Value of Adding Secondary Read Diagnosis Codes to Identify Breast Cancer and Bladder Cancer in the Clinical Practice Research Datalink: A Pilot Validation Study

Lia Gutiérrez,[1] Heather E. Danysh,[2] Jaume Aguado,[1] Phillip R. Hunt,[3] James A. Kaye,[2] Xabier García de Albéniz Martinez,[1] Alicia Gilsenan[4]

[1]RTI Health Solutions, Barcelona, Spain; [2]RTI Health Solutions, Waltham, MA, United States; [3]AstraZeneca, Gaithersburg, MD, United States; [4]Research Triangle Park, NC, United States

DISCLOSURES:

This work was performed within the context of a study funded by AstraZeneca through a contract with RTI Health Solutions (RTI-HS). The contract provides the research team independent publication rights.

Lia Gutiérrez, Heather E. Danysh, Jaume Aguado, James A. Kaye, Xabier García de Albéniz Martinez, and Alicia Gilsenan are full-time employees of RTI Health Solutions, an independent nonprofit research organization that does work for government agencies and pharmaceutical companies.

Phillip R. Hunt is a full-time employee of AstraZeneca, the marketing authorization holder of the study drug.

BACKGROUND

- Previous studies validating cancer outcomes in the United Kingdom's Clinical Practice Research Datalink (CPRD) evaluated algorithms using "primary" Read clinical and International Classification of Diseases, 10th Revision (ICD-10) diagnosis codes.[1]
- In an ongoing postauthorization drug safety study, we conducted a pilot validation in CPRD to evaluate the positive predictive performance of algorithms that also included "secondary" Read morphology codes to identify breast and bladder cancer outcomes among individuals with type 2 diabetes mellitus.
- The European Union Electronic Register of Post-Authorisation Studies (EU PAS Register) number for this study is 12116.

OBJECTIVE

- To estimate the positive predictive values (PPVs) of algorithms combining primary Read clinical, secondary Read morphology, and ICD-10 diagnosis codes versus algorithms omitting secondary Read morphology codes for female invasive breast cancer and in situ or invasive bladder cancer.

METHODS

- We identified outcomes among individuals aged ≥ 40 years initiating an antidiabetic drug.
- Provisional cases of female invasive breast cancer and in situ or invasive bladder cancer in the study population were identified from November 2012 through July 2017 using predefined algorithms (Figure 1).
- For each outcome, a sample of provisional cases was reviewed for clinical case validation using information from clinical patient profiles, general practitioner questionnaires, and prespecified case definitions (Figure 2).
- The PPV and 95% confidence interval (CI), estimated as the proportion of confirmed cases among all provisional cases included in the validation sample, were calculated for all codes combined, separately with and without secondary Read morphology codes, and then also separately for each code type.

Figure 1. CPRD Algorithms for Female Invasive Breast Cancer and In Situ or Invasive Bladder Cancer

Female Breast Cancer

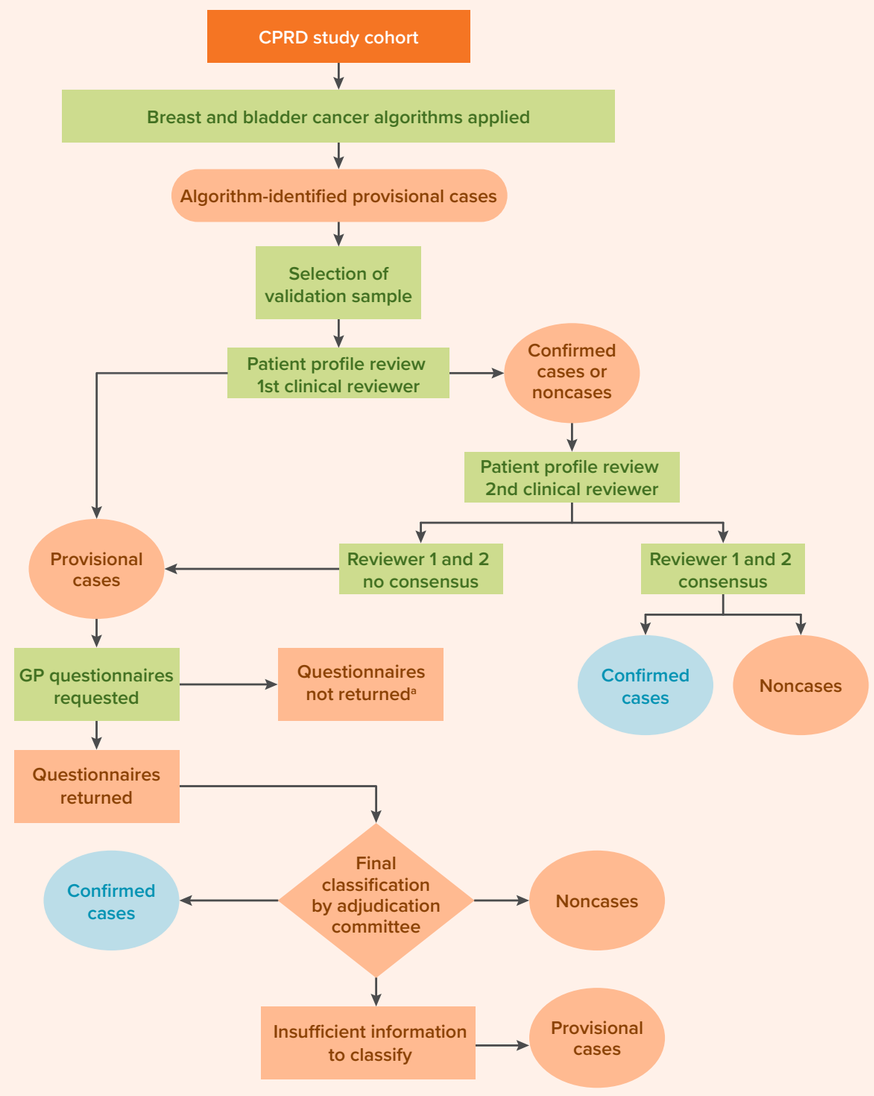
- In female patients, at least one recorded diagnosis code for invasive breast cancer:
  - **In CPRD GOLD data**
    - Primary Read clinical codes B34*.00 from the medical code listing
    - Secondary Read morphology codes from the BB**.00 medical code listing
  - **In HES linkable patients**
    - ICD-10 C50.*

Bladder Cancer

- At least one recorded diagnosis code for in situ or invasive bladder cancer:
  - **In CPRD GOLD data**
    - Primary Read clinical codes B49*.00 or B837.00 from the medical code listing
    - Secondary Read morphology codes from the BB**.00 medical code listing
  - **In HES linkable patients**
    - ICD-10 C67.* or D09.0

GOLD = General Practitioner Online Database of the CPRD; HES = Hospital Episodes Statistics.

Figure 2. Validation Process in the CPRD



[a]Retained as provisional cases for final analysis.

RESULTS

Breast Cancer

- Of the 110 provisional female breast cancer cases identified through the electronic algorithm in CPRD, 89% were confirmed after validation (Table 1).
  - Most provisional cases were identified through primary Read clinical diagnosis codes.
  - Including secondary Read morphology codes did not identify additional cases.

Bladder Cancer

- Of the 74 provisional bladder cancer cases identified through the electronic algorithm in CPRD, 76% were confirmed after validation (Table 2).
  - Most provisional cases were identified through **primary Read** clinical and ICD-10 diagnosis codes.
  - Including secondary Read morphology codes to identify bladder cancer in CPRD lowered the PPV for the bladder cancer algorithm.

Table 1. Positive Predictive Value of CPRD Algorithms: Breast Cancer

BREAST CANCER	Algorithm-Identified Provisional Cases Selected for Validation, N	Confirmed Cases, N	PPV, % (95% CI)
Any primary Read, secondary Read morphology, or ICD-10 code	110	98	89 (82-94)
Subset of all ICD-10 codes[a] (individual codes with counts ≥ 5)[b]	21	15	71 (48-89)
C50.9 Malignant neoplasm: breast, unspecified	13	9	69 (39-91)
Subset of all primary Read clinical codes[a] (individual codes with counts ≥ 5)[b]	91	85	93 (86-98)
B34..00 Malignant neoplasm of female breast	78	73	94 (86-98)
B34z.00 Malignant neoplasm of female breast NOS	6	6	100 (54-100)
Subset of secondary Read morphology codes only	0	0	

[a] Subset of codes: i.e., Read (upper hierarchy of specific neoplasm together and [M] codes separately) in CPRD GOLD data and ICD-10 codes in HES inpatient and HES outpatient data. Please note that multiple codes for a single cancer diagnosis may have been identified by the algorithm on the same date.

[b] According to CPRD policy, any cell with a value of 1 to 4 or any cell that allows a value of 1 to 4 to be derived from other reported cells or information cannot be reported.

Table 2. Positive Predictive Value of CPRD Algorithms: Bladder Cancer

BLADDER CANCER	Algorithm-Identified Provisional Cases Selected for Validation, N	Confirmed Cases, N	PPV, % (95% CI)
Any primary Read, secondary Read morphology, or ICD-10 code	74	56	76 (64-85)
Subset of all ICD-10 codes[a] (individual codes with counts ≥ 5)[b]	30	27	90 (74-98)
C67.9 Malignant neoplasm: bladder, unspecified	21	19	91 (70-99)
Subset of all primary Read codes[a] (individual codes with counts ≥ 5)[b]	37	31	84 (68-94)
B49..00 Malignant neoplasm of urinary bladder	28	24	86 (67-96)
Subset of secondary Read morphology codes only[a,b]	11	NR	< 20 (NR-NR)
ICD-10 and primary Read clinical codes combined (excluding secondary Read morphology codes)	**63**	**54**	**86 (75-93)**

NR = Not reported (number masked due to CPRD data protection rules).

[a] Subset of codes: i.e., Read (upper hierarchy of specific neoplasm together and [M] codes separately) in CPRD GOLD data and ICD-10 codes in HES inpatient and HES outpatient data. Total number of subsets do not add up because multiple codes for a single cancer diagnosis may have been identified by the algorithm on the same date.

[b] According to CPRD policy, any cell with a value of 1 to 4 or any cell that allows a value of 1 to 4 to be derived from other reported cells or information cannot be reported.

CONCLUSIONS

- Overall, the PPVs of the algorithms for at least one code of any type were high (breast, 89%) or moderately high (bladder, 76%).
- Secondary Read morphology codes contributed no additional cases for the breast cancer algorithm.
- Secondary Read morphology codes identified additional provisional cases of bladder cancer, but few were confirmed. Use of secondary Read morphology codes increased the proportion of misclassified algorithm-identified bladder cancer cases and was detrimental to the overall PPV of the bladder cancer algorithm.
- We recommend not using secondary Read morphology codes to identify breast or bladder cancers in CPRD if maximizing PPV is the goal.

REFERENCE

1. Margulis AV, Fortuny J, Kaye JA, Calingaert B, Reynolds M, Plana E, et al. Validation of cancer cases using primary care, cancer registry, and hospitalization data in the United Kingdom. Epidemiology. 2018 Mar;29(2):308-13.

CONTACT INFORMATION

Lia Gutiérrez, BScN, MPH
Senior Director, Epidemiology

RTI Health Solutions
Av. Diagonal, 605, 9-1
08028 Barcelona Spain

O: +34. 93 241 7764
E-mail: lgutierrez@rti.org