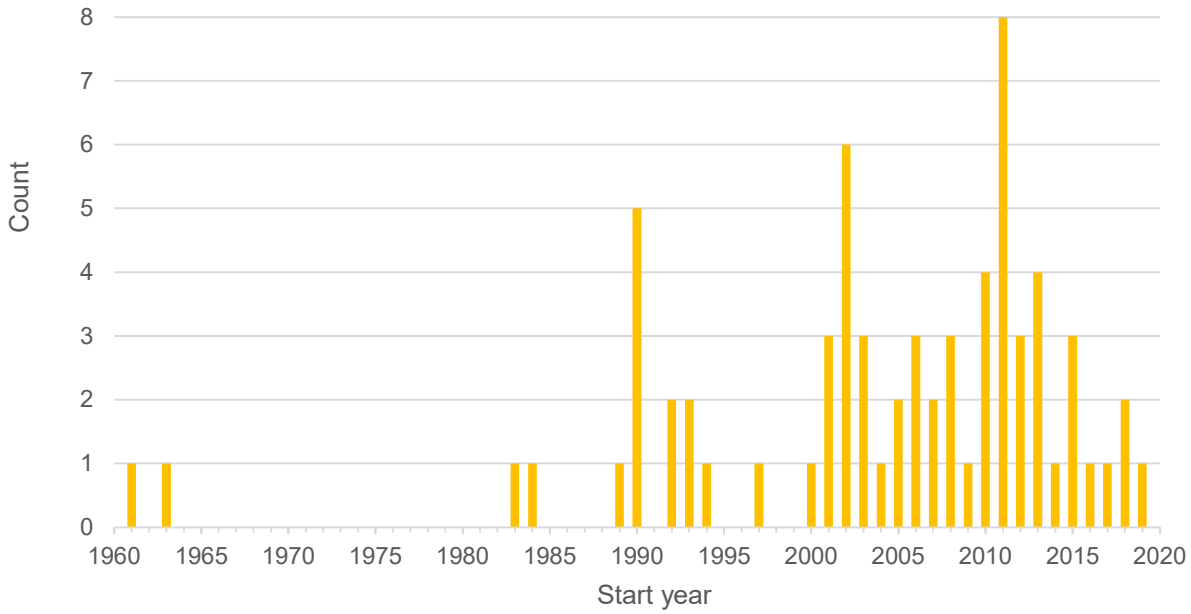
Figure 1.2-3 Research Objective (multiple selections, n = 69)



2) Start date of the cohorts

Two cohort studies began in 1961 and 1963. No other cohort studies were registered until the 1980s. Starting in 1983, new studies were launched sporadically. In 1990, five cohort studies were launched. Every year after 2000, new cohort studies were launched and registered. In particular, as many as eight cohorts were launched in 2011.

Figure 1.2-4 Cohorts by start date of initial data collection (n = 69)



3) Cohort follow-ups

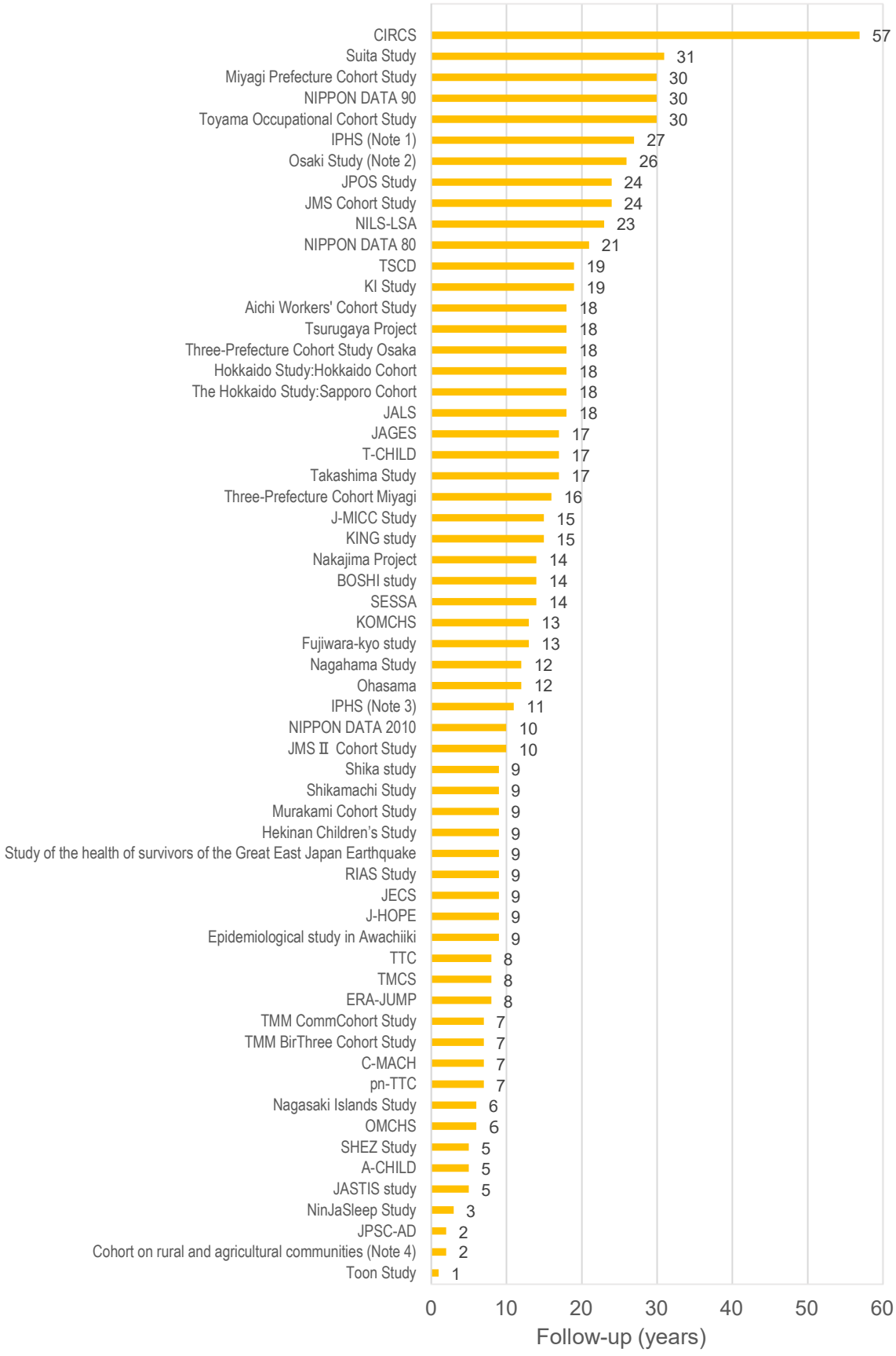
Of the 60 cohorts with consent for information disclosure, ten cohorts had longer follow-up durations than 20 years. Of these, four cohorts followed up their participants for 30 years or longer.

* We defined follow-up duration as the difference between the recruitment start date and the year 2020 for currently ongoing cohorts and as the difference between the recruitment start date and the end date of the last follow-up for cohortss for which follow-up investigations were completed.

* Cohorts were excluded from the current reports where the outcome or termination of follow-up investigations was unknown or where no consent for information disclosure was obtained. We added one cohort to the tabulation because it provided necessary information despite incomplete response to the survey.

- Note 1: Ibaraki Prefectural Health Study (Follow-up study on life prognosis for people receiving medical checkups)
- Note 2: Osaki National Health Insurance Beneficiaries Cohort Study
- Note 3: Ibaraki Prefectural Health Study (Large-scale cohort study of physical well-being, preventive nursing care, and healthcare cost improvements)
- Note 4: Cohort study of preventive nursing care focusing on rural and agricultural communities

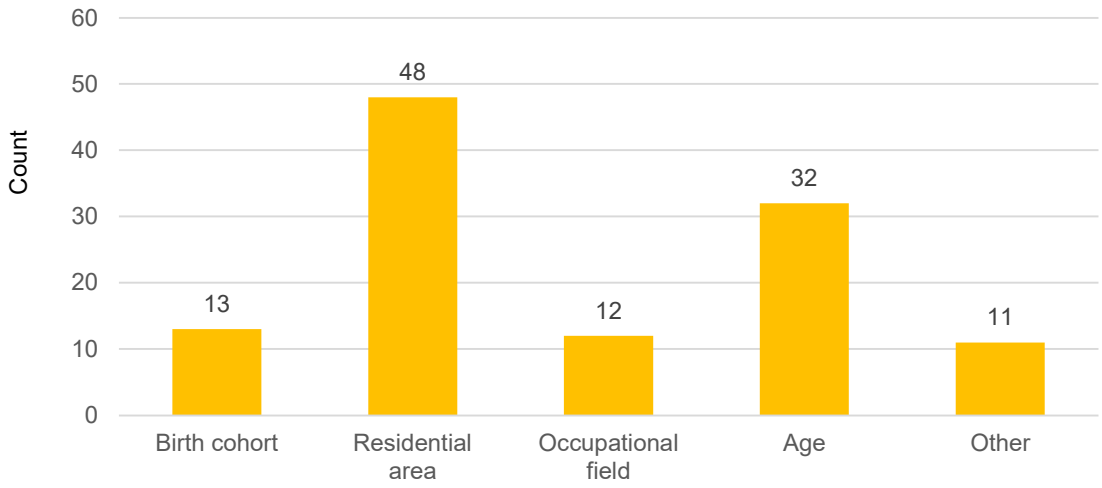
Figure 1.2-5 Length of cohort follow-up (n = 60)



4) Category of target cohorts

The most common recruitment basis for target cohorts was residential area (n = 48), followed by age (n = 32), birth (n = 13), and occupational field (n = 12).

Figure 1.2-6 Cohorts by category (multiple selections, n = 69)

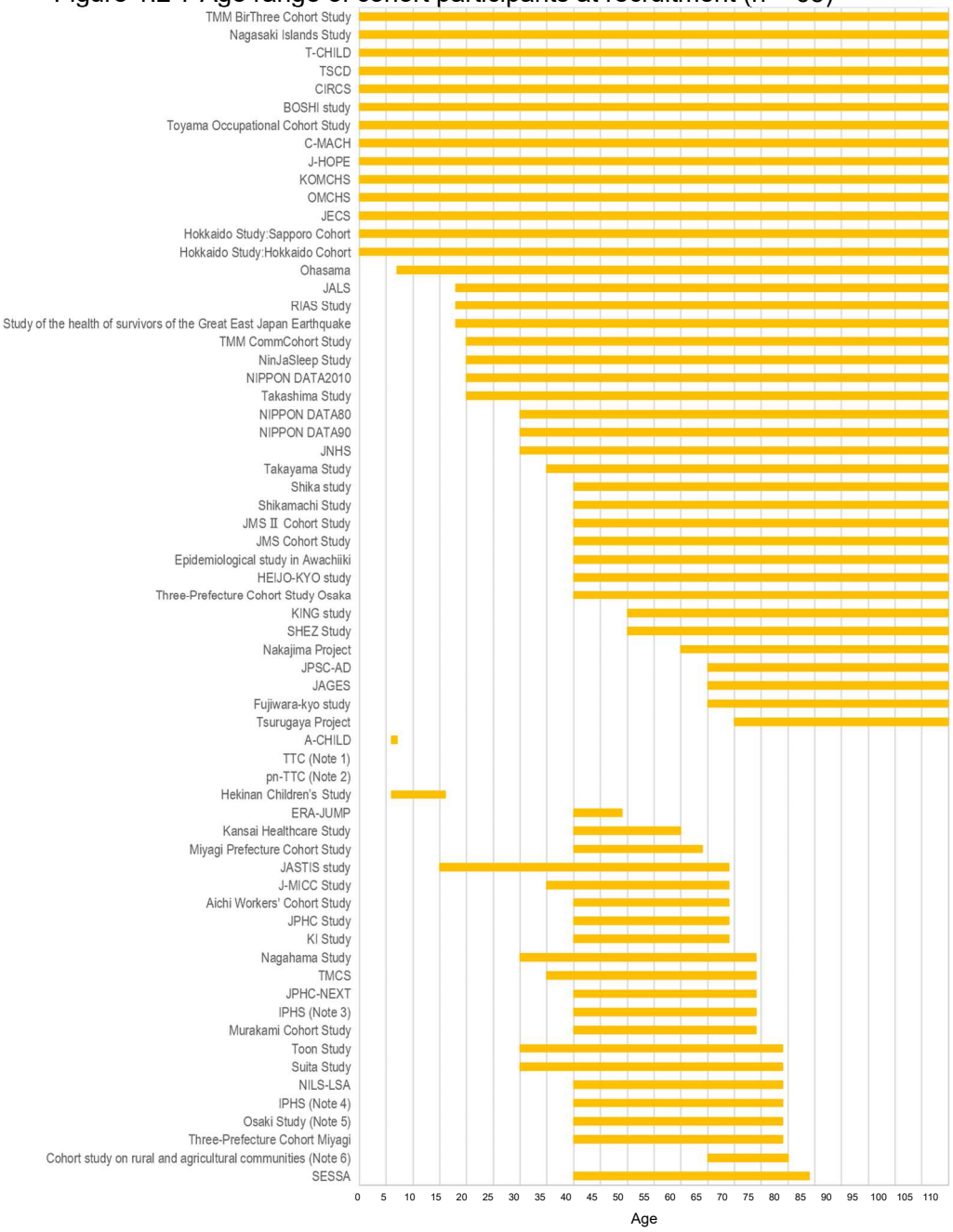


5) Age of the cohort participants at recruitment

About 20% of the cohorts (n = 14) set no lower limit on participant age during recruitment. More than 60% of the cohort studies (n = 40) set no upper limit on participant age.

Forty years was the most common lower age limit (about 32%; n = 21). This can be because many of the cohorts recruited participants who had undergone Specific Health Checkups within the National Health Insurance scheme.

Figure 1.2-7 Age range of cohort participants at recruitment (n = 65)



- Note 1: The age limit for TTC recruitment was 10 years; Note 2: The age limit for pn-TTC recruitment was 11 years; Note 3: Ibaraki Prefectural Health Study (Large-scale cohort study of physical well-being, preventive nursing care, and healthcare cost improvements); Note 4: Ibaraki Prefectural Health Study (Follow-up study on life prognosis for people receiving medical checkups); Note 5: Osaki National Health Insurance Beneficiaries Cohort Study; Note 6: Cohort study of preventive nursing care focusing on rural and agricultural communities

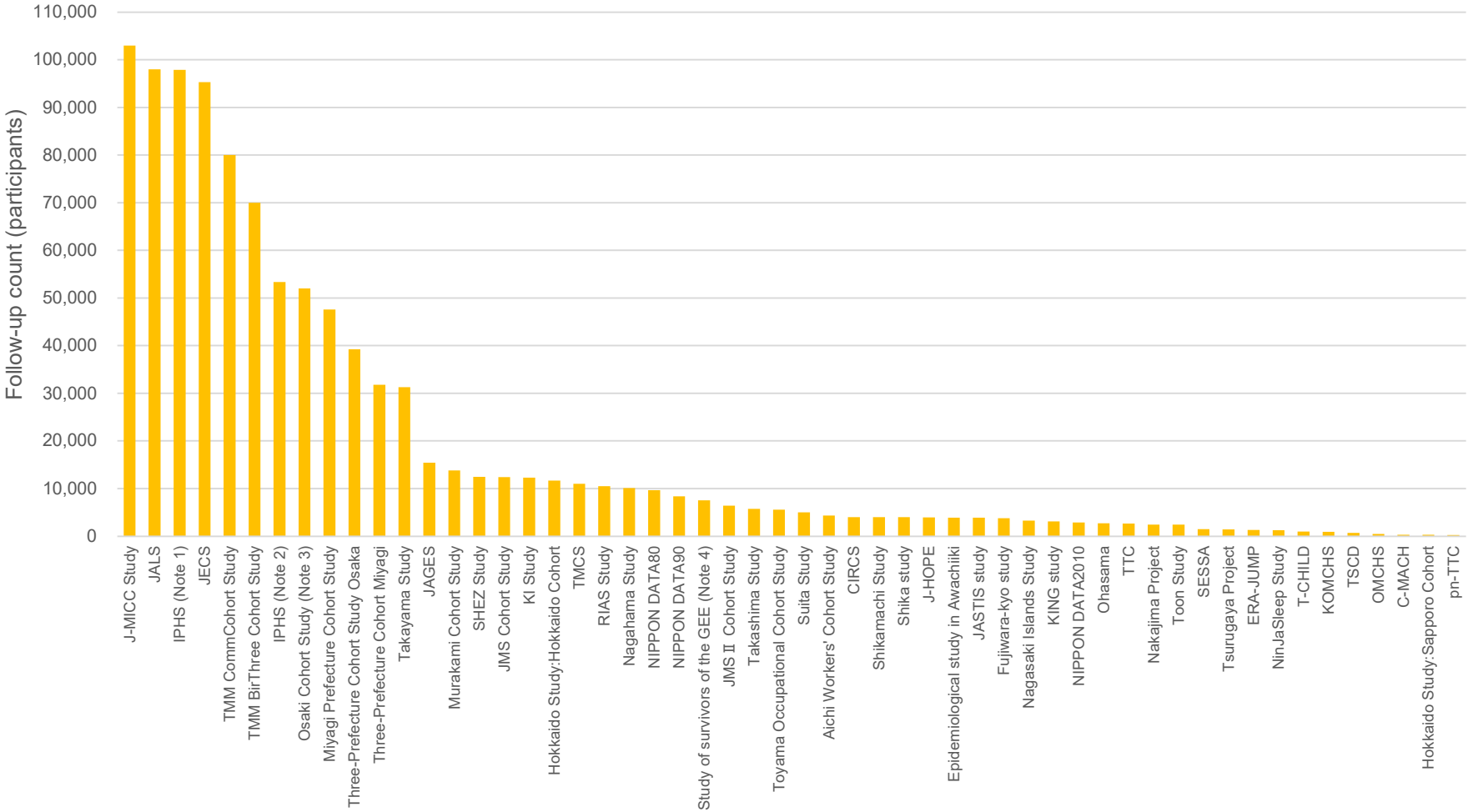
6) Current cohort sample size

The current sample size of each of the cohort studies ranged from 250 for the Tokyo Teen Cohort Population Neuroscience Project (pn-TTC) to more than 100,000 for the Japan Multi-Institutional Collaborative Cohort Study (J-MICC Study).

Twelve cohorts followed more than 30,000 participants; nine cohorts followed between 10,000 and 30,000; 22 cohorts followed between 2,000 and 10,000; and 11 cohorts followed fewer than 2,000.

- Note 1: Ibaraki Prefectural Health Study (Follow-up study on life prognosis for people receiving medical checkups); Note 2: Ibaraki Prefectural Health Study (Large-scale cohort study of physical well-being, preventive nursing care, and healthcare cost improvements); Note 3: Osaki National Health Insurance Beneficiaries Cohort Study; Note 4: Study of the health of survivors of the Great East Japan Earthquake

Figure 1.2-8 Current number of cohort participants (n = 54)

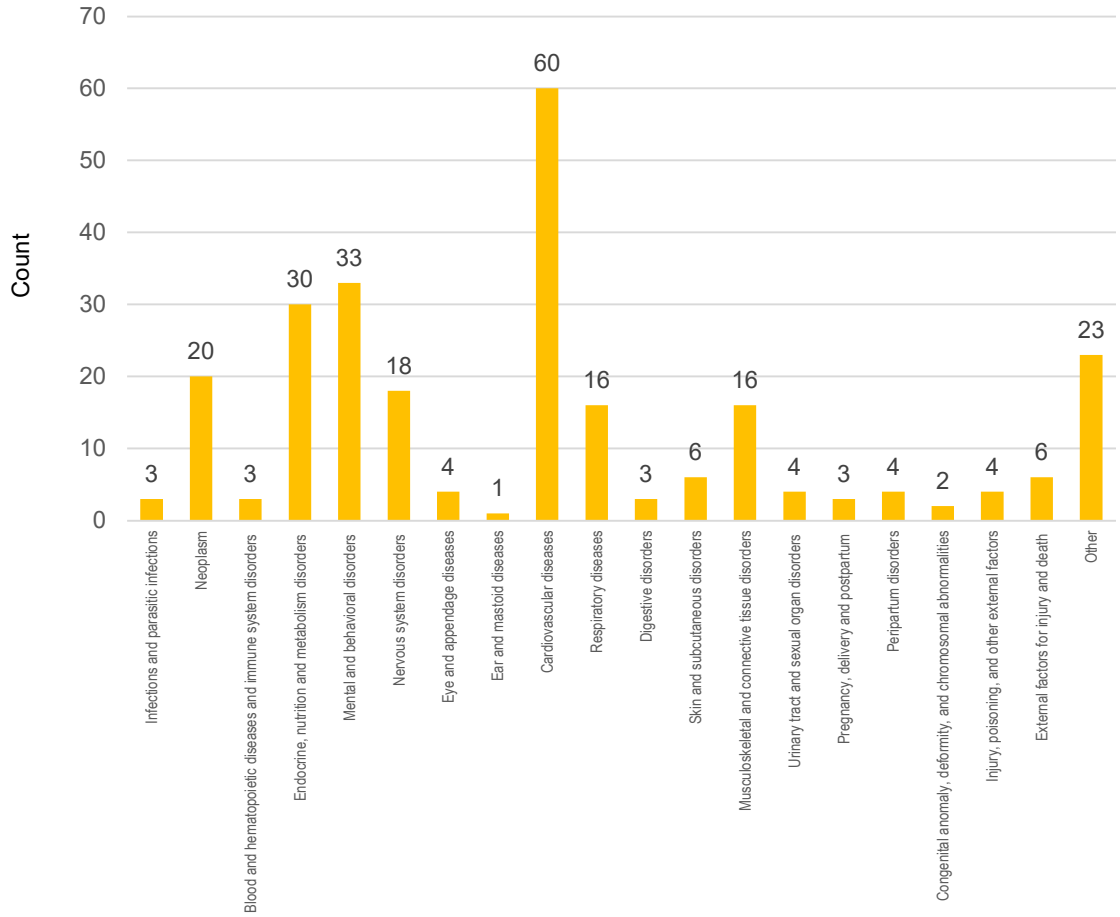


7) Target outcomes

The most common target outcomes were cardiovascular diseases (n = 60), followed by mental and behavioral disorders (n = 33) and endocrine, nutrition and metabolism disorders (n = 30).

Only one cohort study targeted ear and mastoid diseases. In addition, only two studies investigated congenital deformity or chromosome disorders.

Figure 1.2-9 Target outcomes in cohorts (multiple selections, n = 69)



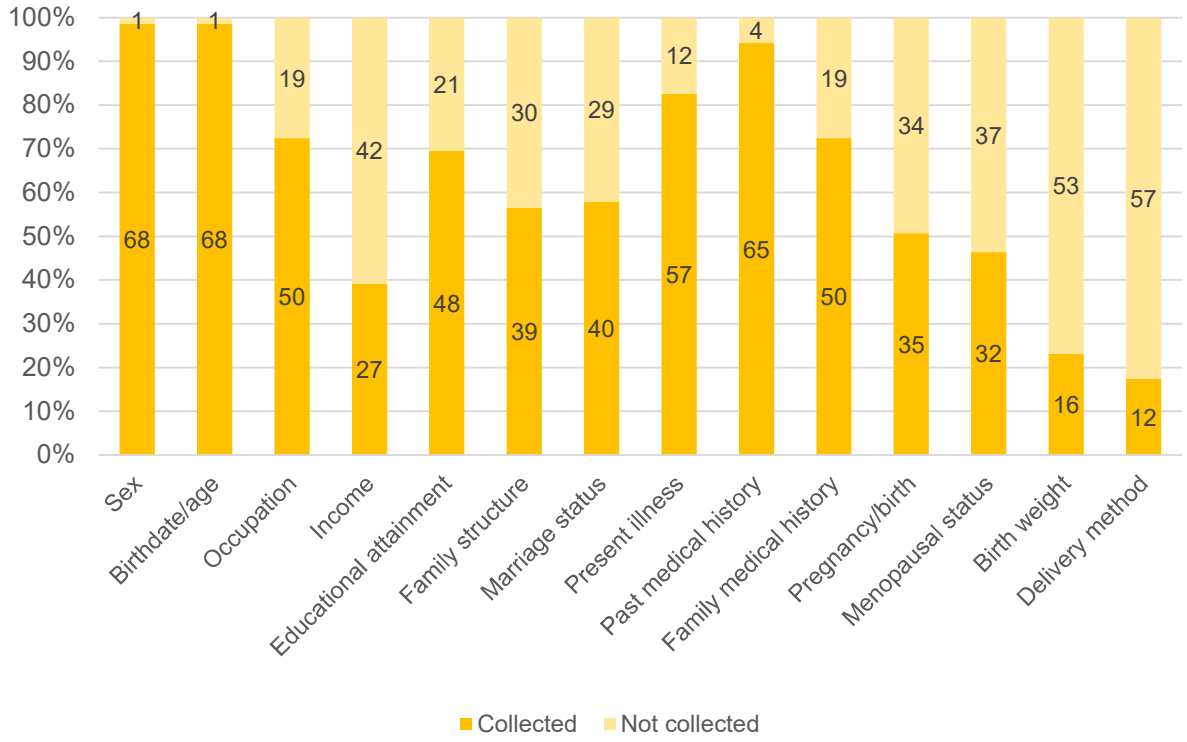
8) Variables collected

◆ General information of participants

For the variables collected by individual studies, almost 100% of cohort studies collected data for gender and birthdate/age (n = 68 each); ≥90% for past medical history (n = 65); about 80% for present illness (n = 57); about 70% for occupation; family history and educational attainment (n = 50, 50, and 48, respectively); and about 60% for family structure and marital status (n = 39 and n = 40, respectively). Only 40% of cohort studies collected data for income (n = 27).

With female participants, about half of the cohorts collected data for history of pregnancy/birth and menopausal status (n = 35 and n = 32, respectively). With infants and children, only about 20% of the cohorts collected data for birth weight and delivery method (n = 16 and 12, respectively).

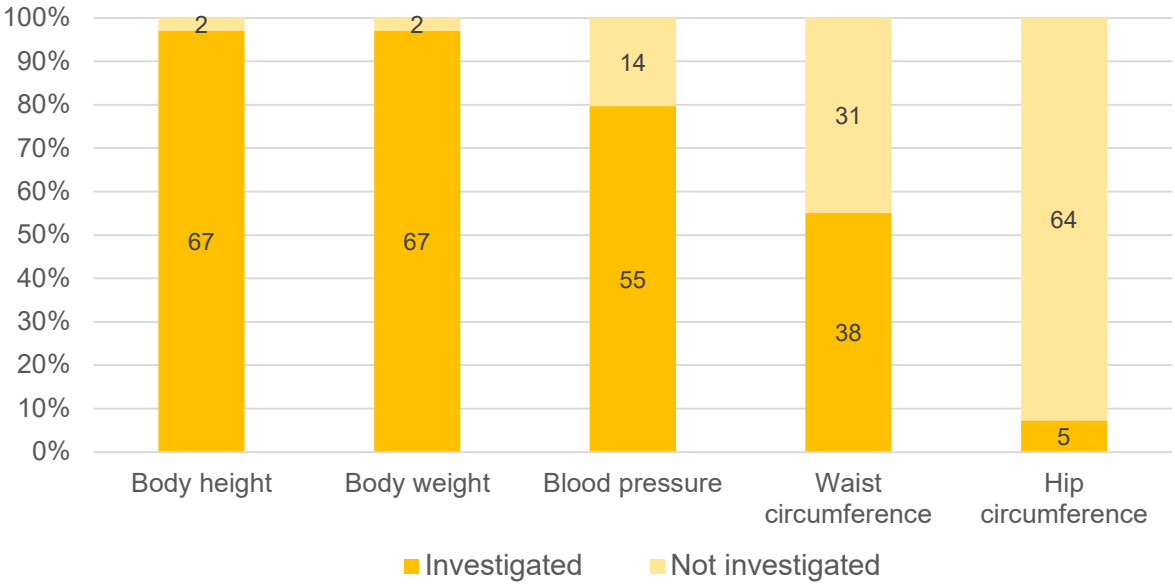
Figure 1.2-10 Basic characteristics information collected by cohorts (n = 69)



◆ Anthropometric and blood pressure measurements

Almost 97% of cohort studies collected data for height and weight (n = 67 each); about 80% measured blood pressure (n = 55); and about 50% for waist circumference (n = 38). Less than 10% of cohort studies collected data for hip circumference (n = 5).

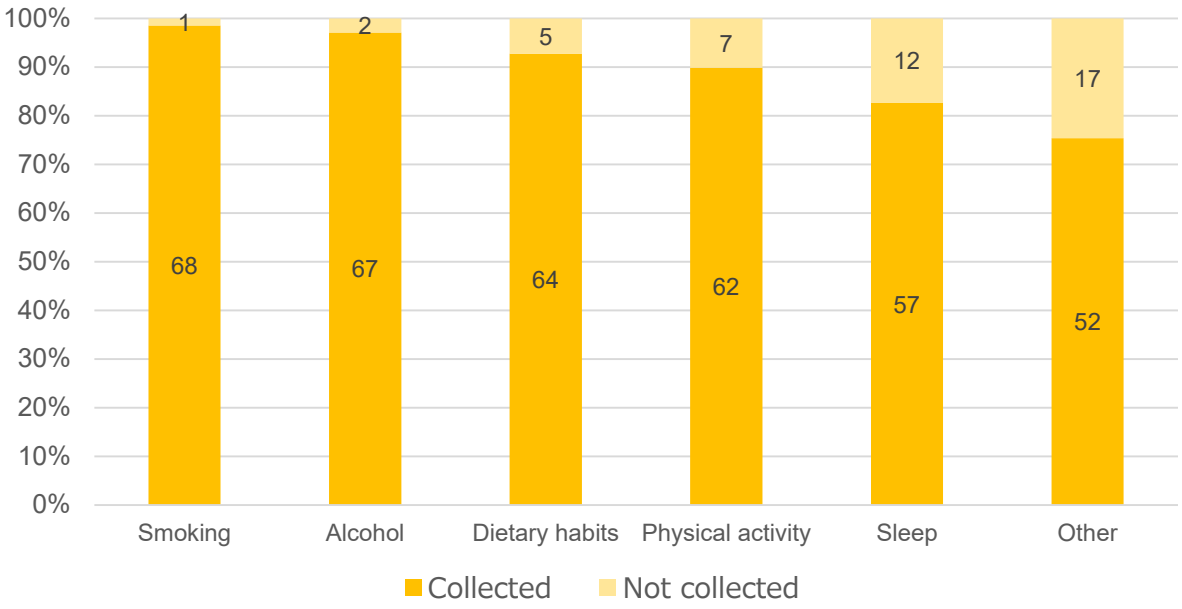
Figure 1.2-11 Proportion of cohorts collecting anthropometric and blood pressure variables (n = 69)



◆ Lifestyle variables

Most cohort studies collected data on lifestyle habits. For all lifestyle categories, more than 80% of studies collected information: about 99% for smoking (n = 68); about 97% for alcohol (n = 67); about 93% for dietary habits (n = 64); about 90% for physical activity (n = 62); and about 83% for sleep (n = 57). About 80% of the cohort studies collected information on other lifestyle habits (n = 52).

Figure 1.2-12 Proportion of cohorts collecting lifestyle variables (n = 69)

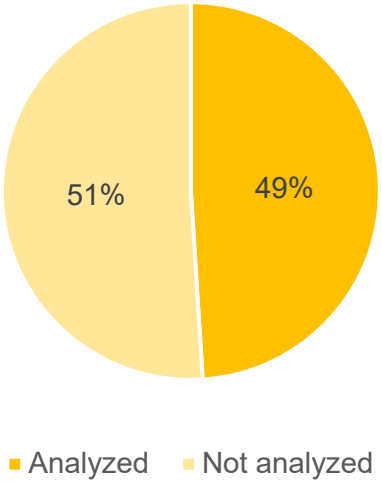


9) Genomic analysis

◆ Cohorts with genomic analysis

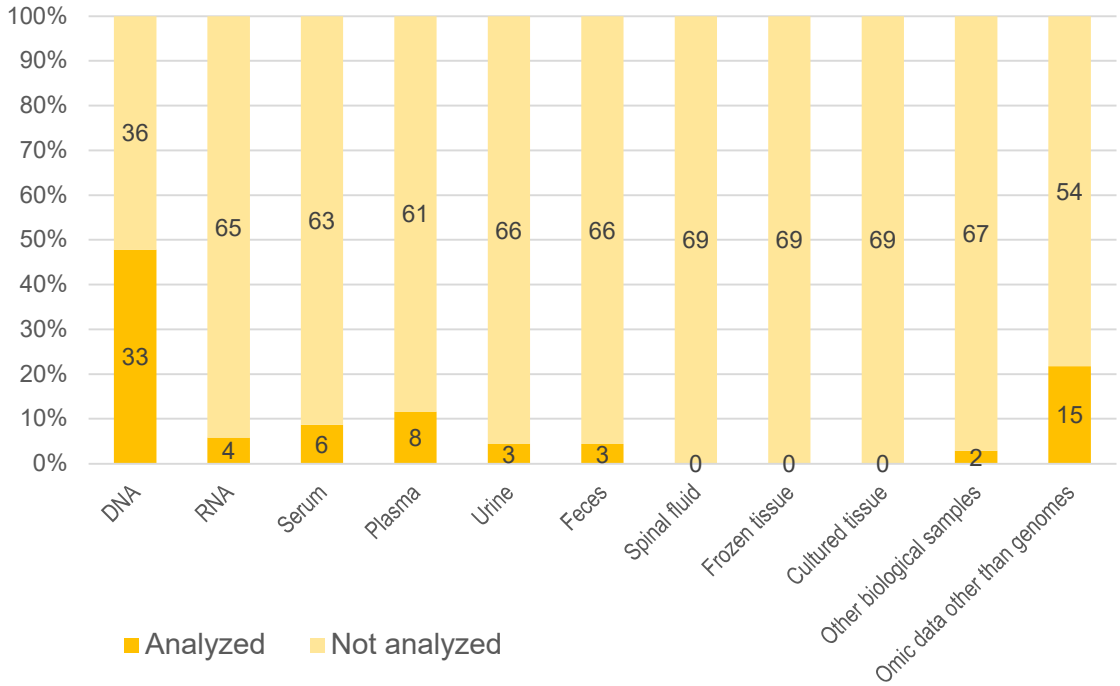
About 49% of the cohort studies collected genomic data (n = 34).

Figure 1.2-132 Analysis of genomic data (n = 69)



The most common sample used for genomic analyses was DNA (about 50%; n = 33).

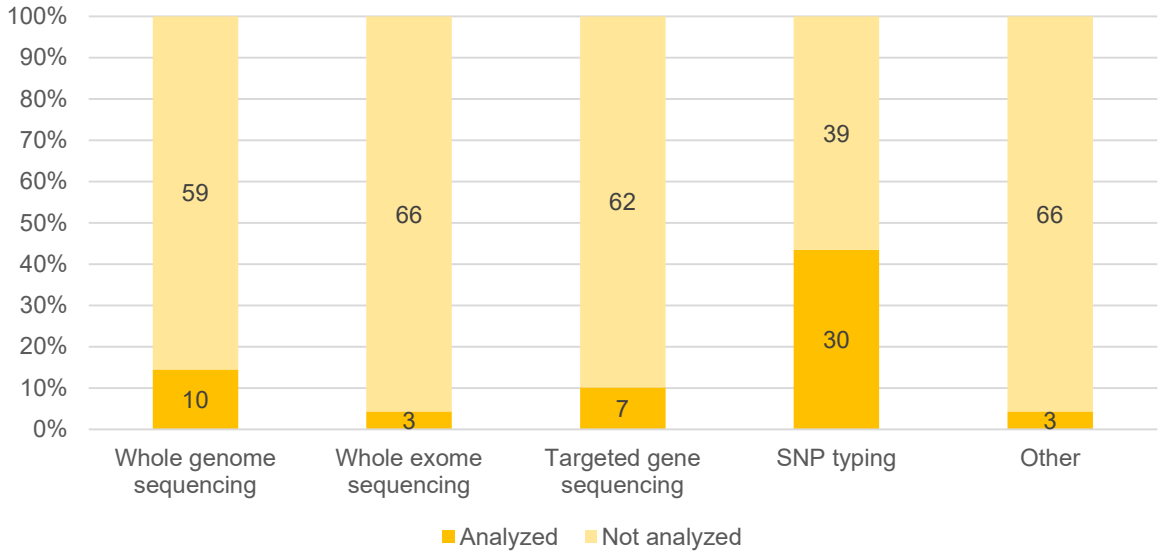
Figure 1.2-14 Samples used for genomic analyses (n = 69)



◆ Genomic analysis methods

The most common method of genomic analysis was SNP typing, performed in about 43% of the cohort studies (n = 30), followed by whole genome sequencing (about 10%; n = 10) and targeted gene sequencing (about 10%; n = 7). Only about 5% of the cohort studies performed whole exome sequencing (n = 3).

Figure 1.2-153 Proportion of cohorts with genomic data analysis, by method (n = 69)

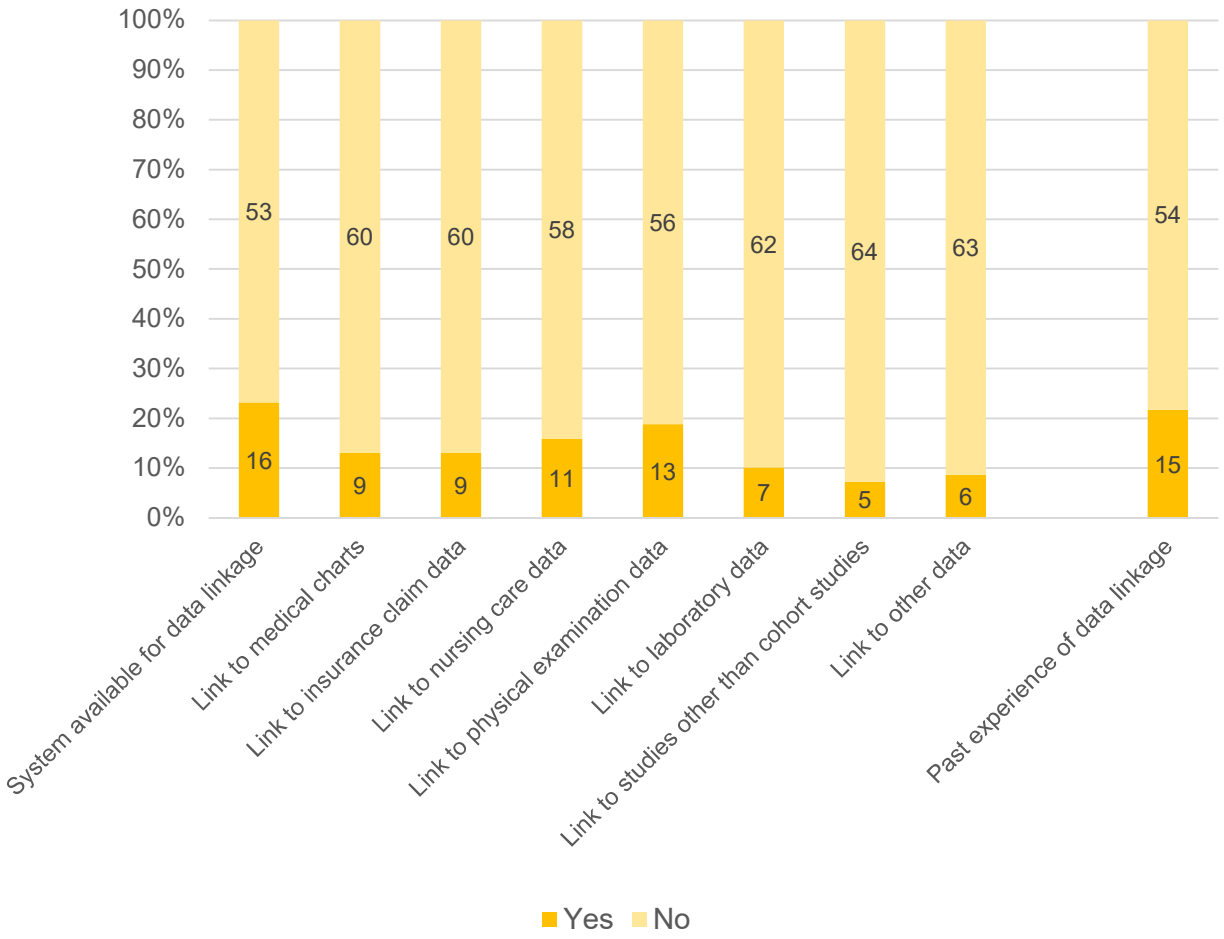


10) Data linkage

Of the 69 cohort studies, 16 studies had prepared for electronical data linkage to other data sources. Of these studies, systems were well established for data linkage to health checkup data (n = 13) and nursing care insurance claims data (n = 11).

Of the 16 cohorts with an established system for data linkage, 15 studies had already linked to different data sources.

Figure 1.2-16 Proportion of cohorts with data linkage (n = 69)

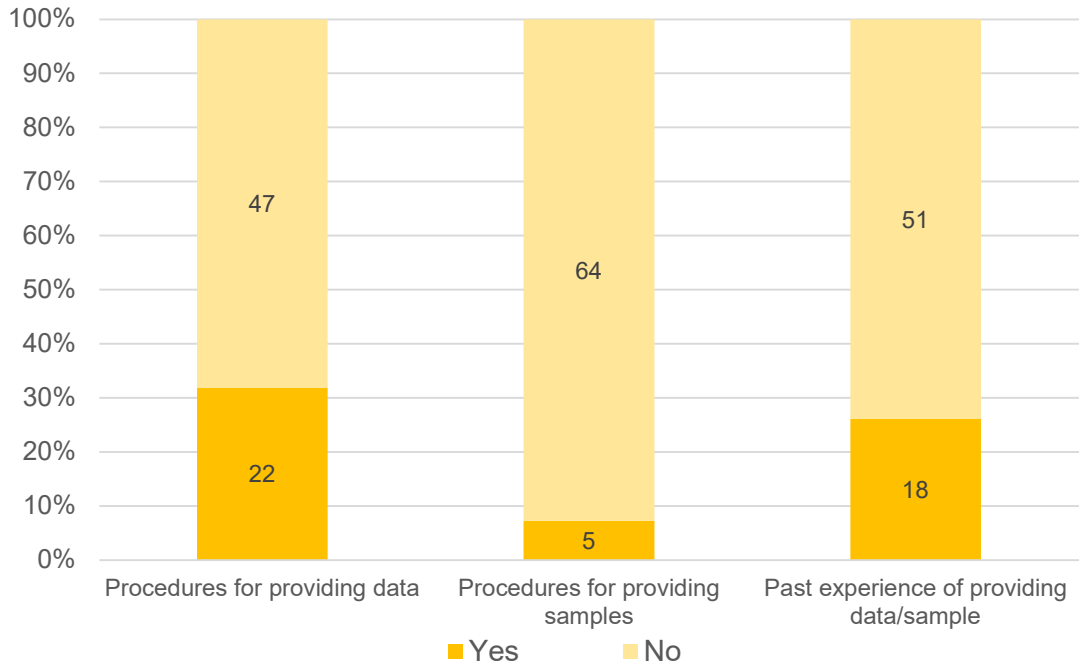


11) Data provision and sharing procedures

Of the 69 cohorts, 22 cohorts had procedures for providing data to collaborators outside the cohort. Five cohorts had procedures for providing samples to collaborators outside the cohort. Of these cohorts with data sharing procedures established, 18 had already experienced providing data and samples to others.

The most common recipients of such data and samples were academic research organizations (n = 18), private companies (n = 4), and government agencies (n = 3).

Figure 1.2-174 Proportion of cohorts with data/sample provision (n = 69)

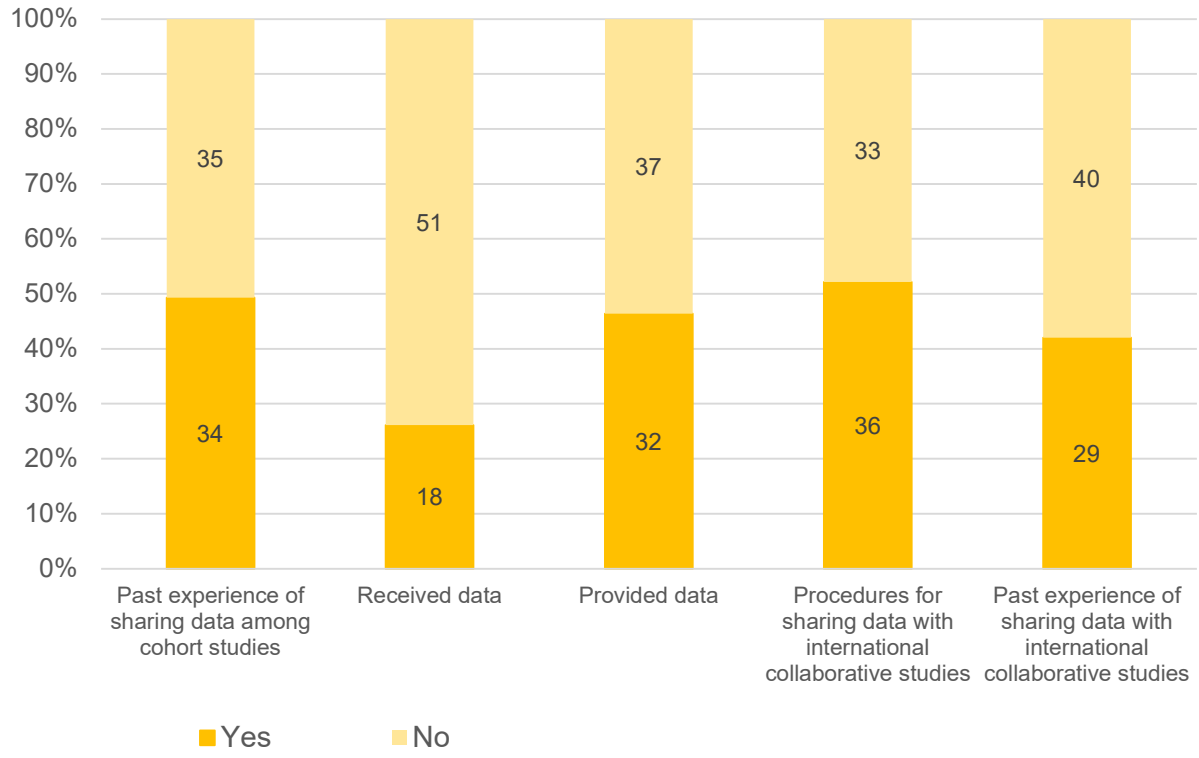


12) Data sharing

Of the 69 cohort studies, 34 shared data with other cohort studies. Of these, 18 studies received data, while 32 studies provided data.

Thirty-six cohort studies prepared for sharing data with international collaborative studies. Of these, 29 studies shared data with international collabolative studies.

Figure 1.2-18 Proportion of cohorts with data sharing (n=69)



1.3 Interviews to cohort investigators

1.3.1 Design

(1) Objective

Of the cohorts identified via the online survey, to document the current status and elucidate challenges, we interviewed researchers mainly involved in the cohorts with genomic data or prospective cohort studies involving data linkage and/or sharing.

(2) Methods

1) Target cohorts of the interview

- The interviews targeted cohorts analyzing genomic/omics samples and data, cohort studies with data linkage and/or data sharing.

2) Data linkage and sharing

- In this survey, data linkage and sharing are defined as follows:
  Data linkage: Information is matched from at least two independent data sources for the same attributes, such as individuals, families, events, and locations (excluding data transfer among sources).
- Data sharing: Information is made available for other users by transferring data from one organization/department to another organization/department (including data transfer among sources).

3) Items investigated

Figure 1.3-1 shows three groups of items investigated by the interviews: collection and management of genomic and omics samples and data, data sharing, and data linkage.

Figure 1.3-1 Interview items

Main items	Analysis of genomic/omics samples and data	Data sharing	Data linkage
	• Study background, objective, and target diseases • Sample storage (location, methods, etc.) • Cost of sample storage • State and cost of sample analysis • Data management methods • Future needs and challenges	• Timeline for establishing study procedures • Informed consent • Study implementation and operation • Steps for providing and sharing data • Contracts with data providers • Providing data for international collaborative studies • Difficulties and issues • Future needs and challenges	• Timeline for establishing study procedures • Summary of achievable project topics • Contents and destination of linked data • Steps for data to be linked • Contracts with data providers • Informed consent

1.3.2 Findings

(1) Interview participants

Of the cohorts interviewed, the following three birth cohorts were included: the Babies and Their Parents' Longitudinal Observation at Suzuki Memorial Hospital in the Intrauterine Period (BOSHI study), the Tohoku Medical Megabank (TMM) BirThree Cohort Study, and the Japan Environment and Children's Study (JECS).

The following eight cohorts of middle- to older-aged population were included: Circulatory Risk in Communities Study (CIRCS); NIPPON DATA 80, 90 & 2010; Japan Public Health Center-based Prospective Study for the Next Generation (JPHC-NEXT); Ibaraki Prefectural Health Study (IPHS): Follow-up study on life prognosis for people receiving medical checkups; Large-scale cohort study of physical well-being, preventive nursing care, and healthcare cost improvements; Japan Multi-Institutional Collaborative Cohort Study (J-MICC study); Tohoku Medical Megabank Community Cohort Study (TMM CommCohort Study); Hisayama Project/Japan Prospective Studies Collaboration for Aging and Dementia (JPSC-AD); Japan Gerontological Evaluation Study (JAGES).

The following nine studies were included as cohorts collecting biological samples: the Babies and Their Parents' Longitudinal Observation at Suzuki Memorial Hospital in the Intrauterine Period (BOSHI study); Tohoku Medical Megabank (TMM) BirThree Cohort Study (TMM BirThree Cohort Study); Japan Environment and Children's Study (JECS); Circulatory Risk in Communities Study (CIRCS); NIPPON DATA 80, 90, & 2010; Japan Public Health Center-based Prospective Study for the Next Generation (JPHC Study, JPHC-NEXT); Japan Multi-Institutional Collaborative Cohort Study (J-MICC study); Tohoku Medical Megabank Community Cohort Study (TMM CommCohort Study); Hisayama Project/Japan Prospective Studies Collaboration for Aging and Dementia (JPSC-AD).

Table 1.3-1 Cohorts interviewed and their features

No	Cohort	Main features
1	Babies and Their Parents' Longitudinal Observation in Suzuki Memorial Hospital in the Intrauterine Period (BOSHI study)	Involved pregnant women and following up from fetal period (currently focused on linking to school health checkup data). At the baseline, the study included participants across three generations (fetus, parents, and grandparents).
2	Tohoku Medical Megabank (TMM) BirThree Cohort Study	World's first three-generation cohort involving pregnant woman and gathering information on fetuses, parents, and grandparents; investigated data on heritable traits from three generations by comprehensively analyzing study-specific items, infant health examinations, school health checkups, childhood chronic disease registry, designated intractable disease data, cancer registry, and insurance claims data.
3	Japan Environment and Children's Study (JECS)	At 15 sites in Japan, involved pregnant women and following up from the fetal period to investigate the relationship between environmental factors and health. Plans call for establishing a biobank for study data and biological samples.
4	Circulatory Risk in Communities Study (CIRCS)	At five sites in Japan, investigating cardiovascular disease as the main outcome by analyzing health examination data and medical records from medical institutions.
5	NIPPON DATA 80, 90, & 2010	The cohort consists of participants in the National Cardiovascular Survey and the National Health and Nutrition Survay (former the National Nutrition Survey) conducted by the national government; the study analyzes data by linking information from the national surveys and follow-up information specific to the cohort.
6	Japan Public Health Center-based Prospective Study for the Next Generation (JPHC Study, JPHC-NEXT)	Subscribers of the National Health Insurance in five or six areas in Japan were recruited and thus the study has been maintained in collaboration with municipalities. Data integration of genomic data has been conducted in different cohorts.
7	Ibaraki Prefectural Health Study (IPHS) - Follow-up study on life prognosis for people receiving medical checkups - Large-scale cohort study of physical well-being, preventive nursing care, and healthcare cost improvement	Ibaraki Prefecture launched this cohort. Data analysis is conducted by linking health examination, nursing care, and insurance claims data.

8	Japan Multi-Institutional Collaborative Cohort Study (J-MICC study)	Started in 2005 as a multiple-site study with the main focus on cancer; collected biological samples of about 100,000 participants and analyzed genomic data.
9	Tohoku Medical Megabank Community Cohort Study (TMM CommCohort Study)	The objective is to find clues for the recovery from the Great East Japan Earthquake and next-generation health, establish a data bank and provide data to other parties. In follow-up examinations, interviews are employed to gather detailed information as well as analyzing health examination, nursing care, and insurance claims data.
10	Hisayama Project/ Japan Prospective Studies Collaboration for Aging and Dementia (JPSC-AD)	Undertaken at eight sites in Japan based on sharing and standardizing know-how and protocols from on the long-running Hisayama Project.
11	Japan Gerontological Evaluation Study (JAGES)	Conducted simultaneously with investigations of preventive nursing care and living environments performed by municipalities every three years.

(2) Genomic samples and data

1) Genomic samples and data in cohorts

- In the cohorts interviewed, the processes of sample collection were thoroughly managed, from collection to storage.
- Collaborative studies have been promoted by combining individual participant data (IPD) among different cohorts given that genomic data analyses require sample sizes in the millions. About research funds for genomic analyses, the costs associated with such analyses have declined. However, since sample maintenance and storage costs are fixed, sample analyses are proceeding while maintenance costs allow.
- Regarding management of samples and data in cohorts, efforts have been made for sample and data management system establishment, quality management and backup system set-up. However, costs were high for building and maintaining such systems.

Box 1.3-1 Genomic sample collection and analyses and data usage and management

Sample collection
◆ **Thorough management of processes from sample collection to storage** • When involving multiple sites, the steps from sample collection to storage are normalized and agreed upon by those involved. • A rough timeline from sample collection to storage is recorded.
Sample analysis and data utilization
◆ **Need for analyses based on data sharing and collaborative study** • Genomic data analyses require sample counts in the millions. Combining samples from major cohorts in Japan provides a pool of some 450,000 samples; this is comparable to major biobanks in other countries, which currently hold data from about 500,000 samples. It is needed to organize Japanese cohort studies in the future.

- Studies have started to integrate their data using IPD among collaborative cohort studies.

◆ **Research funds for sample analyses**

- The cost of analysis has declined since the time cohort studies were started.
- Since sample maintenance and storage costs are fixed, sample analyses are proceeding while maintenance costs allow.

Sample and data management

◆ **Setup of sample and data management systems and quality management**

- To ensure proper temperature management, systems are in place for automatically displaying freezer temperatures and for sounding alarms.
- Security systems are developed such as security cameras.

◆ **Backup system**

- Following collection, biological samples are divided and distributed to sites and the central office to serve as backups.
- Analysis data is also distributed to each site to serve as backups.

◆ **Costs**

- Annual electricity costs for each 700-liter sample freezer amount to several thousand dollars.
- Sample freezers must be replaced every 10 years (15,000 to 20,000 USD/freezer).
- Spaces to store samples are rented; sample storage capacity is shrinking.

2) Challenges and needs

- About the challenges related to sample collection, the difficulty of standardizing the processes for collecting and storing data from multiple study sites was identified. While standardized rules exist for collecting and storing biological samples, it is difficult to completely normalize the process due to varying operational and environmental factors at various study sites.
- About the challenges related to sample analysis and data usage: some interviewees found it difficult to deal with different analysis protocols for data integration. It will be necessary for experts to discuss and determine how genomic data can be integrated among different cohorts with different analytic methods, and then to manage accuracy by taking into account measurement errors attributable to aspects related to measurement by various individuals. As a challenge related to regulations for genomic data, it was suggested to address issues associated with providing genomic data overseas since the revised Act on the Protection of Personal Information defines genomic data as personally identifying information.
- Regarding sample and data management, interviews revealed that there is a widespread need for developing public data/samplng management functionalities.

Box 1.3-2 Challenges and needs related to genomic samples and data

Sample collection
◆ **Difficulty of standardizing processes for collecting and storing data from multiple locations** • While standardized rules exist for collecting and storing biological samples, it is difficult to completely normalize this process due to varying operational and environmental factors among various study sites. There are slightly different limitations in collecting and storing samples at different sampling sites.
Sample analysis and data utilization
◆ **Dealing with the differences in analysis protocols for data integration** • For multi-center studies, implementing the same analysis protocol poses challenges associated with differences in interpretation and policies among institutional review boards. • Even for measurements made with the same equipment, data can vary due to the habits and tendencies of the individuals performing the measurements. Accuracy control is necessary. • When integrating genomic data among cohorts, experts must discuss and determine how genomic data can be integrated among cohorts with different analytic methods. One option would be to normalize measurements and analyses by taking into account study objectives and plans. ◆ **Regulations for handling genomic data** • The revised Act on the Protection of Personal Information defines genomic data as personally identifying information; various issues associated with providing genomic data overseas must be addressed.
Sample and data management
◆ **Development of public data management functions** • Due to the high volumes of data size, storing/managing genomic and omics data requires cost and effort. Ideally, public databases would be used to store and manage genomic and omics data. * Separate discussion should be required before disclosing and making data available publicly. ◆ **Development of public sampling management functions** • It could be possible for public institutions to install and operate sample storage and for researchers to pay for the services. This would be particularly beneficial for small cohorts for which developing a new sample management system is challenging. • When a study is no longer funded, its samples are typically destroyed. It could be possible for public institutions to retain such samples with or without fee to allow effective reuse of resources.

(3) State of data sharing

1) Informed consent

- Informed consent from cohort participants is essential to provide and share data with other parties. The interview found that the Act on the Protection of Personal Information changed how informed consent was obtained.
- Following the enactment of the Act on the Protection of Personal Information, informed consent was obtained differently for adults and for children. Studies that established a data/sample bank obtained informed consent to provide data and samples with other parties.

Table 1.3-2 Informed consent for study participation

Start of study	Study type	Methods to obtain informed consent
Study launched before the Act on the Protection of Personal Information		• Informed consent was not obtained. • Following the Act on the Protection of Personal Information coming into force, data have been made available for collaborative studies by employing opt-out.
Study launched after the Act on the Protection of Personal Information	Studies involving children	• Informed consent is not obtained directly from children, but child-friendly documents are used to explain about the research. Based on best practices for birth cohort studies in other countries, the parties involved are in the process of discussing how such studies can improve consent materials for children. • For studies of participants aged eight and older, plans are formulated to obtain individual informed consent from both children and their parents.
	Studies of adults	• Individual informed consent is obtained (a written informed consent is obtained when linking data to a public database or when a genomic analysis is conducted). • How informed consent is obtained for a potential usage of data among collaborative studies varied among cohorts. • Participants can opt out for events such as establishing a new procedure for linking data for collaborative cohort studies.
	Studies with data/bio-bank functions and share samples and information	• Comprehensive informed consent, including sample and information provision, is obtained.

2) Implementation of data sharing

- Within a framework of collaboration, many cohort studies shared data and samples with collaborators.
- Data and samples were provided to other parties only in two Tohoku Medical Megabank cohorts. The Japan Environment and Children's Study was in the process of setting up a data and sample bank.
- None of the cohorts have provided data and samples for use in overseas studies.

Table 1.3-3 State of data sharing

Type of data sharing	Data sharing cohort studies
Sharing data and samples within collaborative studies	• Japan Gerontological Evaluation Study (no biological samples) • Ibaraki Prefectural Health Study • NIPPON DATA • JPHC-NEXT • Japan Multi-Institutional Collaborative Cohort Study • Circulatory Risk in Communities Study • Babies and Their Parents' Longitudinal Observation in Suzuki Memorial Hospital in the Intrauterine Period • Hisayama Project
Providing data and samples with researchers in Japan	• Tohoku Medical Megabank Community Cohort Study • Tohoku Medical Megabank (TMM) BirThree Cohort Study • Japan Environment and Children's Study (currently in the process of digitizing data and samples)
Providing data and samples outside of Japan	* None of the cohorts had actually provided data outside of Japan while informed consent allowed providing data overseas.

3) Contact information for data sharing

Box 1.3-3 Current state of data delivery, collaborative studies, and data sharing

Current state of data provision, collaborative studies, and data sharing
◆ **Main collaborators** ・ Many cohorts collaborated with research organizations. Some cohorts collaborated with private companies. • **Collaboration with private companies** ・ Since informed consent was obtained for the purpose of academic research, contracts are signed with private companies within the framework permitted. Except for special cases (e.g., using AI), companies usually send their researchers to the lab in the university (the main site of the cohort); thus, the data cannot be transferred to the companies involved. ・ Grants are available for collaborative studies as part of academic-industrial alliances. ・ In collaborations with private companies, cohorts carefully assess potential conflicts of interest and define specific purpose of opportunities in the studies, such as the availability of special technical resources. ・ Few pharmaceutical companies have sought collaboration due to the nature of population-based studies of local residents.

4) Challenges for promoting data sharing

・ We assessed the challenges in promoting data sharing from the following seven perspectives: purpose and definition of data sharing; informed consent; balance with study policies; scientific and epidemiological challenges; study quality management; investigator network and relationships; and research funding sources.

Box 1.3-4 Challenges in promoting data sharing

Purpose and definition of data sharing
◆ **Need to clarify the purpose of data sharing and inform investigators** • Investigators in cohorts may be hesitant and unwilling to share their data because the purpose and definition of data sharing are unclear. They need clear definitions and understanding of data sharing objectives. ◆ **Need to define data sharing** • It is needed to define what the data sharing aims to achieve 1) Research/analysis: Framework for collaborative studies and for providing data and samples to third parties 2) Type of data provided: IPD data, analysis data (for meta-analyses)
Informed consent
• **Current approach to obtaining informed consent** 1) When enrolling participants: Individual informed consent ・ No comprehensive consent for providing data and samples to third parties 2) In the event of changes (e.g., a new study objective) following informed consent at recruitment: Participants have the option of opt-out. 3) When a study anticipates subdividing data and samples at the planning stage: Comprehensive informed consent • **Future approach to obtaining informed consent** • Dynamic informed consent enables ongoing engagement from initial to subsequent studies.[2)] This approach, when using ICT, makes consent easier to obtain. However, since middle aged and elderly participants respond poorly to online surveys, such approach may be difficult to implement in Japan at this time.

[2)] Mizuki Morita: Patient-centered medical and health information management and dynamic informed consent, Journal of Information Processing and Management, Vol. 57, No. 1, pp. 3-11, 2014.

Balance with study policies

- Concerns regarding data provision include the possibility that certain studies could publish results in ways departing from the intentions of the cohort study that owns the original data. The investigators in cohorts are rather hesitant about data provision because they are supported by municipalities or othe public organizations in conducting their studies.
- As is done in some countries, such issues could be determined by data access committees. However, no such system is currently in place in Japan.

Scientific and epidemiological challenges

◆ **Investigating data integration objectives and methods**

- Linking data among ongoing cohorts may result in the merging of heterogenous data, creating large data volumes with numerous errors. While there are some examples that minimized the effects of errors by increasing sample sizes, integrating data haphazardly would not lead to innovative study outcomes.
- Even with small data volumes, meaningful results can be obtained if the error variance is minimal.
- Integrating heterogenous data from different cohorts could lead to biased results. The challenge is to investigate how to integrate data in an effective way.

◆ **Data integration based on least common denominator approach**

- One way to integrate data from ongoing cohort studies is the least common denominator approach, whereby only common items are integrated. With this approach, it is important to address issues such as lack of representativeness of the population and biases in individual study design.

Study quality management

◆ **Quality management of studies**

- The quality of data analysis in collaborating studies is managed by reviewing results from multiple angles by the investigators involved and by confirming reproducibility.
- The quality of the study must be maintained when providing data to third parties. One reason data may be withheld from third parties is that the data provider cannot be responsible for study findings published without their input on data analysis.

◆ **Understanding data characteristics**

- IPD is provided only if the involvement of an investigator who understands the data properties can be assured. As such, data is provided within the framework of collaborative studies.

Investigator network and relationships

◆ **Effective utilization of data**

- Many investigators welcome a broader range of use of data within the framework of collaborative studies. They think that more utilization of data should be encouraged.
- However, without incentives, investigators are hesitant to provide to third parties their data obtained with significant time and effort to refine accuracy or associated with high cost of sample maintenance and management.

◆ **Incentives to share data**

- Each cohort places efforts on strictly maitaining the accuracy and quality of data. No clear benefits accrue from data linkage with another cohort. Investigators are concerned that linking heterogenous data could negatively impact data quality.

◆ **Trust**

Cohort studies are established based on mutual trust among a range of various stakeholders, including participants, collaborating organizations and municipalities. This relationship must be extended when providing data to third parties. Providing data without this trust may lead to loss of integrity.
Research funding sources
• Research funds tend to be compartmentalized at different ministries and departments. Linkage of cohorts with the same funding source is often highlighted, but cross-sectional linkage across different funding sources needs to be promoted to encourage more research activities.

(4) Data linkage

1) Data linkage to public databases used in cohort studies

- Many cohorts have linked their data to public databases owned or managed by the national government, municipalities, or academic societies.
- Infant health examination and school health checkup data is not digitized; these data are analyzed manually. Since each municipality manages Specific Health Checkup data, each municipality must be contacted and data to cohort study databases linked manually. The analysis of medical records gathers data such as the conditions associated with disease onset and diagnostic images. Even if these data were gathered electronically, only the data on disease development could be used.
- Many investigators have found that negotiating and obtaining approval to obtain data from data sources takes considerable time and effort.

Table 1.3-4 Data linkage to public databases used by cohorts

Data type	**Contracts with data management and provision bodies**	**Data linkage**
Infant health examination	By signing an agreement with each municipality for each study.	・ Information from maternal handbooks provided by participants is transcribed. * Infant health examinations are not digitized; electric linkage is not possible.
School health checkup	By signing an agreement with each municipality for each study, then contact each school board.	・ After signing a contract with each municipality, school health checkup data is transcribed. * School health checkups are not digitized; electric linkage is not possible.
Specific Health Checkup (“Metabo” health examinations)	By signing an agreement with each municipality.	・ Data is received electronically (e.g., in csv format) from each municipality via National Health Insurance. For each municipality, the data received is manually integrated with each study database.
Medical records	By contacting each municipality to issue a request	・ Ascertain disease development from patients through medical history-taking,

	for study participation to medical institutions.	questionnaires, and telephone interviews; obtain information related to disease development (e.g., diagnostic images) from medical institutions. * This data is not digitized; electric linkage is not possible. ・ Developing a system that extracts participant data by matching participants to annual admission lists is now in progress at large hospitals. While the system is ready to collect digital medical chart information, the data is currently used only to determine disease diagnosis.
Removal from local resident registries (relocation, death)	By obtaining approval/consent from each participating municipality.	・ After signing a contract with each municipality, data required for the study is obtained from an agency (designated by a prefecture) contracted to manage resident registry data.
Vital statistics (death, cause of death)	By requesting approval from the national government.	・ For cohort participants removed from the resident registry, a request is made to access information related to death and cause of death. After receiving digital data (in csv format), the data is linked automatically based on dates of birth and death using special data linkage software.
Medical data, insurance claims data, and certification of nursing care	By signing a contract with each municipality every three years.	・ Electronic data (in csv format) received from each municipality via the National Health Insurance is linked automatically based on participant insurance ID and research ID (encrypted). Special data linkage software has been developed.
National Health and Nutrition Survey	-	・ Linkage with baseline cohort study data
Cancer registry	By applying to use data.	・ After obtaining individual consent, data is linked based on name, gender, date of birth, and address.

2) Study challenges

◆ Technical challenges

- We examined the various technical challenges associated with data linkages from three viewpoints: data reception, data integration, and data management.
- For data reception, data must be managed manually due to incomplete or insufficient digitization and normalization. For data integration, anonymized data is matched using multiple data points; however, a certain proportion of the data cannot be linked. For data management, data volumes and types are immense. Many investigators stated that data had not been sufficiently organized or utilized.

Table 1.3-5 Technical challenges for data linkage

Data reception	• Infant health examination and school health checkup data is generally not digitized. It is labor intensive for researchers to physically go to where data is stored and manually transcribe target data. • Nursing care and insurance claims data can be received in KDB files as defined by each municipality, but data for medical records and local initiatives related to preventive care have not been standardized and are delivered in different formats. Thus, it takes a long time to clean and organize data with unmatched units (e.g., test data). • In ongoing cohorts of residents, some 70 to 80% of the labor involves setting up a system to collaborate with municipalities and other data providers. If this data were available in public databases, investigators could focus more time on actual research activities.
Data integration	• When linking death data (anonymized vital statistic data provided by the national government) and nursing insurance data received from municipalities, multiple items, such as dates of death, are used. For large municipalities, linking is not possible for several percent of the cases due to overlapping identifier data. • A system for electronically collected digitized medical records has been developed, but this has mostly been used to determine diagnoses due to the lack of other useful information.
Data management	• When linking death data (anonymized vital statistic data provided by the national government) and nursing insurance data received from municipalities, multiple items, such as dates of death, are used. For large municipalities, linking is not possible for several percent of the cases due to overlapping identifier data. • The volume and range of the data makes it difficult to organize and maximize its utilization.

◆ Procedural challenges

- We examined procedural challenges associated with data linkage from three viewpoints: data reception, data integration, and data management.
- For data reception, it takes a long time to receive the data after applying to use it. Contacting each data source, such as a municipality, is labor intensive.
- For data integration, it is vital to obtain informed consent for linking to databases that would be built and maintained in the future.
- For data management, while informed consent was obtained for data usage, rules must be established on providing and sharing insurance claims data due to the range of information found in the claims data.

Table 1.3-6 Procedural challenges for linking information with cohorts

Data reception	• Fees apply when requesting copies of residence certificates (per certificate). • After making an application, it takes a long time to receive death data based on national vital statistics (about one year). • It is labor intensive to contact and work with each municipal school board to receive school health checkup data. Even if a municipality approves a cohort study, it is necessary to contact the school board separately. If more than one cohort study is underway at a municipality, it saves time and effort for studies to collaborate in contacting and dealing with the municipality.
Data integration	• For new cohort studies, it is preferable to obtain informed consent addressing the possibility of linking to databases that may be built in the future. However, this may not be approved by ethical review boards. One option is to obtain a separate consent after the start of the study, but this approach lowers the consent rate and should be avoided.
Data management	• Informed consent for data usage is obtained from participants. But because insurance claims data has many items and contains very sensitive information, the decision must be made whether all insurance claims data should be provided upon request. (For example, based on how insurance claims data is analyzed, it will be possible to assess and compare the quality of medical care among hospitals.)

◆ Challenges associated with the accuracy of data in public databases

・ More medical institutions have recently become less willing to provide their medical information related to disease diagnoses. Some investigators suggested that data in insurance claims could be better used for cohort studies if it could be shown that diagnoses in insurance claims data is epidemiologically valid, although diagnoses listed in insurance claims are clinical diagnoses and do not necessarily match standard epidemiological diagnoses.

(5) Other challenges and the future of cohort studies

・ We examined future challenges and the future of cohort studies from three viewpoints: development of investigators, future of cohort studies, and research funds and environment.

Figure 1.3-7 Other challenges and future of cohort studies

Development of human resources	• Conducting a cohort study requires human resources and long-term positions for researchers. • At present, project researchers in funded projects cannot participate in other research projects under the current rules. The broader involvement of young researchers in a wide range of research activities in other projects would offer them perspectives and experience which potentially advances their careers. • Data scientist training in epidemiology would benefit from targeting people with medical backgrounds, such as medical doctors and healthcare professionals. • Once big data stored in public databases becomes available, anyone will be able to access the data stored. However, study findings with reverse causality or biases could be published by those lacking epidemiologic expertise. Studies must involve not just statistical experts and data scientists, but also clinicians.
Future of cohort studies	◆ **Cross-disciplinary study** • Studies must draw on specialties within both basic and epidemiological research in a balanced way. ◆ **Future direction and practical applications** • Direct-to-consumer (DTC) genetic tests are being offered by private health examination organizations. Still, it is imperative for Japanese epidemiologists to analyze the data and samples currently available and to publish epidemiologic evidence for the Japanese public. • Conducting genomic analyses and comparisons across different groups of disease patterns may help find the disease mechanism, which may contribute to promoting precision public health. The goal is to achieve public health even targeting individuals, not large general populations as a whole.
Research funding and environment	• Small studies with high quality data and analysis can yield useful findings. Research funds should go to both small and large studies based on their strengths. • With the advances in data linkage, research funding should account for the

	cost of data cleansing. • Cohort studies would benefit from the establishment of an agency to oversee contracts and other operational tasks.

1.4 Summary of cohorts in Japan

1.4.1 Genomics/omics

- Around half of the responding cohorts in this survey collected genomic samples and data.
- In the future, for studies integrating genomic data from different cohorts, challenges were identified associated with maintaining measurement accuracy, managing errors, and integrating data analyzed by different methods. Experts will need to discuss and establish rules for integrating genomic data across the country.
- Building and maintaining sample and data management systems at each research organization requires financial and human resources. Many investigators have expressed the need to build public systems for managing genomic samples and data.

1.4.2 Data sharing

- Twenty-two cohorts established procedures for providing data to third parties outside the study or their collaborators. Thirty-four cohorts demonstrated data sharing among different cohorts. In response to questions, many investigators expressed positive opinions on data sharing and utilization but stated concerns regarding providing IPD from their cohorts to a third party or linking data with different cohorts.
- To further promote data sharing, the following issues should be addressed:
  1) The purpose and definition of data sharing should be discussed and determined so that investigators of cohorts accept the concept.
  2) The process of informed consent can be discussed so that consent can be obtained not just for specific objectives (as is the case with many existing cohorts), but for broader use of data and samples.
  3) It should be promoted to establish organizations/committees to organize study goals and quality across different data users in the process of data sharing.
  4) It would be helpful to establish a consortium consisting of expert epidemiologists to discuss methodologies for data matching among cohorts from multiple perspectives before implementing data sharing.

1.4.3 Data linkage

- Many cohort studies had started linking public databases with established systems for entering data, including health examination, nursing care, and insurance claims data. However, there are still some databases where data are transferred manually, such as infant health examinations and school health checkups.
- Plans call for allowing access to infant health examinations and metabolism checkups via the

MyNumber Portal. [3] If this system is made available for epidemiological studies, it may be possible to obtain and effectively utilize IPD through various life stages managed separately to date by different ministries/departments and regulations.

References

- Japan Prospective Studies Collaboration for Aging and Dementia (https://www.eph.med.kyushu-u.ac.jp/jpsc/)
- Mizuki Morita: Patient-centered medical and health information management and dynamic informed consent, Journal of Information Processing and Management, Vol. 57, No. 1, pp. 3-11, 2014.
- Ministry of Health, Labour and Welfare, Committee for Effective Usage of Health, Medical and Nursing Care Information (https://www.mhlw.go.jp/stf/newpage_09958.html)

[3] Ministry of Health, Labour and Welfare, Committee for Effective Usage of Health, Medical and Nursing Care Information (https://www.mhlw.go.jp/stf/newpage_09958.html)

Chapter 2. Survey of technical elements for information linkage in Japan and abroad

2.1 Design

2.1.1 Objective

As stated in Chapter 1, various environmental factors, including undigitized records, have hindered data linkages in existing cohorts. Chapter 2 discusses certain procedural elements, such as national ID systems, that promote information linking in some other countries.

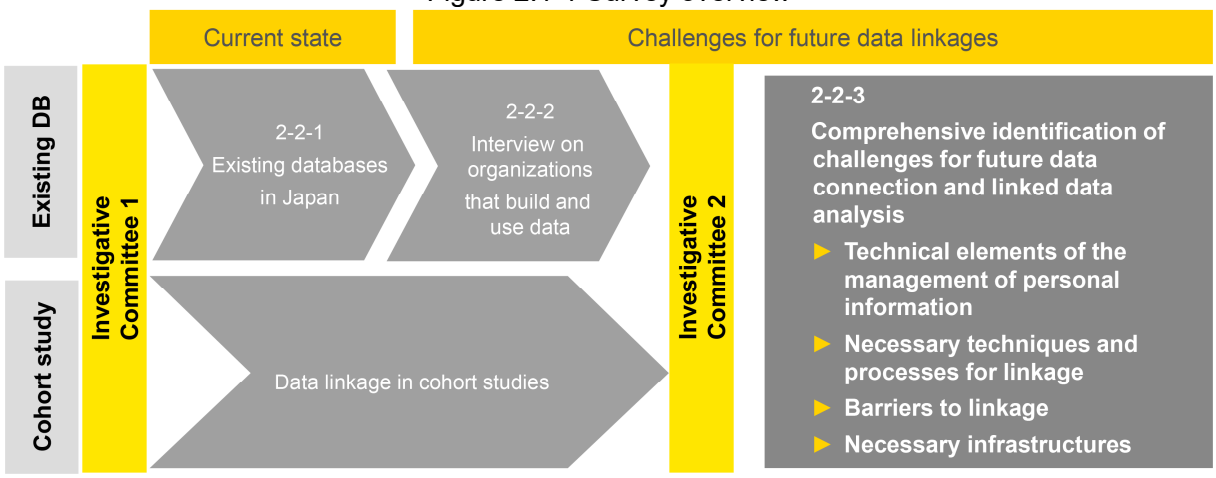
These factors are not independent. For example, national IDs enable data to match and link with other data more easily. Adopting certain procedural framework can lower technical barriers.

This chapter examines the technical elements of information linking from both technological and procedural aspects and clarifies the current and future state of information linkage in Japan.

2.1.2 Methods

When linking healthcare and medical information to cohort studies, not only cohort studies but also the structure of the target databases matter; therefore, this survey investigates both aspects.

Figure 2.1-1 Survey overview



This survey identified existing databases in Japan (2-2-1); interviewed representatives at organizations that develop and use data (2-2-2); then examined the state of information linkage in cohort studies to elucidate challenges on a comprehensive basis for future data connections and linked data analysis (2-2-3).

(1) Existing databases in Japan

Japan currently has or is building several public and private healthcare and medical databases, including the NDB and MID-NET.

This survey summarizes the types of information stored in the databases, potential identifiers for data linkage, and data usage. This survey design was confirmed by an oversight committee.

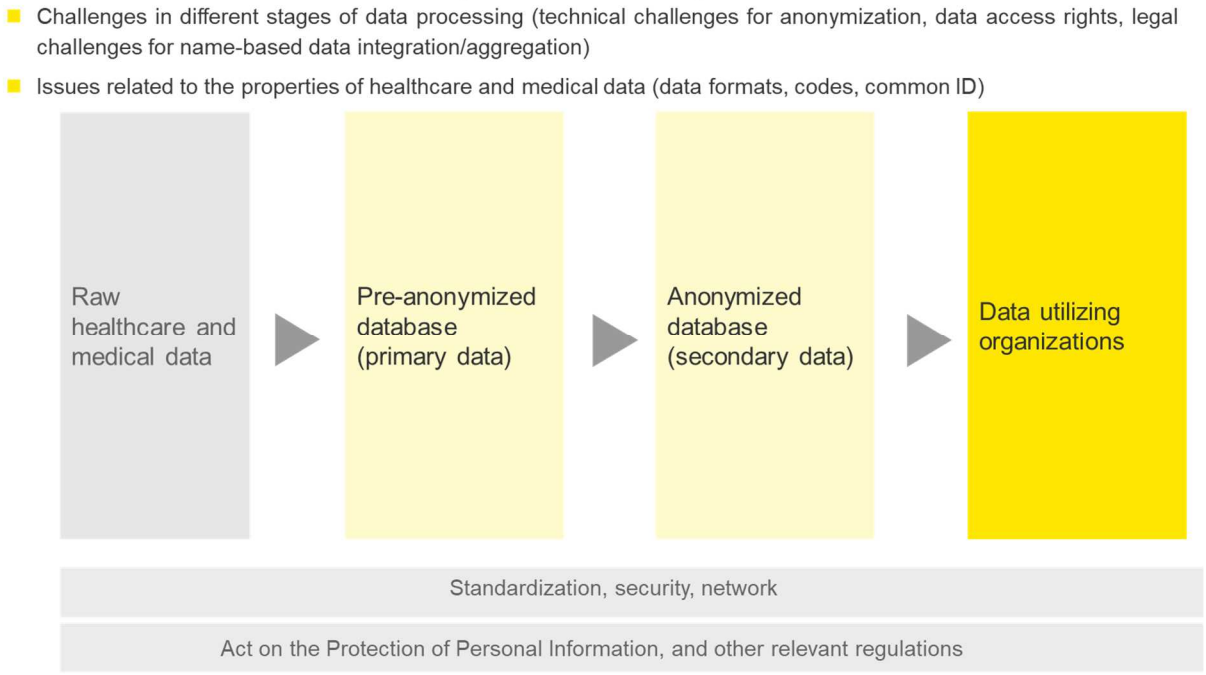
(2) Interviews of organizations that build and utilize data

Based on the results of these investigations of Japanese databases, we interviewed representatives from several organizations that build and use data.

1) Interview targets

First, we examined how raw healthcare and medicine data is collected, processed, and analyzed at each data handling stage, then elucidated technical and procedural challenges. We interviewed individuals with knowledge of and familiarity with the technical and procedural challenges to linking healthcare and medical data from the aspects of data building, usage, and legal framework.

Figure 2.1-2 Flow of healthcare and medical data processing, and technical and procedural challenges



Source: Prepared by EY

◆ Interview participants (in alphabetical order without titles)

- Eiji Ebina (Department of Health and Welfare, Tochigi Prefecture)
- Health Insurance Bureau (divisions in charge of NDB and DPC), Ministry of Health, Labour and Welfare of Japan *for email inquiries and document searches
- Masako Mizumachi (Miyauchi & Mizumachi IT Law Office)

- Tomohiro Sawa (Teikyo University)
- Hiroyuki Yoshihara (Life Data Initiative)
- Yoshiaki Uyama (Director, Division of MID-NET Operations, Office of Medical Informatics and Epidemiology, Pharmaceuticals & Medical Devices Agency (PMDA) of Japan)

2.2 Findings

2.2.1 Existing databases in Japan

Tables 2.2-1 and 2.2-2 show the databases in Japan identified in this survey. These are public databases discussed and considered for data linkage in recent years, as well as a private database (Life Data Initiative: LDI) built as a government-approved system for promoting the delivery of healthcare and medical information to third parties as part of the Next Generation Medical Infrastructure Law.

As shown in the “Relevant law and authority” column, public databases are governed by different laws and regulations. Pre-anonymized data is owned by different parties. This data is anonymized by different agencies. Subsequently, common identifiers are needed when linking databases with unique IDs. Many discussions on promoting information linkage have focused on utilizing insurance subscriber numbers. As part of ongoing discussion on data linkage, building databases with insurance subscriber numbers should make it possible in theory to link databases.

Each existing database has its own guidelines. At the time of this survey, public databases have been designed for governmental and research purposes, with prohibition of data linkage with personally identifying information. In other words, public databases have not not been designed for the purpose of data usage. Additionally, aggregation is not possible if both the insurer and name change.

In contrast, the LDI database has been built to provide data to third parties and makes it possible to gather information about an individual from different medical organizations. However, as with public databases, this process of gathering information about an individual from multiple sources is not automatic. It is currently conducted manually.

Without a national ID system or some method to identify each individual, it is difficult to match data about a certain individual using existing databases under the current regulatory environment.

Following legal revisions, some public databases are in the process of developing and achieving data linkages. With regard to identifier-related issues, a partial revision of the Health Insurance Act for proper and efficient operation of the medical insurance system plans to use subscriber numbers as personal identifiers. As an example, when a request is made for the archival records for a specific subscriber number in a public database, the system provides an answer concerning whether the information is ascribed to the same individual. This approach simply flags subscriber numbers as confirmed for personal identification; by taking privacy into consideration, subscriber numbers are matched every time an inquiry is made without creating and assigning a unique ID.

Merely linking data does not provide access to life course data. From birth to death, people undergo a wide range of healthcare examinations and checkups. When analyzing the relationship of these data with health and disease, various types of information must be linked. Table 2.2-3 summarizes major life course data, including common health examinations in Japan, as reference information for investigating the data that needs to be linked to allow effective use, including existing databases.

Table 2.2-1 Existing databases that could be linked with cohort study data (1 of 2)

DB	Relevant law and authority	Raw data	Anonymizing agency	Common identifier	Subscriber number as raw data	Data containing personal names	Notes
MID-NET	PMDA regulations	Medical charts, insurance claims, DPC	Each medical organization	Hospital patient ID (reception number)	Yes	No	
NDB	Act on Assurance of Medical Care for Elderly People	Insurance claims data, health examination and guidance data	Insurer or adjuster	ID1 = hash value based on Subscriber No: Card No + gender + birthday, ID2 = hash value based on name + gender + birthday	Yes	No	
MHLW DPC	Notice (Calculation of costs for treatments at wards of hospitals approved by the Minister of Health, Labour and Welfare)	DPC	Each medical institution	Hospital patient ID (reception number)	Yes	No	Anonymization methods may vary among medical institutions. No unified method has been proposed as standardization is difficult due to the nature of data.
Nursing care DB	Long-term care insurance act	Nursing care insurance claims data, nursing care certification data	Municipality or national health insurance organization	Nursing care subscriber number	Yes (nursing care subscriber number)	No	
Cancer registry DB	Act for promoting cancer registries	Specified data for registration, death information	National Cancer Research Center	Unique ID	No	Yes	
Designated intractable disease DB	—	Individual clinical chart	MHLW	Unique ID	No	Yes	
Specific pediatric chronic disease DB	—	Medical opinion data	MHLW	Unique ID	No	Yes	
Life Data Initiative DB	Next Generation Medical Infrastructure Law	Medical information, insurance claims data, DPC, employee health examination data, prescription data	Life Data Initiative (approved medical data anonymization agency)	Unique ID	Yes	Yes	Information about a particular person is obtained from different medical institutions (matched manually based on names in kanji characters, names in hiragana characters, gender, and birth dates)

Source: Publicly available information for each DB prepared based on interview findings

Table 2.2-2 Existing databases that could be linked with cohort study data (2 of 2)

DB	Stored data	Third party recipient	Data usage guidelines and regulations	Notes
MID-NET	Test data, prescription, and injection data	Yes (after 2018)	"Guidelines on the effective usage of MID-NET (Medical Information Database Network)" Able to provide to third parties	Distributed database (data containing personal names stored at medical institutions and anonymized data at MID-NET)
NDB	(Insurance claims) diagnosis, medications and tests (Health examinations) lab tests, health guidance	Yes (after 2013)	"Guidelines for providing health insurance claims and specific health checkups and specific health guidance data" Provide to researchers, and provision of access points. At present, unable to link to other databases.	With a legal revision, plan to connect to the nursing care DB in October 2020
MHLW DPC	Simplified medical information, institution information	Yes (after 2017)	"Guidelines for providing DPC (Diagnosis Procedure Combination) data" Only provide spreadsheets At present, unable to link personally identifying data, except for a special case.	With a legal revision, plan to connect to NDB in April 2022
Nursing care DB	(Insurance claims) nursing care service types, units, nursing care level (Required nursing care information) primary and secondary nursing care certification	Yes (after 2018)	"Guidelines for providing information on certification of needed long-term care and long-term care insurance claims data" Have only used by government agencies. A future plan is to make data available to third parties.	With a legal revision, plan to connect to NDB in October 2020
Cancer registry DB	Name, gender, birthday, hospital name, treatments, survival information	Yes (after 2018)	Cancer Registry Promotion Act and "Manual for providing data from the national cancer registry information" Limited to provision for study objective and survival information to submitting hospitals	—
Designated intractable disease DB	Name, gender, family history, lifestyle habits, diagnosis	Yes (after 2019)	"Guidelines for providing designated intractable disease patient data and specific pediatric chronic disease patient data" Limited to provision for study objective.	—
Specific pediatric chronic disease DB	Name, disease, diagnosis	Yes (after 2019)	"Guidelines for providing designated intractable disease patient data and specific pediatric chronic disease patient data" Limited to provision for study objective.	—
Life Data Initiative DB	Treatments, tests, insurance claims, employee health examination, prescription claims	Yes (after 2019)	Next Generation Medical Infrastructure Law	Approved medical data anonymization agency in accordance with Next Generation Medical Infrastructure Law

Source: Publicly available information for each DB prepared by EY based on interview findings

Table 2.2-3 Summary of non-digitized data that could be connected to cohort study data

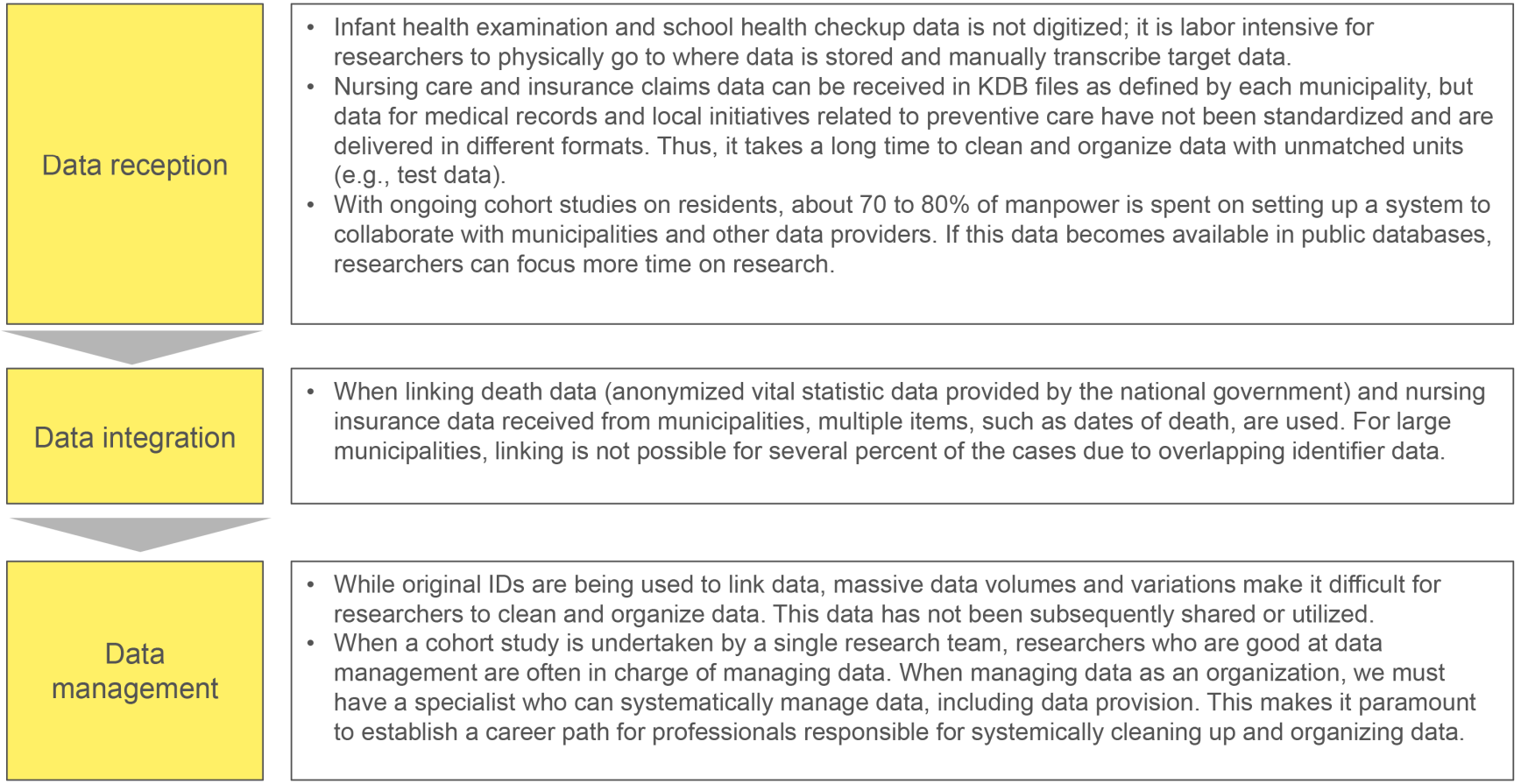
Raw data	Relevant law and authority	Responsible agency	Stored data	Data aggregation	Notes
Prenatal health examinations (maternal and child health handbook)	Maternal and Child Health Act	Municipality	Uterine fundus length, abdominal circumference, blood pressure, edema, urine test, body weight (blood and other tests as needed)	—	
Infant health examinations	Maternal and Child Health Act	Municipality	Development, nutrition, illness, vaccination	—	
School health checkups	School Health and Safety Act	School (municipal school board before entering school)	Height, body weight, disease, urine	—	
Periodic, employment, and specific health examinations	Industrial Safety and Health Act	Employer	Past medical history, smoking, medication history, employment history, blood pressure, urine, blood	—	A randomized investigation is undertaken by the Ministry of Health, Labour and Welfare.
Metabolism examinations	Act on Assurance of Medical Care for Elderly People National Insurance Act	Insurer (national health insurance organizations, health insurance associations, municipal national health organization)	Physical measurements, blood pressure, blood lipids, liver function, blood sugar, urine test	Aggregated by the Health Insurance Claims Review & Reimbursement Services Association of National Health Insurance Organizations DB	In the process of developing an integrated system for linking data with 2020 metabolism examination data and accessing data through the MyNumber Portal.
Cancer screening	Health Promotion Act	Municipality	Stomach cancer, cervical cancer, lung cancer, breast cancer, colon cancer	Aggregated by the survey regarding the implementation of cancer screening among municipalities, Ministry of Health, Labour and Welfare	Medical institutions enter information into Cancer registries
Thorough physical health examinations	— (included in health examinations as specified in the Industrial Safety and Health Act)	Insurer, employer	Physical measurements, blood pressure, EKG, chest X-ray, abdominal ultrasound	—	
Health examinations for the older senior citizens	Act on Assurance of Medical Care for Elderly People	Association of medical care services for older senior citizens	Physical measurements, blood pressure, blood, urine, medications, smoking	Aggregated by the survey on subscribers of Medical Care for Elderly People, Ministry of Health, Labour and Welfare	
Death certificate (simplified)	Rules for reporting stillbirth, Census Registration Act	Municipality (public health center) *Provided to MHLW via prefecture	Gender, birthday, cause of death, hospital	Aggregated by the vital statistics, Ministry of Health, Labour and Welfare	Used for linking death data and cancer registries

Source: Prepared by EY from various information sources

2.2.2 Cohort study data linkages

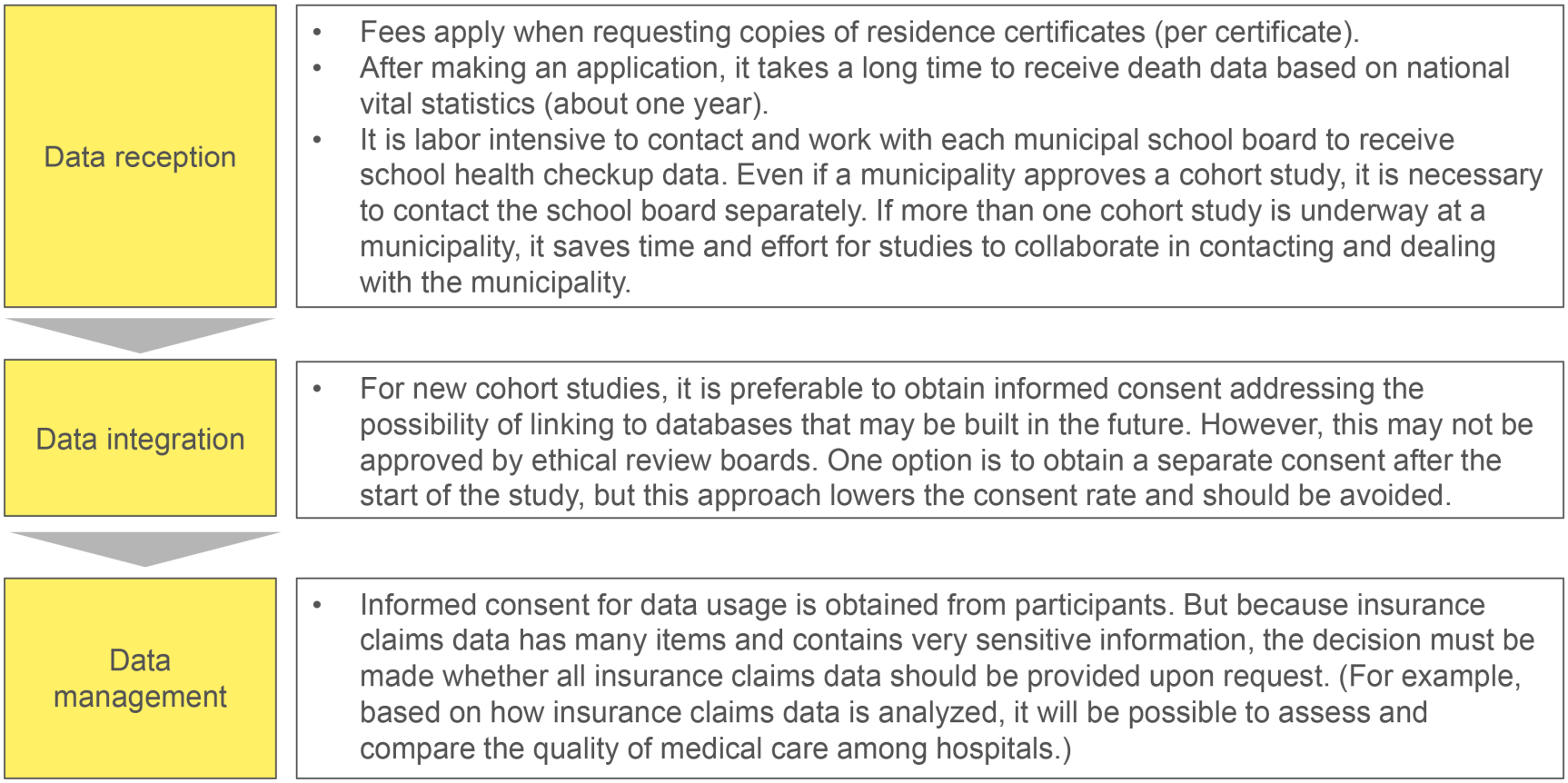
Based on interviews of principal investigators of cohort studies, Figure 2.2-1 shows the technical challenges encountered when seeking to link data with cohort studies. Figure 2.2-2 lists procedural challenges. We examined data handling challenges from three viewpoints: data reception, data integration, and data management.

Figure 2.2-1 Technical challenges for linking information with cohorts



Source: Based on interview findings

Figure 2.2-2 Procedural challenges for linking information with cohorts

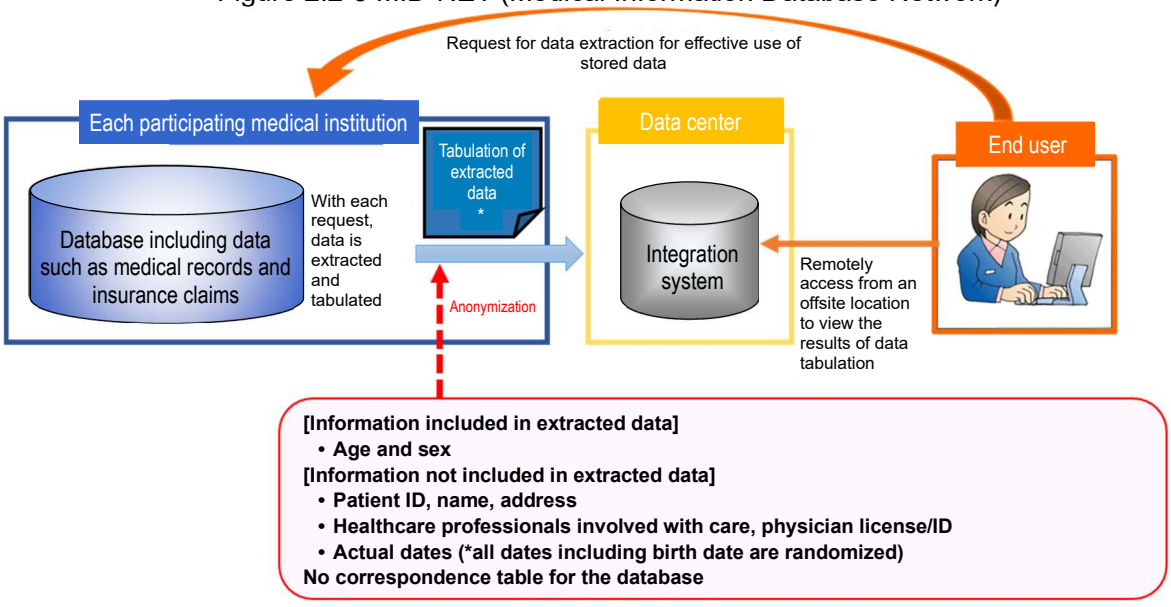


Source: Based on interview findings

2.2.3 Information linking to existing databases

We carefully examined the following three existing databases via interviews and document reviews: the Medical Information Database Network (MID-NET) (public database, Figure 2.2-3), the National Database (NDB) (public database, Figure 2.24-), and the Life Data Initiative Database (LDI-DB) (private database, Figure 2.2-5). Figure 2.2-6 presents the technical challenges to linking data with other databases. Figure 2.2-7 presents procedural challenges. These challenges were examined from the three aspects of data handling: data reception, data integration, and data management. Figure 2.2-8 presents overall procedural challenges that could not be classified by the above-mentioned three aspects of data handling: developing human resources in different fields and coorperative efforts; educating citizens; and regulations.

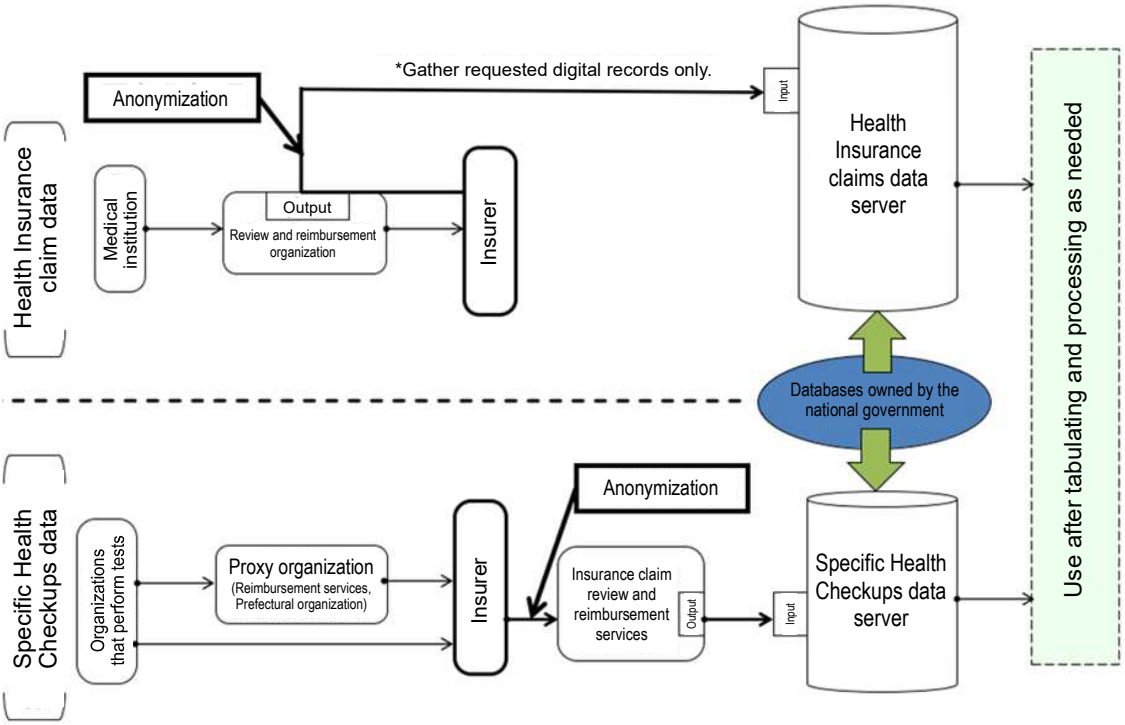
Figure 2.2-3 MID-NET (Medical Information Database Network)



*Data is extracted only upon individual request for data usage. In other cases, only tabulated results are produced. “Tabulated results” are data produced by statistically analyzing extracted data.

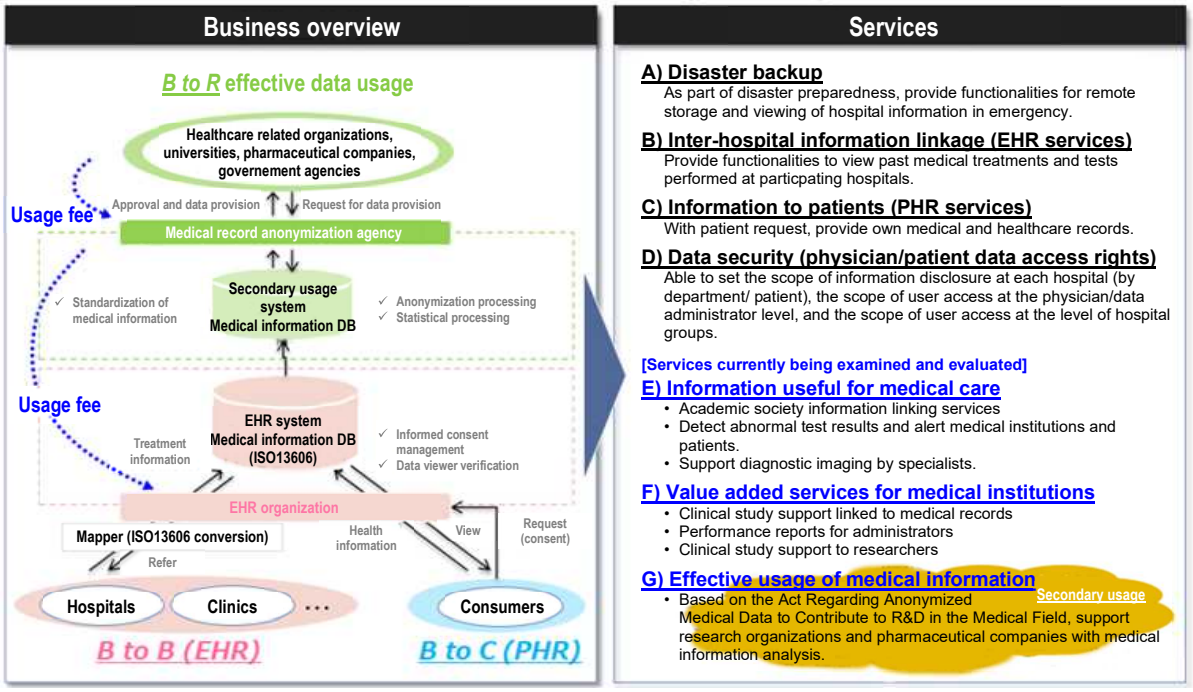
Source: PMDA website

Figure 2.2-4 National Database (NDB Japan)



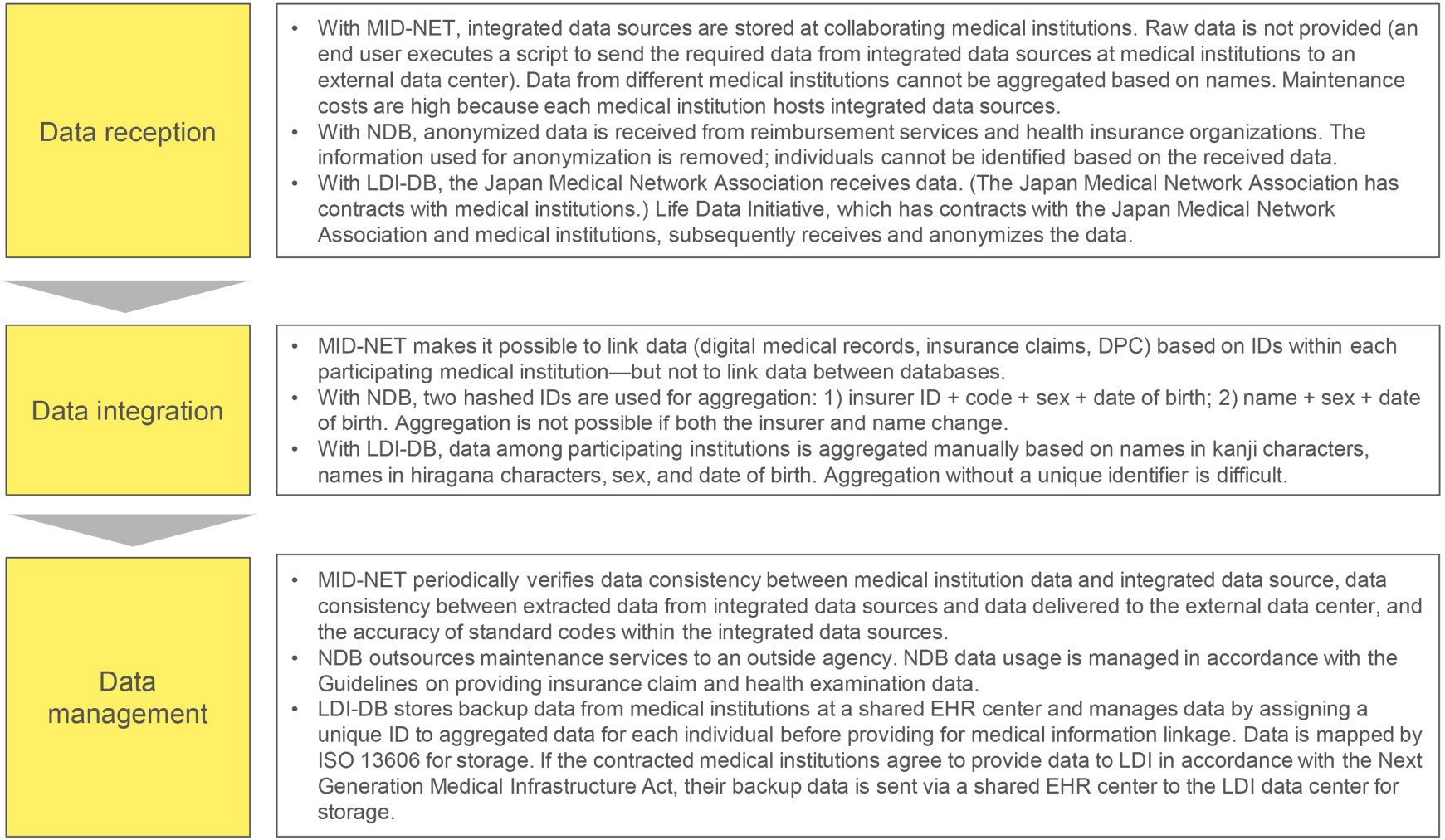
Source: Ministry of Health, Labour and Welfare, "Regarding the National Database (NDB) of Health Insurance Claims and Specific Health Checkups" (2014)

Figure 2.2-5 Life Data Initiative Database



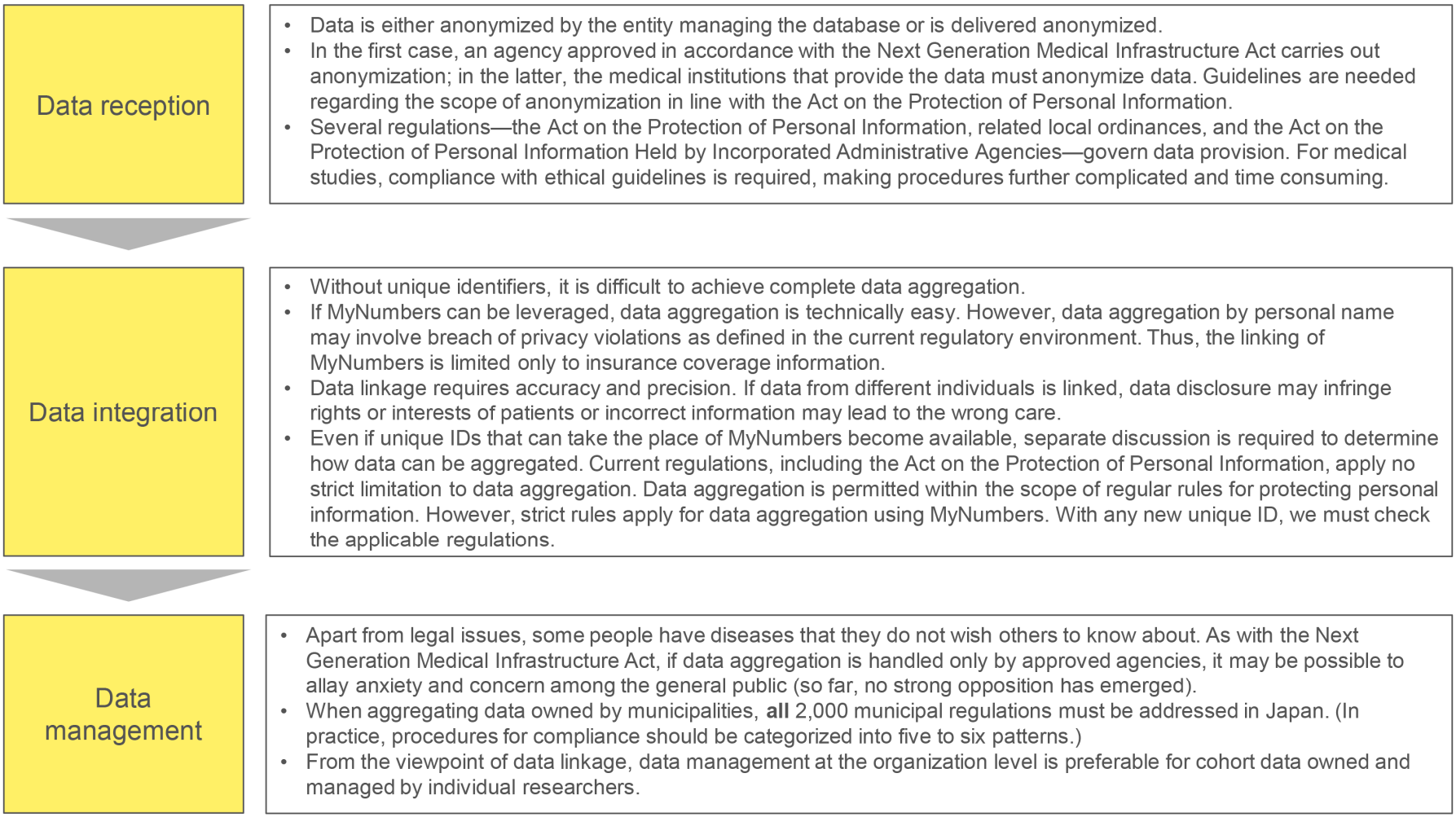
Source: Hiroyuki Yoshihara, "Summary and current state of the millennial medical record project" (2018)

Figure 2.2-6 Technical challenges for linking data with existing databases



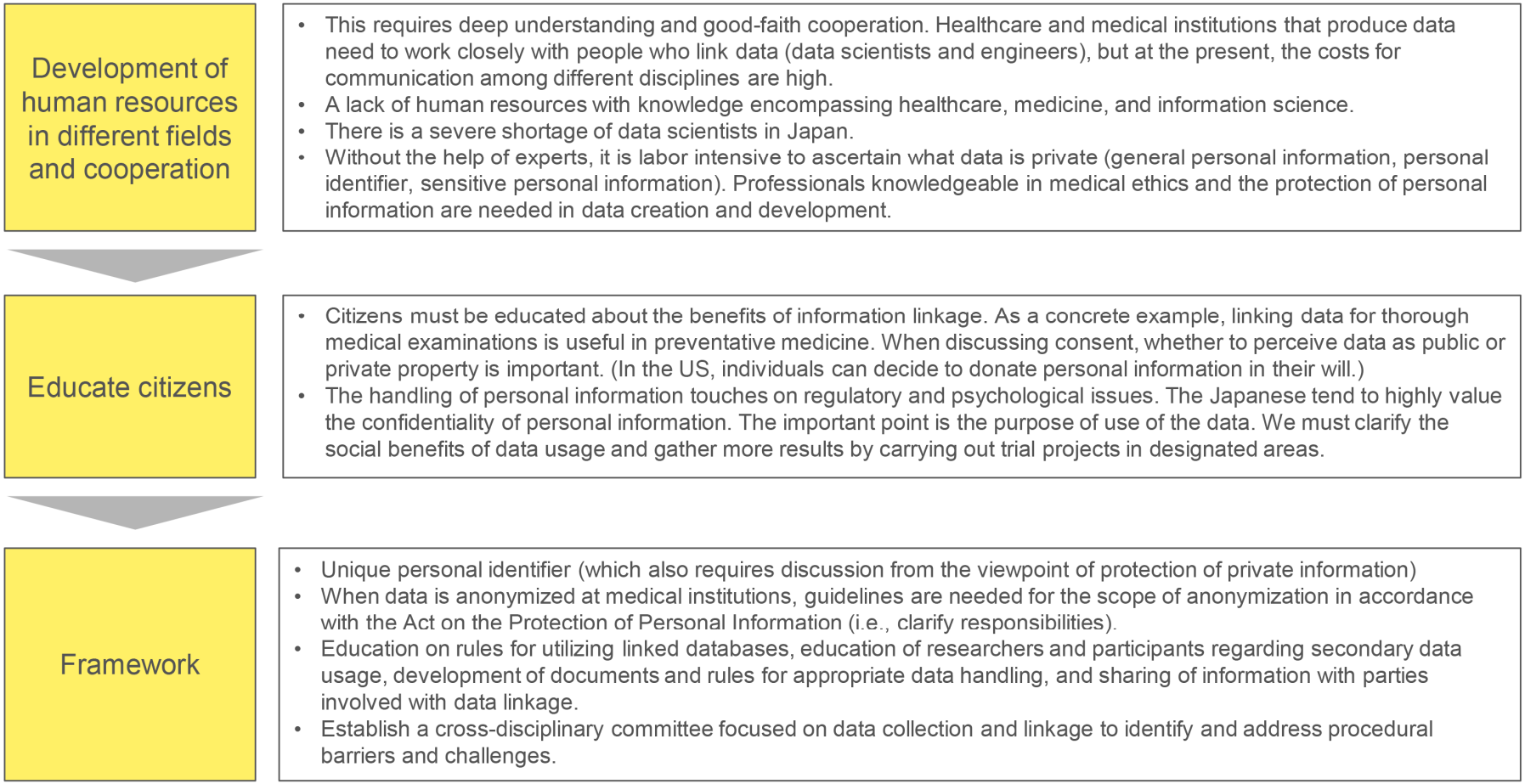
Source: Based on interview findings

Figure 2.2-7 Procedural challenges for linking data with existing databases (1 of 2)



Source: Based on interview findings

Figure 2.2-8 Procedural challenges for linking data with existing databases (2 of 2)



Source: Based on interview findings

2.2.4 Thorough examination and identification of challenges to future data linking and linked data analysis

To thoroughly identify the challenges to linking healthcare and medical data and analyzing linked data, interviews were conducted on the participants of the investigation in Section 2.1.2 1) “Interviews of organizations that build and utilize data”. The process of identifying individuals and linking different information was examined from the viewpoints of 1) technical elements and 2) procedural elements.

(1) Technical elements

Table 2.2-1 summarizes the techniques and processes required for linking; Table 2.2-2 the technical elements for managing personal information and the barriers to linking; and Table 2.2-3 the current state of infrastructures, challenges, and solutions.

Table 2.2-1 Techniques and processes required for linking data

		Data linkage from different sources for a target individual
Techniques and processes	Premise	• Study designs determine data linkage (for what objective data is collected). If a study design is limited to certain diseases and patients, data can theoretically be linked using medical institute IDs and chart numbers. • When linking data, data must be gathered from both vertical and horizontal aspects. • Data linkage is meaningless if the quality of data in the database is low. Discussions must address the goals and objectives of data linkage.
	Identifiers for database linkage	• IDs that uniquely identify individuals are needed for thorough data aggregation. • Aggregation based on names in kanji characters, names in hiragana characters, gender, and birth dates is possible but not applicable in certain cases. • Individual (healthy person/patient) “identifier," “date” and “informed consent.” • Multimodal data for cross-sectional analysis of diseases (genome, transcriptome, metabolome, cytokine, blood test data and diagnostic imaging). • Metadata with definitions (e.g., objective/background for building databases, meaning of data managed in the database, type (numbers, characters, Boolean), length)
	Data preprocessing	• High time and labor costs • Uniform character encoding, processing of unnecessary two-byte spaces, type definition, and data fairing. Preferable to be able to analyze using SQL, Python (e.g., Pandas) or R (e.g., dplyr).
	Standardization	• High time and labor costs • Specifications for databases installed at hospitals differ according to vendor. In particular, non-open-source software and standards pose challenges. • MHLW standards (e.g., ICD10 disease codes) may be useful for medical data management. There is no issue as long as electric medical records are entered as per the specifications, but in practice, data is entered outside the specifications, potentially leading to linking errors. • While progress is being made on data standardization, we need not sit idly by. The level of difficulty in data standardization varies. It is important to discuss standardization by clarifying objectives. • Standardization is important because data to be linked is obtained by different measurement conditions (protocol, human, environment), analysis equipment (device specification, version, settings) and measurement kits. (Examples of standards: AGD, DDBJ, JGA and other Japanese data standardization format, Fast Healthcare Interoperability Resources (FHIR)) • Data structure (e.g., JSON and EML) • Outputting structured data from electric medical record systems must be mandatory.

Source: Based on interview findings

Table 2.2-2 Technical elements of managing personal information, and barriers for linking (technical elements)

		Data linkage from different sources for a target individual
Personal information management	Communication and security	• Medical institutions need to establish communications (responses such as HTTP) and security (cloud service security). • Data need not be aggregated at one location. Obtaining target information is the goal. One important option in protecting information is to extract only the necessary information from decentralized data. • Anonymization processing, verification of data access, access right management, communication security, data/database encryption, logging, and incident response are required.
Linkage barriers	Assurance of data consistency	• The equivalence of data transfer between systems is key to achieving complete matching of data before and after transfer. • With EHR, aggregation of data from different individuals must not occur. On the other hand, when the average data for a certain group is required, in certain cases, a certain degree of fluctuation may be permissible. • Even among hospitals using electric medical record systems from the same vendor, the system is often customized, making it difficult to link data.
	Challenges for existing databases	• Registries owned by academic societies are often anonymized. At the moment, the use of existing data is limited to specific research objectives. • With existing databases, the data has not been sufficiently cleaned for analysis. For example, data is separated into sheets and tables; ideally, medication information would be linked to anonymized IDs.

Source: Based on interview findings

Table 2.2-3 Current state of necessary infrastructures, challenges, and solutions (technical elements)

		Data linkage from different sources for a target individual
Necessary infrastructures and solutions	Environment	• Few clinics use electric medical record systems. As a result, not many institutions can provide digitized data. • It takes significant time and cost to build a database by collecting and standardizing unintegrated data. Ideally, within the framework of effective data usage, standardization should occur upstream of data input, not downstream. However, electric medical records are used as part of routine medical practice, not for data linkage. Thus, the data must be formatted to avoid hindering routine medical care. • Governors have expressed the need for building infrastructures where common healthcare and medical data is managed by the national government and municipalities can access the necessary data for analysis. However, different municipalities face different issues. Municipalities must conduct their own data analysis.
	Human resources	• At many medical institutions, electronic systems are managed by internal administrative staff or outsourced to a local vendor. In-house specialists who can manage medical systems are needed, but such specialists cannot be hired due to payroll issues. • When securing professionals who can manage medical data, either software engineers can learn medicine or physicians can learn programing skills. The latter is more practical. • Municipalities identify issues through communications with prefectural and municipal residents and use data to solve these issues, not start with data analyses to find problems. However, it is unrealistic to assume each municipality can employ such highly specialized experts. The national government should assist and aid municipalities in collaborating with organizations with data analysis experts who understand healthcare and medical data.

Source: Based on interview findings

(2) Procedural elements

Table 2.2-4 summarizes the barriers to linkages. Table 2.2-5 presents the current state of infrastructures, challenges, and solutions.

Figure 2.2-4 Barriers to linkage (procedural elements)

		Data linkage from different sources for a target individual
Procedural barriers	Personal identification method	• Data linkage requires accuracy and precision. If data from different individuals is linked, data disclosure may infringe rights or interests of patients or incorrect information may lead to the wrong care. • At present, subscriber numbers are used as standard IDs. This makes it difficult to achieve complete matching and aggregation (speed and accuracy with which subscriber information can be cross-checked). Unique IDs should be used for personal identification. • If MyNumbers can be leveraged, data aggregation is technically easy. However, data aggregation by personal name may involve privacy violations as defined in the current regulatory environment. Thus, the linking of MyNumbers is limited only to insurance coverage information, not to medical information. • Currently in Japan, Basic Resident Register codes are the only universally available IDs.
	Law and regulation diversity	• When aggregating data owned by municipalities, all 2,000 municipal regulations must be addressed in Japan. (In practice, procedures for compliance may be categorized into five to six patterns.) • Several regulations—the Act on the Protection of Personal Information, related local ordinances, and the Act on the Protection of Personal Information Held by Incorporated Administrative Agencies—govern data provision. For medical studies, compliance with ethical guidelines is required, making administrative procedures further complicated and time consuming. • Several regulations—the Act on the Protection of Personal Information and PMDA Personal Information regulations—govern data provision. For healthcare studies, data reception is time consuming and complicated by ethical guidelines. • With regard to the interpretation of consent in research, more collaboration and dialogue between the office in charge of national ethical guidelines and the Personal Information Protection Commission are required. We must fill the gaps between ethical guidelines and commission guidance regarding how study objectives are explained or informed consent is expressed and obtained, and which university-industry alliance activities do not require consent. • Coordinate rules for protecting personal information at each municipality. • Establish measures to protect personal information matching study objective or data sharing.
	Link with cohort study data	• Build a system in which study data is managed at the level of the study organization, not the individual researcher • However, we cannot make this a requirement as major research funding (grants in aid for scientific research and MLHW research funds) is granted to individual researchers. Since most Japanese studies are funded by the Ministry of Education, Culture, Sports, Science and Technology, changing this system would be difficult.

Source: Based on interview findings

Figure 2.2-5 Current state of necessary infrastructures, challenges and solutions (procedural elements)

		Data linkage from different sources for a target individual
Necessary infrastructures and solutions	Development of human resources in different fields and cooperation	• This requires deep understanding and good-faith cooperation. Healthcare and medical institutions that produce data need to work closely with people who link data (data scientists and engineers), but at the present, the costs for communication among different disciplines are high. • A lack of human resources with knowledge encompassing healthcare, medicine, and information science. • There is a severe shortage of data scientists in Japan. • Without the help of experts, it is labor intensive to ascertain what data is private (general personal information, personal identifier, sensitive personal information). Professionals knowledgeable in medical ethics and the protection of personal information are needed in data creation and development.
	Educate citizens	• Apart from legal issues, some people have diseases that they do not wish others to know about. As with the Next Generation Medical Infrastructure Act, if data aggregation is handled only by approved agencies, it may be possible to allay anxiety and concern among Japanese citizens (so far, no strong opposition has emerged). • Citizens must be educated about the benefits of information linkage. As a concrete example, linking data for thorough medical examinations is useful in preventative medicine. When discussing consent, whether to perceive data as public or private property is important. (In the US, individuals can decide to donate personal information in their will.) • The handling of personal information touches on regulatory and psychological issues. The Japanese tend to highly value the confidentiality of personal information. The important point is the purpose of use of the data. We must clarify the social benefits of data usage and gather more results by carrying out trial projects in designated areas.
	Framework	• Unique personal identifier (need to examine from the viewpoint of protecting privacy information) • Even if unique IDs become available, separate discussion is required to determine how data can be aggregated. Current regulations, including the Act on the Protection of Personal Information, apply no strict limitation to data aggregation. Data aggregation is permitted within the scope of regular rules for protecting personal information. However, strict rules apply for data aggregation using MyNumbers. • When data is anonymized at medical institutions, guidelines are needed for the scope of anonymization in accordance with the Act on the Protection of Personal Information (i.e., clarify responsibilities). • Education on rules for utilizing linked databases, education of researchers and participants regarding secondary data usage, development of documents and rules for appropriate data handling, and sharing of information with parties involved with data linkage • Establish a cross-disciplinary committee focused on data collection and linkage to identify and address procedural barriers and challenges. • More experience and sharing of success stories will be important in revising regulations and easing citizen concerns.

Source: Based on interview findings