

Databases available for pharmacoepidemiology researches in Japan (information obtained from survey answers as of August 2026)

	Hospital-based																
	NHO NCDA/MIA	DPC database	MID-NET	Tokushukai	LDI	DATuM IDEA®	FAST-HDJ_DB	4DIN	MDV	JMDC	JMDC(EMR)	TXP Medical Database	OnelCU	MEI (KDDI Grp.)	Flatiron Health	J-CONNECT	JAMDAS
Database administrator	National Hospital Organization	DPCRG	Pharmaceuticals and Medical Devices Agency	General Incorporated Association Tokushukai	Life Data Initiative (LDI)	TOPPAN Holdings Inc.	Fair and safe use of anonymized health data of Japan	4DIN Ltd.	Medical Data Vision Co., Ltd.	JMDC.Inc	HCEI/JMDC.Inc	TXP	MeDICU, Inc.	Medical Engineering Institute, Inc.	Flatiron Health K.K.	Prime Research Institute for Medical RWD, Inc.	m3 inc.
Database name	NHO Clinical Data Archives / Medical Information Analysis databank	DPC database	Medical Information Database Network (MID-NET)	Tokushukai Medical Database	Millennial Medical Record	DATuM IDEA®	NGMIL database	4DIN Research Network	EBM Provider®	JMDC Hospital Database	JMDC Electronic medical record database	TXP Medical Database	OnelCU	MEI Database	Flatiron Health Research Database	J-CONNECT	Japan Medical Data Survey
Contact information		kfushimi.hci@tmd.ac.jp	wakaru-midnet@pmda.go.jp		contact@ldi.or.jp	https://datumidea.jp/contact/	query@fast-hdj.org	info@4din.com	rwd_sales@mdv.co.jp	mdbhelp@jmdc.co.jp	mdbhelp@jmdc.co.jp	medical-data-service@txpmedical.com	info@medicu.co.jp	rwd@meiz.co.jp	inquiry-japan@flatiron.com	pr-ml-primer@ntt.com	https://docs.google.com/forms/d/e/1FAIpQLSeLU3AaTktKr2KbIV1r5OJUJgDC3NrTKfgyFTE5exIWt3OJHw/viewform
Web site URL	https://nho.hosp.go.jp/cnt1-1_000070.html	None	https://www.pmda.go.jp/safety/mid-net/0001.html	https://www.tokushukai-is.com/service/tmd.php	https://www.ldi.or.jp	https://datumidea.jp/	https://www.fast-hdj.org	https://4din.com/	https://www.mdv.co.jp/	https://www.phm-jmdc.com/	https://www.phm-jmdc.com/	https://medical-dataservice.txpmedical.jp/	https://www.medicu.co.jp/	https://www.meiz.co.jp/rwd/	https://flatiron.co.jp/	https://prime-r.inc/rdu-support/	https://corporate.m3.com/
Published article about the database (DOI)			10.1002/pds.4777 10.1002/pds.4879		https://doi.org/10.3820/jipe.27.3					http://dx.doi.org/10.1002/igt2.367	https://doi.org/10.37737/ace.24009	https://doi.org/10.1002/ams2.554			DOI: 10.1016/j.esmow.2025.100113	https://link.springer.com/article/10.1007/s10147-026-03011-4	
Database Overview																	
General description of the database	A database in which electronic health record (EHR) information from 87 hospitals of the NHO, together with DPC and medical receipt data from all hospitals, are collected and stored at the headquarters.	DPC data from DPC hospital	The medical information Database network system for utilizing in safety assessment managed by PMDA under the Act on the Pharmaceuticals and Medical Devices Agency	The Tokushukai Medical Group operates approximately 80 hospitals nationwide and currently manages medical data for over 17.4 million patients in its database (TMD). Tokushukai is also engaged in research and AI development in collaboration with corporate partners, and can extract hospital medical data, management data, and clinical data for these purposes.	A database certified under Japan's Next Generation Medical Infrastructure Act, primarily comprising electronic medical records (EMRs), medical claims data, and DPC data from large hospitals, including advanced treatment hospitals. Unstructured text from EMRs enables more in-depth research, including the capture of clinical outcomes and granular patient stratification. Anonymized medical imaging data and pseudonymized medical information are also available. Data linkage with public databases, including the National Database of Health Insurance Claims and Specific Health Checkups of Japan (NDB), is also available.	We handle anonymously processed medical information obtained from certified nextgeneration medical infrastructure providers. As approximately half of the facilities covered are regional cancer center hospitals, our electronic medical record database contains a large amount of treatment data for cancer and rare diseases.	Based on the Next-Generation Medical Infrastructure Law, this database collects medical information from cooperating medical institutions, focusing on objective data from electronic medical records. It includes DPC-related files and information related to insurance claims.	Network to collect, deidentify and summarize various data from medical institutions	Administrative database for inpatient and outpatient consists of 949 acute (mainly considered as "advanced treatment hospitals") hospitals in Japan	A case database sourced from receipts, electronic medical records (clinical laboratory values), and DPC survey data collected from more than 1,000 DPC/non-DPC hospitals nationwide. Includes abundant information on elderly and hospitalized patients.	Integrated database of medical information such as electronic medical records, DPC data, and claims	The database centers on university hospitals and regional core hospitals with an average of 650+ beds. It contains structured emergency department and ICU data, as well as bedside vital signs, clinical notes, bacterial culture results, blood product usage history, and 900 lab test items for all inpatients. Covering acute-phase conditions as well as cancer and rare diseases, it provides aggregated statistical data supplied by participating institutions.	A multicenter ICU database integrating critical care data from multiple institutions, structuring diagnoses, vitals, labs, SOFA score, interventions (infusion, blood transfusion, ventilation, RRT, ECMO), and outcomes with minute-level resolution.	A database comprising electronic health record (EHR) data and medical claims/billing data collected primarily from large hospitals and regional core hospitals across Japan. The database consists of anonymized personal information and pseudonymized personal information. Clinical outcomes, including various assessment scores, genetic testing results, and pathology findings, can be derived from unstructured EHR text data. With data available with a lag of approximately one month, the database enables near real-time prescription analyses, including evaluations conducted immediately following label expansions or updates to clinical practice guidelines.	A patient-level longitudinal oncology real-world database containing clinical information critical for cancer research (e.g. stage, histology, genetic mutations, lines of therapy, and clinical outcomes including death and disease progression), curated with Flatiron Health's proprietary data processing methodology, based on structured and unstructured information extracted and de-identified from electronic medical records and other healthcare information systems at domestic cancer centers and university hospitals in Japan	A cancer real-world database derived from electronic health records (EHRs) and built on the National Database of Hospital-Based Cancer Registries. Patient characteristics, treatments, and clinical outcomes are captured with high accuracy using structured data, natural language processing (NLP), and proprietary algorithms. Additional collection and registration of biomarker information from pathology reports and other missing data are also supported. The database is available for secondary use in academic research registries.	The Japan Medical Data Survey (JAMDAS) is a real-world database that provides an accurate and real-time understanding of what is happening in medical settings throughout Japan.

	Hospital-based																
	NHO NCDA/MIA	DPC database	MID-NET	Tokushukai	LDI	DATuM IDEA®	FAST-HDJ_DB	4DIN	MDV	JMDC	JMDC(EMR)	TXP Medical Database	OnelCU	MEI (KDDI Grp.)	Flatiron Health	J-CONNECT	JAMDAS
Data source																	
DPC data (except for DPC claim)	Yes	Limited	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	No	Yes	No
Medical/DPC claim	Yes	No	Yes	Yes	Yes	Yes	Yes	Limited	Yes	Yes	Yes	Yes	No	Yes	No	Yes	No
Dental claim	No	No	No	No	No	No	Yes	No	No	No	No	No	No	No	No	No	No
Pharmacy claim	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No
Electronic medical chart	Yes	No	Yes	Yes	Yes	Yes	Limited	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Lab test results	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	Limited	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Ordering		No	Yes	Yes	Yes	Yes	Yes	Limited	No	No	Yes	Yes	Yes	Yes	Yes	No	Yes
Health checkup	Yes	No	No	No	No	No	Limited	No	No	No	Limited	No	No	Limited	No	No	No
Others	No	No	No	Yes	Yes	Limited	Yes	Limited	Yes	No	No	Yes	Yes	Yes	Yes		Yes
Others, notes					DPC data includes Form 1, Form 3, D-file, EF-file, Outpatient EF-file, H-file, and K-file. Electronic medical records also include unstructured text data such as Progress notes, Clinical summary /Discharge summary and Medical reference letter etc. In addition, Anonymized medical images are available.	Deidentified analytic free-text data, such as physician-authored clinical notes, and structured information and other derived outputs generated from or based on the text.	Prescription, Medical Image data		DPC (Dfile) /DPC claims/blood and urine test data *Blood and urine test data is limited to some hospitals			TXP Medical provides healthcare institutions with specialized departmental systems such as NEXT Stage ER and NEXT Stage ICU. Through these systems, structured clinical data is captured in real time during routine ER and ICU workflows using a unified multi-institutional format. The database also fully incorporates DPC (Diagnosis Procedure Combination) data, including Form 1 as well as inpatient and outpatient EF files.	Minute-level time-series data including patient demographics, clinical diagnoses, outcomes, vital signs, consciousness assessment scores (GCS, JCS, CAM-ICU, RASS), fluid balance, laboratory and microbiological test results, medication records (infusions, injections, transfusions, oral drugs), and detailed device-based treatment parameters (mechanical ventilation, renal replacement therapy, ECMO, IABP, Impella)	A healthcare database integrating information from electronic medical records (EMRs) and administrative systems, developed based on insights gained through 20 years of providing data warehouse solutions to healthcare institutions. In addition to structured data derived from EMRs, the database includes unstructured free-text data such as physicians' and nurses' clinical notes (e.g., progress notes and observation findings) as well as pathology reports. DPC and claims data are also available.	Flatiron Health extracts the following data from electronic health record (EHR) and related hospital information systems: - Structured data (patient demographic characteristics, visit dates, test values etc., which are stored in table format or similar within the EHR) - Unstructured data (progress notes, pathology reports, referral letters, etc., which are stored in text or report format within the EHR or departmental systems)		Medical Questionnaire
Data collected from	Medical institution	Medical institution	Medical institution	Medical institution	Medical institution	Medical institution	Medical institution	Medical institution	Medical institution	Medical institution	Medical institution	Medical institution	Medical institution	Medical institution	Medical institution	Medical institution	Medical institution
Others, notes																	
Number of organizations	140		31	80	77	63	12		949	1400	295	52	50	54	5	16	10000
Data period																	
Starting year	2016	2010	2009	2010	2015	2019	2016	Before 2000	2008	Before 2000	Before 2000	2018	2011	2015	2011	2018	2015
Latest year	2026	2024	2026	2026	2026	2025	2026	2026	2026	2026	2026	2026	2026	2026	2026	2026	2026
Notes				The starting year varies by facility.	The database includes patients who visited participating hospitals on or after July 2020, excluding those who requested that their data not be provided for secondary use. For eligible patients, data from before July 2020 may also be available.		It varies depending on the medical institution	By participating hospitals	*Mainly from April 2010			Data derived from NEXT Stage ER and ICU systems is available starting from the implementation year at each institution. Electronic medical record (EMR) data is accessible from the year of EMR adoption, providing an average of approximately 15 years of historical clinical records. DPC data covers the period from January 2018 to March 2026, with a 3-month lag.		- The start year varies by participating healthcare institution. - Data acquisition will continue on an ongoing basis, and the most recent data will be added continuously.	Patients with a confirmed diagnosis of the tumor type of interest (diagnosis date: 1 Jan 2011 onwards) and from whom appropriate consent has been obtained will be included in the database. As the database is refreshed on a quarterly basis, data recency is maintained at 90 days (e.g. datasets delivered in June 2026 have a data cutoff date of 31 March 2026).	The data availability period varies by data element.	
Laws and regulations related to the data provision to third parties	Act on the Protection of Personal Information, Ethical Guidelines for Medical Research, The Next Generation Medical Infrastructure Act	Ethical Guidelines for Medical Research	Other		The Next Generation Medical Infrastructure Act	The Next Generation Medical Infrastructure Act	Act on the Protection of Personal Information, The Next Generation Medical Infrastructure Act	Act on the Protection of Personal Information	Other	Act on the Protection of Personal Information	Act on the Protection of Personal Information	Act on the Protection of Personal Information, Ethical Guidelines for Medical Research, Other	Act on the Protection of Personal Information, Ethical Guidelines for Medical Research	Act on the Protection of Personal Information	Act on the Protection of Personal Information, Ethical Guidelines for Medical Research	Act on the Protection of Personal Information	
Notes			We collect data and provide it to third parties based on Act on Pharmaceuticals and Medical Devices Agency, also with attention to the persons whose electronic medical records was acquired.						We do not provide or receive special care-required personal information to or from third parties.			This database is provided to third parties as aggregated statistical information. Statistical information refers to data generated by extracting and aggregating common data points from multiple individuals. Because it contains no identifiable individual data, it does not constitute "personal information" and falls outside the scope of Japan's Act on the Protection of Personal Information (APPI).			N/A		

	Hospital-based																
	NHO NCDA/MIA	DPC database	MID-NET	Tokushukai	LDI	DATuM IDEA®	FAST-HDJ_DB	4DIN	MDV	JMDC	JMDC(EMR)	TXP Medical Database	OnelCU	MEI (KDDI Grp.)	Flatiron Health	J-CONNECT	JAMDAS
Handling anonymously processed information	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	No	No	Yes
Compliance with the laws and regulations on third-party data provision	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Size																	
Number of unique identifiers [approx., 10K people]	500	11000	870	1741	371	193	28	1422	6370	6131	2490	2600	26	700	2	10	8700
Notes			As of December,2025		The number of Person IDs represents patients from hospitals for which data linkage has been completed among the 77 hospitals contracted for secondary data use.							Calculated using publicly reported hospital metrics (e.g., annual inpatient and outpatient volumes) for each contracted facility.	26	Patient population covered by institutions with established data integration, among the 87 institutions that have entered into data secondary-use agreements.	As of 2026 Q4, the dataset will include approximately 20,000 patients in total for the target tumor types provided (breast cancer, colorectal cancer, gastric cancer, and non-small cell lung cancer). Other tumor types can also be provided as custom datasets. Flatiron Health is actively expanding both cohort counts and the range of tumor types available.	All patients who received chemotherapy among those registered in the hospital-based cancer registry at participating institutions are included in the database.	
Annual number of unique identifiers (latest year) [approx., 10K people]		700		46	34.3	47	10	94	1871	1501	427	120	3	100	1	1.2	2800
The latest year	2025	2024		2025	2025	2025	2026	2025	2025	2025	2025	2026	2025	2025	2025	2026	2025
Notes				Patients with basic medical fees (Ministry of Health code A) recorded in accounting, excluding dental care. Reference: 1.4 million patients in 2019.			It varies depending on the medical institution		April 2025 - March 2026	April 2025 - March 2026	April 2025 - March 2026			The number of healthcare institutions participating under secondary-use data agreements is expanding rapidly, leveraging 20 years of experience in data collaboration with healthcare providers.	Curation of the Flatiron Health Research Database began in 2023. Data from new patients and the latest data for enrolled patients are being added to the database every three months. As of 2026, the database contains data from approximately 20,000 patients accumulated over the past 3 years, with plans to further accelerate expansion in the future.		
% coverage of target population					8.86% The database covers 35 of the 395 DPC hospitals with 400 or more beds nationwide (based on estimates as of June 1, 2026, from the Central Social Insurance Medical Council).	It covers 1.57% of Japan's population (according to the Population Estimates, December 2025 report by the Statistics Bureau of the Ministry of Internal Affairs and Communications). It covers 41 out of 729 hospitals in Japan with more than 400 beds (5.6%) (according to the Hospital Report from the 2024 Survey of Medical Facilities [Static and Dynamic] by the Ministry of Health, Labour and Welfare).			About 15.2% of Japanese population	about 12% of Japanese population	about 3.6% of Japanese population					Includes approximately 3% of all cases in the National Database of Hospital-Based Cancer Registries and maintains a distribution that closely reflects the national database.	23% coverage of Japan's population

	Hospital-based																
	NHO NCDA/MIA	DPC database	MID-NET	Tokushukai	LDI	DATuM IDEA®	FAST-HDJ_DB	4DIN	MDV	JMDC	JMDC(EMR)	TXP Medical Database	OnelCU	MEI (KDDI Grp.)	Flatiron Health	J-CONNECT	JAMDAS
Annual N of outpatients (latest year) [approx., 10K people]				157	123.7	52.7	5	90	1790	1419	407	35	0	80		10	2800
Annual N of inpatients (latest year) [approx., 10K people]				29	120.9	10.6	1	17	371	306	84	100	26	20			0
Annual N of medicated diabetes patients (latest year) [approx., 10K people]				7	6.9	4.6		7	135	105	30	0.15	1	10			
Definition				Patients with diabetes medications in their outpatient accounting records in 2025 (no duplicates).	Number of patients with an ICD-10 diagnosis code of E10–E14 (excluding suspected cases) who were prescribed medications classified under ATC code A10.	The number of patients with registered diabetes (E10-E14) who will be prescribed A10 diabetes medication by 2025.		Number of patients diagnosed with ICD-10 codes E10–E14 (excluding suspected diagnoses) and prescribed ATC:A10 drugs	• Target disease: ICD10: E10-E14 (exclude suspected cases) • Target medication: ATC (EphMRA): A10 • Condition for target patients: Target medication was prescribed in the same month as the diagnosis of the target disease.	Number of patients with ICD-10:E10-E14 disease (excluding suspected) and prescribed ATC:A10 drugs	Number of patients with ICD-10:E10-E14 disease (excluding suspected) and prescribed ATC:A10 drugs	• Calculated from ICD-10 codes: E10–E14 • Based on the number of patients at the 10 facilities where data were obtained, the number of patients at the 52 facilities with data use agreements was estimated.	ICD-10 codes E10–E14 diagnoses (excluding suspected diagnoses)	Number of patients with a diagnosis of diabetes mellitus (ICD-10 codes E10-E14, excluding suspected diagnoses) and at least one prescription or administration of antidiabetic medication (ATC: A10).	As this is an oncology database, detailed information on diabetes treatment are not provided.	As this database is dedicated to oncology, detailed information on pharmacotherapy for non-oncology diseases is not available.	
Annual N of influenza diagnosed patients (latest year) [approx., 10K people]				7	1	0.49		0.1	32	31	10	0.65	0.05	0.8			
Definition				In 2025, the number of unique patients newly assigned ICD-10 codes J09–J11 in their insurance claims (excluding suspected diagnoses). Reference: 80,000 in 2019; 200 in 2021.	Number of patients with an ICD-10 diagnosis code of J09–J11 (excluding suspected cases).	The data was compiled for cases registered in 2025 with ICD-10 codes J09 to J11.		Number of patients assigned ICD-10 codes J09 to J11 (excluding suspected diagnoses)	ICD10 : J10, J11	Number of patients assigned ICD-10:J09-J11 disease	Number of patients assigned ICD-10:J09-J11 disease	• Calculated from ICD-10 codes: J09–J11 • Based on the number of patients at the 10 facilities where data were obtained, the number of patients at the 40 facilities with data use agreements was estimated.	ICD-10 codes J09–J11 diagnoses (excluding suspected diagnoses)	Number of patients with a diagnosis of influenza (ICD-10 codes J09-J11) and available healthcare encounter data, excluding suspected diagnoses.	As this is an oncology database, detailed information on influenza diagnoses are not provided.	As this database is dedicated to oncology, detailed information on pharmacotherapy for non-oncology diseases is not available.	
Number of patients followed																	
≥1 year [approx., 10K people]				62	220.5	93.9			2650	2596	1140			300	1.1	10	
≥3 year [approx., 10K people]					154.8	57.9			1620	1530	791			200	0.7	8	
≥5 year [approx., 10K people]					105.3	27.9			1112	942	557			100	0.5	4	
Median [approx., year]					1.8	0.95			2.2	0.41	0.6			1.9	3.5		
Notes				Patients who had basic medical fees (Ministry of Health code A) recorded in their accounting records in both 2025 and 2023.					Average	The above median represents all patients. When patients with only one visit are excluded, the median is 0.67.	The above median represents all patients. When patients with only one visit are excluded, the median is 1.6.			The median excludes patients with only a single recorded visit. For reference, the median observation period was 0.3 years when these patients were included.	The median observation period varies by tumor type. Also, for reference, in addition to the median follow-up period shared above, the average observation period is 4.4 years. This follow-up period will continue to be extended to enable longer follow-up as the data is regularly refreshed going forward.	Longitudinal patient-level data are available from the first recorded visit through the latest available follow-up. Observation periods vary by patient and data element.	
Age Distribution																	
0-14 years old [approx., %]				10	13.4	9		9.3	13.5	9	8	10	8.5	6.9	0	0.2	14.5
15-64 years old [approx., %]				47	44.9	39		54.7	48.9	39	40	42	34.2	47.4	52.2	40.3	64.4
65-74 years old [approx., %]				15	15.3	19		13.2	16.1	17	17	20	23.2	17.5	30.5	33.7	9
75 years old or older [approx., %]				28	26.3	33		22.8	21.5	35	35	28	34	28.2	17.2	25.8	12.1
Notes				Using data from January to December 2025.	Patient age was calculated based on the difference between the latest month of observation for each patient and the patient's year and month of birth.	Data collected from January to December 2025						Distribution of 10 facilities that have already acquired data out of 52 contracted facilities with available data		Patient age distribution was calculated based on age at the time of clinical encounter.	As this database only includes patients aged 18 and above, there are no patients aged <18 in the 15-64 years category.		

	Hospital-based																
	NHO NCDA/MIA	DPC database	MID-NET	Tokushukai	LDI	DATuM IDEA®	FAST-HDJ_DB	4DIN	MDV	JMDC	JMDC(EMR)	TXP Medical Database	OneICU	MEI (KDDI Grp.)	Flatiron Health	J-CONNECT	JAMDAS
Prescriptions (data availability)																	
Dispensing (in-hospital)	Yes	Yes		Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Dispensing (out-of-hospital)	Yes	No		Yes	Limited	Yes	No	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	No	Yes
Date of prescription	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Date of dispensing	Yes	No	No		Limited		No	Limited	Limited	Limited	Limited	Limited	Yes	Yes	Yes	Yes	Yes
Dosage of prescription	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Days of supply	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Administration	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	No	Limited	Limited	Yes	Yes	Yes	Yes	Yes	Yes
Injection	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Vaccine	Limited	No	Limited	Yes	Yes	Limited	Yes	Limited	No	Limited	Limited	Yes	Limited	Yes	No	No	Yes
Drug master																	
Any available master	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes
ATC (EphMRA)	No	No	No	No	No	Yes	No	Yes	Yes	Yes	Yes	Yes		Yes	No	No	Yes
ATC (WHO)	No	No	Yes	No	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes		Yes	No	Yes	Yes
YJ code	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes		Yes	No	Yes	Yes
NHI drug price list code	No	Yes		Yes	Yes		No	No	Yes	Yes	Yes	Yes		Yes	No	Yes	No
HOT code	Yes	No	Yes	Yes	No	Yes	Yes	Yes	No	No	No	Yes		Yes	No	Yes	No
Reimbursement code		Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes		Yes	No	Yes	Yes
Hospital-original code	Yes	No	Yes	Yes	No		No	Yes	No	No	Yes			Yes	No	No	No
Others	No	No			Yes		No	No		No	No			No	Yes		
Others, notes					Ingredient quantity master				Titer						Drug codes are harmonized using RxNorm across the Japan, US, UK and German Flatiron Health Databases to enable multinational evidence generation.		
Procedures (data availability)																	
Procedure name	Yes	Yes	Yes		Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Date of procedure	Yes	Yes	Yes		Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Procedure master																	
Any available master	Yes	Yes	Yes		Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes
Reimbursement category	Yes	Yes	Yes		Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	No
Reimbursement code	Yes	Yes	Yes		Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	No	Yes	Yes
Others	No	No			Limited		No	No	No	No	No	Yes	No	Limited	No		
Others, notes					In addition to information obtainable from claims data, detailed procedure information documented in electronic medical record text is also available.							Minute-level time-series records of clinical procedures in the ER/ICU (available at select facilities only)		In addition to administrative billing data, the database captures clinical information from free-text electronic medical record (EMR) data, including patient characteristics, clinical findings, laboratory values, and various clinical assessment scores associated with procedures.	In line with Flatiron Health's data model customized for each tumor type, details specific to surgery in the oncology setting are provided (e.g. surgical approach, date of surgery, reasons for not conducting surgery)		
Hospitalization (data availability)																	
Date of hospital admission	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No
Date of hospital discharge	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No
Medication while hospitalization	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No
Diagnosis (data availability)																	
Diagnosis name	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Suspicious diagnosis or not	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Diagnosis master																	
Any available master	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Standard diagnosis code (Reimbursement code)	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	No	No	Yes
ICD10 code (2013)	Limited	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes		Yes
ICD10 code (other than 2013)	Yes	No	Yes		No	Limited	No	No	Yes	Yes	Yes		Yes	Yes	Yes	Yes	No
Others		No			Limited		No	No	No	No	No	Yes	No	Limited	No	Yes	No
Others, notes					Detailed disease subtypes and classifications that cannot be defined by diagnosis codes in claims data can be extracted and defined from electronic medical record text.							Clinical diagnoses, medical history, and underlying conditions determined by ER/ICU physicians are standardized using ICD-10 codes, with terminology variations fully normalized.		Clinical information, including pathology findings, biomarkers, disease severity, and comorbidities, can be derived from free-text electronic medical record (EMR) data. These data can be used to define patient characteristics relevant to disease subtypes and treatment selection.	In addition to ICD-10 codes, the Flatiron Health Research Database contains granular information regarding cancer diagnosis, such as date of initial diagnosis, clinical / pathological group stage and TNM classification, histology, locoregional/distant recurrence, date of metastatic diagnosis.	ICD-O-3-M, ICD-O-3-T	

	Hospital-based																
	NHO NCDA/MIA	DPC database	MID-NET	Tokushukai	LDI	DATuM IDEA®	FAST-HDJ_DB	4DIN	MDV	JMDC	JMDC(EMR)	TXP Medical Database	OneICU	MEI (KDDI Grp.)	Flatiron Health	J-CONNECT	JAMDAS
Laboratory test (data availability)																	
Lab test (ordering)	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Limited	Yes	Yes	Yes
Lab test results																	
Any available test results	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Lab test for sample (e.g., blood, urine, microbe)	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Physiological test (e.g., respiratory function, brain waves, supersonic waves)	No	No	No		Limited	Limited	Limited	Limited	No	Limited	Limited	Limited	No	Limited	Yes	No	Limited
Results of diagnostic imaging (e.g., report of imaging interpretation)	No	No	No		Limited	Limited	Limited	Limited	No	Limited	Limited	Limited	No	No	Yes	No	No
Genetic test (pathological tissue)	No	No	No		Limited	Limited	Limited	Limited	No	Limited	Limited	Limited	No	Limited	Yes	Yes	No
Others	No	No			Limited	Limited	No	No	No	No	No			Limited	Yes		
Others, notes					The following test results can be obtained from unstructured text in electronic medical records: • Genetic mutations, chromosomal abnormalities • Qualitative results from antibody, antigen, and PCR tests • Biopsy information etc. Anonymized medical images are also available.	In some cases, we may be able to extract and provide test results contained in medical records, including physician-authored clinical notes.								The database captures all measured laboratory values recorded in EMR laboratory tables, including immunology/serology, clinical chemistry, hematology, and urinalysis/fecal testing results. Furthermore, free-text electronic medical record (EMR) data can be used to derive information on biomarkers, genetic testing results, pathology findings, and disease severity.	Detailed information on laboratory tests (date of test, test name, test results date, test results, units) as well as biomarker tests (biomarker name, date of specimen collection, date of result reporting, sample type, type of test (IHC, NGS etc.), results (positive, negative etc.), detailed results (protein expression, amplification, rearrangement, mutation etc.)		
Lab test master																	
Any available master	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes
Loinc code	No	No	No		No	No	No	No	No	No	No	No		No	Yes	No	No
JLAC code	Yes	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes	Yes	Yes		No	No	No	Limited
Reimbursement code	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes		Yes	No	No	Yes
Hospital-original code	Yes	No	Yes	Yes	Limited	Yes	Yes	Yes	Yes	No	Yes	Yes		Yes	No	Yes	Yes
Others		No			No		No	No	No	No	No			No	No		
Others, notes															Test codes are harmonized using LOINC; this global standard is applied across the US, Germany and UK Flatiron Health Research Databases as well, thereby unlocking critical multinational evidence generation.		
Demographic, Vital (data availability)																	
Birth year	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Birth month	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Limited	No	Yes	Yes
Birth date	Yes	Yes	Yes	Yes	No		Yes	Yes	No	No	No	Yes	Yes	No	No	No	No
Death information	Yes	Limited	Limited	Limited	Limited	Yes	No	Limited	Limited	Limited	Limited	Yes	Yes	Yes	Yes	Yes	Limited
Data source			The date of death is available for inpatients who were discharged due to death.	Mortality information recorded in the electronic medical record. (based on in-hospital deaths and information provided by affiliated hospitals)	Millennial Medical Record database alone captures only in-hospital deaths. Linkage with NDB may enable the capture of deaths occurring outside participating hospitals.	We provide either (1) death information identified by hospital staff and recorded in the electronic medical record or (2) death information recorded as the patient outcome on a medical insurance claim.		Electronic Health Records	Discharge Summary Data (Outcome)	in-hospital death	in-hospital death	Includes both in-hospital mortality and out-of-hospital death data reported back to the institution.	hospital mortality	In-hospital death	Mortality data is extracted from both structured data (record of mortality status, death flag) and unstructured data (letters from post-referral sites, death certificates) documented within the EHR. By including unstructured data as a key data source for mortality information, mortality events that occur both within and outside of Flatiron's partner hospitals can be captured, resulting in an approximately 30% increase in the comprehensive capture of mortality data.	Derived from Hospital-Based Cancer Registries	
Age	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Height	Yes	Yes	Yes	Limited	Yes	Yes	Yes	Yes	Limited	Limited	Limited	Yes	Yes	Yes	Yes	No	Yes
Weight	Yes	Yes	Yes	Limited	Yes	Yes	Yes	Yes	Limited	Limited	Limited	Yes	Yes	Yes	Yes	No	Yes
Blood pressure	Yes	No	No	Limited	Yes	Limited	Limited	Limited	Limited	No	Limited	Yes	Yes	Limited	Yes	No	Yes
Insurers for patients																	
National health insurance	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes		No	Yes	Limited	No	Yes
Japan health insurance association	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes		No	Yes	Limited	No	Yes
Health insurance society	Yes	Yes	Yes	Yes	Yes	Limited	Yes	Yes	Yes	Yes	Yes		No	Yes	Limited	No	Yes
Mutual society of health insurance	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes		No	Yes	Limited	No	Yes
Late-stage medical care system for the elderly	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes		No	Yes	Limited	No	Yes

	Hospital-based																
	NHO NCDA/MIA	DPC database	MID-NET	Tokushukai	LDI	DATuM IDEA®	FAST-HDJ_DB	4DIN	MDV	JMDC	JMDC(EMR)	TXP Medical Database	OnelCU	MEI (KDDI Grp.)	Flatiron Health	J-CONNECT	JAMDAS
Data Access, Others																	
Access to raw data via web	No	No	No		No		No	Yes	Limited	Yes	Yes	No	Yes	Yes	Yes	Limited	No
Access to original source data such as medical charts (e.g., outcome validation, additional data collection)	Yes	No	No	Yes	Limited	Limited	Limited	Limited	Limited	Limited	Limited	Limited	Yes	Yes	Yes	Yes	Yes
Timing of data update	Daily	Annually		Daily	Monthly		Daily	Monthly	Monthly	Monthly		As needed	every 3 months	Monthly	Quarterly		Daily
Others, notes			every week at the earliest			Update every quarter					3 months			As participating healthcare institutions are connected online, data from approximately one month prior can be provided in a timely manner.	The database is refreshed on a quarterly basis; as such, data recency is maintained at 90 days (e.g. datasets delivered in June 2025 have a data cutoff date of 31 March 2025).	Medication and laboratory data connected to the EHR are available in real time, while all other data are updated annually.	
Latest data	the day before		The most recent date	today	Approximately 3 months prior		two month ago	2 months ago	2 months ago	2 month ago	1 month ago	NEXT Stage ER/ICU-derived data is accumulated in near real time, with EMR data following a similar timeline. DPC data, on the other hand, is updated with a 3-month lag.	one month ago	Updated through approximately one month prior	90 days recency	Data through the previous day	latest data available up to 2 days ago
Publication																	
Number of past scientific presentations using database	Multiple	5	128	Many	16	6	4	25	Many	Many	Multiple	43	>10	4	10	Multiple	Few
Number of publication to peer review journal	Multiple	90	Therapeutic Innovation & Regulatory Science, PLOS ONE, Frontiers in Medicine.	Many	9	4	2	53	851 (total number of publications, including those not requiring peer review.)	Many	Many	41 (as TXP Research / Collaboration)	5	None	2	Few	Multiple