RWD103

Characteristics and Applications of Body Size Data in Japanese Claims Databases

Masashi Mikami, Kanae Togo

Access & Value, Pfizer Japan Inc., Tokyo, Japan

Objectives

To describe the characteristics of body weight (BW) and body mass index (BMI) in two major claims databases in Japan, the insurance-based database from JMDC Inc. (JMDC) and the hospital-based database from Medical Data Vision Co., Ltd. (MDV) and demonstrate their applications to medical research.

Conclusions

We clarified the characteristics of JMDC and MDV and demonstrated cases of data application that matched to characteristic of each database. Since the characteristics of body size data differs among databases, it is important to choose the most suitable database according to study objectives.

Introduction

The two major databases in Japan, JMDC[1] and MDV[2], provide body size data as characterized in Table 1. Understanding the characteristics of these databases is essential for selecting appropriate data sources in research involving body size data and its clinical implications.

Methods

We examined the characteristics of body size data from JMDC and MDV between and applied these data to two practical cases.

Study Design

- Retrospective observational study using JMDC and MDV databases.
- Population:
  - JMDC: Individuals enrolled in health insurances (from which JMDC collected data) during January 2005 and September 2024.
  - MDV: Patients hospitalized during April 2008 and December 2024.

Statistical Analysis

Characteristics of body size data:

- Characteristics of the study population were summarized, including those with at least one recorded body size measurement.
- We explored the missing patterns of body size measurements by individual and hospitalization attributes.
- The number of measurements per person during the study period was summarized.
- For individuals with multiple body size measurements, we summarized the follow-up period of the measure and the intervals between measurements.

Practical cases:

- JMDC: BMI data from annual health checkups were tracked from age 30 in 2015 for 10 years. A Bayesian hierarchical model was applied to estimate BMI trajectories. From the overall population, two cohorts of 1000 individuals (male and female) were randomly sampled. Using the hierarchical Bayesian framework, we plotted posterior predictive percentile curves (5th–95th) for BMI distributions.
- MDV: Weight distributions for inpatients aged 0–20 in 2015 were analyzed. The Lambda-Mu-Sigma (LMS) method was applied to estimate national growth standards in Japan. From the overall population, two cohorts of 20,000 patients (male and female) were randomly sampled. Using the LMS framework, we plotted Z-Score curves ranging from -2.5 to 2.5.

Results

Characteristics of individuals/patients (Table 2)

- JMDC primarily included working-age adults, enabling long-term tracking.
- In contrast, MDV covered all age groups, including pediatric and elderly populations. The follow-up period was shorter than JMDC because data collection was linked to hospital visits.

Missing patterns (Table 2 and Figure 1-a, 1-b)

- In JMDC, BMI data were mostly missing for children, primarily because employee health insurance unions do not typically provide health checkups for them. As a result, age and sex distributions differed between the overall JMDC population and those whose BMI data is available (Table 2).
- In MDV, missing BW data were more frequently observed in cases involving emergency or critical conditions—such as emergency admissions, and in-hospital deaths—compared to planned hospitalizations.

Number of body size measurements (Figure 2)

- JMDC had a higher average number of measurements compared to MDV, with approximately 80% of cases having two or more measurements and about 30% having six or more.
- In contrast, in MDV, cases with only one measurement accounted for over 60% of the total.

Duration of body size measurements and interval between body size measurements (Figure 3-a, 3-b)

- In JMDC, health checkups were generally conducted once per year. Therefore, the median interval between body size measurements were one year, regardless of the number of measurement.
- In MDV, body weight (BW) measurements were linked to hospitalizations. Consequently, the median duration of body size records was shorter than in JMDC, and the median interval between measurements remained within one year, regardless of measurement frequency.

Practical cases (Figure 4-a, 4-b)

- JMDC: BMI data from 20,996 individuals were identified, with an average of 7.7 measurements per person over a 10-year period. The Bayesian model showed an increasing trend in BMI among individuals in their 30s.
- MDV: BW data from 163,091 patients were identified, with an average of 1.8 measurements per person over a 10-year period. LMS-based curves showed trends consistent with national growth standards in Japan[3] for individuals aged 0–20. Some outliers were likely attributable to digit mis-entry. Their number was small and did not affect the model's predictive performance.

Reference

1. JMDC. JMDC Website. https://www.jmdc.co.jp/en/. Accessed 19 September 2025.
2. Medical Data Vision. MDV Website. https://en.mdv.co.jp/. Accessed 19 September 2025.
3. Isojima T et al. Growth standard charts for Japanese children with mean and standard deviation (SD) values based on the year 2000 national survey. Clin Pediatr Endocrinol. 2016;25(2):71–6.

Table 1. Characteristics of body size data for two major databases in Japan

Database	JMDC (Insurance-based)	MDV (Hospital-based)
Characteristics	- BMI data from annual health checkups, primarily covering working-age adults	- Admission weight and height measurements across all age groups
Strengthens	- BMI data available for both diseased and non-diseased cases - Longitudinal BMI data over extended periods	- Height and weight measurements available - Includes data for all ages - Measurement can be repeated and the interval may be short depending on hospitalization timing
Limitations	- No height or weight measurements (BMI only) - No data for ages 0–17 and 75+; limited data for 65+ - Long measurement intervals (~1 year due to annual health checkups)	- Measurements limited to hospitalized patients - More than half of patients have only one inpatient weight measurement, limiting longitudinal assessment

Table 2. Characteristics of the study population

Characteristic	JMDC Enrolled individuals (n =22,717,646)	MDV Inpatients (n = 17,150,440)
Study population		
Sex, n (%)		
Male	11,435,459 (50.3)	8,650,206 (50.4)
Female	11,282,187 (49.7)	8,500,234 (49.6)
Age at enrollment/earliest hospitalized, years†		
Median (Min, Max)	31 (0, 75)	67 (0, 100)
n (%)		
0–17	5,160,968 (22.7)	1,798,886 (10.5)
18–74	17,555,177 (77.3)	9,350,220 (54.5)
75+	1501 (0.0)	5,844,888 (34.1)
Unknown	0 (0.0)	156,446 (0.9)
Follow-up period, years‡		
Median (Q1, Q3)	3.4 (1.4, 6.4)	1.9 (0.3, 5.7)
Body size data observation status, n (%)		
Observed: BMI in JMDC; BW in MDV	11,213,241 (49.4)	16,274,624 (94.9)
Missing: Missing in JMDC; Missing or recorded as 0 in MDV	11,504,405 (50.6)	875,816 (5.1)
Among people with body size measurements	**(n = 11,213,241)**	**(n = 16,274,624)**
Sex, n (%)		
Male	6,271,974 (55.9)	8,206,638 (50.4)
Female	4,941,267 (44.1)	8,067,986 (49.6)
Age at earliest measurement, years†		
Median (Min, Max)	42 (18, 75)	67 (0, 100)
n (%)		
0–17	0 (0.0)	1,749,191 (10.7)
18–74	11,212,618 (100.0)	8,964,119 (55.1)
75+	623 (0.0)	5,430,649 (33.4)
Unknown	0 (0.0)	130,665 (0.8)

n, number of individuals/patients, Min, Minimum; Max, Maximum; Q1, 1st quartile; Q3 3rd quartile
† JMDC insurance-based database includes individuals covered by employer-based health insurance under 75 years of age. However, a small number of individuals aged 75 years were included due to administrative timing of insurance withdrawal.
‡ JMDC: Enrollment period; MDV: Follow-up period including both inpatient and outpatient care.

Figure 1-a. Missing data patterns in BMI records in JMDC

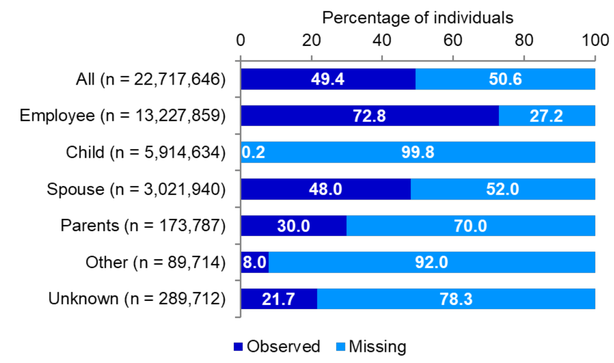
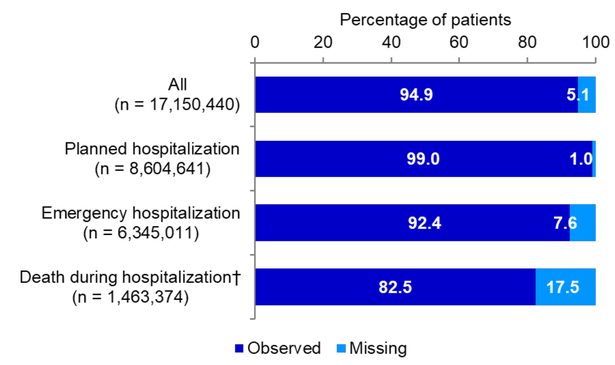
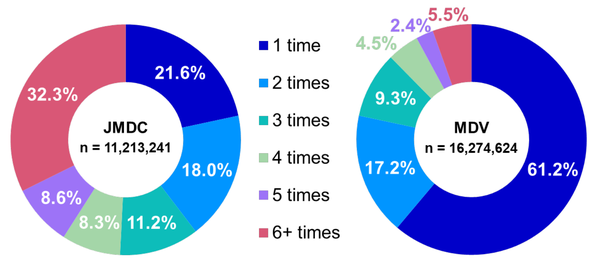


Figure 1-b. Missing data patterns in BW records in MDV



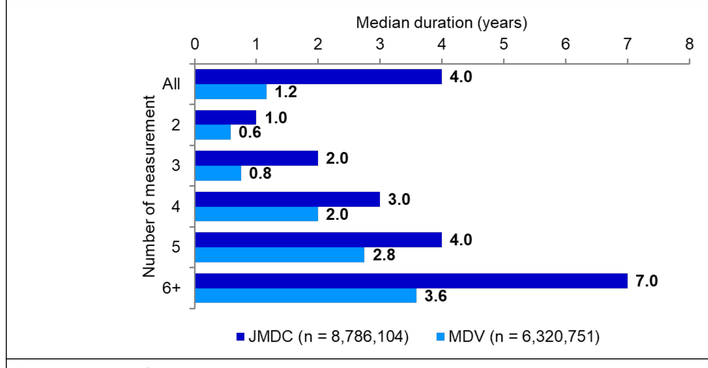
† Death during hospitalization was considered a surrogate for severity at admission.
Note: The analysis includes only patients with claims-based records of admissions and in-hospital death.

Figure 2. Number of body size measurement per person



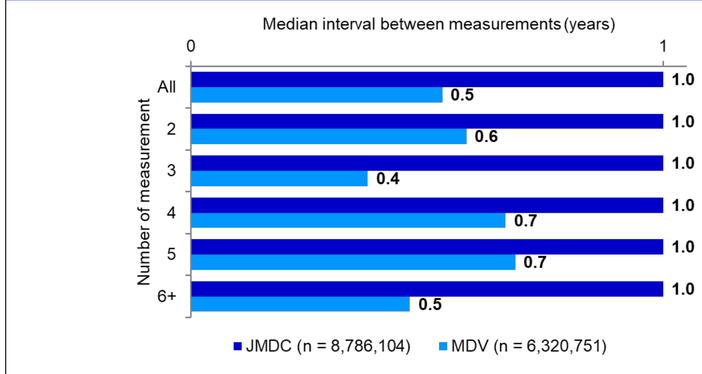
Data are shown for patients who had at least one recorded measurement of body size.

Figure 3-a. Median duration of body size measurements stratified by number of measurements per person



Data are shown for patients with ≥2 body size measurements.
The duration of body size measurement refers to the time span between the earliest and latest recorded measurement dates.

Figure 3-b. Median interval between body size measurements stratified by number of measurements per person



Data are shown for patients with ≥2 body size measurements.
Measurement interval: (Duration of body size measurement)/(Number of measurement - 1)

Figure 4-a. Practical cases using JMDC database: Estimation of BMI trajectories using a Bayesian hierarchical model

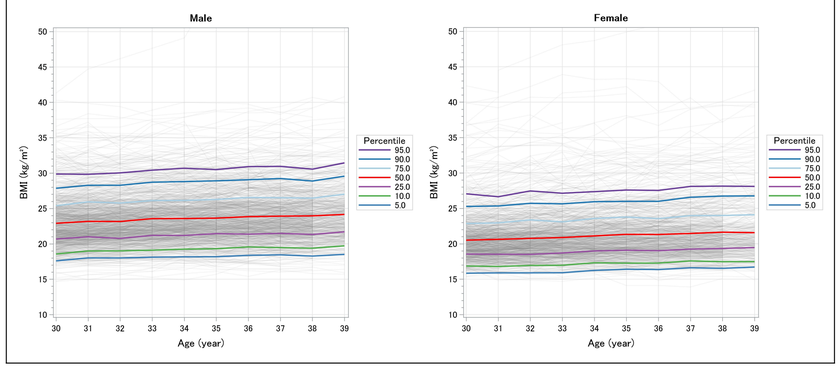


Figure 4-b. Practical cases using MDV database: Estimation of body weight distribution using the LMS method