Advanced Biomedical Engineering
11: 203–217, 2022.

Original Paper

DOI:10.14326/abe.11.203

Development of a New Method to Trace Patient Data Using the National Database in Japan

Tomoya Myojin,[*, **] Tatsuya Noda,[*, #] Shinichiro Kubo,[*] Yuichi Nishioka,[*] Tsuneyuki Higashino,[***]
Tomoaki Imamura[*]

Abstract The National Database of Health Insurance Claims and Specific Health Checkups of Japan (NDB) is a comprehensive database containing health insurance claim information. The structure of the NDB complicates long-term cohorts for two main reasons. First, the NDB data are stored on a per-claim basis. Second, the NDB is a billing-focused record structure. Therefore, the objective of this study was to use ID0 to modify the data structure to allow for long-term cohorts, provided that the data volume is not increased and the runtime per data year is maintained within one month. The NDB uses two primary keys (ID1 and ID2) made from hash values that mask personally identifiable information. ID0 is our recently developed key from ID1 and ID2, which improves patient-matching efficiency with excellent long-term tracing performance. Our study used claim data with filing dates between April 2013 and March 2016 to trace hospitalizations of one month or longer, including outpatient care, in three steps. In Step 1, claims were transferred to a CD-record format. As some diagnosis procedure combination (DPC) claim records contain a mixture of overlapping comprehensive and piece-rate data, we sorted and reorganized them. In Step 2, pharmacy and medical outpatient claims were integrated using the ID0 key, the medical institution code for issuing a prescription, and the prescription issue date. In Step 3, the transferred data were combined and converted from consecutive hospitalization days into sequences based on ID0, the medical institution code, and hospital ward classification. Consequently, the size of the originally extracted comma-separated variable dataset for three years (approximately 10.5 TB) was reduced to an approximately 6 TB main database file that was usable for processing. The process took approximately three months. With similar conventional methods, the data size was 30 times larger, and it took more than seven months to process a year's worth of data. In addition, to demonstrate the application of this method, we conducted a six-year mortality cohort for all Japanese citizens. Our technique makes it easy to perform follow-up and longitudinal cohort surveys while accurately tracing patient data in large-scale medical claims databases.

Keywords: national health insurance, insurance claim, NDB, patient tracing, database.

Adv Biomed Eng. **11**: pp. 203–217, 2022.

1. Introduction

Japan has implemented a compulsory nationwide public health insurance system for all legal Japanese citizens [1], excluding short-term residents (length of stay less than 90 days) [2] and those on welfare. Participants have access to covered medical services when consulting a doctor. Medical institutions submit patient claims to the national insurance program for payment of expenses for the covered services. There are six types of claim: pharmacy, diagnosis procedure combination (DPC), medical hospitalization, medical outpatient, dental hospitalization, and dental outpatient [3]. Each claim is associated by type, month, medical institution, and patient. The DPC is a comprehensive payment system used for the clinical treatment of certain diseases and injuries during a given period, determined by the type of disease or injury. A piece-rate payment system is applied when a

Received on September 10, 2021; revised on November 2, 2021, February 17, 2022, June 6, 2022, and August 16, 2022; accepted on August 22, 2022.

[*] Department of Public Health, Health Management and Policy, Nara Medical University, Nara, Japan.

[**] Department of Diagnostic Pathology, Nara Medical University, Nara, Japan.

[***] Healthcare and Wellness Division, Mitsubishi Research Institute, Inc., Tokyo, Japan.

[#] 840 Shijo-Cho, Kashihara, Nara 634–8512, Japan.
E-mail: noda@naramed-u.ac.jp

claim is above the comprehensive payment limit. Further, when a claim is changed from DPC to piece-rate within a given month, data for that month inherit both types of information.

Since 2009, under relevant legislation, the Japanese Ministry of Health, Labour, and Welfare (MHLW) has retained all insurance claim data, except those related to clinical trials and medical expenses that are not covered by insurance [4] (such as traffic and industrial accidents). The data are stored in the National Database of Health Insurance Claims and Specific Health Checkups of Japan (NDB), which had processed ~20.9 billion claims as of March 2021, making it one of the world's largest health information databases. Other Japanese medical databases exist, including the Japan Medical Data Center Claims database [5–7], the Medical Data Vision Administrative Claims database [8, 9], and databases belonging to municipalities and disparate firms across Japan [10, 11]. The population size of the NDB is markedly the largest.

NDB data include information on the medical institution visited, disease and injury (confirmed/unconfirmed) name per the International Classification of Diseases 10th Revision (ICD-10), medical care received, medical examinations (without results), and prescribed drugs. Drug information includes prescription amount, brand name, dosage, and number of days prescribed. Patients' residential districts are not included.

The NDB is useful for rough aggregations and observational studies, such as those determining medical practice trends and prescription rates [3]. However, there is a paucity of reports from cohort studies using the NDB. Notably, the Health Insurance Bureau of the MHLW provides only the minimum NDB data required for research. Additionally, there are two core issues preventing its use for longitudinal cohort studies: per-claim-based information storage and billing-focused record structure.

1.1 Per-claim-based information storage

NDB data are stored on a per-claim basis. Notably, cohort studies require a per-patient basis. However, owing to privacy restrictions, NDB data are not directly associated with insurance card numbers, patient names, or dates of birth. Instead, owing to the three degrees of freedom of the privacy-sensitive elements, two hash values (SHA-256) are used. The input data for identifier variable type 1 (ID1) include insurer number, insurance card number, gender, and date of birth. Similarly, the input data for identifier variable type 2 (ID2) include patient name, gender, and date of birth. A report shows that in medical outpatient claims captured in any one year period, 20.4% of ID2s are associated with multiple ID1s, and 7.1% of ID1s are associated with multiple ID2s [12]. This complex query scenario is further complicated when the secondary key data feeding the cryptographic hash functions change (including name changes and other corrections). To overcome these shortcomings, in our previous work, we used extant ID1 and ID2 keys to create the ID0 foreign key, which improves patient-matching efficiency with excellent long-term tracing performance. The date of treatment and outcome are applied as composite keys [13].

1.2 Billing-focused record structure

A claim contains information exclusively for billing purposes, and various records are included [i.e., RE refers to patient information, IR refers to the medical institution information, HO refers to cost information, SY and SB refer to the patient's injury or illness (piece-rate and comprehensive, respectively), SI refers to the medical treatment (piece-rate), TO refers to specific equipment and material (piece-rate), IY refers to pharmaceutical information (piece-rate), CZ refers to dispensing information, CD refers to the combination of SI, TO, and IY records (piece-rate and comprehensive), and BU refers to the diagnosis group]. **Table 1** presents these record types, cross-linked to claim types. Note that not every item is recorded for every type of claim. Such independent, many-to-many relationships make it difficult to collect valid data from the current data structure.

1.2.1 Mixture of comprehensive and piece-rate data

In the CD records of the DPC claims, comprehensive and piece-rate payment data are mixed in terms of medical procedure performed, medical equipment used, and prescribed medicine.

1.2.2 Inability to obtain out-of-hospital prescriptions

In the case of outpatient prescriptions, claims are divided into outpatient types assigned by the medical institution and pharmacy types assigned by the pharmacy, making it impossible to trace the prescription status of a given outpatient visit.

1.2.3 Inability to determine hospitalization period

For non-DPC hospitalizations, the dates of admission and discharge are insufficiently documented. The date of the first hospitalization in a given month is recorded in the RE record, but even if the patient is discharged or readmitted within the same month, no information on the date of discharge or readmission is included. Furthermore, if a patient's hospitalization extends into a consecutive month, the original date of hospitalization is not retained. Therefore, it is impossible to determine hospitalization periods lasting longer than a month.

Table 1 Typical information of each record type and the associated claims.

Record	Typical information contained in the record	Type of Claim			
		DPC	Medical hospitalization	Medical outpatients	Pharmacy
RE	Patient information	○	○	○	○
IR	Medical institution information	○	○	○	
YK	Pharmacy information				○
HO	Cost information	○ (Mixed comprehensive and piece-rate)	○	○	○
SY	Patient's injury or illness	○ (Piece-rate only)	○	○	
SB	Patient's injury or illness	○ (Comprehensive only)			
SI	Medical treatment action	○ (Piece-rate only)	○	○	
TO	Specific equipment and material	○ (Piece-rate only)	○	○	○
IY	Pharmaceutical information	○ (Piece-rate only)	○	○	○
CZ	Dispensing information				○
CD	Medical treatment action, specified equipment and material, and pharmaceutical information	○ (Mixed comprehensive and piece-rate)			
BU	Diagnosis group classification information	○ (Comprehensive only)			

○: The record is included in the relevant claim.

The per-claim-based information storage issues described in Section 1.1 were resolved with our previous development of the ID0 foreign key. However, the billing-focused record structure makes cohort studies very difficult. Fujimori et al. [14] developed and operated a method to convert comma-separated-value (CSV) claim data into the DPC EF file format of the main database file (MDF). In 2015, we applied a similar method to medical hospitalization claims for FY 2013 records. However, the conversion increased the data volume by about 30 times, taking seven months to complete. As the annual file size of claims for hospitalization, DPC, outpatient, and pharmacy is approximately 3.5 TB of CSV data, simple calculation shows that the file size after conversion is approximately 100 TB. These file sizes and processing durations are clearly unacceptable.

Rendering a medical data structure suitable for cohort studies is referred to as *patient tracing*. Because this process is exceedingly difficult with the current NDB database, this study seeks to enable a longitudinal cohort study by converting exported NDB data into a patient-tracing-capable dataset while minimizing file size growth and conversion time. Here, the use case is a cohort study collecting data of activities exceeding 500 million person-years. At the request of the Health Policy Bureau of the MHLW, which develops medical care plans, our ultimate goal is to capture one year's worth of NDB patient data, from which the new file size after conversion must not exceed the original CSV-file size, and the conversion must take less than one month.

2. Materials and Methods

2.1 Utilized data

We used NDB data on DPC, hospitalization, outpatient, and pharmacy claims from April 2013 to March 2016, as requested and loaned by the Health Policy Bureau of the MHLW. **Supplementary Figures S1–S3** present entity–relationship (ER) diagrams of the items used in this study for each claim type [15–18]. This study was approved by the Ethics Committee of the Nara University School of Medicine (approval number: 1123).

2.2 Tracing patient data through health insurance claims

A multipronged approach was implemented to trace patient data. First, a verticalized format was applied to the horizontal data, and duplicate records were deleted. Then, outpatient and hospitalization data were traced to health insurance claims. Each step is described in detail in the following subsections.

2.3 Applying a vertical format with duplicate deletion

The SI_IY_TO table (**Table 2A**) of hospitalization, outpatient, and DPC claims contains date information serially arranged in columns (e.g., Day 1, Day 2, ⋯, Day 31). This format is highly inefficient, making it difficult to analyze continuous data. **Table 2A** and the RE table (**Table 2B**) were concatenated using the Serial Sequence number of claims as the join key, and joining the year and month of medical care received column in **Table 2B** and the number of times of each column from Day 1 to Day 31 in **Table 2A** were expanded as vertical records (**Table 2C**).

Recall that the SI_IY_TO table only contains piece-rate records, and CD tables contain all SI/IY/TO records plus comprehensive records with the same attributes. Unfortunately, the CD table does not distinguish between the piece-rate and comprehensive records. For each serial number, date, code, and time attribute (**Table 3A**), matches in the CD records (**Table 3B**) are considered duplicates. Thus, we truncated the CD table by deleting the duplicate records, so that only the comprehensive entries were left (**Table 3C**). Then, we merged the verticalized SI_IY_TO table and the truncated CD table to create a modified SI_IY_TO table (**Table 3D**). Here, an additional attribute “CD flag” is added, in which a value of “1” indicates comprehensive records.

Table 2 Table modification of SI/TO/IY records of medical hospitalization, medical outpatient, and DPC claims.

A. SI_TO_IY table

Serial number of claim	Sequence of record	Code	Usage (IY only)	Score	Total times	Day1 times	Day2 times	…	Day10 times	Day11 times	…	Day29 times	Day30 times	Day31 times
X12dfj7	8	112007410		520	5		1		1	2			1	
X12dfj7	9	112009910		13	4		2			2				
Ghs7dte	12	112009910		13	2							2		
jhgfG7uR	9	120002910		82	4		3						1	

B. RE table

Serial number of claim	ID0	Year and month of medical care received	Gender	Age class
X12dfj7	XXX	201811	1	221
Ghs7dte	YYY	201811	1	218
jhgfG7uR	ZZZ	201811	1	213

C. Verticalized SI_TO_IY table

Serial number of claim	Sequence of record	Date	Code	Usage(IY only)	Score	Times
X12dfj7	8	20181102	112007410		520	1
X12dfj7	8	20181110	112007410		520	1
X12dfj7	8	20181111	112007410		520	2
X12dfj7	8	20181130	112007410		520	1
X12dfj7	9	20181102	112009910		13	2
X12dfj7	9	20181111	112009910		13	2
Ghs7dte	12	20181129	112009910		13	2
jhgfG7uR	9	20181102	120002910		82	3
jhgfG7uR	9	20181130	120002910		82	1

Table 3 Modification of SI_TO_IY table to differentiate piece-rate and comprehensive records.

Before duplicate record deletion

A. Verticalized SI_TO_IY table (piece-rate only)

Serial number of claim	Sequence of record	Date	Code	Score	Times
X12dfj7	*8*	*20181102*	*112007410*	*520*	*1*
X12dfj7	*8*	*20181110*	*112007410*	*520*	*1*
X12dfj7	*8*	*20181111*	*112007410*	*520*	*2*
X12dfj7	*8*	*20181130*	*112007410*	*520*	*1*
X12dfj7	*9*	*20181102*	*112009910*	*13*	*2*
X12dfj7	*9*	*20181111*	*112009910*	*13*	*2*
Ghs7dte	*12*	*20181129*	*112009910*	*13*	*2*
jhgfG7uR	*9*	*20181102*	*120002910*	*82*	*3*
jhgfG7uR	*9*	*20181130*	*120002910*	*82*	*1*

B. CD table (piece-rate and comprehensive)

Serial number of claim	Sequence of record	Date	Code	Times
X12dfj7	87	20181101	112004567	2
X12dfj7	88	20181102	112004567	5
X12dfj7	89	20181102	112007410	1
X12dfj7	90	20181103	112004567	3
X12dfj7	91	20181110	112007410	1
X12dfj7	92	20181110	112008634	1
X12dfj7	93	20181111	112007410	2
X12dfj7	94	20181130	112007410	1
Ghs7dte	54	20181127	112003962	2
Ghs7dte	55	20181128	112003962	1
Ghs7dte	56	20181129	112003962	2
Ghs7dte	57	20181130	112007410	2
jhgfG7uR	112	20181102	120001795	2
jhgfG7uR	113	20181102	120002745	2
jhgfG7uR	114	20181102	120002910	3
jhgfG7uR	115	20181105	120009652	6
jhgfG7uR	116	20181130	120001832	3
jhgfG7uR	117	20181130	120002910	1

After duplicate record deletion

C. Truncated CD table (comprehensive-only)

Serial number of claim	Sequence of record	Date	Code	Times
X12dfj7	87	20181101	112004567	2
X12dfj7	88	20181102	112004567	5
X12dfj7	90	20181103	112004567	3
X12dfj7	92	20181110	112008634	1
Ghs7dte	54	20181127	112003962	2
Ghs7dte	55	20181128	112003962	1
Ghs7dte	57	20181130	112007410	2
jhgfG7uR	113	20181102	120002745	2
jhgfG7uR	114	20181102	120002910	3
jhgfG7uR	115	20181105	120009652	6
jhgfG7uR	116	20181130	120001832	3

After record modification

D. Modified SI_TO_IY table (piece-rate and comprehensive)

Serial number of claim	Sequence of record	Date	Code	Score	Times	CD flag
X12dfj7	87	20181101	112004567		2	1
X12dfj7	88	20181102	112004567		5	1
X12dfj7	*8*	*20181102*	*112007410*	*520*	*1*	
X12dfj7	*9*	*20181102*	*112009910*	*13*	*2*	
X12dfj7	90	20181103	112004567		3	1
X12dfj7	*8*	*20181110*	*112007410*	*520*	*1*	
X12dfj7	92	20181110	112008634		1	1
X12dfj7	*8*	*20181111*	*112007410*	*520*	*2*	
X12dfj7	*9*	*20181111*	*112009910*	*13*	*2*	
X12dfj7	*8*	*20181130*	*112007410*	*520*	*1*	
Ghs7dte	54	20181127	112003962		2	1
Ghs7dte	55	20181128	112003962		1	1
Ghs7dte	*12*	*20181129*	*112009910*	*13*	*2*	
Ghs7dte	57	20181130	112007410		2	1
jhgfG7uR	113	20181102	120002745		2	1
jhgfG7uR	*9*	*20181102*	*120002910*	*82*	*3*	
jhgfG7uR	114	20181102	120002910		3	1
jhgfG7uR	115	20181105	120009652		6	1
jhgfG7uR	116	20181130	120001832		3	1
jhgfG7uR	*9*	*20181130*	*120002910*	*82*	*1*	

2.4 Tracing outpatient data using a health insurance claim

Outpatient data include outpatient visits and out-of-hospital prescriptions. As shown in **Table 4**, we operated on CZ (dispensary), RE (patient), and IY (pharmaceutical) tables to correlate outpatient data to insurance claims. RE records (i.e., serial number, institution code, gender, and age class; **Table 4B**) were joined with CZ records (**Table 4A**), excluding attributes of times and year/month of dispensing. From this, a modified RE_CZ table (**Table 4E**) was created. Using the serial number and number of prescription attributes as keys, an inner join was performed on CZ and IY records (**Table 4C**), forming the modified IY_CZ table (**Table 4F**). The additional score attribute in IY_CZ reflects a multiplication operation on pharmaceutical price (**Table 4D**) and usage (**Table 4C**) and division by 10, where a score of 1 is equivalent to 10 Japanese yen. We intentionally retained prescription and dispensing dates in each record of the modified IY_CZ table. We then performed the same process with the TO records (specific equipment and material) table using the unit price attribute of specific equipment and materials found in the master medical materials table.

For medical outpatient claims (**Table 5**), we partially joined RE (patient; **Table 5B**) and IR (medical institution; **Table 5C**) records using the date attribute from the verticalized SI_IY_TO table (**Table 5A**) to obtain the modified RE_IR table (**Table 5D**), which contains a new “Date of receiving medicine” attribute. Specifically, **Tables 5B and 5C** were inner joined, and the “Date” column of **Table 5A** was joined as the date of receiving medicine. In both cases, “Serial number” of claims was used as the join key.

A combined Out_Pha master table (**Table 6C**) was then created to associate the medical claim number, date of receiving medicine, and pharmacy claim number attributes. For this purpose, records from the modified RE_CZ table (**Table 4E**) were extracted to generate **Table 6B**, and records from the modified RE_IR table (**Table 5D**) were extracted to generate **Table 6A**. ID0 was used as the join key.

As shown in **Figure 1**, we created relationships among the modified RE_IR, modified RE_CZ, and Out_Pha tables. Additionally, verticalized SI, verticalized TO, verticalized IY, original SY, and original HO tables (medical claims) were associated with the modified RE_IR table. YK, modified TO, modified IY, and original HO tables (pharmacy claims) were associated with the modified RE_CZ table. Thus, outpatient and pharmacy claims can be linked. For example, it can be used for the analysis of the combined information on medical treatment in outpatient scenarios and out-of-hospital prescribed medicine.

2.5 Tracing hospitalization data through a health insurance claim

Tracing hospitalization data from insurance claims requires the association of several disparate attributes associated with records among several tables. For our purposes, hospitalization in the same medical institution and the same ward division is deemed one period of hospitalization. The wards were classified into seven categories based on the Medical Service Act: general, mental, tuberculosis, infectious disease, convalescent, disabled persons, and medical clinics with beds. All hospitalizations in the DPC were considered general wards. We did not consider short-stay surgeries.

DPC data reflect limited periods of hospitalization, and they are labeled with “DPC-specific period” attributes. Based on the period, there are three types of DPC claims. Type 1 refers to simple DPC claims filed within a specified period. However, if the specific period is exceeded, the claim would include a generalized Type 2 DPC claim and a generalized Type 3 medical hospitalization claim. Types 2 and 3 claims contain information relevant to the DPC-specific period and the period subsequent to the expiration of the DPC-specific claim period, respectively. Medical hospitalization claims include data related to non-DPC hospitals beyond the DPC-specific period. For example, **Figure 2** shows an explanation of claims in two cases: one in which the hospitalization period is within a specific DPC period and another in which it is not.

For Types 1 and 2 DPC claims, the dates of calculated hospitalization are defined as the period between the date of admission and the date of discharge in the BU record. For Type 3 DPC cases and medical hospitalization claims, the date of calculated hospitalization is defined as the date of the basic or specific hospital charge in the SI record. The ward division of each hospital charge is shown in **Supplementary Table1**. The consecutive days of hospitalization calculated for the same patient, same ward division, and same medical institution are considered as one hospitalization.

To meet the above criteria, we followed the steps listed in **Table 7**. Note that the calculated hospitalization dates included April 3–6, April 29–May 1, and May 22, 2013.

In **Table 7A**, the claim serial number attribute of the RE table, the medical institution code of the IR table, ID0 of the RE table, the ward division of the SI and BU tables, and the day of calculated hospitalization of the SI and BU records are listed in columns, and the data are grouped by serial number, medical institution code, ID0, and ward division attributes. The list is sorted in ascend-

Table 4 CZ, RE, and IY table operations to assess pharmacy claims.

A. CZ table

Serial number of claim	Number of prescriptions	Date of issued prescription	Date of dispensing	Times
Gj7sQs9	1	20181104	20181104	14
Gj7sQs9	1	20181115	20181118	21
Gj7sQs9	2	20181104	20181104	14
Gj7sQs9	3	20181128	20181128	5
cFa598H	1	20181114	20181115	60
8ftuW81	1	20181119	20181120	20

B. RE table

Serial number of claim	ID0	Year and month of dispensing	Medical institution code that issued prescription	Gender	Age class
Gj7sQs9	AAA	201811	z1x2	2	214
cFa598H	AAA	201811	9m8n	2	214
8ftuW81	BBB	201811	3c4v	1	221

C. IY table

Serial number of claim	Number of prescriptions	Medication code	Usage
Gj7sQs9	1	610357924	6
Gj7sQs9	1	621975312	3
Gj7sQs9	1	627246809	3
Gj7sQs9	2	627963009	2
Gj7sQs9	3	627246809	6
cFa598H	1	622123456	1
8ftuW81	1	613572460	3
8ftuW81	1	615846213	2

D. Master of medicine table

Medication code	Pharmaceutical pricing
610357924	6.2
621975312	14.9
627246809	25.6
627963009	128.1
622123456	240.4
613572460	70.1
615846213	8.2

E. Modified RE_CZ table

Serial number of claim	Number of prescriptions	ID0	Date of issued prescription	Date of dispensing	Medical institution code that issued prescription	Gender	Age class
Gj7sQs9	1	AAA	20181104	20181104	z1x2	2	214
Gj7sQs9	1	AAA	20181115	20181118	z1x2	2	214
Gj7sQs9	2	AAA	20181104	20181104	z1x2	2	214
Gj7sQs9	3	AAA	20181128	20181128	z1x2	2	214
cFa598H	1	AAA	20181114	20181115	9m8n	2	214
8ftuW81	1	BBB	20181119	20181120	3c4v	1	221

F. Modified IY_CZ table

Serial number of claim	Number of prescriptions	Medication code	Usage	Score	Date of issued prescription	Date of dispensing	Times
Gj7sQs9	1	610357924	6	3.72	20181104	20181104	14
Gj7sQs9	1	621975312	3	4.47	20181104	20181104	14
Gj7sQs9	1	627246809	3	7.68	20181104	20181104	14
Gj7sQs9	1	610357924	6	3.72	20181115	20181118	21
Gj7sQs9	1	621975312	3	4.47	20181115	20181118	21
Gj7sQs9	1	627246809	3	7.68	20181115	20181118	21
Gj7sQs9	2	627963009	2	25.62	20181104	20181104	14
Gj7sQs9	3	627246809	6	15.36	20181128	20181128	5
cFa598H	1	622123456	1	24.04	20181114	20181115	60
8ftuW81	1	613572460	3	21.03	20181119	20181120	20
8ftuW81	1	615846213	2	1.64	20181119	20181120	20

Table 5 Creation of modified RE_IR table on consultation day in medical outpatient claim.

A. Verticalized SI_TO_IY table

Serial number of claim	Sequence of record	Date	Code	Usage (IY only)	Score	Times
8Ygh6d0	8	20181104	112007410		520	1
8Ygh6d0	8	20181110	112007410		520	1
8Ygh6d0	8	20181115	112007410		520	2
8Ygh6d0	8	20181130	112007410		520	1
8Ygh6d0	9	20181104	112009910		13	2
8Ygh6d0	9	20181115	112009910		13	2
7Hg6fay	12	20181114	112009910		13	2
Hjae47a	9	20181102	120002910		82	3
Hjae47a	9	20181119	120002910		82	1

B. RE table

Serial number of claim	ID0	Year and month of medical care received	Gender	Age class
8Ygh6d0	AAA	201811	2	214
7Hg6fay	AAA	201811	1	201
Hjae47a	BBB	201811	2	213

C. IR table

Serial number of claim	Medical institution code	Secondary medical area	Hospital clinic classification
8Ygh6d0	z1x2	1201	1
7Hg6fay	9m8n	1201	2
Hjae47a	3c4v	1304	1

D. Modified RE_IR table

Serial number of claim	Date of receiving medicine	ID0	Gender	Age class	Medical institution code	Secondary medical area	Hospital clinic classification
8Ygh6d0	20181104	AAA	2	214	z1x2	1201	1
8Ygh6d0	20181110	AAA	2	214	z1x2	1201	1
8Ygh6d0	20181115	AAA	2	214	z1x2	1201	1
8Ygh6d0	20181130	AAA	2	214	z1x2	1201	1
7Hg6fay	20181114	AAA	1	201	9m8n	1201	2
Hjae47a	20181102	BBB	2	213	3c4v	1304	1
Hjae47a	20181119	BBB	2	213	3c4v	1304	1

Table 6 Creation of medical outpatient claim and pharmacy claim master table.

A. Extracted Modified RE_IR table from Table 5D

Serial number of claim	Date of receiving medicine	ID0	Medical institution code
8Ygh6d0	20181104	AAA	z1x2
8Ygh6d0	20181110	AAA	z1x2
8Ygh6d0	20181115	AAA	z1x2
8Ygh6d0	20181130	AAA	z1x2
7Hg6fay	20181114	AAA	9m8n
Hjae47a	20181102	BBB	3c4v
Hjae47a	20181119	BBB	3c4v

B. Extracted Modified RE_CZ table from Table 4E

Serial number of claim	Date of issued prescription	ID0	Medical institution code that issued prescription
Gj7sQs9	20181104	AAA	z1x2
Gj7sQs9	20181115	AAA	z1x2
Gj7sQs9	20181104	AAA	z1x2
Gj7sQs9	20181128	AAA	z1x2
cFa598H	20181114	AAA	9m8n
8ftuW81	20181119	BBB	3c4v

C. Out_Pha master table

Serial number of medical hospitalization claim	Date of receiving medicine	Serial number of pharmacy claim
X12dfj7	20181104	Gj7sQs9
X12dfj7	20181115	Gj7sQs9
Ghs7dte	20181114	cFa598H
jhgfG7uR	20181119	8ftuW81

ing order based on the calculated admission date. For each serial number, medical institution code, ID0, and ward division attribute, the day of calculated hospitalization of the next record is added to "lag" (day of calculated hospitalization), and the day of calculated hospitalization of the previous record is inserted to "lead" (day of calculated hospitalization). When there are no upper or lower values, the value is entered as NULL.

In **Table 7B**, admission and discharge dates with the same claim serial number are included together with medical institution code, ID0, and ward division attributes, where the day of calculated hospitalization minus lag (day of calculated hospitalization) and lead (day of calculated hospitalization) minus day of calculated hospitalization are both set at "1" and deleted. Then, for each claim serial number, medical institution code, ID0, and ward division attribute, the day of calculated hospitalization of the previous record is inserted into lead (2nd day of calculated hospitalization). If there is no lower value, the value is entered as "NULL."

In **Table 7C**, to delete records with discharge dates sharing the same claim serial number, medical institution code, ID0, and ward division attributes, records with the day of calculated hospitalization minus lag (day of calculated hospitalization) being "1" and lead (day of calculated hospitalization) minus the day of calculated hospitalization being anything other than "1" are deleted. Additionally, the day of calculated hospitalization is changed to the date of admission, and the lead (2nd day of calculated hospitalization) is changed to the date of discharge. For records in which the day of calculated hospitalization minus lag (day of calculated hospitalization) is other than "1," and lead (day of calculated hospitalization) minus day of calculated hospitalization is other than "1," which are outpatient hospitalizations, the value of the admission date is entered in the discharge date column.

In **Table 7D**, lag (day of calculated hospitalization) and lead (day of calculated hospitalization) are deleted. The length of hospitalization is inserted as discharge date minus hospitalization date plus 1, and the claim serial number and hospitalization order are assigned the hospitalization serial number.

In **Table 7E**, a parent–subordinate relationship table for the hospitalization serial number is created by self-joining the table in **Table 7D** with the medical institution code, ID0, ward division, date of admission, and date of discharge plus 1.

In **Table 7F**, using the parent–subordinate relationship table, the parent serial number of the hospitalization in **Table 7D** is recursively updated, and the process is repeated until there are no more updated data.

As shown in **Figure 3**, we concatenate the last parent–subordinate relationship records into the Related RE table with the Modified SI_TO_IY (**Table 3D**, respectively), HO, IR, SY, and SB records containing medical hospitalization and DPC claim data.

Finally, the outpatient and hospitalization data obtained from health insurance claims are integrated using the ID0 patient key.

2.6 Analysis

Data were analyzed using Microsoft® SQL server® 2016 Enterprise Edition, a computerized database package with a columnstore index [19]. The hardware used consisted of a DELL PowerEdge R730xd computer with an

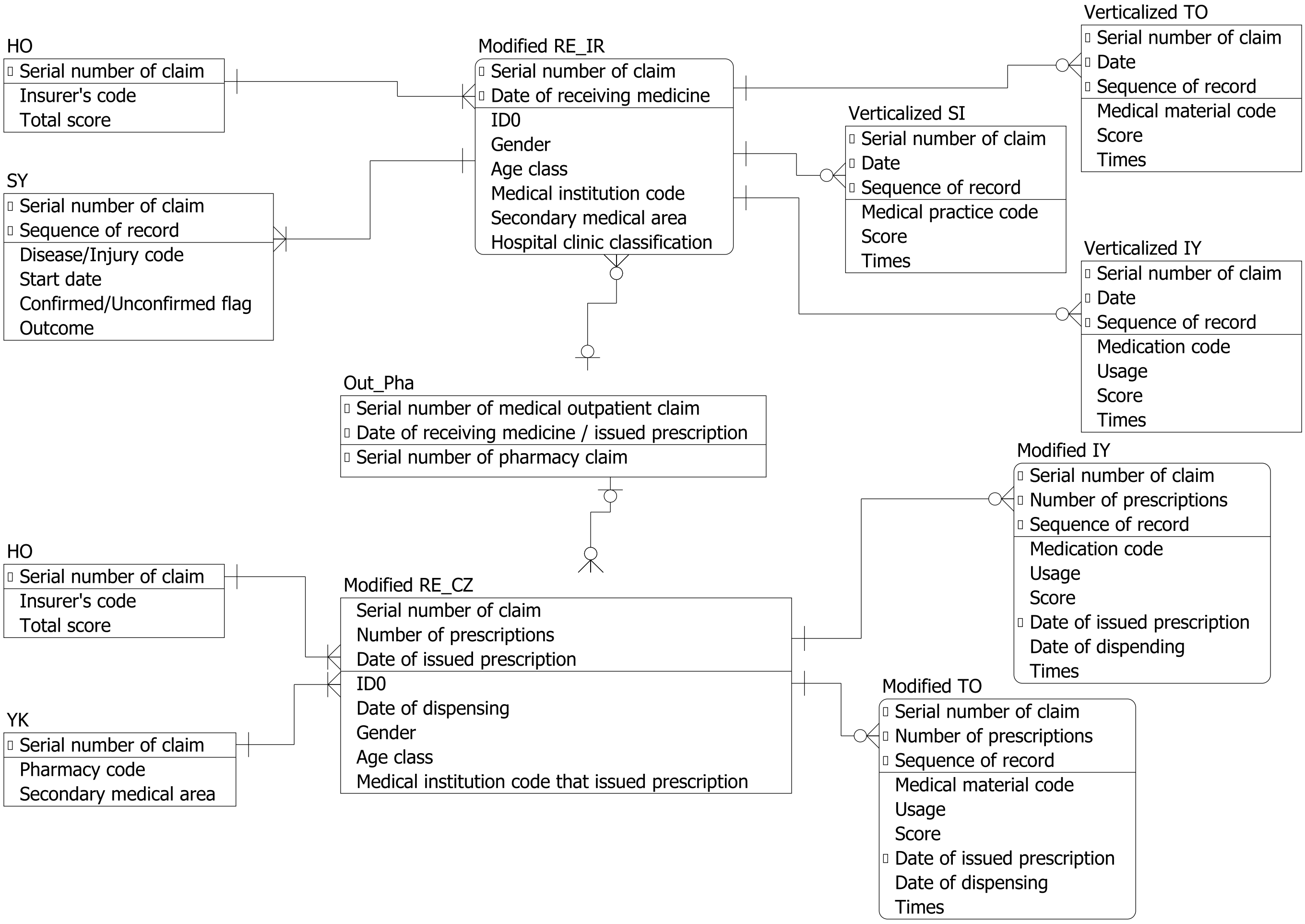

Fig. 1 ER Diagram of outpatient data after merging medical outpatient and pharmacy claims.

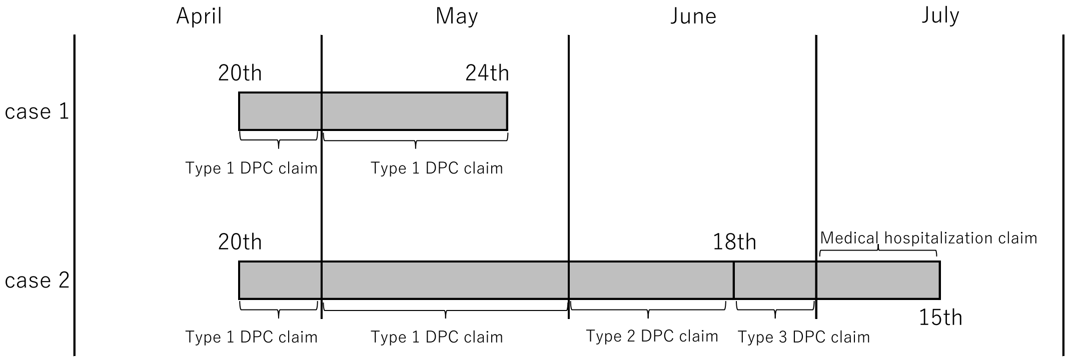

Fig. 2 Type of claim issued during DPC hospitalization (DPC-specific period = 60 days).

Intel XeonR E5-2640 v4 2.4 GHz 10C/20Tx2 CPU, 1536 GB memory (64 GB LRDIMM x24), and a 52 TB disk (2 TB NearLine SAS x26). In this environment, we imported the NDB data of all patients for the periods and items specified in Section 2.1 and processed them as described in Sections 2.2–2.5. Then, we compared file sizes before and after execution and measured execution time.

2.7 Application of analysis

To conduct a large cohort study using the NDB, having the non-target group as the entire population excluding the target group is desirable. We considered that a follow-up period longer than the 3-year period of this study would be ideal. In the NDB, there are approximately 137 million patients (ID0) who received medical service outside the 3-year study period, and a simple calculation

Table 7 Flow of determining the duration of hospitalization.

A. Sort records identified as hospitalization by serial number of claim × medical institution code × ID0 × ward division in order of the date of calculated hospitalization

Serial number of claim	Medical institution code	ID0	Ward division	Day of calculated hospitalization	lag (day of calculated hospitalization)	lead (day of calculated hospitalization)
ABC	1qaz2wsx	hogehoge	1	20130403	NULL	20130404
ABC	1qaz2wsx	hogehoge	1	20130404	20130403	20130405
ABC	1qaz2wsx	hogehoge	1	20130405	20130404	20130406
ABC	1qaz2wsx	hogehoge	1	20130406	20130405	20130429
ABC	1qaz2wsx	hogehoge	1	20130429	20130406	20130430
ABC	1qaz2wsx	hogehoge	1	20130430	20130429	NULL
ZYX	1qaz2wsx	hogehoge	1	20130501	NULL	20130522
ZYX	1qaz2wsx	hogehoge	1	20130522	20130501	NULL

B. Delete records other than admission and discharge dates and add discharge date

Serial number of claim	Medical institution code	ID0	Ward division	Day of calculated hospitalization	lag (day of calculated hospitalization)	lead (day of calculated hospitalization)	lead (2nd day of calculated hospitalization)
ABC	1qaz2wsx	hogehoge	1	20130403	NULL	20130404	20130406
ABC	1qaz2wsx	hogehoge	1	20130404	20130403	20130405	
ABC	1qaz2wsx	hogehoge	1	20130405	20130404	20130406	
ABC	1qaz2wsx	hogehoge	1	20130406	20130405	20130429	20130429
ABC	1qaz2wsx	hogehoge	1	20130429	20130406	20130430	20130430
ABC	1qaz2wsx	hogehoge	1	20130430	20130429	NULL	NULL
ZYX	1qaz2wsx	hogehoge	1	20130501	NULL	20130522	20130522
ZYX	1qaz2wsx	hogehoge	1	20130522	20130501	NULL	NULL

C. Delete discharge date record and set admission date to discharge date if one-day hospitalization

Serial number of claim	Medical institution code	ID0	Ward division	Date of admission	lag (day of calculated hospitalization)	lead (day of calculated hospitalization)	Date of discharge
ABC	1qaz2wsx	hogehoge	1	20130403	NULL	20130404	20130406
ABC	1qaz2wsx	hogehoge	1	20130404	20130403	20130405	
ABC	1qaz2wsx	hogehoge	1	20130405	20130404	20130406	
ABC	1qaz2wsx	hogehoge	1	20130406	20130405	20130429	20130429
ABC	1qaz2wsx	hogehoge	1	20130429	20130406	20130430	20130430
ABC	1qaz2wsx	hogehoge	1	20130430	20130429	NULL	NULL
ZYX	1qaz2wsx	hogehoge	1	20130501	NULL	20130522	20130501
ZYX	1qaz2wsx	hogehoge	1	20130522	20130501	NULL	20130522

D. Set hospitalization serial number using serial number of claim and order of hospitalization

Serial number of claim	Medical institution code	ID0	Ward division	Date of admission	Date of discharge	Length of hospitalization	Serial number of hospitalization
ABC	1qaz2wsx	hogehoge	1	20130403	20130406	4	A01
ABC	1qaz2wsx	hogehoge	1	20130404			
ABC	1qaz2wsx	hogehoge	1	20130405			
ABC	1qaz2wsx	hogehoge	1	20130406	20130429		
ABC	1qaz2wsx	hogehoge	1	20130429	20130430	2	A02
ABC	1qaz2wsx	hogehoge	1	20130430	NULL		
ZYX	1qaz2wsx	hogehoge	1	20130501	20130501	1	B01
ZYX	1qaz2wsx	hogehoge	1	20130522	20130522	1	B02

E. Create a parent-subordinate relationship table by self-joining Table 7D with ID0, medical institution code, ward division, date of hospitalization, and date of discharge + 1

Subordinate serial number of hospitalization	Parent serial number of hospitalization
B01	A02

F. Using the parent-subordinate relationship table in Table 7E, recursively update the parent serial number of hospitalization in Table 7D and repeat until there is no more updated data

Serial number of claim	Medical institution code	ID0	Ward division	Date of admission	Date of discharge	Length of hospitalization	Serial number of hospitalization	Parent serial number of hospitalization
ABC	1qaz2wsx	hogehoge	1	20130403	20130406	4	A01	A01
ABC	1qaz2wsx	hogehoge	1	20130429	20130430	2	A02	A02
ZYX	1qaz2wsx	hogehoge	1	20130501	20130501	1	B01	A02
ZYX	1qaz2wsx	hogehoge	1	20130522	20130522	1	B02	B02

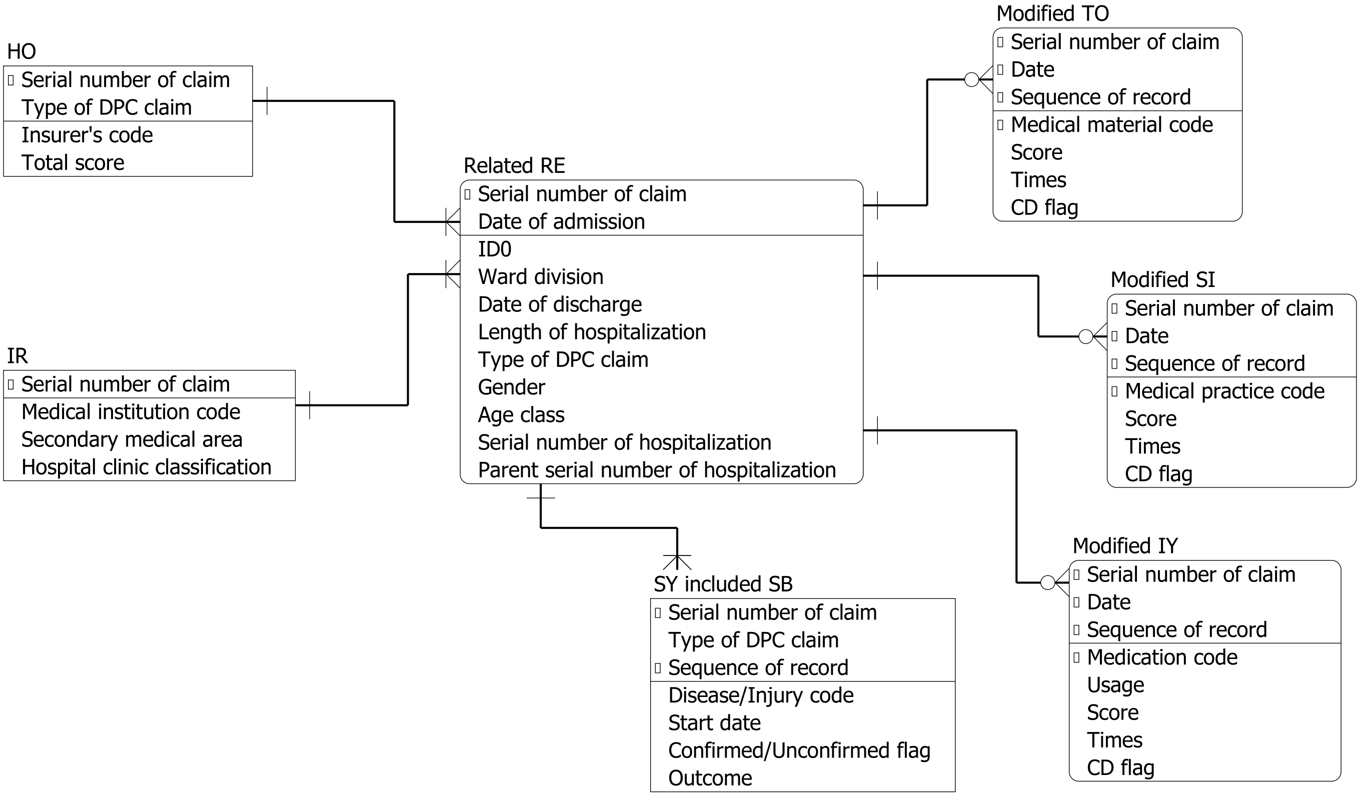

Fig. 3 ER Diagram of hospitalization data after merging medical hospitalization and DPC claims.

yields an approximately 411 million person-year cohort. After validating this method using the NDB data from April 2013 to March 2016, we applied the method to conduct a cohort study on over 500 million person-years using NDB. and published the results [20].

3. Results

We imported the NDB data of all patients for the periods and items specified in Section 2.1 and processed them as described in Sections 2.2 to 2.5. Note that the original CSV file size containing the data extracted with dates from April 2013 to March 2016 before execution was approximately 10.5 TB (CSV), and the file size after execution was approximately 6.0 TB (MDF). The execution time was approximately 3 months.

As mentioned in Section 2.7, this method was applied to a 6-year mortality cohort from April 2013 to March 2019 in all Japanese citizens with or without diabetes and published a paper as a result [20]. The study used NDB data to follow people with or without diabetes and compared their ages at death. The population included 142,797,986 people (7,448,962 females and 6,770,024 males) over a six-year period; among them, 4,647,016 females and 6,507,817 males had diabetes. During the six-year period, 2,786,071 females and 2,975,876 males died; among them, 652,699 females and 954,655 males had diabetes. The mean age of death in diabetic patients was 2.6 years shorter than the mean age of death in non-diabetic patients.

4. Discussion

In this study, we converted the DB structure of NDB to a suitable format for cohort studies, reducing the file size of the MDF format after execution compared with the file size of the CSV format before execution, and the processing time per year of data was approximately one month. Specifically, we unpivoted date information from the SI_IY_TO table and compared the entries to CD table records to separately tag piece-rate and comprehensive DPC claims. Additionally, medical hospitalization claims and DPC claims were linked so that a series of episodes during hospitalization could be identified. Outpatient and pharmacy claims were linked to track drugs prescribed in outpatient consultations. Finally, we accomplished this process by minimizing the conversion time without increasing the data size. This effort does not meet the goal in terms of data processing time, but the improvements made toward reaching it are impressive. This made it far easier to conduct both cross-sectional and longitudinal cohort studies using NDB data.

The processing time shown is the execution time after data loading. Processing time depends on CPU, memory, and hard disk input/output (Disk i/o). In this study, Disk i/o was considered to be the rate-limiting factor.

Therefore, there is a possibility that the processing time could be further reduced by changing the disk type to SSD. The increase in file size was suppressed, probably due largely to the use of the columnstore index, which is a method for storing, retrieving, and managing data using a columnar data format (columnstore), in addition to the fact that Japanese names, which are double-byte characters, were not included.

A past study converted claims to EF file format [14]. The EF file format contained the names of medical practices in Japanese, which may have increased the data volume while lengthening the processing time. The EF file format is good for understanding medical treatments and scores from the same medical institution and the same patient. However, linking a patient to different medical institutions and hospitalizations is difficult because of insufficient normalization of the EF file format. In this study, we primarily focused on patient tracing, and it was easy to combine claims from different medical institutions using our modified database.

Our process also contributed to the development of a death determination logic. This logic [21] was applied to concatenate claims and mortality information for a limited geographic area on an individual basis to generate a condition that occurs from the claim at the time of death. The conditions were then extrapolated to the NDB, and mortalities were confirmed on the NDB. This research and the death determination logic contributed to the retrospective cohort study of over 500 million person-years, which is the most notable accomplishment of this research to date.

The MHLW provides 24 types of aggregated purpose-specific datasets in easy-to-use formats for different purposes [22]. All datasets are simple unpivots of multiple records and do not allow patient tracing. Furthermore, purpose-specific datasets are not linked to medical hospitalization, DPC, outpatient, or pharmacy attribute.

The conduct of this research had several limitations. First, the NDB data did not include claims that were fully paid by the public (such as medical assistance and independence support). Such claims include medical cost information in a KO record instead of an HO record. Notably, the structures of the HO and KO tables are analogous. Thus, this research method can be easily extended.

Second, the definition of a single case of hospitalization was not conclusive. We defined hospitalization in the same medical institution and the same ward division as one period. Thus, the following cases were considered as one hospitalization:

- Readmissions on the same or next day following discharge
- Several consecutive one-day hospitalizations at the same medical institution
- Readmission with the same injury or disease within seven days of discharge in the DPC claim

Notably, it was difficult to determine the number of discharges from the NDB data. Hence, the average length of hospital stay applied during validation may have been overestimated.

Furthermore, short-stay surgeries were not considered in the determination of hospitalization duration. Short-stay surgeries normally have hospitalization of one to five days, and only the day of surgery is calculated with an obvious code. Therefore, it is difficult to determine whether the patient was hospitalized before or after surgery. Again, the average length of hospital stay may have been overestimated. These limitations may be overcome and the hospitalization durations distinguished by carefully investigating the medical practice codes individually.

Finally, the NDB includes only electronic claims after a review of the healthcare bill by the paying organization. Therefore, billing and re-billing are done on paper, and the results are not digitized. However, 98.8% of initial claims are filed electronically [23]. Although the rate of paper claims is not disclosed, the MHLW has indicated that it intends to convert all returned claims to electronic forms by March 2023 [24].

Rather than an exploratory study in the research phase, the present study was conducted for the MHLW to develop medical care plans, and NDB-analysis is expected to be incorporated as a part of the MHLW routine operations. It is also expected to lead to the establishment of a highly effective healthcare delivery system.

5. Conclusion

Our report presents a new method of tracing patient data through health insurance claims using the NDB and provides an example of the analysis.

Author's contribution

All authors participated in this review. Tomoya Myojin gathered the relevant articles and drafted the manuscript. Tsuneyuki Higashino analyzed the data. Tatsuya Noda and Tomoaki Imamura contributed to the interpretation of this review. The other authors edited the manuscript. All authors approved the final manuscript before submission.

Conflict of Interest

The authors declare no conflicts of interest associated with this manuscript.

Acknowledgement

We would like to thank Editage (www.editage.com) for

English language editing.

This research was funded by a grant from the Japan Agency for Medical Research and Development [Grant Number: JP16lk1310001h0001], Development and Japan Society for the Promotion of Science KAKENHI (Grant number: 18H04126), and the Health Science and Labour Research Grant [Grant number: H30-Iryou-Ippan-013 and 21IA1006] from the Ministry of Health, Labour and Welfare, Japan.

Abbreviations

CSV: Comma-Separated Values
DPC: Diagnosis Procedure Combination
ER: Entity–Relationship
ICD-10: International Statistical Classification of Diseases and Related Health Problems 10th Revision
MDF: Main Database File
MHLW: Ministry of Health, Labour and Welfare
NDB: The National Database of Health Insurance Claims and Specific Health Checkups of Japan

References

1. Shibuya K, Hashimoto H, Ikegami N, Nishi A, Tanimoto T, Miyata H, et al.: Future of Japan's system of good health at low cost with equity: beyond universal coverage. Lancet. **378**, 1265–1273, 2011.
2. Institute for Molecular Science. https://www.ims.ac.jp/en/life/immigration.html [accessed on 1 June, 2022].
3. NDB open data. Ministry of Health, Labour and Welfare. https://www.mhlw.go.jp/content/12400000/000821378.pdf [accessed on 1 June, 2022].
4. Okamoto E.: Linkage rate between data from health checks and health insurance claims in the Japan National Database. J Epidemiol. **24**(1), 77–83, 2014.
5. Mizuno K, Takeuchi M, Kishimoto Y, Kawakami K, Omori K.: Indications and outcomes of pediatric tracheotomy: a descriptive study using a Japanese claims database. BMJ Open. **9**, e031816, 2019.
6. Ono S, Ono Y, Koide D, Yasunaga H.: Association between routine nephropathy monitoring and subsequent change in estimated glomerular filtration rate in patients with diabetes mellitus: a Japanese non-elderly cohort study. J Epidemiol. **30**(8), 326–331, 2020.
7. Hosomi K, Fujimoto M, Ushio K, Mao L, Kato J, Takada M.: An integrative approach using real-world data to identify alternative therapeutic uses of existing drugs. PLoS One. **13**(10), e0204648, 2018.
8. Horii T, Oikawa Y, Kunisada N, Shimada A, Atsuda K.: Real-world risk of hypoglycemia-related hospitalization in Japanese patients with type 2 diabetes using SGLT2 inhibitors: a nationwide cohort study. BMJ Open Diabetes Res Care. **8**, e001856, 2020.
9. Nakamura M, Yamashita T, Hayakawa A, Matsumoto T, Takita A, Hasegawa C, et al.: Bleeding risks associated with anticoagulant therapies after percutaneous coronary intervention in Japanese patients with ischemic heart disease complicated by atrial fibrillation: A comparative study. J Cardiol. **77**(2), 186–194, 2021.
10. Nagai K, Iseki C, Iseki K, Kondo M, Asahi K, Saito C, et al.: Higher medical costs for CKD patients with a rapid decline in eGFR: A cohort study from the Japanese general population. PLoS One. **14**(5), e0216432, 2019.
11. Tanaka H, Nakamura F, Higashi T, Kobayashi Y.: Cancer treatment situation in Japan with regard to the type of medical facility using medical claim data of Health Insurance Societies. Nihon Koshu Eisei Zasshi. **62**(1), 28–38, 2015. (in Japanese)
12. Noda T, Kubo S, Myojin M, Nishioka Y, Higashino T, Matsui H, et al.: A new algorithm for patient matching in the National Database of Health Insurance Claims and Specific Health Checkups of Japan (NDB). J Health Welfare Stat. (Kousei-no-Shihyou). **64**(12), 3–15, 2017. (In Japanese)
13. Kubo S, Noda T, Myojin T, Nishioka Y, Higashino T, Matsui H, et al.: National Database of Health Insurance Claims and Specific Health Checkups of Japan (NDB): Outline and Patient-Matching Technique. Biorxiv https://www.biorxiv.org/content/10.1101/280008v1.full.pdf [accessed on 1 June, 2022].
14. Hokkaido University and HQF Co., Ltd. Electronic claim data conversion program and electronic claim data conversion system. P2011-118538A. 16 June, 2011.
15. Health Insurance Claims Review & Reimbursement services. https://www.ssk.or.jp/seikyushiharai/rezept/iryokikan/iryokikan_02.files/jiki_d01.pdf (in Japanese) [accessed on 1 June, 2022].
16. Health Insurance Claims Review & Reimbursement services. https://www.ssk.or.jp/seikyushiharai/rezept/iryokikan/iryokikan_02.files/jiki_i01.pdf (in Japanese) [accessed on 1 June, 2022].
17. Health Insurance Claims Review & Reimbursement services. https://www.ssk.or.jp/seikyushiharai/rezept/iryokikan/iryokikan_02.files/jiki_t01.pdf (in Japanese) [accessed on 1 June, 2022].
18. Ministry of Health, Labour and Welfare. https://www.mhlw.go.jp/content/12400000/000679975.xlsx (in Japanese) [accessed on 1 June, 2022].
19. Microsoft. https://docs.microsoft.com/en-us/sql/relational-databases/indexes/columnstore-indexes-overview?view=sql-server-ver16 [accessed on 1 June, 2022].
20. Nishioka Y, Kubo S, Okada S, Myojin T, Higashino T, Imai K, et al.: The age of death in Japanese patients with type 2 and type 1 diabetes: A descriptive epidemiological study. J Diabetes Investig. **13**(8), 1316–1320, 2022.
21. Kubo S, Noda T, Nishioka Y, Myojin T, Nakanishi Y, Furihata S, et al.: Mortality tracking using the National Database of Health Insurance Claims and Specific Health Checkups of Japan (NDB). Jpn J Med Inform. **40**, 319–335, 2021. (in Japanese)
22. Ministry of Health, Labour and Welfare. Patient Survey. Homepage about data provision of the National Database of Health Insurance Claims and Specific Health Checkups of Japan https://www.mhlw.go.jp/stf/seisakunitsuite/bunya/kenkou_iryou/iryouhoken/reseputo/index.html (in Japanese) [accessed on 18 October, 2021].
23. Health Insurance Claims Review & Reimbursement services. https://www.ssk.or.jp/tokeijoho/tokeijoho_rezept/tokeijoho_rezept_r03.files/seikyu_0402.pdf (in Japanese) [accessed on 1 June, 2022].
24. Ministry of Health, Labour and Welfare. https://www.mhlw.go.jp/content/12401000/000761330.pdf (in Japanese) [accessed on

1 June, 2022].

Tomoya Myojin

Tomoya Myojin received the M.E. in environmental hygiene from Kyoto University in 2009, and received M.D. from Hamamatsu University School of Medicine in 2016. He is currently an Assistant Professor in Nara Medical University after working at Nomura Research Institute as a system engineer, at hospital as a pathologist, and at the Ministry of Health, Labour and Welfare as a technician. His major research interest is public health about health care insurance and medical informatics. He is a member of the Japanese Society of Pathology, the Japanese Society of Public Health and the Japan Association for Medical Informatics.

Tatsuya Noda

Tatsuya Noda received the M.D. and Ph.D. from Kyushu University in 2001 and 2012 respectively. He is currently an Associate Professor in the Department of Public Health, Health Management and Policy, Nara Medical University.

Shinichiro Kubo

Shinichiro Kubo received the M.MS. ans Ph.D. from Nara Medical University in 2018 and 2022 respectively. He is currently a postdoctoral researcher in Nara Medical University and a nursing technician at the Ministry of Health, Labour and Welfare.

Yuichi Nishioka

Yuichi Nishioka received the M.D. and Ph.D. from Nara Medical University in 2015 and 2020 respectively. He is currently an Assistant Professor in the Department of Public Health, Health Management and Policy, Nara Medical University.

Tsuneyuki Higashino

Tsuneyuki Higashino joined Mitsubishi Research Institute after receiving his B.E. degree. He works on promoting claims analysis through database construction, normalization, and bigdata analysis environment development.

Tomoaki Imamura

Tomoaki Imamura received the M.D. from Kansai Medical University in 1988 and received the Ph.D. from the University of Tokyo in 1993. He is currently a Professor in the Department of Public Health, Health Management and Policy, Nara Medical University after working at the Ministry of Health, Labour and Welfare as a technician, at the University of Tokyo Hospital as a financial manager and so on.