


ORIGINAL ARTICLE - GASTROENTEROLOGY (CLINICAL)

Validation of a claims-based algorithm to identify cases of ulcerative colitis in Japan

Haruei Ogino,* Hiromu Morikubo,[†] Keita Fukaura,[‡] Tasuku Okui,[§] Sean Gardiner,[¶] Naonobu Sugiyama,** Noritoshi Yoshii,** Tsutomu Kawaguchi,** Haoqian Chen,^{††} Edward Nonnenmacher,^{††} Soko Setoguchi,^{††,‡‡} Naoki Nakashima[§] and Taku Kobayashi[†] 

*Department of Medicine and Bioregulatory Science, Graduate School of Medical Sciences, Kyushu University, [†]Department of gastroenterology, Saiseikai Futsukaichi Hospital, [§]Medical information center, Kyushu University Hospital, Fukuoka, [‡]Center for Advanced IBD Research and Treatment, Kitasato University Kitasato Institute Hospital, **Inflammation & Immunology, Medical Affairs, Pfizer Japan Inc, Tokyo, Japan; [¶]Inflammation and Immunology, Pfizer Inc, New York, New York, ^{††}Center for Pharmacoepidemiology and Treatment Science, Rutgers Institute for Health, Health Care Policy and Aging Research ^{‡‡}Department of Medicine, Rutgers Robert Wood Johnson Medical School and Institute for Health, New Brunswick, New Jersey, USA

Key words

algorithm, claims database, ulcerative colitis.

Accepted for publication 31 October 2021.

Correspondence

Dr Taku Kobayashi, Center for Advanced IBD Research and Treatment, Kitasato University Kitasato Institute Hospital, 5-9-1 Shirokane, Minato-ku, Tokyo 108-8642, Japan.
Email: kobataku@insti.kitasato-u.ac.jp

Financial support: This work was sponsored by Pfizer Japan Inc.

Conflict of interest: HO, KF, TO, and NN have received funding from Pfizer Inc to attend study-related meetings. HM has received funding from Pfizer Inc to attend study-related meetings and has received research funding from the Japan Foundation of Applied Enzymology. SG, NS, and NY are employees and stockholders of Pfizer Inc. TKA was formerly employed by Pfizer Inc. HC and EN have no potential conflict of interest to declare. SS has served as a member of a US FDA advisory committee; has served as a consultant for Medtronic, Merck, and Pfizer Inc; and has received research funding from Bristol-Myers Squibb, the Cystic Fibrosis Foundation, Daiichi Sankyo, the National Institutes of Health, the Patient-Centered Outcomes Research Institute, Pfizer Inc, and Pfizer Japan. TKo has received personal fees from AbbVie GK, Ajinomoto Pharma, Alfresa Pharma, Asahi Kasei Medical, Astellas, Celltrion, Covidien,

Abstract

Background and Aim: The prevalence of ulcerative colitis (UC) is increasing in Japan. Validated claims-based definitions are required to investigate the epidemiology of UC and its treatment and disease course in clinical practice. This study aimed to develop a claims-based algorithm for UC in Japan.

Methods: A committee of epidemiologists, gastroenterologists, and internal medicine physicians developed a claims-based definition for UC, based on diagnostic codes and claims for UC treatments, procedures (cytapheresis), or surgery (postoperative claims). Claims data and medical records for a random sample of 200 cases per site at two large tertiary care academic centers in Japan were used to calculate the positive predictive value (PPV) of the algorithm for three gold standards of diagnosis, defined as physician diagnosis in the medical records, adjudicated cases, or registration in the Japanese Intractable Disease Registry (IDR).

Results: Overall, 1139 claims-defined UC cases were identified. Among 393 randomly sampled cases (mean age 44; 48% female), 94% had received ≥ 1 systemic treatment (immunosuppressants, tumor necrosis factor inhibitors, corticosteroids, or antidiarrheals), 7% had cytapheresis, and 7% had postoperative claims. When physician diagnosis was used as a gold standard, PPV was 90.6% (95% confidence interval [CI]: 87.7–93.5). PPV with expert adjudication was also 90.6% (95% CI: 87.7–93.5). PPVs with enrollment in the IDR as gold standard were lower at 41.5% (95% CI: 36.6–46.3) due to incomplete case registration.

Conclusions: The claims-based algorithm developed for use in Japan is likely to identify UC cases with high PPV for clinical studies using administrative claims databases.

EA Pharma Co Ltd, Eisai Co Ltd, Eli Lilly, Ferring Pharmaceuticals, Gilead Sciences, Janssen, JIMRO Co Ltd, Kyorin Pharmaceutical Co Ltd, Mitsubishi Tanabe Pharma, Mochida Pharmaceutical, Nippon Kyaku, Pfizer Inc, Takeda Pharmaceutical, Thermo Scientific, and ZERIA; grants from AbbVie GK, Alfresa Pharma, Asahi Kasei Medical, EA Pharma Co Ltd, Kyorin Pharmaceutical Co Ltd, Mochida Pharmaceutical, Nippon Kyaku, Otsuka Holdings Co Ltd, Thermo Fisher Scientific, and ZERIA; and funding from Pfizer Inc to attend study-related meetings.

Author contributions: All authors participated in the conception and development of this article. HO, HM, KF, TO, HC, EN, SS, NN, and TKo were involved in study design, data collection, and analysis. All authors critically reviewed the manuscript and approved the final draft prior to submission.

Introduction

Ulcerative colitis (UC) is a chronic inflammatory condition of the colonic mucosa, characterized by increased frequency and urgency of bowel movements, bloody diarrhea, and abdominal pain.^{1,2} UC

has an unpredictable disease course,¹ and the symptoms are associated with substantial impairment of a patient's health-related quality of life and work productivity.^{3,4} Approximately 10–15% of patients with UC will experience an aggressive course,⁵ and

primarily in patients with moderate-to-severe disease activity, the 5-year and 10-year cumulative risk of colectomy is 10–15%.⁶

The prevalence of UC in the Japanese population is increasing.^{7,8} Epidemiology data show an increase from 63.6 per 100 000 persons in 2005⁷ to 172.9 per 100 000 persons in 2014.⁸ However, there is a general lack of recent data from nationwide epidemiology studies on UC in Japan. In Japan, a national database has been established for the promotion of research into intractable diseases (rare diseases of unknown etiology for which there are limited treatment options), and registered patients receive a medical expense subsidy.⁹ However, only a proportion of patients with UC are currently enrolled by their physicians into this national registry as a result of the recent downgrading of enrollment to include only moderate-to-severe cases of UC.¹⁰ Guidance from the Japanese Society of Gastroenterology highlights a need for further epidemiologic research to determine the etiology of UC in Japan and improve understanding of the differences in presentation *versus* Western countries.¹¹

Administrative claims databases provide a valuable source of population-based and longitudinal data to study epidemiology, treatments, and outcomes for clinically recognizable diseases, and their use is encouraged by regulatory agencies in Japan for post-marketing surveillance. Data derived from administrative claims databases provide useful real-world data, especially for those diseases for which the prevalence is relatively low, including inflammatory bowel disease (IBD). Validated claims-based disease

definitions may also be used to identify study populations for clinical studies. The use of administrative claims data for research purposes is at an early stage in Japan.^{12–15} In contrast, claims-based definitions of diseases have been developed elsewhere, including validated algorithms for identifying patients with IBD in administrative databases in Canada, Israel, and the United States.^{16–18} It is unknown if these can be used with data from Japan, due to unique features of Japanese claims data and differences in clinical practice.

The objective of this study was to develop a claims-based algorithm for UC in Japan and to validate it *versus* gold standards of diagnosis, defined as physician diagnosis in the medical records, adjudicated cases, or registration in the Japanese Intractable Disease Registry.

Methods

Study design. This was a cross-sectional, retrospective review of hospital claims data, medical records, and registry data from two large, tertiary care academic medical centers: Kitasato University Kitasato Institute Hospital, Tokyo, Japan, and Kyushu University Hospital, Fukuoka, Japan. An overview of the study is shown in Figure 1.

A claims-based definition for UC was developed by a steering committee comprising epidemiologists, UC specialists, and

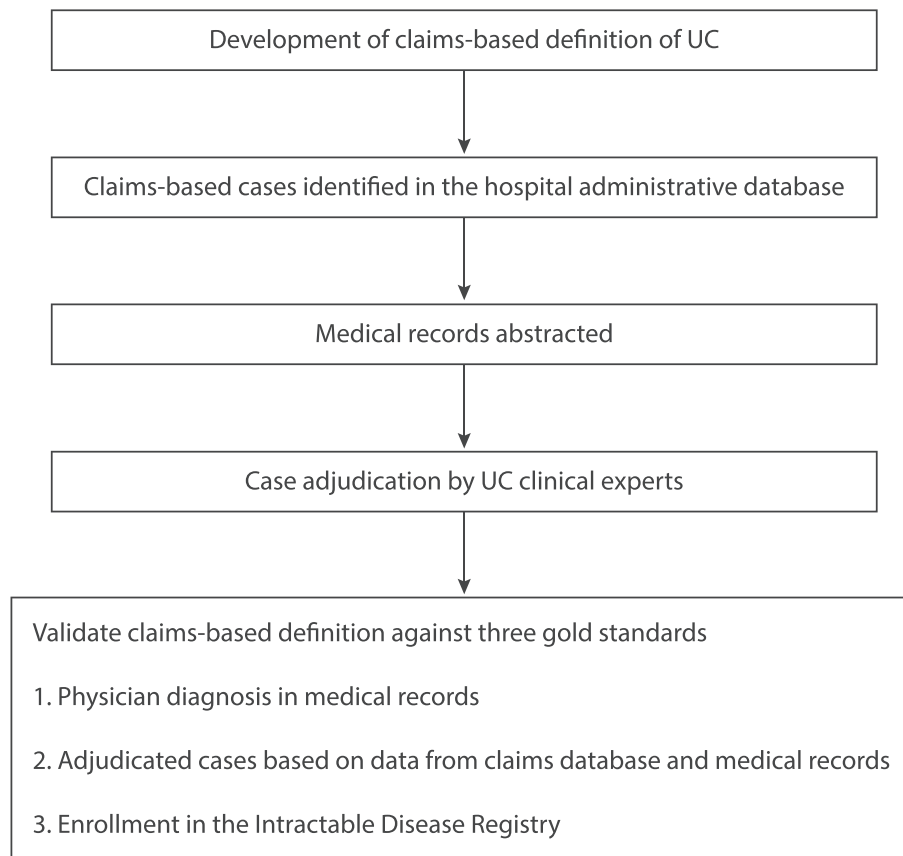


Figure 1 Study flow chart. UC, ulcerative colitis.

internal medicine physicians, in accordance with Japanese clinical practice. Algorithms based on combinations of diagnostic, prescription, procedural, and surgical claims codes were used to identify cases meeting the definition for UC in the hospital databases for data abstraction. A pilot study was conducted on a random sample of 20 claims-based cases per center, and modifications were made to the claims-based definition and data abstraction form before conducting a validation study with a random sample of 200 cases per center. The algorithm use in the validation study was based on a physician diagnosis and a drug, procedural, or surgery claim (Table 1).

Trained abstractors reviewed the hospital medical records for each individual claims-based case to extract the physician diagnosis, as well as the clinical and laboratory data, and to determine whether the patient was enrolled in the Japanese Intractable Disease Registry. Two independent abstractors cross-abstracted data from a sample of 20 cases per center in the pilot study, and inter-abstractor variability was assessed.

An adjudication committee of UC clinical experts independently reviewed the abstracted medical chart data and selected claims data for each case to make a final determination of the UC diagnosis as definite, probable (definition of confirmed UC not met but information suggests the case is likely to be UC), or not UC (documented misdiagnosis or alternative diagnosis). Inter-rater variability between adjudicators was also assessed, based on the number of adjudicated cases for which there was agreement divided by the total number of adjudicated cases.

The study protocol was approved by the Research Ethics Committee of Kitasato University Kitasato Institute Hospital (Protocol number 18086, 20040) and the Research Ethics Committee of Kyushu University Hospital (Protocol number 30-539) and at Rutgers University. The study was conducted in accordance with Guidelines for Good Pharmacoepidemiology Practices issued by the International Society for Pharmacoepidemiology, Good Epidemiological Practice guidelines issued by the International

Epidemiological Association, and Good Practices for Outcomes Research issued by the International Society for Pharmacoepidemiology and Outcomes Research. Patients identified in the claims databases were not required to provide consent for their data to be included in the study; however, patients were provided with the option of opting out of study inclusion.

Identification of cases in the hospital databases.

An algorithm based on combinations of International Classification of Diseases, 10th Revision (ICD-10) diagnosis codes for UC and claims codes for relevant therapies, procedures, and surgery was used to identify cases in the hospital databases that were treated and met the claims-based criteria (Table 1) between January 01, 2012 and March 31, 2018 (Kitasato University Kitasato Institute Hospital) and between October 01, 2010 and March 31, 2018 (Kyushu University Hospital). The cases selected had a primary diagnosis of UC with an inpatient or outpatient claim, with an extended ICD-10 code K51 and Japanese claim code (a full list of the ICD-10 and Japanese claims codes used in the algorithm is shown in Table S1), and had to meet ≥ 1 of the following criteria: prescription for a systemic therapy (tumor necrosis factor inhibitor, immunomodulatory/immunosuppressant drug, corticosteroids, or antidiarrheal) within the same month or ± 1 month (Table S2); procedural code for leukocytapheresis or granulocyte monocyte apheresis within the same month or ± 1 month; or surgery, identified by a diagnosis and claims/receipt code for a postoperative inpatient or outpatient claim within the same month or ± 1 month (Table 1). Cases were excluded from the analysis cohorts if they had a claim for suspected UC as the primary diagnosis (unique to Japanese claims code), a claims diagnosis code for Crohn's disease (CD) and/or Behçet's disease within the same month or ± 1 month, or if the patient opted not to be included in the study.

Table 1 Criteria for claims-based and gold standard definitions of UC

Claims-based definition [†]	Gold standard definition
1) Primary diagnosis of UC in outpatient or inpatient claim ○ Excluding primary diagnosis of suspected UC, definite CD, or Behçet's disease AND 2) Treatment: use of systemic agents [‡] OR 3) Procedure: recorded use of leukocytapheresis or granulocyte monocyte apheresis OR 4) Evidence of prior surgery: patients with total colectomy plus antidiarrheal medication either in inpatient or outpatient claims [§]	1) Physician diagnosis of UC from the medical records 2) Adjudicated cases (definite/probable) based on physician diagnosis from the medical records and information from medical records on: a Core clinical symptoms b Positive colonoscopy 3) Registration in the Japanese Intractable Disease Registry

[†]Patients had to meet criteria 1) and fulfill one or more of criteria 2), 3), or 4) within the same claim month or ± 1 claim months (3-month period due to monthly reporting).

[‡]Immunomodulators/immunosuppressants (sulfasalazine, sulphapyridine, mesalazine, tacrolimus, azathioprine and 6-mercaptopurine, and cyclosporine); TNFi (infliximab, adalimumab, and golimumab); corticosteroids (prednisolone, budesonide, betamethasone, and methylprednisolone); and antidiarrheal (loperamide hydrochloride).

[§]Clinical treatment unique to Japan. Patients with total colectomy are frequently treated postoperatively with antidiarrheal medication (loperamide). CD, Crohn's disease; TNFi, tumor necrosis factor inhibitors; UC, ulcerative colitis.

Data extraction. Data extracted from the claims database included age, gender, date of diagnosis, alternative diagnosis codes, drug codes, daily dose, and procedure codes. Data extracted from the medical records included: physician diagnosis for UC; enrollment in the Japanese Intractable Disease Registry; core symptoms (stool frequency, bloody stools, and abdominal pain); colonoscopy score; total and partial Mayo scores (where available); laboratory tests (hemoglobin, white blood cells, C-reactive protein, erythrocyte sedimentation rate, fecal calprotectin, and cytomegalovirus antigenemia assay); total/partial bowel colectomy; height; weight; and differential diagnosis (CD, amoebic enteritis). Total or partial Mayo scores were not available for all patients due to medical records capturing patients' symptoms as opposed to Mayo scores in some cases. Data were abstracted from the claims database and medical records onto data abstraction forms by trained abstractors, and anonymized data were converted to an electronic database.

Statistical analysis. Demographic and disease characteristics were summarized using descriptive statistics.

Positive predictive value (PPV) for the claims-based algorithm was calculated as the number of cases meeting the claims-based definition that were confirmed by gold standard diagnosis divided by the total number of cases meeting the claims-based definition. The 95% confidence intervals (CIs) for the PPV were calculated using the normal approximation of the binomial distribution. Three definitions of gold standard diagnosis were used separately for the calculation of PPV: (1) physician diagnosis in the medical records, (2) adjudicated cases (definite/probable) based on physician diagnosis and information on core clinical symptoms and colonoscopy from the medical records, and (3) enrollment in the Japanese Intractable Disease Registry (Table 1). PPVs and 95% CIs were also generated using data stratified by gender and age.

It was estimated that a sample of 400 cases would provide a confidence limit of < 10%, assuming a PPV of 85%.

Results

Cases identified in the claims database. A total of 705 cases meeting the claims-based definition for UC were

identified in the hospital databases of Kitasato University Kitasato Institute Hospital (mean age 44 years; 43% female) and 434 from Kyushu University Hospital (mean age 45 years; 49% female) (Table 2). A random sample of 393 cases (200 per site selected; seven cases were excluded during inter-site variability assessment due to patients being admitted for non-UC visits) was used for validation of the claims-based algorithm. Demographics, clinical and laboratory characteristics, and treatments and procedures for the abstracted cases are shown in Table 3.

The mean age (44 years) and proportion of female patients (48%) for the sample of cases was similar to that of the overall cohort of claims-based cases. The time since diagnosis of the patients sampled ranged from < 1 to ≥ 7 years. The majority of patients had used 5-aminosalicylates (5-ASA; 79.4%), and 12.2% had received a tumor necrosis factor inhibitor. The proportion of patients receiving 5-ASA varied between the sites, with a higher proportion of patients receiving 5-ASA at Kitasato University Kitasato Institute Hospital (76% of patients were receiving mesalazine, and 17% of patients were receiving sulfasalazine) compared with Kyushu University Hospital (64% of patients were receiving mesalazine, and 4% of patients were receiving sulfasalazine). Topical treatments are included within the "Immunosuppressants" and "Corticosteroids" categories. Additionally, topical treatment with 5-ASA was included in the treatment codes for 5-ASA.

A small proportion (< 8%) of the abstracted cases had documented UC procedures or surgery in the medical records. Most patients (63%) had bloody stools, 47% had abdominal pain, 24% had ≥ 10 bowel movements/day, and 88% had a positive colonoscopy. Available laboratory data showed that approximately half of patients had low hemoglobin or elevated white blood cells, and around two-thirds had high C-reactive protein or erythrocyte sedimentation rate.

Validation of the claims-based algorithm. Of the total of 393 abstracted cases identified using the final algorithm, 356 had a physician diagnosis in the patient's medical records, and 356 cases were adjudicated as definite/probable UC. Thirty-seven abstracted cases were adjudicated and determined not to be definite/probable UC; that is, the information available suggested that these cases were not UC. The overall PPV was

Table 2 Overall patient disposition

	All	Kitasato University Kitasato Institute Hospital	Kyushu University Hospital
Patients with at least one UC diagnosis code, <i>n</i>	1960	844	1116
Reason for exclusion, <i>n</i>			
1. Diagnosis code for CD	19	10	9
2. Diagnosis code for Behçet's disease	7	4	3
3. Absence of drug use for the treatment of UC, cytapheresis, or postoperative UC	742	125	617
4. Reasons 1 and 2	2	0	2
5. Reasons 1 and 3	30	0	30
6. Reasons 2 and 3	19	0	19
7. Reasons 1, 2, and 3	2	0	2
Analysis population, <i>n</i>	1139	705	434

CD, Crohn's disease; UC, ulcerative colitis.

Table 3 Demographics and disease characteristics of the claims-based cases of UC used for validation of the claims-based algorithm

	Cases meeting claims-based definition of UC (N = 393)
Age, years, mean (SD)	43.8 (17.6)
Female, n (%)	187 (47.6)
Year of diagnosis, n (%)	
2011 [†]	43 (10.9)
2012	56 (14.2)
2013	71 (18.1)
2014	52 (13.2)
2015	61 (15.5)
2016	56 (14.2)
2017	49 (12.5)
2018	5 (1.3)
UC therapies, [‡] n (%)	
5-ASA [§]	312 (79.4)
Mesalazine	278 (70.7)
Sulfasalazine	42 (10.7)
Immunosuppressants [¶]	99 (25.2)
Azathioprine	61 (15.5)
Tacrolimus	28 (7.1)
Corticosteroids ^{**}	57 (14.5)
Prednisolone	49 (12.5)
Antidiarrheal ^{††}	22 (5.6)
TNFi ^{§§}	48 (12.2)
Infliximab	34 (8.7)
UC procedures (cytapheresis), [‡] n (%)	29 (7.4)
Diagnosis for past UC surgery, [‡] n (%)	28 (7.1)
Symptoms, n (%)	
Stool frequency ≥ 10 times/day	76 (24.2) ^{¶¶}
Bloody stool	216 (62.8) ^{¶¶}
Abdominal pain	159 (46.9) ^{¶¶}
Positive colonoscopy	317 (87.6) ^{¶¶}
Abnormal laboratory results, n (%)	
Low Hb ^{¶¶}	220 (57.7) ^{¶¶}
High WBC ^{¶¶}	208 (54.6) ^{¶¶}
High CRP ^{¶¶}	233 (61.6) ^{¶¶}
High ESR ^{¶¶}	156 (66.4) ^{¶¶}

[†]Includes initial diagnosis prior to 2011.

[‡]From medical records.

[§]Mesalazine or sulfasalazine.

[¶]Azathioprine, tacrolimus, mercaptopurine, or cyclosporine.

^{**}Prednisolone, betamethasone, or budesonide.

^{††}Clinical treatment unique to Japan. Patients with total colectomy are frequently treated postoperatively with antidiarrheal medication (loperamide).

^{§§}Infliximab, adalimumab, or golimumab.

^{¶¶}n = 314.

^{¶¶}n = 344.

^{¶¶}n = 339.

^{¶¶}n = 362.

^{¶¶}Low Hb defined as < 14.0–17.0 or < 13.7–16.8 g/dL for male patients, and < 11.0–15.0 or < 11.6–14.8 g/dL for female patients at Kitasato/Kyushu.

^{¶¶}n = 381.

^{¶¶}High WBC defined as > 3.60–8.00 × 10³ or > 3.30–8.60 × 10³ at Kitasato/Kyushu.

^{¶¶}High CRP defined as > ULN 0.30 mg/dL or > ULN 0.14 mg/dL at Kitasato/Kyushu.

^{¶¶}n = 378.

^{¶¶}High ESR defined as > 10.0 mm for male patients and > 15.0 mm for female patients at both hospitals.

^{¶¶}n = 235.

5-ASA, 5-aminosalicylates; CRP, C-reactive protein; ESR, erythrocyte sedimentation rate; Hb, hemoglobin; SD, standard deviation; TNFi, tumor necrosis factor inhibitors; UC, ulcerative colitis; ULN, upper limit of normal; WBC, white blood cells.

91% when determined using either the physician diagnosis from the medical records or the expert clinical diagnosis (adjudicated cases) as the gold standard (Table 4). The PPVs for the individual hospitals were generally consistent. A total of 163 of the UC cases with a physician diagnosis in the medical records were also enrolled in the Japanese Intractable Disease Registry. The PPV determined using this criterion as a gold standard was 42% overall, and 30% and 53% for Kyushu University Hospital and Kitasato University Kitasato Institute Hospital, respectively.

When stratified by gender, the PPV when using either the physician diagnosis from the medical records or the expert clinical diagnosis (adjudicated cases) as the gold standard was 87% and 94% for female and male patients, respectively (Table 5). The PPV determined using the Japanese Intractable Disease Registry as the gold standard was 40% for female patients and 43% for male patients.

When stratified by age, the PPV when using the physician diagnosis from the medical records as the gold standard was 93%, 97%, 91%, and 78% for patients aged < 20, 20–39, 40–59, and ≥ 60 years, respectively (Table 6). When using the expert clinical diagnosis (adjudicated cases) as the gold standard, the PPV was 93%, 96%, 92%, and 78% for patients aged < 20, 20–39, 40–59, and ≥ 60 years, respectively. Lastly, when using the Japanese Intractable Disease Registry as the gold standard, the PPV was 17%, 51%, 45%, and 28% for patients aged < 20, 20–39, 40–59, and ≥ 60 years, respectively.

Discussion

This cross-sectional, retrospective study of hospital claims data in Japan validated a new claims-based definition of UC. The claims-based algorithm, developed with expert input from epidemiologists, gastroenterologists, and internal medicine physicians, was based on a combination of diagnostic codes and claims codes for UC systemic drugs, procedures, or surgery, and identified a cohort of UC cases from hospital administrative databases. The signs and symptoms, and treatment history of the cohort identified were as expected for UC. The PPV of the claims-based UC definition *versus* gold standard diagnoses, based on physician diagnosis from medical records or adjudication by independent expert gastroenterologists, was 91%. When stratified by gender, the PPV remained similar, at 87% for female patients and 94% for male patients. However, when using the Japanese Intractable Disease Registry as the gold standard, the PPV was 42% overall, 40% for female patients, and 43% for male patients. When stratified by age, the PPV for all age categories < 60 years ranged from 91% to 97% when using physician diagnosis from medical records or adjudication by independent expert gastroenterologists as the gold standard. The PPV was 78% in the ≥ 60 years age category when using

Table 4 PPV (95% CI) of a claims-based algorithm for identifying cases of UC confirmed by gold standard diagnoses

Gold standard criteria	All (<i>N</i> = 393) [†]		Kitasato University Kitasato Institute Hospital (<i>N</i> = 200)		Kyushu University Hospital (<i>N</i> = 193) [†]	
	<i>N</i>	PPV (95% CI)	<i>N</i>	PPV (95% CI)	<i>N</i>	PPV (95% CI)
Physician diagnosis [‡]	356	90.6% (87.7–93.5)	186	93.0% (89.5–96.5)	170	88.1% (83.5–92.7)
Adjudicated cases [§]	356	90.6% (87.7–93.5)	187	93.5% (90.1–96.9)	169	87.6% (82.9–92.2)
Enrollment in Japanese Intractable Disease Registry	163	41.5% (36.6–46.3)	105	52.5% (45.6–59.4)	58	30.1% (23.6–36.5)

[†]Seven cases removed during inter-site variability assessment to exclude patients admitted for non-UC visits.

[‡]From medical records.

[§]Based on physician diagnosis from medical records and information on core clinical symptoms and colonoscopy from the medical records. CI, confidence interval; *N*, number of cases meeting gold standard definition; PPV, positive predictive value; UC, ulcerative colitis.

Table 5 PPV (95% CI) of a claims-based algorithm for identifying cases of UC confirmed by gold standard diagnoses, stratified by gender

Gold standard criteria	Female (<i>N</i> = 187)		Male (<i>N</i> = 206)	
	<i>N</i>	PPV (95% CI)	<i>N</i>	PPV (95% CI)
Physician diagnosis [†]	162	86.6% (81.8–91.5)	194	94.2% (91.0–97.4)
Adjudicated cases [‡]	162	86.6% (81.8–91.5)	194	94.2% (91.0–97.4)
Enrollment in Japanese Intractable Disease Registry	75	40.1% (33.1–47.1)	88	42.7% (36.0–49.5)

[†]From medical records.

[‡]Based on physician diagnosis from medical records and information on core clinical symptoms and colonoscopy from the medical records. CI, confidence interval; *N*, number of cases meeting gold standard definition; PPV, positive predictive value; UC, ulcerative colitis.

Table 6 PPV (95% CI) of a claims-based algorithm for identifying cases of UC confirmed by gold standard diagnoses, stratified by age

Gold standard criteria	< 20 years (<i>N</i> = 30)		20–39 years (<i>N</i> = 143)		40–59 years (<i>N</i> = 137)		≥ 60 years (<i>N</i> = 83)	
	<i>N</i>	PPV (95% CI)	<i>N</i>	PPV (95% CI)	<i>N</i>	PPV (95% CI)	<i>N</i>	PPV (95% CI)
Physician diagnosis [†]	28	93.3% (84.4–99.9)	138	96.5% (93.5–99.5)	125	91.2% (86.5–96.0)	65	78.3% (69.4–87.2)
Adjudicated cases [‡]	28	93.3% (84.4–99.9)	137	95.8% (92.5–99.1)	126	92.0% (87.4–96.5)	65	78.3% (69.4–87.2)
Enrollment in Japanese Intractable Disease Registry	5	16.7% (3.3–30.0)	73	51.0% (42.9–59.2)	62	45.3% (36.9–53.6)	23	27.7% (18.1–37.3)

[†]From medical records.

[‡]Based on physician diagnosis from medical records and information on core clinical symptoms and colonoscopy from the medical records. CI, confidence interval; *N*, number of cases meeting gold standard definition; PPV, positive predictive value; UC, ulcerative colitis.

physician diagnosis from medical records or adjudication by independent expert gastroenterologists as the gold standard.

Reported PPVs for other claims-based algorithms for UC or IBD which have been validated in Canada, Europe, Israel, and the United States are comparable with our results.^{16–20} For example, a United States study using algorithms based only on diagnostic codes identified cases of UC in the Veterans Affairs Health Care Systems with PPVs of 66% to 89%.¹⁷ An additional study reported a PPV of 81.4% for distinguishing UC from CD cases in a Canadian health administrative database.¹⁹

Less than half of the claims-based cases identified at the two participating hospitals in our study were registered in the Japanese Intractable Disease Registry. PPV for the algorithm using enrollment in this registry as the gold standard diagnosis was lower than that for the other gold standard criteria (42%), indicating that UC registry enrollment status cannot be used for validation purposes. Case registration in the Japanese Intractable Disease Registry is

likely to be incomplete for a number of reasons. In 2018, there was a change to the enrollment criteria to only accept moderate-to-severe cases of UC or patients with high medical costs, meaning that access to advanced treatment for patients with less severe disease is limited. In addition, the registration process through which physicians apply on behalf of their patients is cumbersome and must be repeated annually. Patient relocation or referral to different medical centers may also have resulted in incomplete registration of cases.

Reporting guidelines for validation studies of claims-based algorithms caution that use of diagnostic codes alone in claims-based algorithms may not achieve adequate accuracy.²¹ A strength of our study was the inclusion of drug codes, as well as claims codes for UC procedures and surgery (postoperative claims), in combination with the diagnostic codes. The algorithms were developed by a group of multidisciplinary experts and were relevant for the Japanese health-care system. Of the patients with

a diagnosis code for UC, a large proportion (821 of 1960 cases) were excluded due to the lack of prescription/procedure codes, although it is still possible that there were patients with UC who did not require medical treatment. This study is based on each institution's Diagnosis Procedure Combination receipt and Medical receipt; therefore, only in-hospital prescription data were included in the model. Use of out-of-hospital prescriptions is common in those patients attending Kyushu University Hospital; therefore, it was possible for patients with UC attending Kyushu University Hospital to be coded without any prescriptions. In addition, we evaluated the use of multiple definitions of gold standard diagnosis for validation of the claims-based criteria.

This study may serve as a model for the construction of population-based cohorts for the study of numerous clinical aspects of UC, such as biologic use during pregnancy in patients with UC and treatment patterns among patients with UC treated with corticosteroids in Japan. The use of administrative data for such purposes is still in its early stages in Japan; however, several national databases, such as the Japan Medical Data Center and Medical Data Vision, are now available to researchers. Recently, two studies in Japan reported to have used the Japan Medical Data Center database to investigate rates of persistence with biologic UC therapies¹³ and explore the prescription pattern and economic burden²² in patients with UC identified using a combination of diagnostic codes and prescription data. A further study has investigated the use of ICD-10 codes to identify patients with UC and CD in the Medical Data Vision database in order to determine the incidence of nonmelanoma skin cancers and non-Hodgkin lymphoma, and their relationship with use of IBD treatments.¹⁴ This study may serve as a model for further validation studies in Japan, in accordance with Pharmaceuticals and Medical Devices Agency recommendations for studies to support the credibility of claims database research for the purposes of post-marketing surveillance.²³ Validated claims-based algorithms will be useful for analysis of real-world drug use, treatment persistence, and clinical course (e.g. rates of adverse events of special interest, hospitalization, and surgery) in UC and other inflammatory diseases in Japan. The validity of disease-identifying algorithms based on combinations of diagnostic codes and drug prescriptions has also been investigated using hospital claims data in The Validity of Algorithms in Large Databases: Infectious Diseases, Rheumatoid Arthritis, and Tumor Evaluation in Japan (VALIDATE-J) studies.^{24,25}

Limitations of this study include the use of only two participating centers, which were not randomly selected and may not be representative of the wider Japanese population. Nonetheless, up to 60% of UC patients are seen by specialists at tertiary care centers such as the two hospitals involved in this study, and it is likely that our claims-based algorithm will identify most UC cases. Moreover, comparison of the patients with UC registered in the clinical databases at each site during the study period with the number of patients extracted using the algorithm revealed that almost all patients might be extracted (personal communication Haruei Ogino and Taku Kobayashi). Additionally, the participating centers differ in their use of in-hospital *versus* out-of-hospital prescription systems, which may have contributed to the disparities in the numbers of patients recorded as receiving UC treatment between the centers. A further limitation of this study is that patients treated only with topical therapies were not extracted by the algorithm.

The limited number of extracted patients who were postoperative is also a limitation. Further validation studies for UC involving different types of hospitals are needed to assess/confirm the generalizability of our results. In addition, negative predictive values and sensitivity/specificity were not estimated in the study.

In conclusion, the PPV of the claims-based UC definition using physician diagnosis and medical record-based gold standards was 91%, suggesting that this claims-based algorithm is likely to identify UC cases with high PPV for clinical studies including Japanese post-marketing studies using administrative databases.

Acknowledgments

Medical writing support, under the guidance of the authors, was provided by Helen Findlow, PhD, CMC Connect, McCann Health Medical Communications, and Kirsteen Munn, PhD, on behalf of CMC Connect, and was funded by Pfizer Japan Inc, Tokyo, Japan in accordance with Good Publication Practice (GPP3) guidelines (Ann Intern Med 2015; 163: 461–464). The authors thank Dr Takehiro Torisu (Department of Medical and Clinical Science, Graduate School of Medical Sciences, Kyushu University) for the use of the data generated within his department.

Data availability statement

The datasets generated and/or analyzed during the current study are available from the corresponding author on reasonable request.

References

- Ordás I, Eckmann L, Talamini M, Baumgart D, Sandborn W. Ulcerative colitis. *Lancet* 2012; **380**: 1606–19.
- Rubin DT, Ananthakrishnan AN, Siegel CA, Sauer BG, Long MD. ACG clinical guideline: ulcerative colitis in adults. *Am. J. Gastroenterol.* 2019; **114**: 384–413.
- Ghosh S, Mitchell R. Impact of inflammatory bowel disease on quality of life: results of the European Federation of Crohn's and Ulcerative Colitis Associations (EFCCA) patient survey. *J. Crohns Colitis* 2007; **1**: 10–20.
- Aniwan S, Bruining DH, Park SH *et al.* The combination of patient-reported clinical symptoms and an endoscopic score correlates well with health-related quality of life in patients with ulcerative colitis. *J. Clin. Med.* 2019; **8**: 1171.
- Fumery M, Singh S, Dulai PS, Gower-Rousseau C, Peyrin-Biroulet L, Sandborn WJ. Natural history of adult ulcerative colitis in population-based cohorts: a systematic review. *Clin. Gastroenterol. Hepatol.* 2018; **16**: 343–56.e3.
- Fuehrstein JD, Isaacs KL, Schneider Y, Siddique SM, Falck-Ytter Y, Singh S. AGA Clinical Practice Guidelines on the management of moderate to severe ulcerative colitis. *Gastroenterology* 2020; **158**: 1450–61.
- Asakura K, Nishiwaki Y, Inoue N, Hibi T, Watanabe M, Takebayashi T. Prevalence of ulcerative colitis and Crohn's disease in Japan. *J. Gastroenterol.* 2009; **44**: 659–65.
- Murakami Y, Nishiwaki Y, Oba MS *et al.* Estimated prevalence of ulcerative colitis and Crohn's disease in Japan in 2014: an analysis of a nationwide survey. *J. Gastroenterol.* 2019; **54**: 1070–7.
- Kanatani Y, Tomita N, Sato Y, Eto A, Omoe H, Mizushima H. National registry of designated intractable diseases in Japan: present status and future prospects. *Neurol. Med. Chir. [Tokyo]* 2017; **57**: 1–7.

- 10 Japan Intractable Disease Information Center. 2020. Available from URL: <https://www.nanbyou.or.jp/>
- 11 Matsuoka K, Kobayashi T, Ueno F *et al.* Evidence-based clinical practice guidelines for inflammatory bowel disease. *J. Gastroenterol.* 2018; **53**: 305–53.
- 12 Koram N, Delgado M, Stark JH, Setoguchi S, de Luise C. Validation studies of claims data in the Asia-Pacific region: a comprehensive review. *Pharmacoepidemiol. Drug Saf.* 2019; **28**: 156–70.
- 13 Kobayashi T, Udagawa E, Uda A, Hibi T, Hisamatsu T. Impact of immunomodulator use on treatment persistence in patients with ulcerative colitis: a claims database analysis. *J. Gastroenterol. Hepatol.* 2020; **35**: 225–32.
- 14 Kobayashi T, Uda A, Udagawa E, Hibi T. Lack of increased risk of lymphoma by thiopurines or biologics in Japanese patients with inflammatory bowel disease: a large-scale administrative database analysis. *J. Crohns Colitis* 2020; **14**: 617–23.
- 15 Matsuoka K, Igarashi A, Sato N *et al.* Trends in corticosteroid prescriptions for ulcerative colitis and factors associated with long-term corticosteroid use: analysis using Japanese claims data from 2006 to 2016. *J. Crohns Colitis* 2021; **15**: 358–66.
- 16 Friedman MY, Leventer-Roberts M, Rosenblum J *et al.* Development and validation of novel algorithms to identify patients with inflammatory bowel diseases in Israel: an epi-IIRN group study. *Clin. Epidemiol.* 2018; **10**: 671–81.
- 17 Hou JK, Tan M, Stidham RW *et al.* Accuracy of diagnostic codes for identifying patients with ulcerative colitis and Crohn's disease in the Veterans Affairs Health Care System. *Dig. Dis. Sci.* 2014; **59**: 2406–10.
- 18 Rezaie A, Quan H, Fedorak RN, Panaccione R, Hilsden RJ. Development and validation of an administrative case definition for inflammatory bowel diseases. *Can. J. Gastroenterol.* 2012; **26**: 711–7.
- 19 Benchimol EI, Guttman A, Mack DR *et al.* Validation of international algorithms to identify adults with inflammatory bowel disease in health administrative data from Ontario, Canada. *J. Clin. Epidemiol.* 2014; **67**: 887–96.
- 20 Jakobsson GL, Sternegård E, Olén O *et al.* Validating inflammatory bowel disease (IBD) in the Swedish National Patient Register and the Swedish Quality Register for IBD (SWIBREG). *Scand. J. Gastroenterol.* 2017; **52**: 216–21.
- 21 Benchimol EI, Manuel DG, To T, Griffiths AM, Rabeneck L, Guttman A. Development and use of reporting guidelines for assessing the quality of validation studies of health administrative data. *J. Clin. Epidemiol.* 2011; **64**: 821–9.
- 22 Miyazaki C, Sakashita T, Jung W, Kato S. Real-world prescription pattern and healthcare cost among patients with ulcerative colitis in Japan: a retrospective claims data analysis. *Adv. Ther.* 2021; **38**: 2229–47.
- 23 Ando T. Recent trend on utilization of real world data—challenges in Japan. 2018. Available from URL: <https://www.pmda.go.jp/files/000226214.pdf>
- 24 Setoguchi S, Sugiyama N, Chen H *et al.* VALIDATE J: a validation study of rheumatoid arthritis in Japanese claims data. 2018. Presented at: *Asian Conference on Pharmacoepidemiology*. October 27–29, 2018. Xi'an, China.
- 25 de Luise C, Sugiyama N, Morishima T *et al.* Validity of claims-based algorithms for selected cancers in Japan: results from the VALIDATE-J study. *Pharmacoepidemiol. Drug Saf.* 2021; **30**: 1153–61.

Supporting information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Table S1. Extended ICD-10, procedural, and surgical codes, and corresponding Japanese claim codes, for claims-based definition of ulcerative colitis.

Table S2. Japanese claim codes for drugs used in claims-based definition.