SEMINAR

Updated information on the Diagnosis Procedure Combination data

Hideo Yasunaga

Department of Clinical Epidemiology and Health Economics, School of Public Health, The University of Tokyo, Tokyo, Japan

ABSTRACT

The Ministry of Health, Labor, and Welfare, Japan, launched the Diagnosis Procedure Combination system in 2002. Detailed information on the Diagnosis Procedure Combination data was reported in *Annals of Clinical Epidemiology* in 2019. In this report, I provide updated information on the Diagnosis Procedure Combination. The data included the discharge abstracts and administrative claims data for each inpatient. Several entities (including the Ministry, academic groups, and private companies) independently collected anonymized Diagnosis Procedure Combination data. The advantages of Diagnosis Procedure Combination data include detailed process and clinical data, which enable researchers to conduct clinical epidemiology and health services research. Diagnoses are recorded using the International Classification of Diseases-10th Revision codes, and several indices based on these codes can be used. Several clinical measures are available for specific diseases including stroke, respiratory failure, heart failure, pneumonia, liver cirrhosis, pancreatitis, burns, and multiple organ failure. Scores for consciousness, activities of daily living, functional independence, and dementia are also available. Studies that use Diagnosis Procedure Combination data are interdisciplinary and include clinical medicine, epidemiology, statistics, and medical informatics.

INTRODUCTION

I previously reported information on the Japanese Diagnosis Procedure Combination (DPC) data in 2019 in *Annals of Clinical Epidemiology*[1]. Five years later, the DPC data has changed. In this report, I have detailed the updated information on the DPC data.

Real-world data are defined as electronic data on patient health status and healthcare delivery that are routinely collected from various sources. Real-world data include administrative claims, electronic medical records, and patient registry data. Real-world data can be utilized in clinical, epidemiological, health services, health economics, and policy research.

DPC data are currently the most widely utilized for research purposes and have the greatest impact on healthcare among the various real-world data in Japan[2]. This report briefly describes the DPC system in Japan, explains updated information on DPC data, and provides an overview of the use of DPC data for research purposes.

THE DPC SYSTEM

In 2002, the Ministry of Health, Labor and Welfare (MHLW) in Japan originally developed a case-mix patient classification system called the DPC system. The DPC system is linked to a per-diem lump-sum payment system for inpatients. Specifically, the DPC refers to approximately 4,500 diagnostic groups classified according to age, severity, surgery/procedures, and comorbidities, based on approximately 500 underlying diseases in 18 Major Diagnostic Categories. Each diagnostic group classification has three levels of hospitalization, and a per-diem payment is set for each period.

All 82 university hospitals adopted the DPC system; however, adoption by community hospitals was voluntary. As of April 2024, 1,786 acute care hospitals have adopted the DPC system among approximately 8,000 hospitals in Japan, which are called DPC hospitals.

All DPC hospitals are obliged to create and submit DPC data to the MHLW. The DPC data included the discharge abstracts and administrative claims data for each inpatient. The MHLW uses DPC data to track national trends in healthcare utilization in acute care hospitals for health policy planning.

Aggregated summaries of the data are disclosed on the MHLW website, which show the number of patients and the average length of stay for each of the approximately 4500 diagnostic groups in each hospital. This information can be utilized by patients to select hospitals based on their clinical information and improve their clinical practice.

DPC Databases

In fact, there are various "DPC databases". Each DPC hospital has its own data. Several entities (including MHLW, academic groups, and private companies) independently collected anonymized DPC data from DPC hospitals to create a secondary database and provide datasets for researchers. Data recipients can use the DPC data to identify, track, and analyze healthcare utilization, access, quality, outcomes, and costs in acute care hospitals. Furthermore, DPC data can be utilized for clinical epidemiology and health services research because they contain clinical data as well as detailed process data.
(i) The MHLW DPC database.
The MHLW collects anonymized DPC data from DPC hospitals nationwide and establishes and operates the DPC Database (DPCDB). In 2019, the MHLW began providing aggregated data from the database upon request from researchers. In April 2022, the MHLW began providing individual-level data to researchers and

private companies. Applicants must apply for data use in accordance with "the guideline for providing the DPC data"[3].

In addition, consolidated data from the DPCDB and National Database of Health Insurance Claims and Specific Health Checkups of Japan (NDB) are provided[4].
(ii) DPC databases created by academic groups.
Several academic groups independently collect anonymized DPC data from the DPC hospitals.

The Japanese Registry of All Cardiac and Vascular Diseases (JROAD) was launched in 2004 by the Japanese Circulation Society to assess clinical activity for treating cardiovascular diseases at each hospital. Since 2014, DPC data on patients with cardiovascular disease have been collected voluntarily from more than 900 hospitals[5].

The DPC Study Group, a government-funded academic group, voluntarily collects anonymized DPC data from more than 1,000 hospitals, independent of the MHLW, for research purposes[1]. The number of patients in the database is approximately 7 million per year. The coverage rate of inpatients in the DPC database for all acute care inpatients in Japan has exceeded 50%. In particular, the database covers approximately 90% of all tertiary care emergency hospitals, 44% of the institutions certified by the Japanese Surgical Society, and 80% of the institutions certified by the Japanese Association for Infectious Diseases to train board specialists.
(iii) Commercially available databases
Several private companies, including Medical Data Vision Co., Ltd. (https://en.mdv.co.jp/) and JMDC Inc. (https://www.jmdc.co.jp/en/), collected anonymized DPC data from DPC hospitals and built relatively small databases. They provide DPC datasets to researchers and private companies for a fee.

Recorded Diagnoses in the DPC Data

In DPC data, diagnoses are recorded using text data in Japanese and International Classification of Diseases-10th Revision (ICD-10) codes[6]. The comorbidities at admission and complications after admission are recorded separately. This is an advantage of the DPC data that is not found in other administrative claims databases in Japan and other countries.

Using the ICD-10 codes of comorbidities, researchers can calculate the Charlson Comorbidity Index[7]. The Charlson Comorbidity Index helps to predict the risk of death by weighting or classification.

Several ICD-10-based scores can also be calculated, including the ICD-10-based trauma mortality prediction scoring system[8] and Hospital Frailty Risk Score[9].

A previous study showed that the validity of the diagnoses and procedure records in the DPC database was generally high. The sensitivity and specificity of the primary diagnosis were 78.9% and 93.2%, respectively[10].

The number of validation studies using Japanese administrative claims data has been increasing. According to a review article on validation studies, of the 36 validation studies published through March 2022, 29 used other data sources such as electronic medical records or patient registries (cancer registries, stroke registries) as the gold standard. Several validation studies have been conducted on the disease names of cancers, heart diseases, cerebrovascular diseases, and diabetes. In particular, cancer disease names in the DPC data have high sensitivity and specificity. Other validation studies have been conducted on individual disease names, such as postoperative infection, gastrointestinal perforation, Takotsubo syndrome, age-related macular degeneration, congenital malformations, febrile neutropenia, hemophilia, rheumatoid arthritis, Crohn's disease, ulcerative colitis, and sepsis[11].

Data Items in the DPC Data

Table 1 lists the items in the DPC data. The DPC data include discharge abstract data (Format 1) and administrative claims data (EF files).

(i) Format 1

Format 1 of DPC data includes various clinical measures that can be utilized in clinical studies. Several severity indices are available for specific diseases, including the modified Rankin scale for stroke, the Hugh–Jones classification for respiratory diseases, the New York Heart Association classification for heart failure, the Canadian Cardiovascular Society classification for angina pectoris, the Killip classification for myocardial infarction, the A-DROP scoring system for pneumonia, Child–Pugh classification for liver cirrhosis, the Japanese severity classification for acute pancreatitis, the burn index, and Sequential Organ Failure Assessment score. Items regarding rehabilitation included the Barthel Index for Activities of Daily Living and the Functional Independence Measure.

The Japan Coma Scale (JCS) is a commonly used method for evaluating the level of consciousness in Japan and can be converted into the Glasgow Coma Scale[12]. JCS 0 represents alert consciousness, JCS 1–3 represents wakefulness without stimuli but not fully alert, JCS 10–30 represents arousal with stimuli, and JCS 100–300 represents coma.

Items regarding frailty included the level of independence in daily living of elderly people with dementia[13];

care-need level (support levels 1–2 and care-need levels 1–5)[14]; malnutrition, eating and swallowing dysfunction, tube and intravenous feeding, and falls. A previous study showed that care-need level was highly associated with the Barthel index[15].

(ii) EF files

In the EF files, the dates of procedures and the dates of using drugs and devices are all recorded; thus, the interval between the start and end of any process can be calculated (e.g., duration of mechanical ventilation, duration of chest tube drainage, etc.).

The database also includes the estimated total costs based on reference prices in the Japanese national fee schedule, which determines item-by-item prices for surgical, pharmaceutical, laboratory, and other inpatient services.

The following examples show how to use DPC data on prescriptions, procedures, and surgeries for research.

1) The sensitivity of the disease name "hypertension" is low. However, the prescription data for antihypertensive drugs have high sensitivity and specificity. Therefore, the existence of "hypertension" is identified by prescription data, not by the name of the disease.

2) Prescription data can indicate which antibiotic was administered after surgery, in what volume, and for how many days.

3) Data on blood loss are unavailable. However, the volume of blood transfusion can be determined from the prescription data.

4) The degree of intra-abdominal contamination caused by gastrointestinal tract perforation remains unknown. However, the amount of saline solution used to clean the intra-abdominal cavity during surgery is known from prescription data.

5) The sensitivity of the disease name 'acute heart failure' is low. Severe heart failure requiring catecholamine administration, intra-aortic balloon pumping, and extracorporeal membranous oxygenation can be identified using procedural data.

6) The sensitivity of "acute respiratory failure" is low. However, data on procedures, such as "endotracheal intubation" and "mechanical ventilation", have high sensitivity and specificity. Mild respiratory failure cannot be identified, but severe respiratory failure requiring intubation and ventilation can be identified using procedural data.

7) The sensitivity of the disease name "renal failure" is low. However, severe renal failure requiring renal replacement therapy can be identified using procedural data.

Table 1	Data items in the DPC data
	Data
Format 1 (Patient basic information)	
Hospital data	unique identifiers of the hospitals, location of the hospitals
Patient demographics	anonymized patient identifier; patients' age and sex; zip codes of patients' residing area
Diagnoses	main diagnosis, admission precipitating diagnosis, most resource consuming diagnosis, second most resource consuming diagnosis, comorbidities present on admission, complications arising after admission
Admission and discharge information	days of admission and discharge, type of admission (urgent or elective), type of psychiatric admission (voluntary or involuntary), ambulance service use, dates of admission and discharge, discharge status (discharged to home, discharged to other facility, inhospital death)
Surgical information	day of surgery, surgical codes, name of operation, anesthesia, preventive antimicrobial injection
Clinical data	(1) body weight and height,
	(2) smoking index (pack years)
	(3) pregnancy; labor, amount of bleeding during labor
	(4) birth weight, gestational age at admission, gestational age at birth
	(5) decubitus
	(6) Japan Coma Scale
	(7) Tumor-Node-Metastasis classification, cancer stage, chemotherapy
	(8) modified Rankin scale for stroke, date of stroke onset
	(9) Hugh–Jones classifications for respiratory diseases
	(10) New York Heart Association classification for heart failure
	(11) Canadian Cardiovascular Society classification for angina pectoris
	(12) Killip classification for myocardial infarction
	(13) A-DROP scoring system for pneumonia
	(14) Child–Pugh classification for liver cirrhosis;
	(15) Japanese severity classification for acute pancreatitis
	(16) burn index
	(17) Barthel index, Functional Independence Measure
	(18) level of independence in daily living of elderly people with dementia; care need level; malnutrition; eating and swallowing dysfunction; tube and intravenous feeding; falls
	(19) Global Assessment of Functioning Scale
	(20) type of hospitalization, number of days of seclusion, and number of days of physical restraint under the Mental Health and Welfare Act
	(21) aortic dissection Stanford A or B
	(22) P/F ratio, FiO_2, oxygenation, respiratory support, systolic blood pressure, use of circulatory agents on emergency visit and admission to ward
	(23) Sequential Organ Failure Assessment score
	(24) gamma-globulin for Kawasaki disease
EF file (Administrative claims data)	
	anesthesia, surgery, rehabilitation and other procedures; duration of anesthesia; volume of blood transfusion; dates of procedures
	pharmaceuticals and devices used; dates of using drugs and devices
	estimated costs

8) There are no data on the operation time. However, data on the duration of anesthesia are available. Anesthesia reimbursement was added every 30 min.

9) Rehabilitation time can be identified by the number of minutes spent on rehabilitation each day because reimbursement for rehabilitation is added every 20 min.

10) The duration of the chest tube drainage is known because of the specific costs incurred for chest tube insertion and daily management. When the tube is removed, management costs are no longer reimbursed.

11) The use of elastic stockings or intermittent pneumatic compression devices for the prevention of pulmonary thromboembolism can be determined by the fee for the prevention of pulmonary thromboembolism.

CONCLUSION

This report introduces the details of the DPC data. The advantages of DPC data include detailed process and clinical data, which enable researchers to create real-world evidence. To utilize DPC data, the following skills are required: (i) ability to generate research questions, (ii) epidemiological competence to construct study designs, (iii) medical informatics skills to handle data, and (iv) statistical ability to analyze retrospective observational data. Studies that use DPC data are interdisciplinary and include clinical medicine, epidemiology, statistics, and medical informatics. It is essential to develop research groups in which researchers from multiple fields collaborate.

REFERENCE

1. Yasunaga H. Real World Data in Japan: Chapter II The Diagnosis Procedure Combination Database. *Ann Clin Epidemiol*. 2019; 1:76–9.

2. Sato S, Yasunaga H. A Review of Studies Using Japanese Nationwide Administrative Claims Databases. *Ann Clin Epidemiol*. 2023; 5:58–64.

3. Ministry of Health, Labour, and Welfare, Japan. The guideline for providing DPC data. Second edition. https://www.mhlw.go.jp/content/001164498.pdf

4. Yasunaga H. Updated Information on NDB. *Ann Clin Epidemiol*. 2024;6:73–6.

5. Yasuda S, Miyamoto Y, Ogawa H. Current Status of Cardiovascular Medicine in the Aging Society of Japan. *Circulation*. 2018;138: 965–7.

6. Bramer GR. International statistical classification of diseases and related health problems. Tenth revision. *World Health Stat Q*. 1988;41:32–6.

7. Quan H, Li B, Couris CM, et al. Updating and validating the Charlson comorbidity index and score for risk adjustment in hospital discharge abstracts using data from 6 countries. *Am J Epidemiol*. 2011;173:676–82.

8. Wada T, Yasunaga H, Yamana H, et al. Development and validation of a new ICD-10-based trauma mortality prediction scoring system using a Japanese national inpatient database. *Inj Prev*. 2017;23:263–7.

9. Gilbert T, Neuburger J, Kraindler J, et al. Development and validation of a Hospital Frailty Risk Score focusing on older people in acute care settings using electronic hospital records: an observational study. *Lancet*. 2018;391;1775–1782.

10. Yamana H, Moriwaki M, Horiguchi H, et al. Validity of diagnoses, procedures, and laboratory data in Japanese administrative data. *J Epidemiol*. 2017;27:476–82.

11. Yamana H, Konishi T, Yasunaga H. Validation studies of Japanese administrative health care data: A scoping review. *Pharmacoepidemiol Drug Saf*. 2023;32:705–17.

12. Nakajima M, Okada Y, Sonoo T, et al. Development and Validation of a Novel Method for Converting the Japan Coma Scale to Glasgow Coma Scale. *J Epidemiol*. 2023;33:531–5.

13. Ministry of Health, Labour and Welfare Level of Independence in Daily Living of Elderly People with Dementia. [(accessed on 26 July 2022)]. https://www.mhlw.go.jp/topics/2013/02/dl/tp0215-11-11d.pdf

14. Konishi T, Inokuchi H, Yasunaga H. Services in public long-term care insurance in Japan. *Ann Clin Epidemiol*. 2024;6:1–4.

15. Matsuda T, Iwagami M, Suzuki T, et al. Correlation between the Barthel Index and care need levels in the Japanese long-term care insurance system. *Geriatr Gerontol Int*. 2019;19:1186–7.