ORIGINAL RESEARCH

Characterizing Fit-for-Purpose Real-World Data: An Assessment of a Mother–Infant Linkage in the Japan Medical Data Center Claims Database

Julie Barberio [1,2], Rohini K Hernandez[2], Ashley I Naimi[1], Rachel E Patzer[1,3], Christopher Kim[2], Timothy L Lash [1]

[1]Department of Epidemiology, Emory University, Atlanta, GA, USA; [2]Center for Observational Research, Amgen, Inc, Thousand Oaks, CA, USA; [3]Regenstrief Institute, Indianapolis, IN, USA

Correspondence: Julie Barberio, Department of Epidemiology, Emory University, 1518 Clifton Road, Atlanta, GA, 30322, USA, Tel +1 404 727 3956, Email julie.barberio@emory.edu

Purpose: Observational postapproval safety studies are needed to inform medication safety during pregnancy. Real-world databases can be valuable for supporting such research, but fitness for regulatory purpose must first be vetted. Here, we demonstrate a fit-for-purpose assessment of the Japan Medical Data Center (JMDC) claims database for pregnancy safety regulatory decision-making.

Patients and Methods: The Duke-Margolis framework considers a database's fitness for regulatory purpose based on relevancy (capacity to answer the research question based on variable availability and a sufficiently sized, representative population) and quality (ability to validly answer the research question based on data completeness and accuracy). To assess these considerations, we examined descriptive characteristics of infants and pregnancies among females ages 12–55 years in the JMDC between January 2005 and March 2022.

Results: For relevancy, we determined that critical data fields (maternal medications, infant major congenital malformations, covariates) are available. Family identification codes permitted linkage of 385,295 total mother–infant pairs, 57% of which were continuously enrolled during pregnancy. The prevalence of specific congenital malformation subcategories and maternal medical conditions were representative of the general population, but preterm births were below expectations (3.6% versus 5.6%) in this population. For quality, our methods are expected to accurately identify the complete set of mothers and infants with a shared health insurance plan. However, validity of gestational age information was limited given the high proportion (60%) of missing live birth delivery codes coupled with suppression of infant birth dates and inaccessibility of disease codes with gestational week information.

Conclusion: The JMDC may be well suited for descriptive studies of pregnant people in Japan (eg, comorbidities, medication usage). More work is needed to identify a method to assign pregnancy onset and delivery dates so that in utero medication exposure windows can be defined more precisely as needed for many regulatory postapproval pregnancy safety studies.

Keywords: routine health care data, international databases, database evaluation

Introduction

Exclusion of pregnant people from clinical trials precludes the premarket availability of information regarding medication safety during pregnancy, thus postapproval safety studies are generally required by regulatory agencies when new drug products are expected to be used among persons of childbearing age.[1,2] Such requirements have conventionally been completed via the establishment of pregnancy exposure registries, but there has been increasing interest in real-world database approaches to complement registry-based approaches.[3]

Administrative claims databases allow researchers to construct large, diverse pregnancy cohorts according to the pharmacoepidemiologic target trial framework, enabling causal inference regarding medication safety during pregnancy.[4–10] This resource has several advantages over pregnancy exposure registries, which often suffer from difficulties with representative enrollment and retention.[11,12] Methodological challenges arise, however, because claims

databases do not indicate correspondence between mothers and infants (needed to bridge maternal exposure and infant outcome information) or pregnancy timing (needed to anchor exposure measurement around stages of fetal development).[13,14] This information must therefore be reconstructed by researchers who use the databases.[15–25]

It is important to understand when real-world databases may be "meaningfully, validly, and transparently" used to support in the answering of questions of regulatory interest, henceforth referred to as "fitness for purpose".[26] For this reason, the Duke-Margolis Center for Health Policy, in collaboration with the United States (US) Food and Drug Administration (FDA), has developed a framework for evaluating fitness for purpose based on (1) the regulatory question of interest, (2) clinical context, (3) data considerations, including availability of relevant, high-quality data, and (4) application of sufficient methodological approaches.[26,27]

Recent publications have used a mother–infant linked cohort in the Japan Medical Data Center (JMDC) claims database, but the suitability of this resource for informing regulatory decision-making has yet to be evaluated.[23,24] Despite the establishment of mother–infant linkages in US claims databases, there is value in adding this resource in Japan due to differences in standards of obstetric care, prescription medication recommendations, and healthcare systems.[28–30] Therefore, the objective of this analysis was to apply the Duke-Margolis framework, focusing particularly on the data considerations, to evaluate the fitness for purpose of a mother–infant linked cohort in the JMDC claims database, within the specific regulatory context of estimating infant major congenital malformations associated with in utero exposure to marketed medications. It is important to note that mother–infant claims linkages are inherently limited to pregnancies ending in live births, which may not be an appropriate study population for all future pregnancy safety studies in this database.

Materials and Methods

Study Population

The JMDC is compiled from over 1400 private companies belonging to the Health Insurance Association, one of five payer organizations of Japan's National Health Insurance System, and includes inpatient, outpatient, and pharmacy claims. The national insurance covers most medical services, except for over-the-counter drugs, vaccinations, and cosmetic surgeries. Childbirth is not necessarily covered unless certain procedures are required. Additionally, pregnancy confirmation tests and prenatal health visits are covered by a separate government-subsidized program. A unique family identification code is assigned to the insured individual and all their dependents (any financially supported family member), enabling family linkages.

Data Considerations

The Duke-Margolis framework considers a real-world database to be fit for regulatory decision-making, within a given context, depending on (1) relevancy and (2) quality. Relevancy relates to its capacity to answer the regulatory question, in terms of the availability of critical data fields and a sufficiently sized, representative population. Quality relates to its ability to accurately, reliably, and transparently answer the regulatory question. Here, we considered the fitness of the JMDC claims database to be used to assist in answering regulatory questions related to medication safety during pregnancy in terms of its (1) relevancy and (2) quality of information available for (2A) formation of the mother–infant matches and (2B) estimation of the gestational period.

Statistical Analyses

Two linkage methods were performed to create pairs of mothers (females ages 12–55 years) and infants (dependents born during the study period) between January 2005 and March 2022. Linkage Method A identified pregnancy episodes, according to active pregnancy and delivery codes, and then determined the proportions linked to infants (Appendix 1). Active pregnancy codes were diagnosis codes that indicated a pregnant state; delivery codes indicated occurrence of a live birth (Supplemental Table 1).[31] Linkage Method B identified mother–infant pairs, according to matching family identification codes, and then determined the prevalence of pregnancy and delivery codes (Appendix 2).

Next, using the pairs from Linkage Method B, we evaluated the relationships that the matched individuals had to the insurance holder to inform validity of the assumed mother–infant relationships. We assumed incestuous relationships,

although biologically possible, to be implausible. In an attempt to improve linkage validity, two exclusions were made: (1) invalid/inconclusive pairs, including "possible" pairs without pregnancy or delivery codes, and (2) infants matched to multiple mothers.

Using the valid pairs, we then created annual birth estimates based on (1) Linkage Method A, (2) Linkage Method B, and (3) Linkage Method B restricted to those with pregnancy/delivery codes.

Finally, using the valid pairs from Linkage Method B, we examined descriptive characteristics of (1) all pairs and (2) restricted to those with pregnancy/delivery codes. We proceeded with Linkage Method B for this analysis given that this method allows for direct identification of pairs, without reliance on pregnancy-related codes. Characteristic assessment time periods are provided in Supplemental Figure 1 and variable definitions in Supplemental Tables 2 and 3. Characteristics of unlinked females (delivery code but no linked infant) and infants were also examined. To assess representativeness of the JMDC population relative to the general Japan population, we leveraged publicly available data from the Global Burden of Disease Study (medical conditions among females ages 10–54 years, major congenital malformations among infants <1 year) and Japanese Vital Statistics (maternal age, infant sex, and gestational age).[32,33]

Ethics

Emory University's Institutional Review Board does not require review of studies that do not meet the definitions of human subjects research or clinical investigation, such as this one. Informed consent is not required for this type of study. Permission was obtained from JMDC Inc. to use the database for the purposes of this study. All data acquired were kept anonymized.

Results

There were 5,795,818 females ages 12–55 years and 717,034 infants identified during the study period. Based on Linkage Method A (Figure 1), 643,483 pregnancy episodes were identified via pregnancy codes, 40% of which linked with an
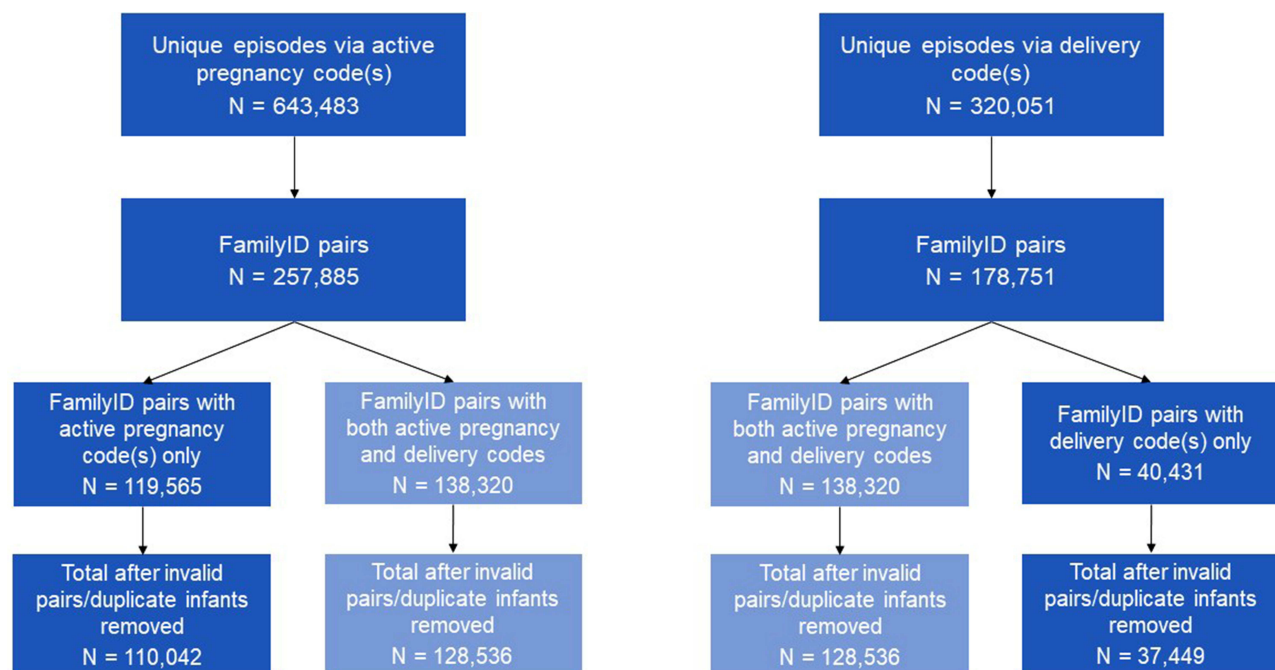

Figure 1 Assessment of mother–infant linkages relative to total unique pregnancy episodes in the JMDC claims database between January 2005 and March 2022 (Linkage Method A).
Notes: Pregnancy episodes identified based on presence of active pregnancy or delivery codes among females ages 12–55 years. Active pregnancy codes were diagnosis codes that indicated a pregnant state; delivery codes indicated occurrence of a live birth. FamilyID pairs formed based on matching family identification codes between females with pregnancy episode and dependents born during the study period and enrolled in the JMDC during their birth month.
Abbreviation: FamilyID, Family identification code.

infant. After removal of invalid pairs (7%), 238,578 remained. Additionally, Linkage Method A identified 320,051 pregnancy episodes via delivery codes, 56% of which linked with an infant. After removal of invalid pairs (7%), 165,985 remained. Comparison of these two groups revealed 276,027 unique pairs. Based on Linkage Method B (Figure 2), 446,441 total unique pairs were identified via family identification codes, 67% of which possessed pregnancy or delivery codes. After removal of invalid pairs (14%), 385,295 remained. These results are discussed in more detail in the context of the Duke-Margolis framework in the following sections.

Data Relevancy
Availability of Key Data Elements
Exposure (Maternal Medication Exposure)
The JMDC contains records of all prescribed medications, with dates of dispensing, prescribed daily dose, and number of days administered.[24] These prescription claims are expected to offer a complete representation of all medications dispensed during pregnancy (ie, high exposure sensitivity) as needed to define the exposure of interest. However, all claims databases have inherent limitations such that a filled prescription does not indicate that the medication was consumed, which could contribute to low exposure positive predictive value.

Outcome (Infant Major Congenital Malformations)
All infants in Japan are enrolled in the National Health Insurance System within one month of birth and the JMDC captures all infant medical visits; therefore, major congenital malformations are expected to be completely captured. However, presence of a diagnosis code on a medical claim is not necessarily indicative of disease, as the code may be incorrectly coded or included as rule-out criteria. Similarly, the absence of a diagnosis code on a medical claim may not necessarily indicate the absence of a disease, for example, due to a missed diagnosis, an error in reporting, or the possibility that a condition present at birth is not clinically manifest until a later age. Validation of a JMDC claims-based algorithm to identify any infant major congenital malformations against gold standard medical records in Japan found the positive predictive value to be 91.5% (95% CI 85.6–95.5%); negative predictive value was not reported.[34]
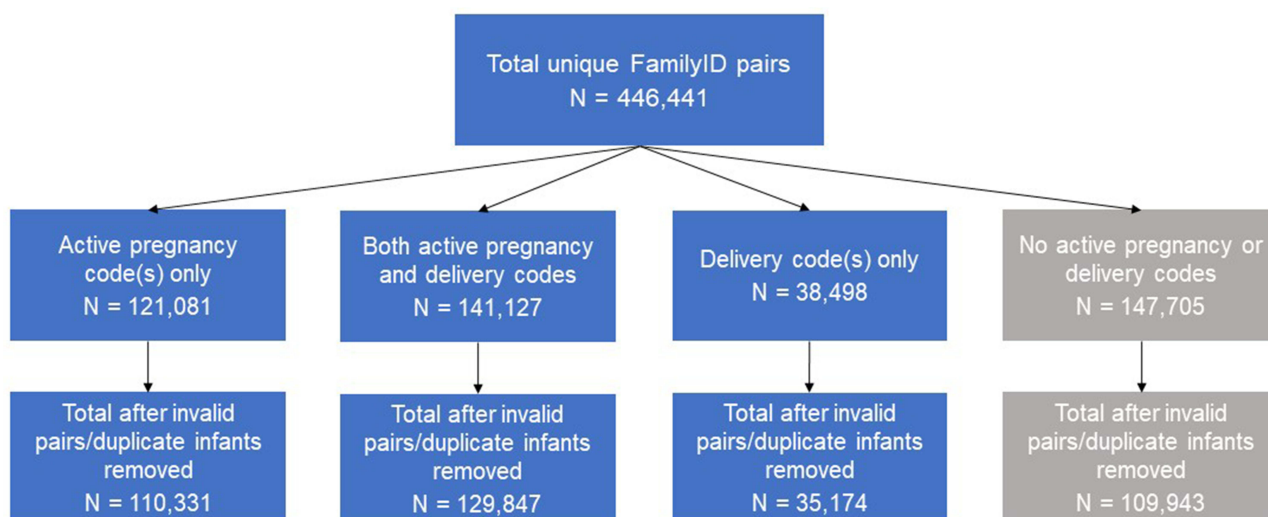

Figure 2 Assessment of pregnancy and delivery coding among mother–infant linkages in the JMDC claims database between January 2005 and March 2022 (Linkage Method B).
Notes: Total unique FamilyID pairs formed based on matching family identification codes between females ages 12–55 years and dependents born during the study period and enrolled in the JMDC during their birth month. Presence of active pregnancy codes assessed in the 294 days before the 15th day of the infant birth month. Presence of delivery codes assessed in the 60 days before and after the 15th day of the infant birth month. Active pregnancy codes were diagnosis codes that indicated a pregnant state; delivery codes indicated occurrence of a live birth.
Abbreviation: FamilyID, Family identification code.

Covariates

We expect to be able to measure demographic characteristics, comorbidities, concomitant medications, and healthcare utilization characteristics as required to address confounding of the exposure-outcome association of interest. As previously mentioned, there are general limitations of using claims databases for research purposes (eg, whether the presence versus absence of a diagnosis code corresponds to the true disease state). Missing data will be a concern mainly when lifestyle (eg, alcohol use, smoking, physical activity), biometrics (eg, body mass index, blood pressure, cholesterol), and pregnancy-related characteristics (eg, reproductive history) are important covariates for a given analysis because these items are not well captured in claims.[25]

Patient-Level Linking (Linkage of Maternal and Infant Records)

A unique family identification code is assigned to the insured individual and their dependents, enabling linkage of maternal and infant records. Further detail is discussed as part of the data quality considerations.

Representativeness

There were no differences between the JMDC and general Japan populations in terms of maternal age or infant sex distribution (Table 1). Preterm birth occurred in 3.6% of all JMDC pairs versus 5.6% of the general population. Comparison of the linked and unlinked populations did not reveal any differences in maternal age, infant sex, or preterm birth.

Table 1 Prevalence of Maternal Medical Conditions, Medication Use, and Infant Major Congenital Malformations in the Japan Medical Data Center (JMDC) Claims Database Linked and Unlinked Populations, Compared to the General Japan Population, 2005–2022

	General Japan Population	JMDC Linked, All Valid Pairs	JMDC Linked, Valid Pairs with Codes	JMDC Unlinked[a]
Female Characteristics[b]		**N = 287,444**	**N = 207,681**	**N = 119,676**
Demographic Characteristics[c]				
Age, years (mean)	31.7	32.3 (SD 5.0)	32.6 (SD 5.0)	32.6 (SD 5.0)
Medical Conditions[d]				
Alcohol use disorders	0.5%	0.1%	0.1%	0.2%
Anxiety disorders	3.5%	2.9%	3.5%	4.7%
Asthma	4.1%	11.4%	13.5%	19.8%
Breast cancer	0.6%	0.1%	0.1%	0.2%
Cervical cancer	0.1%	0.1%	0.1%	0.2%
Diabetes mellitus	1.9%	1.0%	1.3%	1.3%
Idiopathic epilepsy	0.2%	0.6%	0.7%	0.6%
Major depressive disorder	2.4%	2.2%	2.6%	3.5%
Migraine	20.2%	3.6%	4.3%	5.7%
Substance use disorders	1.5%	0.2%	0.2%	0.5%
Medication Use				
ACE-Is	Not reported	0.0%	0.0%	0.0%
Anti-diabetics	Not reported	1.2%	1.6%	1.2%
Anticonvulsants	Not reported	0.5%	0.7%	0.7%
Methotrexate	Not reported	0.0%	0.0%	0.0%
Mycophenolate	Not reported	0.0%	0.0%	0.0%
SSRIs	Not reported	0.5%	0.7%	0.6%
Statins	Not reported	0.1%	0.1%	0.1%
Warfarin	Not reported	0.0%	0.0%	0.0%

(*Continued*)

Table 1 (Continued).

	General Japan Population	JMDC Linked, All Valid Pairs	JMDC Linked, Valid Pairs with Codes	JMDC Unlinked[a]
Infant Characteristics[e]		**N = 385,295**	**N = 275,352**	**N = 410,179**
Demographic Characteristics[c]				
Preterm versus term				
Preterm (<37 weeks)	5.6%	3.6%	4.1%	3.3%
Term	94.3%	96.4%	95.9%	96.7%
Sex				
Male	51.3%	51.2%	51.1%	51.3%
Female	48.7%	48.8%	48.9%	48.7%
Major Congenital Malformations[d]				
Any major congenital malformation[f]	5.3%	10.8%	11.1%	11.2%
Congenital heart anomalies	1.2%	2.8%	3.0%	2.7%
Congenital musculoskeletal/limb anomalies	2.1%	1.7%	1.7%	1.8%
Digestive congenital anomalies	0.5%	0.5%	0.5%	0.5%
Down syndrome	0.1%	0.2%	0.2%	0.1%
Klinefelter syndrome	0.0%	0.0%	0.0%	0.0%
Neural tube defects	0.0%	0.2%	0.2%	0.2%
Orofacial clefts	0.2%	0.2%	0.2%	0.2%
Other chromosomal abnormalities	0.7%	0.1%	0.2%	0.1%
Other congenital birth defects	0.1%	6.2%	6.3%	6.7%
Turner syndrome	0.1%	0.0%	0.0%	0.0%
Urogenital congenital anomalies	0.9%	0.3%	0.4%	0.4%

Notes: [a]Unlinked population represents two separate cohorts: (1) females with delivery codes without a paired infant and (2) infants without a paired mother. [b]Female characteristics assessed among total number of unique females. If a female had more than one delivery during the study period, a single delivery was randomly selected. [c]General Japan population data obtained from Japanese Vital Statistics. [d]General Japan population data obtained from the Global Burden of Disease Study. [e]Infant characteristics assessed among total number of unique infants (equivalent to total number of unique mother–infant pairs). [f]Does not equate to a simple sum of the individual major congenital malformation categories due to some infants possessing >1 major congenital malformation.
Abbreviations: ACE-Is, Angiotensin-converting enzyme inhibitors; SSRIs, Selective serotonin reuptake inhibitors.

The JMDC linked females had a similar prevalence of anxiety disorders, diabetes mellitus, major depressive disorder, and cervical cancer as same-aged females in Japan. Alcohol and substance use disorders were less prevalent in the JMDC versus general population, but these conditions are generally under-captured in claims.[35,36] Migraine was also less prevalent in the JMDC versus general population (3.6% versus 20.2%), potentially due to claims representing the severe migraine population.[37–39] Asthma was more common in the JMDC compared to the general population (11.4% versus 4.1%), potentially representing a more prevalent lifetime history of asthma, rather than current asthma.[40,41] Restriction of the linked population to those with pregnancy or delivery codes increased the prevalence of chronic conditions, perhaps due to a greater likelihood of healthcare utilization. The unlinked mothers were more likely than the linked mothers to have most of the medical conditions examined. The prevalence of medication use was similar across the total linked, the linked with pregnancy or delivery codes, and the unlinked populations for all medication classes examined.

Total infant major congenital malformations were more prevalent among JMDC linked infants (10.8%) than the general Japan population (5.3%), but the prevalence of each specific congenital malformation subtype (ie, the defined subcategories, less the catch-all category for "other" congenital birth defects) was the same or similar between the two populations. Given that it is unlikely that future pregnancy safety studies would specify the primary outcome as "other" congenital birth defects, but rather would use specific subtypes, we do not consider the apparently heightened prevalence of "other" congenital birth defects to be a major flaw of the JMDC claims database. Restricting to pairs with pregnancy or delivery codes did not result in substantial changes of congenital birth defect prevalence, suggesting the missingness of these codes to be non-differential with respect to outcome.

36

DovePress

Clinical Epidemiology 2024:16

Sufficient Subjects

The JMDC captured, on average, around 4% of live births in Japan from 2005 to 2022 (Table 2). The percentage of live births captured in the JMDC increased over the study period, such that the rates were 9–11% in 2018 to 2021. Fifty-four percent of all infants were validly linked to a mother via Linkage Method B, creating 385,295 total pairs (Table 2). Among these pairs, about 41,000 infant major congenital malformations were observed within one year of birth (Supplemental Table 4). Congenital heart anomalies were the most common subtype (N = 10,934). All additional subtypes had greater than 500 events, except for Klinefelter (N = 9) and Turner (N = 13) syndromes, which are often diagnosed later in life. There were 3395 mothers exposed to anti-diabetics in the year before delivery and about 1500 each exposed to anticonvulsants and SSRIs; other medication exposures occurred more rarely (Supplemental Table 5). Rothman and Greenland have derived a method to plan the size of an epidemiologic study based on precision (ie, desired width of the 95% confidence interval for the effect estimates), rather than power, to divert focus from statistical significance testing.[42] Under the null (equal outcome event rates among exposed and unexposed), the associations between anti-diabetics, anticonvulsants, or SSRIs with congenital heart anomalies, congenital musculoskeletal and limb anomalies, or total congenital anomalies (both with and without "other" congenital birth defects included) were the only exposure–outcome combinations with expected precision less than or equal to 2.2, measured as the ratio of the expected upper and lower confidence limits. All other associations resulted in insufficient precision estimates. These results suggest that the JMDC mother–infant linked population may have sufficient subjects to study some, but not all, in utero medication exposures and infant congenital malformations.

Longitudinality

At least one year of infant follow-up (a generous follow-up period given that 94% congenital malformations are recorded within 90 days) was available for 86% of valid pairs (Table 3).[43] Additionally, 57% of mothers were continuously enrolled in their health plan throughout pregnancy. Restriction of valid pairs to those with pregnancy or delivery codes

Table 2 Summary of Annual Births in the Japan Medical Data Center (JMDC) Claims Database Among Linked Mother–Infant Pairs, Compared to the General Japan Population

Year	Total Annual Births in Japan	Total Number of JMDC Covered at Birth	Total Valid Pairs, Linkage Method A	Total Valid Pairs, Linkage Method B	Total Valid Pairs with Codes, Linkage Method B[a]
2005	1,062,530	4167	1002	1714	981
2006	1,092,674	4585	1267	1893	1253
2007	1,089,818	4765	1410	2015	1398
2008	1,091,156	6062	1960	2901	1934
2009	1,070,036	8903	3380	4857	3336
2010	1,071,305	14,614	6068	9170	6000
2011	1,050,807	18,518	8212	11,913	8181
2012	1,037,232	24,735	10,819	15,791	10,770
2013	1,029,817	34,879	15,344	22,465	15,301
2014	1,003,609	38,265	16,790	23,784	16,767
2015	1,005,721	52,451	21,433	30,824	21,408
2016	977,242	64,121	26,168	37,098	26,154
2017	946,146	78,747	31,502	44,359	31,500
2018	918,400	86,014	33,335	46,191	33,327
2019	865,239	88,590	32,913	45,301	32,842
2020	840,835	89,644	31,797	42,741	31,724
2021	811,622	86,897	29,017	37,619	28,889
2022[b]	799,728	11,077	3610	4659	3587
Total	**17,763,917**	**717,034**	**276,027**	**385,295**	**275,352**

Notes: [a]Linked population from Linkage Method B, restricted to those with codes indicating an active pregnancy (assessed in the 294 days before the 15th day of the infant birth month) or delivery (assessed in the 60 days before and after the 15th day of the infant birth month). Active pregnancy codes were diagnosis codes that indicated a pregnant state; delivery codes indicated occurrence of a live birth. [b]Data availability ended in March 2022.

Table 3 Maternal and Infant Enrollment Characteristics Among Linked Mother–Infant Pairs in the Japan Medical Data Center (JMDC) Claims Database, 2005–2022

	All Valid Pairs[a] (N = 385,295) N (%)	Valid Pairs with Codes[b] (N = 275,352) N (%)
Maternal continuous enrollment during pregnancy	218,560 (56.7%)	186,676 (67.8%)
Length of pre-pregnancy continuous enrollment		
Not enrolled at pregnancy onset[c]	166,735 (43.3%)	88,676 (32.2%)
≤30 days	7227 (1.9%)	6251 (2.3%)
31–90 days	13,556 (3.5%)	11,702 (4.2%)
91–180 days	18,913 (4.9%)	16,209 (5.9%)
>180 days	178,864 (46.4%)	152,514 (55.4%)
Length of infant follow-up		
≤30 days	0 (0.0%)	0 (0.0%)
31–90 days	9207 (2.4%)	7154 (2.6%)
91–180 days	14,649 (3.8%)	11,251 (4.1%)
181–365 days	28,798 (7.5%)	21,936 (8.0%)
≥366 days	332,641 (86.3%)	235,011 (85.4%)
Both maternal continuous enrollment during pregnancy and at least 365 days of infant follow-up	186,826 (48.5%)	158,838 (57.7%)

Notes: [a]Linked population from Linkage Method B. [b]Linked population from Linkage Method B, restricted to those with codes indicating an active pregnancy (assessed in the 294 days before the 15th day of the infant birth month) or delivery (assessed in the 60 days before and after the 15th day of the infant birth month). Active pregnancy codes were diagnosis codes that indicated a pregnant state; delivery codes indicated occurrence of a live birth. [c]Only females continuously enrolled during pregnancy were enrolled at the onset of pregnancy and therefore were eligible for assessment of pre-pregnancy continuous enrollment.

increased this proportion to 68%, perhaps due to a greater likelihood to engage with the healthcare system. Given that all individuals in Japan are covered by the National Health Insurance and pregnancy-related care is subsidized, we do not suspect the low proportion with continuous enrollment to be reflective of gaps in care received during pregnancy, but rather may be a result of changes in the contracts between the JMDC and the companies that provide claims to the database. Regardless, future pregnancy safety studies in the JMDC requiring continuous enrollment during pregnancy may suffer from limited sample sizes and a potentially high proportion of censoring during the follow-up period.

Data Quality: Mother–Infant Matches
Completeness
Our methods are expected to capture all mothers and infants enrolled in a shared health insurance plan. Successful linkage occurred for 56% of pregnancy episodes with live birth delivery codes (Figure 1) and 54% of infants (Table 2). These linkage rates are expected given the likelihood of infants to be covered under their mother's, versus other parent's, insurance plan. Although only 40% of pregnancy episodes identified via active pregnancy codes linked to an infant (Figure 1), this group likely disproportionately includes pregnancies ending in non-live birth outcome.

Accuracy
Validity
Unfortunately, the JMDC does not provide documentation related to the validity of the family identification variable. However, given that these codes are generated and used for billing purposes, they are expected to correctly indicate correspondence between insured individuals and their dependents.

Conformance
The family identification codes are always assigned, regardless of the number of dependents, and database creation and maintenance does not disrupt the variable. This information is therefore expected to remain congruent with the standard.

Mother's Relationship to Insurance Holder	Infant's Relationship to Insurance Holder								
	Adopted child	Brother/Sister	Child	Child's Spouse	Grandson	Nephew/Niece	Spouse's child	Other	Missing
Adopted child	0	0	1116	0	65	0	0	3	0
Brother/Sister	0	0	79	0	0	3	0	0	0
Child	11	0	7307	0	1992	0	1	5	0
Child's spouse	0	0	3	0	141	0	0	0	0
Common-law wife	0	0	93	0	2	0	1	3	0
Foster parents	0	0	23	0	0	0	0	0	0
Grandson	0	0	0	0	5	0	0	0	0
Insured	2	0	19,299	0	353	5	0	1	152
Nephew/niece	0	0	11	0	0	0	0	0	0
Sister/brother-in-law	0	0	12	0	0	0	0	0	0
Spouse	16	1	381,333	1	746	1	4	30	2
Spouse's child	0	0	165	0	11	0	0	4	0
Spouse's parents	0	0	288	0	0	0	0	0	0
Younger brother/sister	0	0	317	0	1	13	0	0	0
Younger sister-/brother-in-law	0	0	45	0	0	0	0	0	0
Parents	0	0	827	0	0	2	0	0	0
Other	0	0	368	0	57	0	0	12	0
Missing	0	0	14	0	0	0	0	0	31,465

	N	%
Assumed valid mother–infant pairs	400,730	89.8
Possible mother–infant pairs	2225	0.5
Improbable mother–infant pairs	11,340	2.5
Unknown/missing relationships	32,116	7.2
Total	446,411	

Figure 3 Value of variable indicating relationship to insurance holder for all linked pairs of females of childbearing age with members enrolled at birth who share family identification codes.

Notes: Mother–infant pairs assumed to be valid pairs were those in which the mother was the insured individual or the spouse of the insured individual and the infant was indicated to be the child. Possible mother–infant pairs were those in which a mother–child relationship was possible but not confirmable without additional family tree information (eg, aunt and nephew/niece, child and grandchild). Improbable mother–infant pairs were those in which, assuming incestuous relationships to be improbable, an alternative familial relationship was more likely (eg, siblings, grandparent and grandchild). Although some of the "adopted child" infant relationships could be valid relationships between mother and child, these are not valid mother–infant pairs for the purposes of studying in utero exposures.

Logical Plausibility

Of 446,411 potential mother–infant pairs with matching family identification codes, examination of relationships to the insurance holder indicated 90% to be mother–child relationships (Figure 3). An additional 0.5% had possible, but not definite, relationships (eg, aunt–niece/nephew relationship). In total, 2.5% were implausible mother–infant pairs (eg, siblings) and 7% were missing this information. Restriction to pairs with pregnancy or delivery codes nearly eliminated implausible pairs (Supplemental Figure 2). Overall, use of the relationship variable supports an increased confidence in the validity of the mother–infant pairs.

Consistency

Unfortunately, our assessment of consistency was limited by the lack of detailed documentation within the JDMC regarding creation and use of the family identification variable. Useful information would be the date of inception of the current family identification coding version, which would indicate whether assignment of codes is consistent across all cohort members.

Transparency of Data Processing

Mother–infant pairs may be created in the JMDC via transparent data processing, with no restrictions on the publishing of linkage algorithms.

Provenance

The JMDC does not perform any transformations to the family identification codes after receiving the data from the payers.

Data Quality: Gestational Period

Completeness

As neither date of delivery nor infant date of birth (suppressed to month and year to avoid re-identification) are available in the JMDC, this information must be estimated according to claims that indicate delivery.[44] Fifty-seven percent of valid pairs

lacked delivery codes (Figure 2), likely due to the health insurance processes in Japan. In this case, an estimated date (eg, 15th day of infant birth month) would need to be imputed. Following delivery date estimation, gestational age would need to be estimated to back calculate the pregnancy onset date (to allow for designation of critical exposure windows). Claims-based gestational age algorithms often employ codes related to pregnancy milestones, which may be limited in this cohort due to missing active pregnancy codes in 38% of valid pairs (Figure 2).[15,21,31] When pregnancy milestone codes are missing, the gestational length of a term pregnancy may be imputed (although this may compromise accuracy, as described in more detail below). Dual imputation of delivery date and gestational age would be required for 29% of valid pairs.

Accuracy
Validity
Inserting an estimated delivery date and/or gestational age for each mother–infant pair is expected to result in misclassification, such that the assigned in utero medication exposure windows will not match the truth for a subset of pairs. Imputing a full-term gestational length shifts the estimated exposure window earlier in time for preterm deliveries (Supplemental Figure 3E–H) and later in time for post-term deliveries (Supplemental Figure 3I–L), the degree of which depends on the true gestational age. It is possible that misclassification of the exposure window could be associated with infant congenital malformation status (ie, differential exposure misclassification) if there are shared mechanisms for congenital malformations and preterm birth.[45,46] Reassuringly, for term deliveries (Supplemental Figure 3A–D), which represent 94% of deliveries in Japan, exposure window misclassification is expected to be minor.

Conformance
The JMDC does not support the fifth-level digit of the ICD-10 code (where gestational week information is available), which is inconsistent with standard availability in other databases and limits the ability to finely estimate gestational age (Supplemental Table 6).

Logical Plausibility
As previously mentioned, published algorithms could be used to insert an estimated gestational age for each mother–infant pair (eg, fixed durations of 35 weeks for preterm births and 40 weeks for non-preterm births).[15,21,31] Because of this, no logically implausible gestational ages can occur (ie, greater than post-term or less than viable).

Consistency
We expect the use of pregnancy and delivery codes, as needed to estimate gestational age, to be consistent across all mothers in the linked cohort because the current ICD-10 coding scheme has been in use in Japan since 1990 and therefore covers our entire study period.

Transparency of Data Processing
Gestational age may be estimated in the JMDC via transparent data processing, with no restrictions on the publishing of algorithms.

Provenance
The way in which the JMDC transforms the collected data into the version available in the database involves translation of standard Japanese disease names to ICD-10 diagnosis codes, which limits the maximum unit of ICD-10 information to the fourth level. This data transformation is expected to impact the quality of the gestational age estimation.

Discussion
We have applied the Duke-Margolis framework to assess whether a mother–infant linked cohort in the JMDC claims database is fit for regulatory use within the context of estimating infant outcomes associated with in utero exposure to marketed medications.[26] A summary of our assessment is provided in Supplemental Table 7. Although complete and accurate identification of mothers and their liveborn infants who share a health insurance plan was possible, the limited gestational age information may impede valid assignment of pregnancy onset and delivery dates as needed to define

critical in utero exposure windows. Future researchers considering the use of a mother–infant linked cohort in the JMDC claims database for pregnancy safety studies should understand the implications of a population restricted to live birth outcomes, which depend on the research question of interest.[45]

In terms of data relevancy, we determined that critical fields (maternal medication exposures, infant major congenital malformations, covariates) were available. A total of 385,295 mother–infant pairs were identified, representing about 2% of live births in Japan during the study period. A sufficiently sized population may be available to study associations between some maternal medication exposures and major congenital malformations that occurred more commonly in this population. Fifty-seven percent of pairs were continuously enrolled during pregnancy, which is expected to be reflective of how care is captured by the database, rather than how care is received in this population. Comparison to publicly available data suggested that the distribution of maternal characteristics and the occurrence of specific major congenital malformation subtypes were mostly consistent with the general population of Japan. Preterm births occurred less often than expected (3.6% versus 5.6%) in this population, which could be due to coding practices or a true lower prevalence in this privately employed population.

In terms of quality, our methods were expected to accurately identify the complete set of mothers and infants in the JMDC enrolled in a shared health insurance plan. Infants and females with evidence of a live birth delivery both had linkage rates of about 50%, which aligns with expectations of infant insurance coverage under the mother's, versus other parent's, plan. Exclusion of invalid and indeterminate relationships was intended to improve validity. Missing delivery date and pregnancy timing information, coupled with suppression of infant birth dates and inaccessibility of ICD-10 codes with fifth-level digits (where gestational week information would have been available), limit the ability to finely estimate gestational timing as needed for regulatory pregnancy safety studies. We suspect that for the majority (94%) of pregnancies, which involve term deliveries, missing delivery and gestational timing information may result in only minor misclassification of the estimated exposure window. However, for preterm births without gestational age information, in utero exposure measurement could erroneously occur before pregnancy start date. This mismeasurement may be problematic as many people cease or change medication use upon becoming pregnant.[46,47] The magnitude and directionality of bias introduced due to misalignment of the estimated and true exposure windows depends on the specific research question of interest.

This analysis is not without limitations. First, it is important to note that the Duke-Margolis framework sets forth a list of considerations for assessing fitness for purpose of real-world data but does not specify analysis plans nor provide quantitative thresholds for determining whether the relevancy and quality dimensions are met. Our assessment was therefore based on our translation of the framework into descriptive epidemiologic research questions, which was informed by experts in the field of pregnancy safety, as well as those familiar with Japan's healthcare system and the JMDC. Next, our analysis identified pregnancy and delivery episodes based on ICD-10 diagnosis codes alone. Future studies may examine how incorporation of medical and surgical procedure codes may improve delivery date estimation for the nearly 60% of this cohort that was missing live birth diagnosis codes. Finally, our mother–infant pairs were created according to values indicated for the family identification codes and relationships to the insurance holder. Unfortunately, there is a scarcity of detailed information available from the JMDC data vendor regarding these variables, including potential for miscoding and reasons for missingness. Given that the family identification codes are used to identify relations between insurance plan holders and dependents for the purposes of billing, we expect these values (which are always non-missing) to be correctly coded. The accuracy of the relationship variable, however, is less clear given that insurance benefits are provided to all dependents, regardless of specific familial relationship, so there is less motivation from the billing viewpoint for this variable to be valid and non-missing.

Conclusion

Overall, results suggest that the JMDC claims database may be well suited for descriptive studies of pregnant people in Japan (eg, comorbidities, medication usage). However, before the database can be considered fit for supporting regulatory decision-making, more work is needed to identify a method to assign pregnancy onset and delivery dates so that in utero exposure windows can be defined more precisely as needed for many regulatory postapproval pregnancy safety studies.

Acknowledgments

Disclosure

References

1. Brandon AR, Shivakumar G, Lee SC, Inrig SJ, Sadler JZ. Ethical issues in perinatal mental health research. *Curr Opin Psychiatry.* 2009;22 (6):601–606. doi:10.1097/YCO.0b013e3283318e6f
2. Allesee L, Gallagher CM. Pregnancy and protection: the ethics of limiting a pregnant woman's participation in clinical trials. *J Clin Res Bioeth.* 2011;2(108):1000108. doi:10.4172/2155-9627.1000108
3. Margulis AV, Anthony M, Rivero-Ferrer E. Drug safety in pregnancy: review of study approaches requested by regulatory agencies. *Curr Epidemiol Rep.* 2019;6(3):380–389. doi:10.1007/s40471-019-00212-6
4. Hernán MA, Robins JM. Using big data to emulate a target trial when a randomized trial is not available. *Am J Epidemiol.* 2016;183(8):758–764. doi:10.1093/aje/kwv254
5. Hernán MA, Alonso A, Logan R, et al. Observational studies analyzed like randomized experiments: an application to postmenopausal hormone therapy and coronary heart disease. *Epidemiology.* 2008;19(6):766–779. doi:10.1097/EDE.0b013e3181875e61
6. Hernán MA, Sauer BC, Hernández-Díaz S, Platt R, Shrier I. Specifying a target trial prevents immortal time bias and other self-inflicted injuries in observational analyses. *J Clin Epidemiol.* 2016;79:70–75. doi:10.1016/j.jclinepi.2016.04.014
7. Daw JR, Hanley GE, Greyson DL, Morgan SG. Prescription drug use during pregnancy in developed countries: a systematic review. *Pharmacoepidemiol Drug Saf.* 2011;20(9):895–902. doi:10.1002/pds.2184
8. Feldman Y, Koren G, Mattice K, Shear H, Pellegrini E, MacLeod SM. Determinants of recall and recall bias in studying drug and chemical exposure in pregnancy. *Teratology.* 1989;40(1):37–45. doi:10.1002/tera.1420400106
9. Agency for Healthcare Research and Quality (US). AHRQ methods for effective health care. In: Gliklich RE, Dreyer NA, Leavy MB, editors. *Registries for Evaluating Patient Outcomes: A User's Guide.* Agency for Healthcare Research and Quality (US); 2014.
10. Schneeweiss S, Avorn J. A review of uses of health care utilization databases for epidemiologic research on therapeutics. *J Clin Epidemiol.* 2005;58 (4):323–337. doi:10.1016/j.jclinepi.2004.10.012
11. Krueger WS, Anthony MS, Saltus CW, et al. Evaluating the safety of medication exposures during pregnancy: a case study of study designs and data sources in multiple sclerosis. *Drugs Real World Outcomes.* 2017;4(3):139–149. doi:10.1007/s40801-017-0114-9
12. Margulis AV, Andrews EB. The safety of medications in pregnant women: an opportunity to use database studies. *Pediatrics.* 2017;140(1): e20164194. doi:10.1542/peds.2016-4194
13. Hernández-Díaz S, Huybrechts KF, Chiu YH, Yland JJ, Bateman BT, Hernán MA. Emulating a target trial of interventions initiated during pregnancy with healthcare databases: the example of COVID-19 vaccination. *Epidemiology.* 2023;34(2):238–246. doi:10.1097/ ede.0000000000001562
14. MotherToBaby. Critical periods of development. Organization of Teratology Information Specialists. Available from: https://mothertobaby.org/fact-sheets/critical-periods-development/pdf/. Accessed December, 2022.
15. Palmsten K, Huybrechts KF, Mogun H, et al. Harnessing the Medicaid Analytic eXtract (MAX) to evaluate medications in pregnancy: design considerations. *PLoS One.* 2013;8(6):e67405. doi:10.1371/journal.pone.0067405
16. Garbe E, Suling M, Kloss S, Lindemann C, Schmid U. Linkage of mother–baby pairs in the German pharmacoepidemiological research database. *Pharmacoepidemiol Drug Saf.* 2011;20(3):258–264. doi:10.1002/pds.2038
17. Taylor LG, Thelus Jean R, Gordon G, Fram D, Coster T. Development of a mother–child database for drug exposure and adverse event detection in the military health system. *Pharmacoepidemiol Drug Saf.* 2015;24(5):510–517. doi:10.1002/pds.3759
18. Andrade SE, Toh S, Houstoun M, et al. Surveillance of medication use during pregnancy in the mini-sentinel program. *Maternal Child Health J.* 2016;20(4):895–903. doi:10.1007/s10995-015-1878-8
19. Andrade SE, Davis RL, Cheetham TC, et al. Medication exposure in pregnancy risk evaluation program. *Matern Child Health J.* 2012;16 (7):1349–1354. doi:10.1007/s10995-011-0902-x
20. Moore Simas TA, Huang M-Y, Packnett ER, Zimmerman NM, Moynihan M, Eldar-Lissai A. Matched cohort study of healthcare resource utilization and costs in young children of mothers with postpartum depression in the United States. *J Med Econ.* 2020;23(2):174–183. doi:10.1080/ 13696998.2019.1679157
21. Margulis AV, Setoguchi S, Mittleman MA, Glynn RJ, Dormuth CR, Hernández-Díaz S. Algorithms to estimate the beginning of pregnancy in administrative databases. *Pharmacoepidemiol Drug Saf.* 2013;22(1):16–24. doi:10.1002/pds.3284
22. Yusuf A, Chia V, Xue F, Mikol DD, Bollinger L, Cangialose C. Use of existing electronic health care databases to evaluate medication safety in pregnancy: triptan exposure in pregnancy as a case study. *Pharmacoepidemiol Drug Saf.* 2018;27(12):1309–1315. doi:10.1002/pds.4658

23. Ishikawa T, Obara T, Jin K, et al. Folic acid prescribed to prenatal and postpartum women who are also prescribed antiepileptic drugs in Japan: data from a health administrative database. *Birth Defects Res*. 2020. doi:10.1002/bdr2.1748

24. Ishikawa T, Obara T, Nishigori H, et al. Antihypertensives prescribed for pregnant women in Japan: prevalence and timing determined from a database of health insurance claims. *Pharmacoepidemiol Drug Saf*. 2018;27(12):1325–1334. doi:10.1002/pds.4654

25. Andrade SE, Bérard A, Nordeng HME, Wood ME, van Gelder MMHJ, Toh S. Administrative claims data versus augmented pregnancy data for the study of pharmaceutical treatments in pregnancy. *Curr Epidemiol Rep*. 2017;4(2):106–116. doi:10.1007/s40471-017-0104-1

26. Daniel G, Silcox C, Bryan J, McClellan M, Romine M, Frank K. Characterizing RWD quality and relevancy for regulatory purposes; 2022.

27. Berger M, Daniel G, Frank K, et al. A framework for regulatory use of real-world evidence. White paper prepared by the Duke Margolis Center for Health Policy. 2017:6.

28. Noh Y, Yoon D, Song I, Jeong HE, Bae JH, Shin J-Y. Discrepancies in the evidence and recommendation levels of pregnancy information in prescription drug labeling in the United States, United Kingdom, Japan, and Korea. *J Womens Health*. 2018;27(9):1086–1092. doi:10.1089/jwh.2017.6792

29. Little SH, Motohara S, Plegue M, Medaugh C, Sen A, Ruffin MT. Japanese women's concerns and satisfaction with pregnancy care in the United States. *J Perinat Educ*. 2020;29(3):152–160. doi:10.1891/j-pe-d-19-00009

30. Ruggles BM, Xiong A, Kyle B. Healthcare coverage in the US and Japan: a comparison. *Nursing*. 2019;49(4):56–60. doi:10.1097/01.NURSE.0000553277.03472.d8

31. Hornbrook MC, Whitlock EP, Berg CJ, et al. Development of an algorithm to identify pregnancy episodes in an integrated health care delivery system. *Health Serv Res*. 2007;42(2):908–927. doi:10.1111/j.1475-6773.2006.00635.x

32. Institute for Health Metrics and Evaluation. *Protocol for the Global Burden of Diseases, Injuries, and Risk Factors Study (GBD)*; 2020.

33. Japan Ministry of Health Labour and Welfare. Vital statistics of Japan, final data natality. Available from: https://www.e-stat.go.jp/en. Accessed October, 2021.

34. Ishikawa T, Oyanagi G, Obara T, et al. Validity of congenital malformation diagnoses in healthcare claims from a University Hospital in Japan. *Pharmacoepidemiol Drug Saf*. 2021;30(7):975–978. doi:10.1002/pds.5244

35. Kim HM, Smith EG, Stano CM, et al. Validation of key behaviourally based mental health diagnoses in administrative data: suicide attempt, alcohol abuse, illicit drug abuse and tobacco use. *BMC Health Serv Res*. 2012;12(1):18. doi:10.1186/1472-6963-12-18

36. Desai RJ, Solomon DH, Shadick N, Iannaccone C, Kim SC. Identification of smoking using medicare data — a validation study of claims-based algorithms. *Pharmacoepidemiol Drug Saf*. 2016;25(4):472–475. doi:10.1002/pds.3953

37. Meyers JL, Davis KL, Lenz RA, Sakai F, Xue F. Treatment patterns and characteristics of patients with migraine in Japan: a retrospective analysis of health insurance claims data. *Cephalalgia*. 2019;39(12):1518–1534. doi:10.1177/0333102419851855

38. Sakai F, Igarashi H. Prevalence of migraine in Japan: a nationwide survey. *Cephalalgia*. 1997;17(1):15–22. doi:10.1046/j.1468-2982.1997.1701015.x

39. Sakai F, Hirata K, Igarashi H, et al. A study to investigate the prevalence of headache disorders and migraine among people registered in a health insurance association in Japan. *J Headache Pain*. 2022;23(1):70. doi:10.1186/s10194-022-01439-3

40. Kusunoki T, Morimoto T, Nishikomori R, et al. Changing prevalence and severity of childhood allergic diseases in Kyoto, Japan, from 1996 to 2006. *Allergol Int*. 2009;58(4):543–548. doi:10.2332/allergolint.09-OA-0085

41. Fukutomi Y, Taniguchi M, Watanabe J, et al. Time trend in the prevalence of adult asthma in Japan: findings from population-based surveys in Fujieda City in 1985, 1999, and 2006. *Allergol Int*. 2011;60(4):443–448. doi:10.2332/allergolint.10-OA-0282

42. Rothman KJ, Greenland S. Planning study size based on precision rather than power. *Epidemiology*. 2018;29(5):599–603. doi:10.1097/ede.0000000000000876

43. Cooper WO, Hernandez-Diaz S, Gideon P, et al. Positive predictive value of computerized records for major congenital malformations. *Pharmacoepidemiol Drug Saf*. 2008;17(5):455–460. doi:10.1002/pds.1534

44. Ishikawa T, Obara T, Nishigori H, et al. Development of algorithms to determine the onset of pregnancy and delivery date using health care administrative data in a University Hospital in Japan. *Pharmacoepidemiol Drug Saf*. 2018;27(7):751–762. doi:10.1002/pds.4444

45. Khoury MJ, Flanders WD, James LM, Erickson JD. Human teratogens, prenatal mortality, and selection bias. *Am J Epidemiol*. 1989;130(2):361–370. PMID: 2750731. doi:10.1093/oxfordjournals.aje.a115342

46. Grzeskowiak LE, Gilbert AL, Morrison JL. Exposed or not exposed? Exploring exposure classification in studies using administrative data to investigate outcomes following medication use during pregnancy. *Eur J Clin Pharmacol*. 2012;68(5):459–467. doi:10.1007/s00228-011-1154-9

47. Alwan S, Reefhuis J, Rasmussen SA, Friedman JM; Study NBDP. Patterns of antidepressant medication use among pregnant women in a United States population. *J Clin Pharmacol*. 2011;51(2):264–270. doi:10.1177/0091270010373928