Received: 14 December 2022 | Revised: 18 July 2023 | Accepted: 16 August 2023

DOI: 10.1002/pds.5690

ORIGINAL ARTICLE

WILEY

Validation of an ICD-10 case-finding algorithm for endometrial cancer in US insurance claims

Djeneba Audrey Djibo[1] | Andrea V. Margulis[2] | Cheryl N. McMahill-Walraven[1] | Catherine W. Saltus[3] | Patricia Shuminski[1] | James A. Kaye[3] | Catherine B. Johannes[3] | Mark Libertin[4] | Shelli Graham[5]

[1]Safety, Surveillance & Collaboration, CVS Health, Blue Bell, Pennsylvania, USA

[2]Epidemiology, RTI Health Solutions, Barcelona, Spain

[3]Epidemiology, RTI Health Solutions, Waltham, Massachusetts, USA

[4]Medical Policy Operations, Aetna, CVS Health, Cleveland, Ohio, USA

[5]TherapeuticsMD, Boca Raton, Florida, USA

Correspondence
Djeneba Audrey Djibo, Safety, Surveillance & Collaboration, CVS Health, Blue Bell, PA, USA.
Email: audrey.djibo@cvshealth.com

Funding information
TherapeuticsMD

Abstract

Purpose: To evaluate the positive predictive value (PPV) of an endometrial cancer case finding algorithm using International Classification of Disease 10th revision Clinical Modification (ICD-10-CM) diagnosis codes from US insurance claims for implementation in a planned post-marketing safety study. Two algorithm variants were evaluated.

Methods: Provisional incident endometrial cancer cases were identified from 2016 through 2020 among women aged ≥50 years. One algorithm variant used diagnosis codes for malignant neoplasms of uterine sites (C54.x), excluding C54.2 (malignant neoplasm of myometrium); the other used only C54.1 (malignant neoplasm of endometrium). A random sample of medical records of recent incident provisional cases (2018–2020) was requested for adjudication. Confirmed cases showed biopsy evidence of endometrial cancer, documentation of cancer staging, or hysterectomy following diagnosis. We estimated the PPV of the variants with 95% confidence intervals (CI) excluding cases that had insufficient information.

Results: Of 294 provisional cases adjudicated, 85% were from outpatient settings ($n = 249$). Mean age at diagnosis was 69.3 years. Among the 294 adjudicated cases (identified with the broader algorithm variant), the same 223 were confirmed endometrial cancer cases by both algorithm variants. The PPV (95% CI) for the broader algorithm variant was 84.2% (79.2% and 88.3%), and for the variant using only C54.1 was 85.8% (80.9% and 89.8%).

Conclusion: We developed and validated an algorithm using ICD-10-CM diagnosis codes to identify endometrial cancer cases in health insurance claims with a sufficiently high PPV to use in a planned post-marketing safety study.

KEYWORDS
cancer, endometrial, medical records, PPV, validity

Key Points

- Validating rare outcomes, such as endometrial cancer, in administrative claims is important for generating high quality real-world evidence
- Two algorithm variants were assessed; both required one inpatient or two outpatient encounters; the broader variant used several codes in the C54 ICD-10-CM chapter (malignant neoplasms of uterine sites) and the narrower variant used only the C54.1 (malignant neoplasm of endometrium) code
- Both algorithm variants performed well to identify endometrial cancer among women ages 50 years and older
- The narrower variant that included only the C54.1 ICD-10-CM diagnosis code identified all confirmed cases and had fewer false positive cases than the broader variant

Plain Language Summary

Our goal was to validate an algorithm, a relevant set of criteria and diagnosis codes, to identify patients with endometrial cancer from insurance claims collected at healthcare visits. Because the claims are collected primarily for billing and reporting purposes, the accuracy of the clinical information may not be uniform and must be verified. The information surrounding the cancer diagnoses available in medical records is considered the gold standard; therefore, the algorithm's validity was reviewed against that gold standard. A group of clinicians reviewed 294 medical records, and 223 endometrial cancer cases were confirmed. Overall, the performance of the algorithm was high. We concluded that the algorithm could be used to identify endometrial cancer cases confidently in future studies such as drug safety studies.

1 | INTRODUCTION

Using real-world data sources, such as insurance claims to conduct health outcomes studies has become a valued practice to determine the occurrence of rare outcomes.[1] However, administrative claims are collected for financial and programmatic rather than medical purposes. Therefore, the accuracy of diagnoses listed in claims needs to be validated to increase confidence in the inferences generated from such data.[2] Using validated algorithms for case identification has been considered acceptable within the United States Food and Drug Administration's Sentinel Initiative, a nationwide collaborative program evaluating the safety of drugs, vaccines, and other medical products.[1,3] By using a prespecified, validated algorithm, researchers can improve the quality of information inferred from observational data.[4]

Endometrial cancer, the most common form of uterine cancer, represented 3.4% of all new cancer cases in 2022 in the United States.[5] Women have a 3.1% chance of being diagnosed with endometrial cancer over their lifetime, most often between the ages of 55 and 64.[6] Despite the incidence of endometrial cancer, only one previous validation study of such cases in real world data from US insurance claims has been published, and it used *International Classification of Diseases, 9th Revision, Clinical Modification* (ICD-9-CM) codes.[7] To date, no validation of the currently standard ICD-10-CM diagnosis codes for endometrial cancer in US-based insurance claims has been published. Since ICD-9-CM codes for uterine cancer distinguished anatomic sites (body, cervix, and other sites) while ICD-10-CM codes can also specify which layer of the uterus is affected, direct mapping of these coding systems is not possible.

The primary purpose of this study was to quantify the validity of two variants of an ICD-10-CM endometrial cancer case-finding algorithm in a US administrative claims database from a single health insurance payer for implementation in a planned post-marketing safety study.

2 | METHODS

2.1 | Study design

Through a retrospective study, we validated electronically identified endometrial cancer cases via medical record adjudication to determine the positive predictive values of two variants of a case-finding algorithm.

2.2 | Data source

Provisional cases were retrospectively identified in two U.S. electronic administrative claims databases from CVS Health: the Aetna Enterprise Data Warehouse (AEDW), which contained the most recent healthcare utilization as reflected by administrative health insurance claims, and the Aetna Sentinel Common Data Model (SCDM), which contained longitudinal follow up since 2008 and updated on a quarterly basis,[8] together referred to as "the database" thereafter. The linked database contained information on any type of healthcare utilization, physician diagnoses, dispensed prescriptions, inpatient and outpatient diagnoses, treatments and procedures, and medical provider contact information.

2.3 | Study population

The study population was comprised of female patients aged 50 years or older with both medical and pharmacy benefits. Those with a history of hysterectomy, endometrial ablation, and/or endometrial cancer before the study period (2016–2020) were excluded using all available database information (which started in 2008).

2.4 | Case-finding algorithm

Based on the results of the published ICD-9-CM validation study,[7] the algorithm we used to identify provisional cases required one inpatient or two outpatient encounters with an ICD-10-CM code of interest and continuous enrollment in a health plan for at least 12 months before the first endometrial cancer diagnosis date. The ICD-10-CM codes were identified using ICD-9-CM to ICD-10-CM forward-backward mapping[9] based on the code correspondence provided by CMS.[10] The definitions were refined using clinical expertise. Code C54.1 was chosen as the stand-alone code for the narrow algorithm because it is the code that specifies malignant neoplasm of the endometrium (typically endometrial carcinoma). The date of diagnosis was the earliest of the date of the inpatient code or the first outpatient code, as appropriate. The ICD-10-CM codes of interest were C54 (malignant neoplasm of corpus uteri), C54.0 (malignant neoplasm of isthmus uteri), C54.1 (malignant neoplasm endometrium), C54.3 (malignant neoplasm fundus uteri), C54.8 (malignant neoplasm of overlapping sites of corpus uteri), and C54.9 (malignant neoplasm of corpus uteri, unspecified). We investigated the performance of two algorithm variants based on these criteria: algorithm A (the broader algorithm) included all codes previously listed, while algorithm B (the narrower algorithm) was a subset of algorithm A using only C54.1 as the identifying diagnosis code.

2.5 | Adjudication process

A random sample of provisional cases with diagnosis dates from July 2018 to December 2020 was selected for medical records requests, with a target of 300 medical records to obtain a narrow 95% confidence interval around the positive predictive value (PPV). This time period for record requests was selected to capture recent clinical and provider details and improve medical-record retrieval likelihood as compared with earlier years. With this sample size, assuming most of the 300 medical records are informative, and 250 of them are adjudicated as confirmed cases, the study would estimate a PPV around 83.3% (78.6%–87.4%).

To choose the appropriate encounter (i.e., a visit at a medical practice or hospital with a provider with an oncology or gynecology specialty) for which to request the medical record that was considered likely to contain clinical information to confirm the diagnosis of endometrial cancer, we created a patient-provider profile. The profile contained the patient identifiers, encounter date range of interest, provider name, address, and contact information. Medical record requests were sent via fax with a follow-up via phone to the providers identified in the patient-provider profiles, and included a request for pathology reports, inpatient admission/history and physical examination, inpatient progress notes, discharge summary, outpatient reports, outpatient progress notes surgery reports for inpatient encounters, laboratory reports, cytology reports, procedure reports, clinical notes, and referrals to or from oncology specialists.

Each medical record received was reviewed and adjudicated by two independent registered nurses with cancer expertise and experience in direct patient care using a structured electronic abstraction form (Supplemental Figure S1) to classify each provisional case as confirmed, probable, non-case, or case remaining provisional (when there was insufficient information for determination). The case definitions were developed with the following considerations: first, the research team anticipated good algorithm performance based on the published article by Esposito et al.[7] Second, because the research team anticipated that some requested records would not be available, and that records obtained might not contain sufficient information to confirm all cases definitively (such as a pathology report from a biopsy or hysterectomy specimen), we intended to retain such cases for sensitivity analyses, unless there was other clear evidence of an alternative diagnosis. Discordant case status findings were reviewed and adjudicated by a third reviewer, a physician with an obstetrics and gynecology specialty (ML). Evidence of one of the following was required to confirm cases: (1) biopsy or surgical pathology report indicating endometrial cancer, (2) radical hysterectomy with salpingectomy and oophorectomy within 30 days before or after the date identified by the electronic algorithm, (3) any hysterectomy followed within 60 days by systemic cytotoxic chemotherapy, or (4) recorded endometrial cancer staging.[11] Provisional cases were classified as probable if only evidence of vaginal brachytherapy or external pelvic radiation was present. To be classified as non-cases, provisional cases had to have evidence of a different diagnosis such as endometrial hyperplasia or another malignancy. Because the exposure of interest for the planned safety study is a vaginal estrogen-containing product, information on use of vaginal estrogen products was removed from medical records prior to case review by adjudicators.

2.6 | Analyses

Descriptive characteristics of the provisional cases and adjudication results were obtained from the database starting from 2008. Other characteristics of interest were age at diagnosis, obesity, type 1 or type 2 diabetes mellitus (DM), endometrial hyperplasia, ever use of estrogen hormonal replacement therapy, and ever use of selective estrogen receptor modifiers (SERMs). We estimated the PPV of the case-finding algorithms with exact 95% confidence intervals (CI), excluding from the numerator and the denominator any cases that remained provisional. As sensitivity analyses, we conducted additional PPV calculations: first, including in the denominator cases that remained provisional (PPV_2), and secondly by including in the numerator and denominator the probable cases (PPV_3). Clinical characterization of the endometrial cancer cases described cancer type (I or II),

cancer grade as low (grade 1 or 2) or high (grade 3), estrogen receptor status, and cancer stage at diagnosis per FIGO classification.[11]

3 | RESULTS

Of the 3143 provisional endometrial cancer cases identified electronically with broad algorithm variant A, medical records were requested for 444 randomly selected women and received for 303 women (68%), see Figure 1.

After exclusions, 294 adjudicated cases comprised the final study sample. Twelve received medical records were excluded due to not describing an incident endometrial cancer diagnosis ($n = 7$), prior hysterectomy not captured in claims ($n = 1$), duplicate records ($n = 3$), or not having a study diagnosis code ($n = 1$). Among these 294 provisional cases with adjudicated case status from algorithm variant A, the mean age at diagnosis was 69.3 (standard deviation: 9.4) years; 49.0% ($n = 144$) were obese, 42.5% ($n = 125$) had type 1 or type 2 diabetes mellitus, and 27.6% ($n = 81$) had endometrial hyperplasia (Table 1). Most women (90.1%) were not using estrogens or SERMs. The characteristics of adjudicated cases ($n = 288$) from algorithm variant B (restricted to diagnosis code C54.1) were overall similar to those from algorithm variant A (Table 1). Moreover, the characteristics of adjudicated cases with either variant were similar to the random sample selected for medical record retrieval (Table 1).

The results of the adjudication included 223 confirmed cases and 40 non-cases using algorithm variant A; and 223 confirmed cases and 35 non-cases from algorithm variant B. The confirmed cases identified using the two algorithm variants were identical. The crude PPV (95% CI) was 84.2% (79.2%–88.3%) for algorithm variant A, and 85.8% (80.9% and 89.8%) for algorithm variant B, respectively (Table 2). For both algorithm variants, the PPVs were highest among

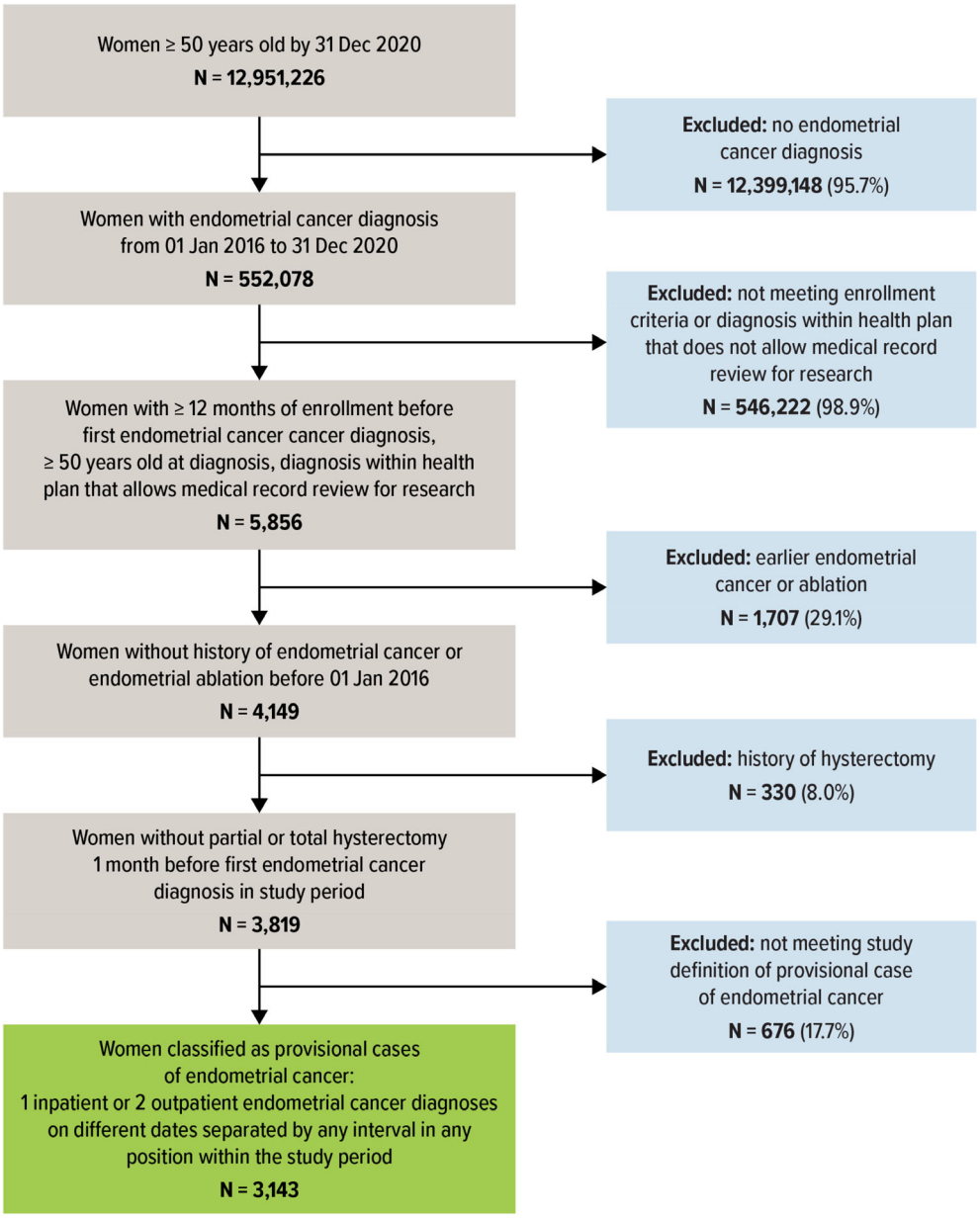

FIGURE 1 Identification of provisional endometrial cancer cases.

TABLE 1 Characteristics of women at date of endometrial cancer diagnosis among randomly selected provisional cases.

Characteristics	Algorithm variant A *n* (%)	Algorithm variant B *n* (%)	Provisional cases *n* (%)
N	294	288	443[a]
Age, years; mean, (SD)	69.3 (9.4)	69.3 (9.5)	69.0 (9.1)
50 to <60 years	53 (18.0%)	53 (18.4%)	78 (17.6%)
60 to <70 years	100 (34.0%)	98 (34.0%)	153 (34.5%)
70 to <80 years	101 (34.4%)	97 (33.7%)	160 (36.1%)
80 years or older	40 (13.6%)	40 (13.9%)	52 (11.7%)
Obesity	144 (49.0%)	144 (50.0%)	220 (49.7%)
Diabetes mellitus (type 1 or type 2)	125 (42.5%)	123 (42.7%)	183 (41.3%)
Endometrial hyperplasia	81 (27.6%)	81 (28.1%)	121 (27.3%)
Estrogen use	21 (7.1%)	21 (7.3%)	26 (5.9%)
No use of estrogen or SERMs	265 (90.1%)	259 (89.9%)	403 (91.0%)

Note: Percentages are column percentages (the denominator is the N in the same column).
Abbreviations: SD, standard deviation; SERMs, selective estrogen receptor modulators.
[a]One record was excluded from the random sample of provisional cases as it had been identified using a nonstudy diagnosis code for endometrial cancer.

TABLE 2 Characteristics of confirmed cases and positive predictive values of algorithm by variant.

	Algorithm variant A		Algorithm Variant B	
Characteristics	Confirmed cases (*n*)	PPV (95% CI)	Confirmed cases (*n*)	PPV (95% CI)
All	223	84.2% (79.2%–88.3%)	223	85.8% (80.9%–89.8%)
Age category (years)				
50 to <60	45	90.0% (78.2%–96.7%)	45	90.0% (78.2%–96.7%)
60 to <70	76	84.4% (75.3%–91.2%)	76	86.4% (77.4%–92.8%)
70 to <80	78	82.1% (72.9%–89.2%)	78	84.8% (75.8%–92.4%)
80 or older	24	80.0% (61.4%–92.3%)	24	80.0% (61.4%–92.3%)
Obesity	120	90.9% (84.7%–95.2%)	120	90.9% (84.7%–95.2%)
Diabetes mellitus	96	85.7% (77.8%–91.6%)	96	87.3% (79.6%–92.9%)
Endometrial hyperplasia	73	96.1% (88.8%–99.2%)	73	96.0% (88.9%–99.2%)
Estrogen use	15	75.0% (50.9%–91.3%)	15	75.0% (50.9%–91.3%)
No use of Estrogen or SERMs	202	84.9% (79.7%–89.2%)	202	86.7% (81.6%–90.8%)

Abbreviations: CI, confidence interval; PPV, positive predictive value; SERMs, selective estrogen receptor modifiers.

the 50- to 60-year-old age group. Among the cases identified with algorithm variant B, the PPVs were 90.9% (95% CI, 84.7%–95.2%) among obese patients, and 96.0% (95% CI, 88.9%–99.2%) among those with endometrial hyperplasia.

The characteristics of the 223 confirmed endometrial cancer cases are presented in Table 3. Information on endometrial cancer type was available for 96.9% of the 223 confirmed cases; most confirmed endometrial cancer cases (78.5%, $n = 175$) were of type I (endometrioid adenocarcinoma), and were found to be low grade (71.7%, $n = 160$). Information on estrogen receptor status was not available for 79.4% ($n = 177$). Using the FIGO surgical staging, 57.0% ($n = 127$) were classified as stage I (i.e., tumor confined to the corpus uteri).

Results of the sensitivity analyses of the PPVs are presented in Table 4. By including in the denominator cases that remained provisional (PPV_2), the positive predictive value was reduced to 75.9% (70.5%–80.6%) and 77.4% (72.2%–82.1%) for algorithm variants A and B respectively. When probable cases were included in the numerator (as well as in the denominator), the PPV (PPV_3) was 84.9% (80.0%–89.0%) for the broad algorithm variant A, and 86.5% (81.8%–90.4%) for the narrow variant B.

4 | DISCUSSION

This validation study of endometrial cancer cases evaluated 2 variants of an ICD-10-CM based algorithm within a national US electronic claims data source. The algorithm required at least 1 inpatient or 2 outpatient claims with different service dates, with ICD-10-CM

TABLE 3 Selected pathology characteristics of confirmed endometrial cancer cases.

Characteristic	Description	Confirmed cases algorithm variant A and variant B [*n* (%)]
Number of confirmed cases		223
Endometrial cancer type[a]		
Type I	Endometrioid adenocarcinomas May arise from complex atypical hyperplasia and are pathogenetically linked to unopposed estrogenic stimulation	175 (78.5%)
Type II	Characterized by clear cell and papillary serous tumor histologies Develops from atrophic endometrium and is not linked to hormonally driven pathogenesis	41 (18.4%)
Cancer grade		
Low (G1 or G2)	Well or moderately differentiated	160 (71.7%)
High (G3)	Poorly or undifferentiated	47 (21.1%)
Unknown	No information available	16 (7.2%)
Estrogen receptor status[b]		
ER+		39 (17.5%)
Unknown		177 (79.4%)
FIGO surgical staging		
I	Tumor confined to the corpus uteri	127 (57.0%)
II	Tumor invades cervical stroma but does not extend beyond the uterus. Endocervical glandular involvement only should be considered as stage I and no longer as stage II.	35 (15.7%)
III or IV	Local and/or regional spread of the tumor or tumor invades bladder and/or bowel mucosa, and/or distant metastases	23 (10.3%)
Unknown FIGO staging	The medical record does not contain information to classify this case in any of the categories above	37 (16.6%)

Note: Percentages are column percentages. The same medical records were confirmed as cases for variant A and variant B.
Abbreviations: ER+, estrogen receptor–positive; FIGO, Fédération Internationale de Gynécologie et d'Obstétrique.
[a]Excluding unknown cancer type.
[b]Excluding non ER+.

TABLE 4 Sensitivity analyses of positive predictive values of algorithm by variant.

	Algorithm variant A		Algorithm variant B	
Characteristics	PPV_2 (95% CI)	PPV_3 (95% CI)	PPV_2 (95% CI)	PPV_3 (95% CI)
All	75.9% (70.5%–80.6%)	84.9% (79.2%–88.3%)	77.4% (72.2%–82.1%)	86.5% (81.8%–90.4%)
Age category (years)				
50 to <60	84.9% (72.4%–93.3%)	90.0% (78.2%–96.7%)	84.9% (72.4%–93.3%)	90.0% (78.2%–96.7%)
60 to <70	76.0% (66.4%–84.0%)	85.6% (76.6%–92.1%)	77.6% (68.0%–85.4%)	87.5% (78.7%–93.6%)
70 to <80	77.2% (67.8%–85.0%)	82.1% (72.9%–89.2%)	80.4% (71.1%–87.8%)	84.8% (75.8%–91.4%)
80 or older	60.0% (43.3%–75.1%)	83.3% (65.3%–94.4%)	60.0% (43.3%–75.1%)	83.3% (65.3%–94.4%)
Obesity	83.3% (76.2%–89.0%)	91.7% (85.6%–95.8%)	83.3% (76.2%–89.0%)	91.7% (85.6%–95.8%)
Diabetes mellitus	76.8% (68.4%–83.9%)	87.5% (79.9%–93.0%)	78.0% (69.7%–85.0%)	89.1% (81.7%–94.2%)
Endometrial hyperplasia	90.1% (81.5%–95.6%)	96.1% (88.9%–99.2%)	90.1% (81.5%–95.6%)	96.1% (88.9%–99.2%)
Estrogen use	71.4% (47.8%–88.7%)	75.0% (50.9%–91.3%)	71.4% (47.8%–88.7%)	75.0% (50.9%–91.3%)
No use of estrogen or SERMs	76.2% (70.6%–81.2%)	85.7% (80.6%–89.9%)	78.0% (72.4%–82.9%)	87.6% (82.6%–91.5%)

Note: PPV_2 accounted for cases that remained provisional in the denominator, PPV_3 included probable cases along with confirmed cases in the numerator.
Abbreviations: CI, confidence interval; PPV, positive predictive value; SERMs, selective estrogen receptor modifiers.

codes C54, C54.0, C54.1, C54.8, and C54.9 (variant A) or C54.1 (variant B). We found that both variants performed adequately, with observed PPV of 84.2% (95% CI, 79.2%–88.3%) for variant A, and 85.8% (95% CI, 80.9%–89.8%) for variant B. Although both variants of the algorithm performed adequately (PPVs higher than the prespecified 80% threshold for the point estimate), the narrower algorithm (variant B) identified the same cases as the broader one (variant A) and variant B also had fewer false positives; therefore, variant B is the preferred algorithm for the planned safety study.

To our knowledge this is the first investigation of the validity of identifying endometrial cancer in US health insurance data using ICD-10-CM diagnosis codes. ICD-10 diagnosis codes to identify recurrent endometrial cancer have been validated in electronic health data in the national Danish health registries,[12] however recurrent endometrial cancer is a different outcome. There are many programmatic differences between a national health registry and insurance claims, including differences in underlying health care systems and in methods of capturing and coding the relevant data.[13,14] For all these reasons, results on recurrent endometrial cancer validation from health registries in Denmark are likely not applicable to newly diagnosed cases in US health insurance claims data. The algorithm evaluated in the present study was designed to identify endometrial cancer cases among women aged 50 years or older in comparable US commercial healthcare claims data sources that use ICD-10-CM codes. Use of the algorithm in other data sources or populations may require an assessment of the prevalence of endometrial cancer and other characteristics of the target population. These parameters may affect the magnitude of the PPV since this parameter depends on the prevalence of the outcome of interest in the source population[14] and potentially on how accurately specific diagnostic codes are used in different care settings.[2]

The study findings have some limitations. First, the database used originated from a single national US health insurer and may not be representative of non-US health claims data sources. Other sources of health insurance claims may have slightly different findings based on their own data specificities. In addition, it is possible that some exclusion criteria, such as prior hysterectomy or endometrial cancer, are not documented in the electronic claims if they occurred before the earliest data available for the study or before the patients' enrollment with the health plan. That is, women who had a hysterectomy or endometrial cancer or ablation before enrolling in Aetna medical insurance plans may not have records for those procedures or diagnoses in Aetna data. We excluded eight women from the analytical sample after clinical review because their medical records clearly documented historical hysterectomies and/or prior endometrial cancer that were not recorded in the electronic claims data. Finally, we did not endeavor to validate the absence of endometrial cancer codes, as true non-cases, to obtain a negative predictive value. Similarly, the practical decision, based on time and resources, to sample provisional cases identified by the algorithm resulted in our inability to estimate sensitivity and specificity. It is possible that some true cases were misclassified as non-cases by the case-finding algorithm and were therefore missed.

Even though the algorithm yielded adequate predictive value, there may be opportunity for refinement, as 12%–14% of provisional cases were found to be noncases, depending on the algorithm variant. Upon further clinical review, medical records contained information on the presence of other types of uterine malignancies or neoplasms in nearby organs in almost half of the medical records from non-cases, while these medical records did not contain critical information on endometrial cancer (Supplemental Table S1). Additionally, there were no frequently-appearing common diagnosis codes related to neoplasms in nearby organs to incorporate in the algorithm (Supplemental Table S2).[2]

In conclusion, both variants of the case-finding algorithm based on ICD-10-CM codes were able to accurately identify endometrial cancer cases in a US healthcare claims data source. The narrow algorithm (variant B) identified the same cases as the broad one (variant A) but with fewer false positives and is therefore preferred for the planned post-marketing safety study.

ACKNOWLEDGMENTS

The authors would like to acknowledge the contributions of Carla Brannan (CVS Health), and Brian Calingaert, MS (RTI Health Solutions) to the study protocol and conduct, and of Kimberly Welder, RN (CVS Health), and Amanda Unger, RN (CVS Health), for clinical review.

CONFLICT OF INTEREST STATEMENT

This study was sponsored by TherapeuticsMD. Catherine B. Johannes, James A. Kaye, Andrea V. Margulis, and Catherine W. Saltus are full-time employees of RTI Health Solutions, a unit of RTI International, an independent, nonprofit organization that conducts work for government, public, and private organizations, including pharmaceutical companies. Djeneba Audrey Djibo, Cheryl N. McMahill-Walraven, Patricia Shuminski, and Mark Libertin are full-time employees of CVS Health and conduct work for government, public, and private organizations, including pharmaceutical companies, as part of their employment. Shelli Graham is a TherapeuticsMD employee.

ETHICS STATEMENT

The study was exempted by an independent Institutional Review Board (study identifier 8947-ADjibo).

PREVIOUS PRESENTATIONS

The results of this study have been presented in part at the International Society for Pharmacoepidemiology (ISPE) 2022 Annual Meeting in Copenhagen, Denmark, and at the North American Menopause Society (NAMS) 2022 Annual Conference, in Atlanta, Georgia, United States.

ORCID

Djeneba Audrey Djibo https://orcid.org/0000-0003-3378-0691
Andrea V. Margulis https://orcid.org/0000-0001-7388-6082
Cheryl N. McMahill-Walraven https://orcid.org/0000-0002-3765-6527
Catherine W. Saltus https://orcid.org/0000-0003-4727-5987
James A. Kaye https://orcid.org/0000-0002-9357-4227
Catherine B. Johannes https://orcid.org/0000-0002-0586-9886

REFERENCES

1. Sherman RE, Anderson SA, Dal Pan GJ, et al. Real-world evidence: what is it and what can it tell us? *N Engl J Med*. 2016;375(23):2293-2297. doi:10.1056/NEJMsb1609216
2. Weinstein EJ, Ritchey ME, Lo Re V *3rd*. Core concepts in Pharmacoepidemiology: validation of health outcomes of interest within real-world healthcare databases. *Pharmacoepidemiol Drug Saf*. 2023;32:1-8. doi:10.1002/pds.5537
3. Schumock G. Alternative Methods for Health Outcomes of Interest Validation. 2013. Accessed September 13, 2022 https://www.sentinelinitiative.org/methods-data-tools/health-outcomes-interest/alternative-methods-health-outcomes-interest-validation
4. Margulis AV, Arana A. Improving the quality of observational data: registries of validated outcomes and other patient events. *Epidemiology*. 2021;32(5):661-663. doi:10.1097/EDE.0000000000001392
5. Society AC. Cancer facts and figures. October 20. 2022 https://www.cancer.org/content/dam/cancer-org/research/cancer-facts-and-statistics/annual-cancer-facts-and-figures/2022/2022-cancer-facts-and-figures.pdf
6. Institute NC. SEER Cancer Stat Facts: Uterine Cancer. 2022 https://seer.cancer.gov/statfacts/html/corp.html
7. Esposito DB, Banerjee G, Yin R, et al. Development and validation of an algorithm to identify endometrial adenocarcinoma in US administrative claims data. *J Cancer Epidemiol*. 2019;2019:1938952. doi:10.1155/2019/1938952
8. Brown JS, Mendelsohn AB, Nam YH, et al. The US Food and Drug Administration sentinel system: a national resource for a learning health system. *J Am Med Inform Assoc*. 2022;29:2191-2200. doi:10.1093/jamia/ocac153
9. Fung KW, Richesson R, Smerek M, et al. Preparing for the ICD-10-CM transition: automated methods for translating ICD codes in clinical phenotype definitions. *EGEMS (Wash DC)*. 2016;4(1):1211. doi:10.13063/2327-9214.1211
10. CMS. ICD-10 CM and GEMs. Accessed January 27, 2021. 2018. https://www.cms.gov/Medicare/Coding/ICD10/2018-ICD-10-CM-and-GEMs.html
11. Amant F, Mirza MR, Koskas M, Creutzberg CL. Cancer of the corpus uteri. *Int J Gynaecol Obstet*. 2018;143(Suppl 2):37-50. doi:10.1002/ijgo.12612
12. Rasmussen LA, Jensen H, Virgilsen LF, et al. Identification of endometrial cancer recurrence: a validated algorithm based on nationwide Danish registries. *Acta Oncol*. 2021;60(4):452-458. doi:10.1080/0284186X.2020.1859133
13. Schmidt M, Schmidt SA, Sandegaard JL, Ehrenstein V, Pedersen L, Sorensen HT. The Danish National Patient Registry: a review of content, data quality, and research potential. *Clin Epidemiol*. 2015;7:449-490. doi:10.2147/CLEP.S91125
14. Lanes S, Brown JS, Haynes K, Pollack MF, Walker AM. Identifying health outcomes in healthcare databases. *Pharmacoepidemiol Drug Saf*. 2015;24(10):1009-1016. doi:10.1002/pds.3856

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: Djibo DA, Margulis AV, McMahill-Walraven CN, et al. Validation of an ICD-10 case-finding algorithm for endometrial cancer in US insurance claims. *Pharmacoepidemiol Drug Saf*. 2023;1-8. doi:10.1002/pds.5690