Received: 31 January 2023 | Revised: 3 August 2023 | Accepted: 13 September 2023

DOI: 10.1002/pds.5702

ORIGINAL ARTICLE

WILEY

Use of clinical classifications software to address ICD coding transition in large healthcare databases analyzed via high-dimensional propensity scores

Lai San Hong[1] | Xabier Garcia-Albeniz[2] | David Friesen[3] | Nadia Foskett[4] | Raphaelle Beau-Lejdstrom[5]

[1]Redsen Limited, Bournemouth, UK

[2]Pharmacoepidemiology and Risk Management, RTI Health Solutions, Barcelona, Spain

[3]Parasol Limited, Warrington, UK

[4]UCB Pharma, Brussels, Belgium

[5]UCB Pharma, Bulle, Switzerland

Correspondence
Lai San Hong, Suite 1 Second Floor Everdene House, Deansleigh Road, Bournemouth BH7 7DU, UK.
Email: corinna.hong@gmail.com

Funding information
UCB Pharma

Abstract

Purpose: The EUPAS26595 study characterized the rate of acute renal failure (ARF) in patients exposed to levetiracetam versus other antiepileptic drugs using healthcare claims data and a high-dimensional propensity score (hd-PS) for confounding adjustment. The data contained several coding systems by design and an update in *International Classification of Diseases* (ICD) coding dictionary. Such coding heterogeneity can affect the performance of hd-PS, and manually coding harmonization is not feasible. Our objective was to explore the impact of code aggregation via Clinical Classifications Software (CCS) on the analysis of a large claims-based database using hd-PS.

Methods: Patients with epilepsy, who were new-users of an antiepileptic drug, were identified from the IBM® MarketScan® Research Databases. We used CCS categories to harmonize coding and compared the results with other alternatives. Incidence rate ratios (IRRs) were computed using modified Poisson regression model with a robust variance estimator.

Results: For January 2008–October 2015 (before ICD update), 34 833 eligible patients initiated levetiracetam and 52 649 initiated a comparator drug; IRR (95% CI) for ARF for the hd-PS analysis was 1.34 (0.72–2.50) without CCS categories and 1.30 (0.71–2.39) with CCS categories. For January 2008–December 2017 (including ICD coding change), 45 672 eligible patients initiated levetiracetam and 64 664 initiated a comparator drug; IRR (95% CI) for the hd-PS analysis was 1.34 (0.78–2.29) without CCS categories and 1.37 (0.80–2.34) with CCS categories.

Conclusions: Using single-level CCS categories to overcome differences in coding provides consistent results and can be used in studies that use large claims data and hd-PS for adjustment.

KEYWORDS
aggregation, claims data, clinical classification software, epidemiology, epilepsy, observational studies, propensity score

Key Points

- Real-world evidence studies using large databases of medical records often face issues regarding multiple coding dictionaries or change of coding system.
- Aggregation of codes through the choice of data granularity is a balance between introducing measurement error and improving confounding control.
- We found the level of adjustment by using CCS diagnosis/procedure categories to feed the hd-PS algorithm was congruent with that by other approaches of aggregation in the current study.
- The proposed aggregation approach can be used in studies that use large, population-based claims data and hd-PS or other automated algorithms for baseline adjustment; and can help solve problems related to multiplicity of coding dictionaries and computing constraints.

Plain Language Summary

Real-world evidence studies using large databases of medical records often face issues regarding multiple coding dictionaries for diagnoses, procedures, and drugs. In 2015, the coding system for US electronic healthcare transactions was changed for both diagnosis and procedure codes from International Classification of Diseases (ICD) 9th to 10th revision. Consistency in coding is key to accurately identify covariates and adjust confounding for any quality pharmacoepidemiological study. We addressed these challenges of coding multiplicity during implementation of EUPAS26595, a voluntary postauthorization safety study that found no evidence of an association of levetiracetam on the risk of acute renal failure (ARF) compared with other antiepileptic drugs. This study used the IBM® MarketScan® Research Databases (2008–2017) and adjusted for baseline confounding using high-dimensional propensity score (hd-PS) method, a data-driven automated algorithm. We used Clinical Classifications Software (CCS) categories to harmonize multiple coding systems and compared the results with other alternatives. Incidence rate ratios (IRRs) were computed using modified Poisson regression model with a robust variance estimator. Results demonstrated that the aggregation of codes using single-level CCS categories provided consistent results, highlighting the utility of this method in studies that use large, population-based claims data and hd-PS or other automated algorithms for baseline adjustment.

1 | INTRODUCTION

Postauthorization safety studies are often tasked with estimating effects in a real-world setting, evaluating rare or long-term events, and studying subpopulations underrepresented in registration trials; thus, they often rely on secondary use of routinely collected patient data. One such data source, healthcare claims, comprehensively captures billable medical transactions between patients and their healthcare providers.[1]

An automated data-driven baseline confounding adjustment algorithm that takes advantage of this wealth of information is the high-dimensional propensity score (hd-PS) method.[2] Briefly, the hd-PS algorithm creates indicators that correlate with potential unmeasured confounders to adjust for baseline confounding using analyst-supplied files containing all available baseline records of different data dimensions (e.g., inpatient/outpatient services, medications). It identifies the *n* most prevalent codes (*n* determined by the analyst) in each data dimension and creates indicators that reflect their recurrence. For each recurrence indicator of the *n* most prevalent codes in each data dimension, the possible amount of confounding each indicator could adjust for is computed using a multiplicative model given a binary exposure and outcome. Then, the *k* indicators (*k* defined by the analyst) that explain the most confounding according to this approach are selected and, with demographic and other variables predefined by the analyst, used to estimate the conditional probability of exposure, that is, the propensity score. As this algorithm stems from the prevalence of every code in the baseline records, 3 factors should be considered when constructing the data dimensions.

- First, aggregation of codes through the choice of data granularity (e.g., the number of characters selected of the diagnoses/procedure codes). Grouping codes by reading fewer characters will increase the prevalence of that code and may introduce measurement error (as granularity is lost), whereas reading more characters can result in lower prevalence within the same hierarchical classification and thus less chances to be chosen by the algorithm. Such aggregation can also affect the covariate-exposure and/or covariate-outcome associations, which also influence the probability of being selected by the hd-PS algorithm. Code aggregation may improve confounding control in settings with a rare outcome or rare exposure.[3]
- Second, the use of multiple coding dictionaries. An assumption in the hd-PS algorithm is that codes are mutually exclusive within each data dimension. If a database contains multiple coding dictionaries

for the same data element, the indicators created by the hd-PS algorithm may be subject to measurement error. For example, procedures are coded using 3 data dictionaries in the IBM® MarketScan® Research Databases— *International Classification of Diseases* (ICD), Healthcare Common Procedure Coding System (HCPCS), and *Current Procedural Terminology, 4th Edition* (CPT-4).
- Finally, changing the coding dictionary during the study period. Because the same event will receive a different code depending on the date, its prevalence will be lower than if a single coding dictionary were used. The coding system for US electronic healthcare transactions was changed for both diagnosis and procedure codes on October 1, 2015, from *ICD, Ninth Revision, Clinical Modification* (ICD-9-CM) to the *Tenth Revision* (ICD-10-CM) and *Procedure Coding System* (ICD-10-PCS). Therefore, the same clinical event received different codes depending on which ICD-era it occurred. Although it is possible to map curated code lists for a limited number of predefined variables, doing so is not feasible at the scale needed for the hd-PS algorithm.

We faced these challenges during implementation of EUPAS26595,[4] a voluntary category 3 postauthorization safety study that found no evidence of an association of levetiracetam on the risk of acute renal failure (ARF) compared with other antiepileptic drugs.[4] This study used the IBM® MarketScan® Research Databases (2008–2017) and adjusted for baseline confounding using hd-PS.

We used Clinical Classifications Software (CCS), a categorization scheme used to harmonize diagnoses and procedures in diverse sources of data,[5–10] to overcome the challenges introduced by the presence of multiple data dictionaries. In this article, we describe this approach and compare it with other analytic alternatives.

2 | METHODS

2.1 | EUPAS26595 overall study design

EUPAS26595 was a cohort study comparing the incidence of ARF associated with the initiation of levetiracetam versus the initiation of other antiepileptic drugs among epileptic patients who either had been antiepileptic drug free for the last year or were currently receiving another baseline antiepileptic drug to which an additional drug (levetiracetam or a comparator) was added.[4] Patients were eligible if they had a new prescription of levetiracetam or a comparator antiepileptic drug (index date), were diagnosed with epilepsy, and had continuous medical and prescription coverage in the year before the index date. Patients were excluded if they initiated 2 antiepileptic drugs concomitantly, if they had preexisting renal disease in the year before index date, rhabdomyolysis or status epilepticus in the 90 days before index date, or a major surgical procedure or use of a radiocontrast agent in the 30 days before index date (see Figure, Supplemental Digital Content 1 for cohort design).

The primary outcome was ARF, defined as the presence of a code for a compatible diagnosis in the primary diagnosis field of at least 1 inpatient or 1 emergency department visit claim in the 30 days after index date.[4] Patients were followed from index date to the first diagnosis of ARF, 30 days after index date, discontinuance of insurance coverage, or end of study period, whichever occurred first. For the current exercise, we evaluated the primary outcome on the cohort of patients who were antiepileptic treatment free in the year before index date.

2.2 | Study data

We used the IBM® MarketScan® Commercial Database, IBM® MarketScan® Medicare Supplemental Database, and IBM® MarketScan® Multi-State Medicaid Database covering January 2008–December 2017. The databases capture information on medical (inpatient, outpatient, and emergency care healthcare encounters, including date; ICD-9-CM, ICD-10-CM, ICD-10-PCS, HCPCS, CPT-4 codes), and pharmacy claims (dispensed medications, including National Drug Code, dispensation date, quantity, and days supplied), plus enrollment history (age, sex, US census region, health insurance payer type, and monthly enrollment status). During our study period, ICD coding changed from 9th to 10th revision coding in October 2015.

2.3 | CCS aggregation approach

The CCS is a diagnosis and procedure categorization scheme for ICD-9-CM, ICD-10-CM/PCS, and CPT-4/HCPCS codes, developed as part of the Healthcare Cost and Utilization Project, sponsored by the Agency for Healthcare Research and Quality. CCS aggregates thousands of diagnosis and procedure codes into fewer, clinically meaningful categories. We used the following CCS categories to group records in data dimensions from different data-coding dictionaries:

- ICD-9-CM diagnoses: We used single-level CCS categories to enable matching of the level of classification system for ICD-9-CM procedures. We used the last version (2015) before the change to ICD-10-CM.[11]
- ICD-9-CM procedures: We used single-level CCS categories to match the only CCS classification system developed for CPT-4 and HCPCS procedures.[11]
- CPT-4 and HCPCS procedures: We used CCS–Services and Procedures 2019.1 version.[12]
- ICD-10-CM diagnoses: We used single-level CCS 2019.1 (beta) version,[13] which has the same set of CCS categories as for ICD-9-CM.
- ICD-10-PCS procedures: We used single-level CCS 2019.1 (beta) version,[14] which has the same set of CCS categories as for ICD-9-CM, CPT-4, and HCPCS.

2.4 | Statistical analysis

Because the ICD coding change from 9th to 10th revision may have impacted characterization of CCS categories themselves, we modelled

each CCS category prevalence by linear regression using monthly prevalence as dependent variable and calendar month plus a binary indicator for dates after October 2015 as independent variables. This model assumed a linear relationship between monthly prevalence and calendar month that could have a step change in October 2015 (see Figure, Supplemental Digital Content 2 for an example). We described the absolute and relative change in the model-predicted prevalence of all CCS categories using standard descriptive statistics and histograms.

The main study outcome, ARF incidence rate, was calculated as the number of new ARF diagnoses divided by the sum of person-months during follow-up period. We computed incidence rate ratios (IRRs) and incidence rate differences of ARF between patients initiating levetiracetam versus those initiating other antiepileptic drugs using a modified Poisson regression model with robust variance estimator.[15]

We adjusted for baseline confounding by weighting the outcome model by the inverse of propensity score. We stabilized the weights with marginal probability of the observed exposure and truncated them at the first and 99th percentiles to avoid undue influence of outliers. We estimated the propensity score using multivariate logistic regression and different sets of covariates depending on the analysis (see below). When we used the hd-PS algorithm, we input subject-matter-knowledge variables (see Table 2) as predefined and exclude those with any category present in <2% or >98% of the participants.[16] We let the hd-PS algorithm construct empirical indicators for adjustment from the following 5 data dimensions: Inpatient/Outpatient diagnoses, Inpatient/Outpatient procedures, and Outpatient pharmacy dispensing. We ran the hd-PS with the following parameters: *n* (number of most prevalent indicators to consider from each dimension of data) = 200, "Bross bias formula" as the ranking method, and k (number of indicators to include in the final propensity score) = 500. Additionally, indicators were created by the hd-PS algorithm to reflect the intensity of healthcare services utilization within each dimension.

For the current exercise, we compared the following analyses to evaluate the added value, if any, of harmonizing clinical information via CCS categories:

1. Unadjusted analysis
2. Analysis based on subject-matter-knowledge: using only predefined variables
3. Analysis using hd-PS without CCS: in addition to the variables in analysis 2, data dimensions were specified as input to the hd-PS algorithm as follows:
   - Inpatient/outpatient diagnoses: first 3 characters of ICD-9-CM diagnoses or ICD-10-CM
   - Inpatient/outpatient procedures: CPT-4, HCPCS, first 3 characters of ICD-9-CM procedures or ICD-10-PCS
   - Outpatient pharmacy dispensing: generic drug name from IBM® RED BOOK®
4. Analysis using hd-PS with CCS: in addition to the variables in analysis 2, data dimensions were specified as input to the hd-PS algorithm as follows
   - Inpatient/outpatient diagnoses: single-level CCS diagnosis categories
   - Inpatient/outpatient procedures: single-level CCS procedure categories
   - Outpatient pharmacy dispensing: generic drug name from IBM® RED BOOK®

These analyses were run both for the full study period and for the period ending at the time of ICD dictionary change, October 2015.

3 | RESULTS

Between 2008 and 2017, there were 110 336 eligible patients, of whom 45 672 initiated levetiracetam and 64 664 initiated a comparator drug (Table 1). Between 2008 and October 2015, 34 833 eligible patients initiated levetiracetam and 52 649 initiated a comparator drug.

Patients in the levetiracetam group were less likely to be women, but there were no relevant differences in other demographic characteristics or in the presence of other conditions. Patients in the levetiracetam group were also less likely to use analgesics, antidepressants, or antipsychotics. There were relevant differences in epilepsy-related health services, with patients in the levetiracetam group more frequently being admitted to hospital or visiting emergency department than patients in the comparator, who more frequently had outpatient visits (Table 1). The proportion of patients censored before reaching the 30-day risk period was 2.2% in both groups. Overall, 39 cases of ARF occurred in the levetiracetam group and 27 in the comparator group.

3.1 | Continuity of clinical codes using CCS

The median absolute change of the CCS category prevalence between the periods before and after October 2015 was −0.18 cases/100000 individuals (minimum, −729.10; Q1, −5.54; Q3, 5.22; maximum, 1241.17 cases/100000 individuals). On a relative scale, the median change was −0.4% (minimum, −98.7%; Q1, −5.6%; Q3, 5.8%; maximum, 209.4%) (Figure 1).

3.2 | Comparison of methods using CCS versus source-coding dictionaries

For January 2008–October 2015 (i.e., before ICD-10-CM/PCS coding was implemented), the IRR for ARF (95% confidence interval [CI]) was 2.02 (1.14–3.55) for the unadjusted analysis, 1.45 (0.79–2.64) for the analysis based on subject-matter-knowledge, and 1.34 (0.72–2.5) for the analysis using hd-PS without CCS (Table 2). In this analysis, the hd-PS algorithm created 217 indicators from ICD-9-CM diagnoses, 163 from CPT-4/HCPCS procedures, 82 from medications, 16 from ICD-9-CM procedures, 10 and 12 indicators of healthcare services utilization based on diagnosis and procedure codes, respectively (Table 3). The analysis using hd-PS with CCS yielded an IRR (95% CI) of 1.3 (0.71–2.39) (Table 2). In this analysis, the hd-PS algorithm created 113 indicators based on medications, 244 and 143 based on CCS

TABLE 1 Baseline characteristics of eligible participants, IBM MarketScan Databases 2008–2017.

Characteristic	Levetiracetam N = 45 672	Comparator N = 64 664
Female, *n* (%)	22 675 (49.6)	34 728 (53.7)
Age, mean (SD), years	30.3 (25.76)	29.7 (21.72)
Database, *n* (%)		
CCAE/MDCR	29 487 (64.6)	40 763 (63.0)
MDCD	16 185 (35.4)	23 901 (37.0)
Region of employee residence, *n* (%)		
North central	7227 (15.8)	9527 (14.7)
Northeast	5777 (12.6)	7186 (11.1)
South	10 925 (23.9)	16 543 (25.6)
West	5228 (11.4)	6956 (10.8)
Unknown	16 515 (36.2)	24 452 (37.8)
Conditions present in the 12 months before index date, *n* (%)		
Any	5724 (12.5)	7551 (11.7)
Burns	72 (0.2)	116 (0.2)
Cardiovascular disease	1049 (2.3)	1048 (1.6)
Cerebrovascular disease	475 (1.0)	399 (0.6)
Chronic obstructive pulmonary disease	798 (1.7)	1139 (1.8)
Diabetes mellitus	931 (2.0)	1240 (1.9)
Diabetic nephropathy	30 (0.1)	19 (0.0)
Heart failure	432 (0.9)	320 (0.5)
Hemolysis	1 (0.0)	1 (0.0)
Hypertension	276 (0.6)	352 (0.5)
Hypotension	171 (0.4)	215 (0.3)
Hypovolemia	274 (0.6)	317 (0.5)
Liver disease	220 (0.5)	283 (0.4)
Myoglobinuria	1 (0.0)	0 (0.0)
Obesity	144 (0.3)	337 (0.5)
Other renal diseases	1380 (3.0)	2212 (3.4)
Peripheral vascular disease	368 (0.8)	404 (0.6)
Proteinuria	12 (0.0)	27 (0.0)
Renal calculi	154 (0.3)	348 (0.5)
Sepsis	422 (0.9)	373 (0.6)
Small kidney of unknown cause	0 (0.0)	0 (0.0)
Systemic lupus erythematosus	66 (0.1)	110 (0.2)
Non-antiepileptic medications used in the 12 months before index date, *n* (%)		
Any	38 949 (85.3)	55 324 (85.6)
Analgesics	13 886 (30.4)	24 563 (38.0)
Antidepressants	6926 (15.2)	15 652 (24.2)
Antimicrobials and antiviral drugs	21 842 (47.8)	31 232 (48.3)
Antineoplastics	127 (0.3)	145 (0.2)
Antipsychotics	1446 (3.2)	4692 (7.3)
Benzodiazepines	6710 (14.7)	9504 (14.7)
Cardiovascular drugs	7553 (16.5)	9173 (14.2)
Immunosuppressants	214 (0.5)	302 (0.5)
Methotrexate	165 (0.4)	189 (0.3)
Other drugs	5994 (13.1)	9027 (14.0)
Steroids	9724 (21.3)	14 219 (22.0)
Epilepsy-related health services utilization in the 12 months before index date, *n* (%)		
Overall outpatient visits	25 299 (55.4)	39 868 (61.7)
Hospitalizations	6503 (14.2)	4726 (7.3)
ED visits	19 972 (43.7)	18 704 (28.9)
Health services utilization in the 12 months before index date not related to epilepsy, *n* (%)		
Overall outpatient visits	43 142 (94.5)	61 504 (95.1)
Hospitalizations	9812 (21.5)	13 060 (20.2)
ED visits	19 965 (43.7)	29 220 (45.2)
Index treatment		
Levetiracetam	45 672 (100)	0 (0.0)
Carbamazepine	0	3914 (6.1)
Ethosuximide	0	3498 (5.4)
Gabapentin	0	8208 (12.7)
Lacosamide	0	676 (1.0)
Lamotrigine	0	9478 (14.7)
Oxcarbazepine	0	9005 (13.9)
Pregabalin	0	962 (1.5)
Phenobarbital	0	1106 (1.7)
Phenytoin	0	7425 (11.5)
Topiramate	0	9287 (14.4)
Valproic acid	0	9442 (14.6)
Zonisamide	0	1663 (2.6)

Abbreviations: CCAE/MDCR, IBM MarketScan® Commercial Database/ IBM MarketScan® Medicare Supplemental Database; ED, emergency department; MDCD, IBM MarketScan® Multi-State Medicaid Database; SD, standard deviation.

diagnosis and procedure categories, respectively (Table 3). For all analyses, the mean of truncated stabilized weights was 0.99 (see Table, Supplemental Digital Content 3 for weight distribution).

For January 2008–December 2017 (i.e., when data up to October 2015 period were coded with ICD-9-CM and later data were coded with ICD-10-CM/PCS), the IRR (95% CI) was 2.05 (1.25–3.34) for the unadjusted analysis, 1.55 (0.92–2.61) for the analysis based on

TABLE 2 Results for the different analyses by study period.

	Study period: January 2008 to October 2015		Study period: January 2008 to December 2017	
	Incidence rate difference[a], cases per 10 000 (95% CI)	Incidence rate ratio[a] (95% CI)	Incidence rate difference[a], cases per 10 000 (95% CI)	Incidence rate ratio[a] (95% CI)
Unadjusted analysis	4.09 (0.63 to 7.56)	2.02 (1.14 to 3.55)	4.41 (1.27 to 7.56)	2.05 (1.25 to 3.34)
Analysis based on subject matter knowledge[b]	2.10 (−1.23 to 5.44)	1.45 (0.79 to 2.64)	2.68 (−0.4 to 5.75)	1.55 (0.92 to 2.61)
Analysis using hd-PS without CCS[c]	1.70 (−1.68 to 5.07)	1.34 (0.72 to 2.5)	1.80 (−1.28 to 4.88)	1.34 (0.78 to 2.29)
Analysis using hd-PS with CCS[d]	1.58 (−1.84 to 4.99)	1.30 (0.71 to 2.39)	2.00 (−1.12 to 5.12)	1.37 (0.8 to 2.34)

Abbreviations: CCS, clinical classifications software; CI, confidence interval; CPT, current procedural terminology, 4th edition; HCPCS, healthcare common procedure coding system; hd-PS, high-dimensional propensity score; ICD-10, international classification of diseases, tenth revision; ICD-9, international classification of diseases, ninth revision.

[a]Incidence rate differences and incidence rate ratios of ARF between patients initiating levetiracetam versus those initiating other antiepileptic drugs.

[b]Analysis adjusted for the following predefined variables: age, sex, region, year of baseline, health plan type, database type, comorbidities (burns, cardiovascular disease, cerebrovascular disease, chronic obstructive pulmonary disease, diabetes mellitus, diabetic nephropathy, heart failure, hemolysis, hypertension, hypotension, hypovolemia, liver disease, myoglobinuria, obesity, other renal diseases, peripheral vascular disease, proteinuria, renal calculi, sepsis, small kidney of unknown cause, and systemic lupus erythematosus), co-medications (categorized as analgesics, antidepressants, antimicrobials and antivirals, antineoplastics, antipsychotics, benzodiazepines, cardiovascular drugs, immunosuppressants, steroids, and others), and healthcare utilization variables (the number of health encounters as inpatients, outpatients, or emergency department classified as related or not to epilepsy care).

[c]Analysis adjusted for predefined variables and for indicators created by the hd-PS algorithm using ICD-9, ICD-10 diagnosis and procedure category codes, CPT-4/HCPCS codes, and medications.

[d]Analysis adjusted for predefined variables and for indicators created by the hd-PS algorithm using CCS diagnosis, CCS procedure categories, and medications.

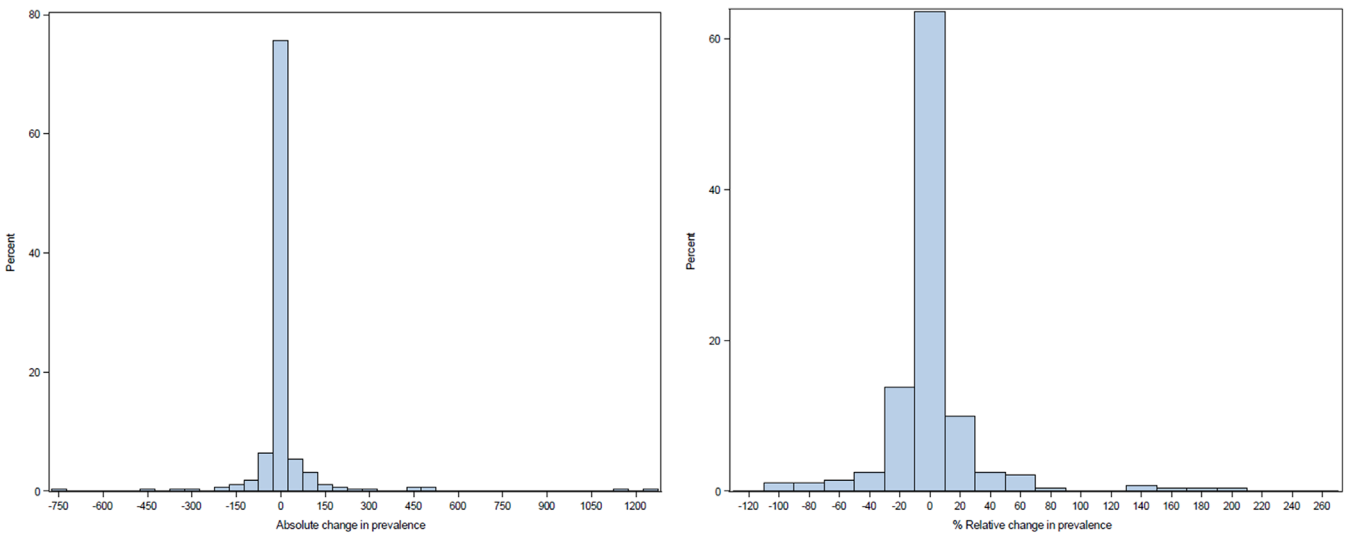

FIGURE 1 Absolute and relative changes in the predicted prevalence in October 2015 of the clinical classifications software categories.

subject-matter-knowledge, and 1.34 (0.78–2.29) for the analysis using hd-PS without CCS (Table 2). The hd-PS algorithm (without CCS) created 173 indicators from ICD-9-CM diagnoses, 169 from CPT-4/HCPCS procedures, 77 from medications, 46 from ICD-10-CM diagnosis codes, 1 from ICD-10-PCS procedure codes, 12 and 10 indicators of healthcare services utilization based on diagnosis and procedure codes, respectively (Table 3). The analysis using hd-PS with CCS yielded an IRR (95% CI) of 1.37 (0.80–2.34) (Table 2), and the hd-PS algorithm created 110 indicators based on medications, 257 and 133 based on CCS diagnosis and procedure categories, respectively (Table 3). For all analyses, the mean of truncated stabilized weights was 0.99 (see Table, Supplemental Digital Content 3).

4 | DISCUSSION

We estimated the difference in the incidence of ARF for initiators of levetiracetam compared with initiators of a comparator antiepileptic

TABLE 3 Covariates used for adjustment in the different analyses by study period.

Variable, *n* (%)	Study period: January 2008 to October 2015			Study period: January 2008 to December 2017		
	Analysis based on subject matter knowledge ($n = 81$)[a]	Analysis using hd-PS without CCS ($n = 581$)[b]	Analysis using hd-PS with CCS ($n = 581$)[c]	Analysis based on subject matter knowledge ($n = 81$)[a]	Analysis using hd-PS without CCS ($n = 581$)[b]	Analysis using hd-PS with CCS ($n = 581$)[c]
Predefined variables	81 (100)	81 (13.9)	81 (13.9)	81 (100)	81 (13.9)	81 (13.9)
ICD-9-CM diagnosis	0	217 (37.3)	0	0	173 (29.8)	0
ICD-10-CM diagnosis	0	0	0	0	46 (7.9)	0
CCS diagnosis	0	0	244 (42.0)	0	0	257 (44.2)
Diagnosis codes–based health services utilization indicators	0	10 (1.7)	0	0	12 (2.1)	0
ICD-9-CM procedure	0	16 (2.8)	0	0	12 (2.1)	0
ICD-10-PCS procedure	0	0	0	0	1 (0.2)	0
CPT-4/HCPCS procedure	0	163 (28.1)	0	0	169 (29.1)	0
CCS procedure	0	0	143 (24.6)	0	0	133 (22.9)
Procedure codes–based health services utilization indicators	0	12 (2.1)	0	0	10 (1.7)	0
Medications	0	82 (14.1)	113 (19.4)	0	77 (13.3)	110 (18.9)

Abbreviations: CCS, clinical classifications software; CPT, current procedural terminology-4th edition; HCPCS, healthcare common procedure coding system; hd-PS, high-dimensional propensity score; ICD, international classification of diseases; ICD-10-CM, international classification of diseases, tenth revision, clinical modification; ICD-9-CM, international classification of diseases, ninth revision, clinical modification.

[a]Analysis adjusted for the following predefined variables: age, sex, region, year of baseline, health plan type, database type, comorbidities (burns, cardiovascular disease, cerebrovascular disease, chronic obstructive pulmonary disease, diabetes mellitus, diabetic nephropathy, heart failure, hemolysis, hypertension, hypotension, hypovolemia, liver disease, myoglobinuria, obesity, other renal diseases, peripheral vascular disease, proteinuria, renal calculi, sepsis, small kidney of unknown cause, and systemic lupus erythematosus), co-medications (categorized as analgesics, antidepressants, antimicrobials and antivirals, antineoplastics, antipsychotics, benzodiazepines, cardiovascular drugs, immunosuppressants, steroids, and others), and healthcare utilization variables (the number of health encounters as inpatients, outpatients, or emergency department classified as related or not to epilepsy care).

[b]Analysis adjusted for predefined variables and for indicators created by the hd-PS algorithm using ICD-9, ICD-10 diagnosis and procedure category codes, CPT-4/HCPCS codes and medications.

[c]Analysis adjusted for predefined variables and for indicators created by the hd-PS algorithm using CCS diagnosis, CCS procedure categories, and medications.

drug in IBM® MarketScan® claims database using different analytical approaches to address the multiplicity of coding dictionaries. The results of this empirical example suggest that the use of CCS categories as a harmonization procedure to aggregate diagnostic and procedure codes yields consistent results in the adjustment of baseline variables using hd-PS and can be useful when facing challenge such as a change in the coding dictionary occurs during the study period or when a data source uses multiple coding dictionaries.

The alternative CCS Refined (CCSR) aggregation tool[17] was not used because it would not have allowed having the same clinical concept from both ICD coding dictionaries under the same CCS category (each ICD-10-CM code can be mapped to multiple CCSR categories). Nevertheless, this comes with some loss of granularity for ICD-10-CM codes. We show in our study that, in the ICD-9 era, the use of CCS diagnosis/procedure categories to feed the hd-PS algorithm (together with predefined variables and medications) provided a level of adjustment that was congruent with the level of adjustment provided by the use of 3-character ICD diagnosis/procedure codes, and CPT-4/HCPCS procedures (together with predefined variables and medications as well). Of note, the analysis using CCS categories gave less weight to procedures than the nonaggregated approach and did not create indicators based on healthcare services utilization. A study that used the same data source reported that using CCS diagnostic categories to aggregate ICD-9-CM diagnosis codes worked well using the hd-PS algorithm under different scenarios of sample size, exposure prevalence, and outcome incidence.[3]

By analyzing the change in the prevalence of CCS categories, we show that, in our study population and corresponding study period, single-level CCS aggregation was not perfect in bridging the change of ICD coding that occurred in October 2015: although the median change of the prevalence, both in relative and absolute level, was close to zero and most of the CCS categories had small changes, there were CCS categories with relevant variations. The alternative to the CCS diagnosis/procedure category aggregation in using the hd-PS adjustment algorithm and the full study period was to feed the hd-PS-algorithm with ICD-9-CM and ICD-10-CM/PCS codes. Because the hd-PS algorithm prioritizes codes by their prevalence, and our study period contained >7 years of data using ICD-9-CM coding and slightly >2 years of ICD-10-CM/PCS coding, the latter have an artificially low prevalence, and individuals for whom their empirical indicators were constructed using data after October 2015 would rarely contribute indicators based on ICD-10-CM/PCS coding (i.e., confounder

misclassification). In our analysis that used unaggregated ICD coding from the full study period, hd-PS indicators based on ICD-10-CM diagnosis were only 8% of all the indicators and only 1 indicator (0.2%) was created using ICD-10-PCS procedures. Despite this likely confounder misclassification, this analysis performed better than adjusting only for predefined variables and yielded a result that was consistent with the analysis using CCS harmonization. These findings are specific to the current study; a different scenario where ICD-9-CM and ICD-10-CM/PCS coding dictionaries across the study period follow a different distribution may yield different results, and the consistency of effect measures between the two distinct ICD-eras as study periods may be explored.

Our study has limitations. We studied levetiracetam and ARF incidence as a case study, and our conclusions might not be generalizable to other exposure-outcome relationships. Our interpretations assume that the true association of levetiracetam on the incidence of ARF is null, or close to null, as previously reported.[18] This assumption is supported by the fact that, with higher levels of adjustment, the point estimates moved toward the null value. In this study, the hd-PS added value to the adjustments based only on predefined variables, as has been observed in other studies.[19] Nevertheless, one limitation of the hd-PS is that it may create indicators that correspond to variables that act as colliders[20] (yielding biased effect estimates) or are not causally related to the outcome (effect estimates may have unnecessarily wide 95% CIs). Additionally, our study evaluated the incidence of ARF in a short time period (30 days), and censoring during follow-up was negligible; therefore, it could be safely assumed that time-varying confounding[21] was not an issue. Nevertheless, in studies with longer follow-up where nonrandom losses to follow-up happen or when the treatment effect under complete adherence is being estimated, researchers will need to adjust for potential time-varying confounding, and the hd-PS can only be used to adjust for baseline variables.

In conclusion, we show that the aggregation of ICD and CPT-4/HCPCS codes using single-level CCS categories provides consistent results; can be used in studies that use large, population-based claims data and hd-PS or other automated algorithms for baseline adjustment; and can help solve problems related to multiplicity of coding dictionaries and computing constraints. Further scenarios with different exposure-outcome relationships and different prevalence of exposure and outcome need to be explored.

AUTHOR CONTRIBUTIONS

Lai San Hong contributed to the design of the study and the interpretation of the data, analyzed the data, revised the manuscript, and approved the final version to be published. Xabier Garcia-Albeniz contributed to the analysis and interpretation of the data, drafted the manuscript, and approved the final version to be published. David Friesen contributed to the analysis of the data and approved the final version of the manuscript to be published. Raphaelle Beau-Lejdstrom contributed to the design and the analysis, interpreted the data, and revised and approved the final version of the manuscript to be published. Nadia Foskett contributed to the design of the study, revised, and approved the final version to be published.

ACKNOWLEDGMENTS

This study was funded by UCB Pharma. Nada Boudiaf (Covance Laboratories Limited, Harrogate, UK) and Kathrin Haeffs (UCB Pharma, Monheim am Rhein, Germany) contributed to the validation of the analysis. The authors acknowledge Kathleen Richards, PhD (UCB Pharma, Smyrna, GA, USA) for publication coordination.

FUNDING INFORMATION

Work on the manuscript was funded by UCB Pharma through a contract with RTI Health Solutions.

CONFLICT OF INTEREST STATEMENT

Lai San Hong was an independent contractor for UCB Pharma. Xabier Garcia-Albeniz is an employee of RTI Health Solutions. David Friesen was an independent contractor for UCB Pharma. Nadia Foskett was an employee of UCB Pharma at the time of the analysis and is currently affiliated with Ferring SAS Gentilly, France. Raphaelle Beau-Lejdstrom is an employee of UCB Pharma.

DATA AVAILABILITY STATEMENT

The dataset (IBM® MarketScan® Research Databases) generated during and/or analyzed during the current study are not publicly available but are available from IBM® on reasonable request.

ETHICS STATEMENT

Not applicable.

ORCID

Lai San Hong https://orcid.org/0000-0001-7846-580X
Xabier Garcia-Albeniz https://orcid.org/0000-0002-9814-2343
Raphaelle Beau-Lejdstrom https://orcid.org/0000-0003-0877-0109

REFERENCES

1. 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.
2. Schneeweiss S, Rassen JA, Glynn RJ, Avorn J, Mogun H, Brookhart MA. High-dimensional propensity score adjustment in studies of treatment effects using health care claims data. *Epidemiology*. 2009;20(4):512-522.
3. Le HV, Poole C, Brookhart MA, et al. Effects of aggregation of drug and diagnostic codes on the performance of the high-dimensional propensity score algorithm: an empirical example. *BMC Med Res Methodol*. 2013;13(1):142.
4. Beau-Lejdstrom R, Hong LS, Garcia de Albeniz X, et al. Incidence of acute renal failure in patients using levetiracetam versus other antiseizure medications: a voluntary post-authorization safety study. *Drug Saf*. 2022;45(7):781-790.
5. Blackburn J, Locher JL, Kilgore ML. Comparison of long-term care in nursing homes versus home health: costs and outcomes in Alabama. *Gerontologist*. 2016;56(2):215-221.
6. Goudie A, Dynan L, Brady PW, Fieldston E, Brilli RJ, Walsh KE. Costs of venous thromboembolism, catheter-associated urinary tract infection, and pressure ulcer. *Pediatrics*. 2015;136(3):432-439.
7. Goudie A, Dynan L, Brady PW, Rettiganti M. Attributable cost and length of stay for central line-associated bloodstream infections. *Pediatrics*. 2014;133(6):e1525-e1532.

8. Shewale AR, Barnes CL, Fischbach LA, Ounpraseuth ST, Painter JT, Martin BC. Comparative effectiveness of intra-articular hyaluronic acid and corticosteroid injections on the time to surgical knee procedures. *J Arthroplast*. 2017;32(12):e24.
9. Adams DJ, Susi A, Nylund CM. Clinical characteristics, risk factors, and outcomes of patients hospitalized in the US military health system with carbapenem-resistant Enterobacteriaceae infection. *Am J Infect Control*. 2020;48(6):644-649.
10. Patel H, Calip GS, DiDomenico RJ, Schumock GT, Suda KJ, Lee TA. Comparison of cardiac events associated with azithromycin vs amoxicillin. *JAMA Netw Open*. 2020;3(9):e2016864.
11. Agency for Healthcare Research and Quality. Clinical Classifications Software (CCS) for ICD-9-CM. Available at: https://www.hcup-us.ahrq.gov/toolssoftware/ccs/ccs.jsp. Accessed Jan 06, 2023
12. Agency for Healthcare Research and Quality. CCS-Services and Procedures Software, v2019.1. Available at: https://www.hcup-us.ahrq.gov/toolssoftware/ccs_svcsproc/ccscpt_downloading.jsp. Accessed Jan 06, 2023
13. Agency for Healthcare Research and Quality. CCS for ICD-10-CM diagnoses, v2019.1 (October 2018) (beta version). Available at: https://hcup-us.ahrq.gov/toolssoftware/ccsr/ccsr_archive.jsp#ccsbeta. Accessed Jan 06, 2023
14. Agency for Healthcare Research and Quality. CCS for ICD-10-PCS procedures, v2019.1 (beta version) (October 2018). Available at: https://www.hcup-us.ahrq.gov/toolssoftware/ccs10/ccs10.jsp. Accessed Jan 06, 2023
15. Zou G. A modified Poisson regression approach to prospective studies with binary data. *Am J Epidemiol*. 2004;159(7):702-706.
16. Girman CJ, Gokhale M, Kou TD, Brodovicz KG, Wyss R, Stürmer T. Assessing the impact of propensity score estimation and implementation on covariate balance and confounding control within and across important subgroups in comparative effectiveness research. *Med Care*. 2014;52(3):280-287.
17. Agency for Healthcare Research and Quality. Clinical Classifications Software Refined (CCSR). Available at: https://www.hcup-us.ahrq.gov/toolssoftware/ccsr/ccs_refined.jsp. Accessed Jan 06, 2023
18. Yau K, Burneo JG, Jandoc R, et al. Population-based study of risk of AKI with levetiracetam. *Clin J Am Soc Nephrol*. 2019;14(1):17-26.
19. Guertin JR, Rahme E, LeLorier J. Performance of the high-dimensional propensity score in adjusting for unmeasured confounders. *Eur J Clin Pharmacol*. 2016;72(12):1497-1505.
20. Hernán MA, Hernández-Díaz S, Robins JM. A structural approach to selection bias. *Epidemiology*. 2004;15(5):615-625.
21. Hernán MA, Robins JM. Time-varying treatments. *Causal Inference: What if*. Chapman & Hall/CRC; 2020:235-246.

SUPPORTING INFORMATION

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

How to cite this article: Hong LS, Garcia-Albeniz X, Friesen D, Foskett N, Beau-Lejdstrom R. Use of clinical classifications software to address ICD coding transition in large healthcare databases analyzed via high-dimensional propensity scores. *Pharmacoepidemiol Drug Saf*. 2023;1-9. doi:10.1002/pds.5702