Development and Optimization of a Bladder Cancer Algorithm Using SEER-Medicare Claims Data

John L. Gore, MD[1] ; Phoebe Wright, PharmD[2] ; Vanessa Shih, PharmD[2]; Nancy N. Chang, PharmD[2] ; Sina Noshad, MD, PhD[3]; Gabriel G. Rey, MS[3]; Steven Wang, MS[3]; and Sujata Narayanan, MBBS, MS[2]

ABSTRACT

PURPOSE Categorizing patients with cancer by their disease stage can be an important tool when conducting administrative claims-based studies. As claims databases frequently do not capture this information, algorithms are increasingly used to define disease stage. To our knowledge, to date, no study has used an algorithm to categorize patients with bladder cancer (BC) by disease stage (non–muscle-invasive BC [NMIBC], muscle-invasive BC [MIBC], or locally advanced/metastatic urothelial carcinoma [la/mUC]) in a US-based health care claims database.

METHODS A claims-based algorithm was developed to categorize patients by disease stage on the basis of the administrative claims portion of the SEER-Medicare linked data. The algorithm was validated against a reference SEER registry, and the algorithm's parameters were iteratively modified to improve its performance. Patients were included if they had an initial diagnosis of BC between January 2016 and December 2017 recorded in SEER registry data. Medicare claims data were available for these patients until December 31, 2019. The algorithm was evaluated by assessing percentage agreement, Cohen's kappa (κ), specificity, positive predictive value (PPV), and negative predictive value (NPV) against the SEER categorization.

RESULTS A total of 15,484 patients with SEER-confirmed BC were included: 10,991 (71.0%) with NMIBC, 3,645 (23.5%) with MIBC, and 848 (5.5%) with la/mUC. After multiple rounds of algorithm optimization, the final algorithm had an agreement of 82.5% with SEER, with a κ of 0.58, a PPV of 87.0% for NMIBC, and 76.8% for MIBC and a high NPV for la/mUC of 98.0%.

CONCLUSION This claims-based algorithm could be a useful approach for researchers conducting claims-based studies categorizing patients with BC at diagnosis.

INTRODUCTION

Bladder cancer (BC) is the sixth most common cancer in the United States with an estimated incidence of 82,290 cases in 2023.[1,2] The most common type of BC in the United States and Western Europe is urothelial carcinoma (UC), which is classified into three stages according to disease location at diagnosis:[3-5] (1) Non–muscle-invasive BC (NMIBC), characterized by restriction of the cancer to the superficial lining of the bladder; (2) muscle-invasive BC (MIBC), characterized by cancer penetration into the muscle layer of the bladder; (3) locally advanced/metastatic UC (la/mUC), characterized by cancer spread to other organs and/or distant lymph nodes.

Patient 5-year relative survival is 77.9% after BC diagnosis but varies considerably by cancer stage, determined clinically using the TNM staging system.[6] In the United States, 5-year relative survival is 70.9% in patients with localized disease versus 8.3% in patients with distant metastases.[7]

Similarly, standard-of-care treatment varies between stages, as patients with NMIBC typically receive transurethral resection of the bladder tumor (TURBT) plus intravesical bacillus Calmette–Guérin therapy, whereas platinum-based, systemic chemotherapy has been the standard-of-care treatment for patients with la/mUC.[3,8]

As prognosis and treatment regimens vary between cancer stages, it is important that researchers can accurately identify patients by their disease stage in real-world claims databases, as this enables them to conduct studies to understand real-world treatment patterns and health care resource use.

Algorithms have been developed to identify the incidence of oncological conditions in health care claims databases, including advanced lung, breast, ovarian, and gastric cancers.[9-12] Although a previous UK-based study developed an algorithm to identify patients with BC and to distinguish

CONTEXT

Key Objective
To develop a claims-based algorithm to classify patients with bladder cancer (BC) by disease stage on the basis of Medicare claims data and to validate and optimize the algorithm using reference SEER registry data.

Knowledge Generated
The algorithm developed correctly categorized more than 80% of patients, with the highest positive predictive values seen for patients with non–muscle-invasive BC (NMIBC) and muscle-invasive BC (MIBC). The final algorithm would be best suited for use in future claims-based studies where NMIBC and MIBC are of interest.

Relevance
The algorithm is able to categorize BC by muscle invasions and stage from Medicare claims data—this information would otherwise not be readily available.

between NMIBC and MIBC in an electronic medical record (EMR) database, to our knowledge, no previous studies have used a claims-based algorithm to identify patients with BC according to their disease stage at a population level or have validated such algorithms against US registry data.[13]

Additionally, novel treatments were recently approved for NMIBC, MIBC, and la/mUC.[14-16] Hence, it is important that researchers can identify patients with BC by disease stage to evaluate how new treatments perform in real-world clinical practice.

As claims databases do not directly capture disease stage, their usefulness to researchers has been limited. This study was conducted to address this limitation by developing and validating a claims-based algorithm using the administrative claims portion of the SEER registry and Centers for Medicare and Medicaid Services Medicare–linked data set (SEER-Medicare). The SEER-Medicare data set was selected as it allowed validation of the staging algorithm against reference clinical staging information in the SEER registry. Additionally, this data set is considered the most data-rich US-based source of comprehensive, population-based cancer staging data, which also incorporates a commonly used health care claims data set.

The purpose of the algorithm was to classify patients with BC as having NMIBC, MIBC, or la/mUC based on Medicare claims data and to validate and optimize this algorithm using the patient's known clinical stage in the reference SEER registry data.

METHODS

Study Population

Patients were included if they had an initial diagnosis of BC between January 2016 and December 2017 recorded in SEER registry data and had tumor, node, and metastasis staging information in the SEER database. Patients were required to have 12 months of continuous enrollment in Medicare parts A and B before their index date (date of their initial diagnosis of BC) and at least 1 month of follow-up after their index date.

Study Design

This study used retrospective data from the administrative claims portion of the SEER-Medicare–linked data set that contains both clinical staging information from SEER and medical claims information from Medicare.[17] Medicare claims data and death status were available for included patients until December 31, 2019.

An initial algorithm was developed to stage patients using data from the Medicare administrative claims database. This staging was then validated against clinical staging information held in the SEER registry, which acted as the reference standard.

Participant Disease Staging

SEER Categorization

A diagnosis of BC was identified using the earliest appearance of the International Classification of Diseases codes C67.X Malignant neoplasm of bladder or 188.X Malignant neoplasm of bladder in the SEER data set.

Information on cancer-directed surgery, disease staging, and histology data from SEER were then used to categorize patients as having NMIBC, MIBC, or la/mUC at diagnosis. This categorization served as the reference standard, which the claims algorithm was evaluated against. The SEER categorization is explained further in Table 1.

Development of Initial Claims-Based Algorithm

After SEER registry categorization, an initial, claims-based algorithm was developed using Medicare administrative

claims data. The goal of the algorithm was to categorize patients with BC by disease stage. As Medicare claims data contain only limited information regarding disease stage apart from secondary malignancy codes, which can indicate advanced disease, the algorithm was designed to distinguish between patients based on the treatments they received. Hence, after clinician input on which treatments were most relevant in determining each disease stage, the following parameters were chosen for the algorithm: (1) receipt of TURBT $\pm$ intravesical chemotherapy; (2) receipt of systemic chemotherapy; (3) receipt of radiation therapy and/or cystectomy; (4) presence of secondary malignancy (metastasis). Patients were categorized on the basis of these parameters and on their timing relative to patient diagnosis date.

Patients who did not have any of the prespecified parameters (ie, who did not receive a listed BC treatment and did not have evidence of secondary malignancy) were recorded as to be determined (TBD). Parameters were assessed on the basis of diagnostic and treatment codes; additional information on codes used is provided in the Data Supplement (Supplementary Materials S1, Supplementary Tables S1-S5). The initial algorithm is described in more detail in the Data Supplement (Supplementary Materials 2, Supplementary Table S6).

Algorithm Validation

After the algorithm categorized patients on the basis of Medicare claims data, this categorization was validated against the patient's corresponding SEER registry categorization to determine the algorithm's accuracy. Accuracy was assessed using the following indices: Cohen's kappa statistic assessed agreement between the algorithm and the SEER categorization in instances where both classified patients into group k (where k refers to patients with NMIBC, MIBC, and la/mUC). A kappa value of <0 indicates no agreement, 0-0.20 slight agreement, 0.21-0.40 fair agreement, 0.41-0.60 moderate agreement, 0.61-0.80 substantial agreement, and 0.81-1.0 almost perfect agreement.[18]

For example, for NMIBC, assuming an observed agreement of 90% between SEER and the algorithm and assuming 70% of patients with BC in SEER have NMIBC (expected agreement by chance alone), the Cohen's kappa will be calculated as 0.5, indicating moderate agreement.

Percentage agreement was defined as the probability that the algorithm and SEER categorization classified a patient as belonging to group k with a higher percentage indicating greater agreement.

TABLE 1. SEER Cancer Categorization

Stage	T	N	M	Surgery	SEER Category
Stage 0a	Ta	N0	M0	No constraint	NMIBC
Stage 0is	Tis	N0	M0	No constraint	NMIBC
Stage I	T1	N0	M0	No constraint	NMIBC
Stage II	T2a	N0	M0	No constraint	MIBC
	T2b	N0	M0	No constraint	MIBC
Stage IIIA	T3a	N0	M0	No constraint	MIBC
	T3b	N0	M0	Cancer-directed surgery or radiation	MIBC
	T4a	N0	M0	Cancer-directed surgery or radiation	MIBC
	T1-T4a	N1	M0	Cancer-directed surgery or radiation	MIBC
Stage IIIB	T1-T4a	N2,N3	M0	Cancer-directed surgery or radiation	MIBC
Stage IVA	T4b	Any N	M0	Cancer-directed surgery or radiation	MIBC
Stage IIIA	T3b	N0	M0	No cancer-directed surgery and no radiation	laUC
	T4a	N0	M0	No cancer-directed surgery and no radiation	laUC
	T1-T4a	N1	M0	No cancer-directed surgery and no radiation	laUC
Stage IIIB	T1-T4a	N2	M0	No cancer-directed surgery and no radiation	laUC
	T1-T4a	N3	M0	No cancer-directed surgery and no radiation	mUC
Stage IVA	T4b	Any N	M0	No cancer-directed surgery and no radiation	laUC
	Any T	Any N	M1a	No constraint	mUC
Stage IVB	Any T	Any N	M1b	No constraint	mUC

NOTE. Cancer stage was determined using the TNM classification system.[6] la/mUC or MIBC status were determined based on treatment received after diagnosis.
Abbreviations: la/mUC, locally advanced/metastatic urothelial carcinoma; laUC, locally advanced urothelial carcinoma; M, metastasis; MIBC, muscle-invasive bladder cancer; mUC, metastatic urothelial carcinoma; N, node; NMIBC, non−muscle-invasive bladder cancer; T, tumor; Ta, noninvasive papillary; Tis, carcinoma in situ.

Sensitivity assessed the probability that the algorithm correctly classified patients as belonging to group *k* among those that the SEER categorization classified as group *k*.

Specificity analyses assessed the probability that the algorithm correctly classified patients as not belonging to group *k* among those that the SEER categorization classified as belonging to group *k*.

Positive predictive value (PPV) was defined as the proportion of patients classified as belonging to group *k* by the algorithm who were also classified as belonging to group *k* by the SEER registry categorization.

Negative predictive value (NPV) was defined as the proportion of patients not classified as belonging to group *k* by the algorithm who were also not classified as belonging to group *k* according to disease staging and histology information by the SEER registry categorization.

Continuous variables were summarized by means and standard deviations (SD), and categorical variables were summarized by counts and percentages.

All analyses were performed using SAS version 9.4 (SAS Institute, Cary, NC). Additional details are provided in the Data Supplement (Supplementary Materials S3, Supplementary Tables S7 and S8).

Algorithm Optimization

After the accuracy of the initial algorithm was compared with the SEER registry categorization, the parameters of the initial algorithm were iteratively modified to improve the algorithm's performance. This approach enables continuous refinements to be made to the algorithm, increasing its accuracy at each iteration. Modifications were driven by patient data and expert clinician opinion, with changes at each iteration focused on correctly classifying the largest group of patients misclassified by the previous iteration. Modifications intended to enhance the accuracy of predictions included changes to parameter definitions, their order in the algorithm, or the time window over which they were considered.

Following each modification, the algorithm was evaluated with input from clinicians to determine whether the change was clinically relevant to avoid overfitting and then its predictive accuracy was assessed against the SEER–Medicare categorization by running the modified algorithm using patient–level data to determine whether the change resulted in improvements in its accuracy. Modifications that improved accuracy were retained. The process was repeated iteratively to reduce the proportion of patients who were misclassified until additional modifications were either not clinically relevant or did not improve performance. At this point, the algorithm was considered final.

RESULTS

Study Cohort

In total 15,484 individuals with BC with data recorded in the SEER-Medicare−linked database were included in the study (Table 2).

Of the 15,484 patients included in the study, 10,991 (71.0%) had NMIBC, 3,645 (23.5%) had MIBC, and 848 (5.5%) had la/mUC according to the SEER database.

Algorithm Optimization

The initial algorithm underwent five rounds of revision. These changes and their impact on the algorithm's performance are shown in Table 3. Revisions that did not result in improvements in accuracy were not retained and are not reported here. The greatest increase in performance was observed between the initial algorithm and iteration 1, with percentage agreement and kappa increasing from 64.9% to 78.4% and from 0.2511 to 0.4169, respectively. Additional results for accuracy indices are shown in the Data Supplement (Supplementary Material S4, Supplementary Figures S1-S4).

Final Algorithm

Compared with the initial algorithm, the final algorithm (Fig 1) included an additional branch, which identified patients who received radiation or cystectomy ≤3 months after diagnosis. This treatment approach is typical for patients with MIBC, enabling patients on this branch to be categorized as having MIBC. Additionally, the TBD category in the initial algorithm was changed to not applicable (NA).

Algorithm Performance

Of the 15,484 patients included in this study, the initial algorithm classified 13,494 patients (87.1%) as having NMIBC, MIBC, or la/mUC (Fig 2). The remaining 1,990 patients (12.9%) did not meet any of the decision parameters and were classified as TBD. When the final algorithm was rerun using data for the 15,484 included patients, 2,702 (17.5%) were categorized as NA. These patients could not be classified, as they either received radiation or cystectomy alone or did not receive any treatment within specified time periods.

When the sample was limited to the remaining 12,782 patients, the final algorithm and the SEER-registry were found to have an overall agreement of 82.5% as 10,551 of 12,782 patients were correctly categorized, yielding an overall kappa of 0.5777 (moderate agreement).

A proportion of patients were miscategorized by the final algorithm (Fig 2), specifically 412 patients with NMIBC were categorized as having MIBC and 203 as having la/mUC; 1,117

TABLE 2. Cohort Attrition for Study Cohort

Inclusion Criteria	No.	Percentage of Previous Step
A bladder cancer diagnosis record in SEER registration data	203,349	100.0
Initial diagnosis of BC from January 2016 to December 2017 (the index date)	38,726	19.0
12 months of continuous enrollment in Medicare parts A and B before their index date and 1 month of follow-up after their index date	32,229	83.2
Available TNM staging information	21,173	65.7
Exclusion Criteria		
No bladder cancer diagnosis record in Medicare claims data from January 2016 to December 2017	16,901	79.8
A bladder cancer diagnosis during the 12-month baseline (prevalent cases)	15,484	91.6

Abbreviation: BC, bladder cancer.

patients with MIBC were categorized as having NMIBC and 263 as having la/mUC; 149 patients with la/mUC were categorized as having NMIBC and 87 as having MIBC.

The initial algorithm reported a PPV for patients with NMIBC of 81.0%, 72.9% for patients with MIBC, and performance was lower for patients with la/mUC, with a PPV of 16.0%, Table 4.

Algorithm optimization improved the accuracy indices of the final algorithm across all disease categories versus the initial algorithm (Table 4): PPV increased from 16.0%-81.0% in the initial algorithm to 47.2%-87.0% in the final algorithm. NPV increased from 44.2%-97.9% in the initial algorithm to 79.7%-98.0% in the final algorithm. Sensitivity increased from 25.2%-76.1% in the initial algorithm to 54.4%-93.2% in the final algorithm. Specificity increased from 51.5%-97.4% in the initial algorithm to 65.6%-96.2% in the final algorithm.

DISCUSSION

This study designed and optimized a four-step algorithm to categorize patients with BC into NMIBC, MIBC, and la/mUC

TABLE 3. Summary of Adopted Changes During Algorithm Refinement and Their Impact on Algorithm Performance

		Overall Algorithm Performance	
Iteration Number	Summary of Major Changes to the Algorithm	Percentage Agreement	Kappa
(Initial algorithm)	—	64.9	0.2511
1	A more conservative definition for metastases was adopted. Previously, presence of at least one metastatic code was considered sufficient. It was modified to require at least one inpatient or two outpatient codes on separate days (with the first code occurring within the first month after the initial BC diagnosis) Initially, TURBT was deemed sufficient to identify NMIBC. On further examination, it was decided to also include + intravesical chemotherapy to align better with clinical guidelines	78.4	0.4169
2	Code lists for systemic therapies were expanded to include pembrolizumab, enfortumab vedotin, erdafitinib, and durvalumab. This change was made to ensure that potential treatment regimens that were off-label during the study period but were still available to clinicians for experimental use were also captured Patients receiving systemic chemotherapy who did not initiate either therapeutic radiation or cystectomy any time after initial BC diagnosis were originally flagged as having la/mUC. Given the high rate of misclassification, they were updated to be flagged as TBD	78.5	0.4184
3	Patients without systemic therapy but with receipt of therapeutic radiation or cystectomy on or up to 3 months after initial BC diagnosis were originally categorized as having NMIBC. Since this therapeutic approach is more consistent with MIBC, they were recategorized as having MIBC	81.0	0.5475
4	Previously, all patients with secondary metastatic diagnoses were assumed to have la/mUC. An exception was added that if patients also had a cystectomy within the first 30 days after their initial diagnosis, they were assumed to have MIBC with secondary malignancy codes more likely to indicate local invasion rather than metastasis to distant organs	81.8	0.5636
5 (final, optimized algorithm)	Previously, to classify patients who had systemic therapy as having MIBC, they were only required to have therapeutic radiation or cystectomy at any time after their initial BC diagnosis. Given the ambiguity of the rule limiting the reproducibility of the algorithm, the timeframe was changed to within 12 months after their initial BC diagnosis	82.5	0.5777

Abbreviations: BC, bladder cancer; la/mUC, locally advanced/metastatic urothelial carcinoma; MIBC, muscle-invasive bladder cancer; NMIBC, non–muscle-invasive bladder cancer; TBD, to be determined; TURBT, transurethral resection of bladder tumor.
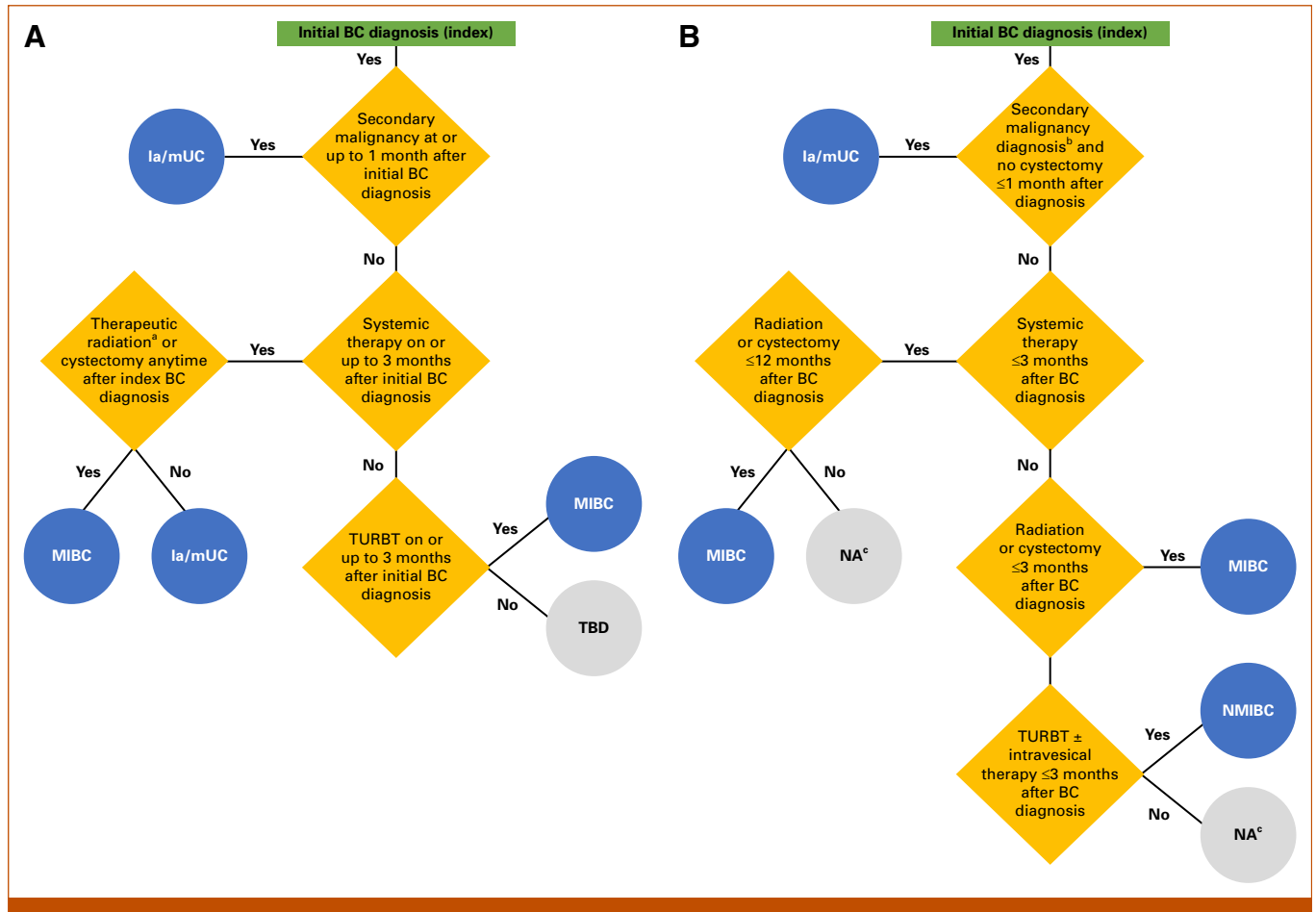
FIG 1. (A) Initial algorithm and (B) final, optimized claims-based algorithm for bladder cancer staging at diagnosis. [a]Therapeutic radiation refers to ≥14 radiation treatments within 120 days before/after first systemic therapy date on or after index BC diagnosis; [b]Secondary malignancy diagnosis with ≥1 inpatient claim or ≥2 outpatient claims with a secondary malignancy on separate days (first claim to be within 1 month of initial BC diagnosis and second claim to be any time after); [c]NA, not applicable as patients did not meet the algorithm criteria. BC, bladder cancer; BCG, Bacillus Calmette-Guerin; la/mUC, locally advanced/metastatic urothelial carcinoma; MIBC, muscle-invasive bladder cancer; NA, not applicable; NMIBC, non−muscle-invasive bladder cancer; TBD, to be determined; TURBT, transurethral resection of bladder tumor.

at diagnosis and evaluated its performance against the reference standard SEER categorization.

The optimized algorithm correctly categorized more than 80% of patients and demonstrated a kappa statistic of 0.5777. Analyses of the algorithm's sensitivity found that the final algorithm correctly staged 93.2% of patients with NMIBC, 63.8% of patients with la/mUC, and 54.4% of patients with MIBC; however, these results should be considered in relation to the observed PPV of between 87.0% and 47.2% depending on disease stage, with the highest PPV seen for patients with NMIBC and MIBC. As a PPV between 70% and 80% is typically considered to be an indication of a high-performing diagnostic algorithm, these results suggest that the final algorithm may be suitable for use by researchers investigating NMIBC and MIBC despite the observed variations in sensitivity outputs.[19,20] Future studies could use this algorithm, or researchers could adapt it for their health care system or

population of interest with confidence that patients are being correctly categorized by disease stage, enabling them to capture real-world treatment patterns and patient outcomes.

These results are consistent with those reported in a study by Esposito et al,[10] which also used data from health care databases, including SEER, to identify patients with a range of metastatic/advanced cancers, and reported a PPV of 78.0% for UC. However, unlike our study, Esposito et al used a predictive model methodology and did not report PPV by disease stage.

The lower PPV shown by the final algorithm for la/mUC is likely due to the lower prevalence of la/mUC, as PPV is mathematically dependent on disease prevalence, with a higher prevalence leading to greater PPV.[21] This caveat should be considered when using the algorithm with external databases.

		SEER-Registry Categorization			
	Category	NMIBC	MIBC	la/mUC	Total
Initial claims-based algorithm categorization	NMIBC	7,509	1,584	172	9,265
	MIBC	231	729	40	1,000
	la/mUC	2,130	582	517	3,229
	Total	9,870	2,895	729	13,494
Final, optimized, claims-based algorithm categorization	NMIBC	8,487	1,117	149	9,753
	MIBC	412	1,648	87	2,147
	la/mUC	203	263	416	882
	Total	9,102	3,028	652	12,782

FIG 2. Assessed concordance between the initial and final claims-based algorithms and SEER-Medicare Database; green values denote where there is concordance between the algorithm and the SEER-Registry categorization; orange values denote where there is discordance between the algorithm and the SEER-Registry categorization. la/mUC, locally advanced/metastatic urothelial carcinoma; MIBC, muscle-invasive bladder cancer; NMIBC, non–muscle-invasive bladder cancer.

Conversely, the final algorithm demonstrated a NPV of 98.0% for la/mUC, so patients with NMIBC or MIBC were not misclassified as having la/mUC by the algorithm with 98% certainty. This suggests that the algorithm may be of particular interest for future studies where NMIBC and MIBC are disease categories of interest.

Machine-learning studies have previously been suggested as an alternative to clinical algorithms for classifying patients by disease stage. However, Brooks et al[9] found that machine-learning algorithms did not lead to significantly greater accuracy than clinical algorithms but resulted in substantially greater algorithmic complexity. Additionally, machine-learning algorithms demonstrated less accurate performance when applied to a cross-validation data set, indicating overfitting. These results support the continued use of clinical algorithms in oncology research.

A previous UK-based study by Mamtani et al used an algorithm based on EMR data to distinguish between NMIBC and MIBC and reported a PPV for MIBC of 70.1%, which is similar to the PPV of 76.8% for MIBC reported here by our final algorithm. However, unlike this study they used a cross-sectional design and validated their algorithm against physician surveys, rather than against patient registry data. This study also stratified patients into la/mUC in addition to the NMIBC and MIBC groupings used by Mamtani et al.[13]

That the final, optimized algorithm was unable to categorize 17.5% of the sample by disease stage, instead categorizing them as NA, is of note as it indicates that a proportion of the study sample was treated conservatively, receiving either radiation or cystectomy alone or no treatment within specified time periods. Considering that the study cohort included patients 65 years and older, this may reflect differences in the treatment of older patients between treatment guidelines and clinical practice because of factors known to reduce the likelihood of receiving systemic treatments. These could include patient preferences, patient characteristics including assessed comorbidities or frailty and that a significant proportion of older patients have been shown to not receive guideline-recommended treatments.[22,23]

The presence of older patients in the study cohort may also help account for the discordance observed between the SEER categorization and the final algorithm, as deviations in treatments received from the guidelines used to develop the algorithm would have resulted in misclassification of these patients. Additionally, a previous study of patients with NMIBC[24] observed that real-world BC treatments frequently differ from treatment guidelines. These differences across all included patients in this study may explain why this guideline-based algorithm could not correctly categorize all included patients.

This study demonstrates and validates a novel, claims-based algorithm that categorized patients with BC into NMIBC, MIBC, and la/mUC at diagnosis. Future studies could be conducted to validate the algorithm in other real-world claims databases, or our methodology could be adapted to develop and validate algorithms in other disease areas. The algorithm may be useful for researchers conducting retrospective claims-based studies in BC as it augments the clinical information available in claims data. Further uses of the algorithm could include treatment landscape studies in which patients are identified by their BC stage to enable researchers to assess which treatments they receive in clinical practice.

TABLE 4. Accuracy Indices for the Initial and Final, Claims-Based Algorithms Stratified by BC Type

	PPV		NPV		Sensitivity		Specificity	
Category	Initial Algorithm (%)	Final Algorithm (%)	Initial Algorithm (%)	Final Algorithm (%)	Initial Algorithm (%)	Final Algorithm (%)	Initial Algorithm (%)	Final Algorithm (%)
NMIBC	81.0	87.0	44.2	79.7	76.1	93.2	51.5	65.6
MIBC	72.9	76.8	82.7	87.0	25.2	54.4	97.4	94.9
la/mUC	16.0	47.2	97.9	98.0	70.9	63.8	78.8	96.2

Abbreviations: BC, bladder cancer; la/mUC, locally advanced/metastatic urothelial carcinoma; MIBC, muscle-invasive bladder cancer; NMIBC, non–muscle invasive bladder cancer; NPV, negative predictive value; PPV, positive predictive value.

This study used claims data from Medicare beneficiaries. Although the Medicare population is considered representative of the overall patient population with BC in the United States, it may not be representative of all patients with BC, including patients with commercial insurance or without health insurance coverage. The study requirement that patients must have had continuous enrollment in Medicare parts A and B for at least 12 months before their index date, and for at least 1 month after that date, excluded patients with intermittent health care coverage, thus limiting the generalizability/representativeness of our sample. However, this is a limitation inherent in all claims-based studies and is less likely to occur with Medicare coverage. Additionally, this study included only patients with an initial diagnosis of BC recorded during the study period, so patients diagnosed with BC before the study period and with disease progression during the study period were not captured.

There were differences between the SEER registry categorization and the final algorithm's output for which the main causes were likely differences between treatment guidelines and treatment received by older patients. An additional cause of discordance between the SEER final algorithm classifications may be the SEER categorization itself as while a patient's initial disease stage is captured, their subsequent pathology events and their impact on patient disease classification are not. For instance, a patient with stage I disease (T1, N0, M0) would be categorized by SEER as having NMIBC. If, after a later resection, their disease was recategorized as stage II disease (T2a, N0, M0), they would be considered by their clinician to have MIBC and would be expected to receive treatment for MIBC. However, this change would not be captured by the SEER categorization, resulting in a difference between the SEER and claims-based algorithm classifications and likely accounting for some of the observed difference between the SEER registry categorization and the final algorithm. Additionally, the completeness of SEER registry data can be affected by patient movement during treatment, possibly resulting in additional discordance between the SEER categorization and the final algorithm's output. SEER registry data may also be subject to some degree of data entry error or misclassification.

As this study was conducted using a US-based data set that reflects US clinical coding and practice, the final algorithm should be validated in other health care systems before being applied outside this setting. Additionally, as the data used in this study predate the COVID-19 pandemic, they do not account for the impact of pandemic-related disruptions to health care systems on treatment patterns or for subsequent changes in treatment access.

The final algorithm reflects contemporary approaches to BC treatment and would need to be adapted if applied to historical data and following future changes to treatment guideline recommendations.

In conclusion, this study demonstrated the utility of a novel claims-based algorithm that categorized patients with BC at diagnosis and demonstrated a high degree of agreement with the SEER database.

Based on the accuracy metrics, the algorithm would be best suited for use in future claims-based studies where NMIBC and MIBC are BC categories of interest.

AFFILIATIONS

[1]University of Washington, Seattle, WA
[2]Pfizer Inc, Bothell, WA
[3]Genesis Research Group, Hoboken, NJ

CORRESPONDING AUTHOR

John L. Gore, MD; e-mail: jlgore@uw.edu.

DATA SHARING STATEMENT

The data supporting the findings of this study are available within the article and its Supplementary Materials.

AUTHOR CONTRIBUTIONS

AUTHORS' DISCLOSURES OF POTENTIAL CONFLICTS OF INTEREST

ACKNOWLEDGMENT

REFERENCES

1. Saginala K, Barsouk A, Aluru JS, et al: Epidemiology of bladder cancer. Med Sci (Basel) 8:15, 2020
2. Siegel RL, Miller KD, Wagle NS, et al: Cancer statistics, 2023. CA Cancer J Clin 73:17-48, 2023
3. Hepp Z, Shah SN, Smoyer K, et al: Epidemiology and treatment patterns for locally advanced or metastatic urothelial carcinoma: A systematic literature review and gap analysis. J Manag Care Spec Pharm 27:240-255, 2021
4. Alimohamed N, Grewal S, Wirtz HS, et al: Understanding treatment patterns and outcomes among patients with de novo unresectable locally advanced or metastatic urothelial cancer: A population-level retrospective analysis from Alberta, Canada. Curr Oncol 29:7587-7597, 2022
5. Said N: Roles of SPARC in urothelial carcinogenesis, progression and metastasis. Oncotarget 7:67574-67585, 2016
6. Rosen RD, Sapra A: TNM Classification. Treasure Island, FL, StatPearls, 2024. Disclosure: Amit Sapra declares no relevant financial relationships with ineligible companies, StatPearls Publishing Copyright © 2024, StatPearls Publishing LLC
7. SEER*Explorer: An Interactive Website for SEER Cancer Statistics. Surveillance Research Program. National Cancer Institute. https://seer.cancer.gov/statistics-network/explorer/
8. Powles T, Bellmunt J, Comperat E, et al: Bladder cancer: ESMO clinical practice guideline for diagnosis, treatment and follow-up. Ann Oncol 33:244-258, 2022
9. Brooks GA, Bergquist SL, Landrum MB, et al: Classifying stage IV lung cancer from health care claims: A comparison of multiple analytic approaches. JCO Clin Cancer Inform 10.1200/CCI.18.00156
10. Esposito DB, Russo L, Oksen D, et al: Development of predictive models to identify advanced-stage cancer patients in a US healthcare claims database. Cancer Epidemiol 61:30-37, 2019
11. Bergquist SL, Brooks GA, Keating NL, et al: Classifying lung cancer severity with ensemble machine learning in health care claims data. Proc Mach Learn Res 68:25-38, 2017
12. Smith GL, Shih YCT, Giordano SH, et al: A method to predict breast cancer stage using Medicare claims. Epidemiol Perspect Innov 7:1, 2010
13. Mamtani R, Haynes K, Boursi B, et al: Validation of a coding algorithm to identify bladder cancer and distinguish stage in an electronic medical records database. Cancer Epidemiol Biomarkers Prev 24:303-307, 2015
14. FDA: FDA approves enfortumab vedotin-ejfv with pembrolizumab for locally advanced or metastatic urothelial cancer. https://www.fda.gov/drugs/resources-information-approved-drugs/fda-approves-enfortumab-vedotin-ejfv-pembrolizumab-locally-advanced-or-metastatic-urothelial-cancer
15. FDA: FDA approves nogapendekin alfa inbakicept-pmln for BCG-unresponsive non-muscle invasive bladder cancer. https://www.fda.gov/drugs/resources-information-approved-drugs/fda-approves-nogapendekin-alfa-inbakicept-pmln-bcg-unresponsive-non-muscle-invasive-bladder-cancer
16. FDA: FDA approves nivolumab in combination with cisplatin and gemcitabine for unresectable or metastatic urothelial carcinoma. https://www.fda.gov/drugs/resources-information-approved-drugs/fda-approves-nivolumab-combination-cisplatin-and-gemcitabine-unresectable-or-metastatic-urothelial
17. National Cancer Institute: SEER-medicare: Brief description of the SEER-medicare database. https://healthcaredelivery.cancer.gov/seermedicare/overview/
18. Landis JR, Koch GG: The measurement of observer agreement for categorical data. Biometrics 33:159-174, 1977
19. Weinstein EJ, Ritchey ME, Lo Re V III: Core concepts in pharmacoepidemiology: Validation of health outcomes of interest within real-world healthcare databases. Pharmacoepidemiol Drug Saf 32:1-8, 2023
20. Lanes S, Beachler DC: Validation to correct for outcome misclassification bias. Pharmacoepidemiol Drug Saf 32:700-703, 2023
21. Monaghan TF, Rahman SN, Agudelo CW, et al: Foundational statistical principles in medical research: Sensitivity, specificity, positive predictive value, and negative predictive value. Medicina (Kaunas) 57:503, 2021
22. Taylor JA III, Kuchel GA: Bladder cancer in the elderly: Clinical outcomes, basic mechanisms, and future research direction. Nat Clin Pract Urol 6:135-144, 2009
23. Skinner EC: Treatment of muscle-invasive bladder cancer in older patients. Am Soc Clin Oncol Educ Book 35:e228-e233, 2016
24. Balakrishnan AS, Washington SL III, Meng MV, et al: Determinants of guideline-based treatment in patients with cT1 bladder cancer. Clin Genitourin Cancer 17:e461-e471, 2019