Published in final edited form as:

Epidemiology. 2019 May ; 30(3): 466–471. doi:10.1097/EDE.0000000000001007.

Validation of an Algorithm for Claims-based Incidence of Prostate Cancer

LE Parlett^a, DC Beachler^b, S Lanes^b, RN Hoover^c, and MB Cook^c

^aHealthCore, Inc., Alexandria, VA

^bHealthCore, Inc., Wilmington, DE

^cNational Cancer Institute, NIH, DHHS, Bethesda MD

Abstract

Background: Prostate cancer is a commonly studied outcome in administrative claims studies, but there is a dearth of validated case-identifying algorithms. The long-term development of the disease increases the difficulty in separating prevalent from incident prostate cancer. The purpose of this validation study was to assess the accuracy of a claims algorithm to identify incident prostate cancer among men in commercial and Medicare Advantage US health plans.

Methods: We identified prostate cancer in claims as a prostate cancer diagnosis within 28 days after a prostate biopsy and compared case ascertainment in the claims with the gold standard results from the Georgia Comprehensive Cancer Registry (GCCR).

Results: We identified 74,008 men from a large health plan claims database for possible linkage with GCCR. Among the 382 prostate cancer cases identified in claims, 312 were also identified in the GCCR (positive predictive value [PPV]=82%). Of the registry cases, 91% (95%CI=88, 94) were correctly identified in claims. Claims and registry diagnosis dates of prostate cancer matched exactly in 254/312 (81%) cases. Nearly half of the false positive cases also had claims for prostate cancer treatment. Thirteen (43%) false negative cases were classified as noncases by virtue of having a biopsy and diagnosis greater than 28 days apart as required by the algorithm. Compared to matches, false negative cases were older men with less aggressive prostate cancer.

Conclusions: Our algorithm demonstrated a PPV of 82% with 92% sensitivity in ascertaining incident PC. Administrative health plan claims can be a valuable and accurate source to identify incident prostate cancer cases.

Keywords

Validation Studies; Prostatic Neoplasms; Male; Registries; Algorithms; Administrative Claims; Healthcare

Corresponding Author: Lauren Parlett, 1925 Ballenger Ave, Suite 540, Alexandria, VA 22134, 703-236-6321, lparlett@healthcore.com.

Obtaining data/code: These data are not available for request due to patient privacy rules.

BACKGROUND

Prostate cancer is the most frequently diagnosed cancer in men in the United States,[1] with a yearly incidence rate of about 120/100,000.[2] Despite its relatively high incidence, large populations are often needed to obtain sufficient numbers of cases. Administrative claims databases offer large real-world populations, but they are not designed for research, so diagnostic codes can be inaccurate.[3] For example, false positives could occur if a provider records a prostate cancer diagnosis as a reason for a screening test or a presumptive diagnosis. Additionally, prostate cancer is often indolent and slow progressing and watchful waiting is a common strategy, thus a diagnosis code could indicate a prevalent prostate cancer. [4]

Previous studies conducted in claims databases have used different algorithms to identify prostate cancer,[5-13] and prior validation studies have been limited.[5, 6] One validation study identified prostate cancer using one diagnosis code.[5] The study observed a sensitivity and positive predictive value (PPV) of 73% and 70% respectively, and did not report the accuracy of the diagnosis date of prostate cancer in claims. Another study used diagnosis and treatment codes for its algorithm, and reported a sensitivity of 88%, but also did not report accuracy of diagnosis date.[6] Both prior validation studies occurred among men 65 or older in the Surveillance, Epidemiology and End Results (SEER)-Medicare environment and were conducted more than 20 years ago. Since then, prostate cancer diagnosis, screening, and treatment guidelines have changed.[14-16]

The purpose of this validation study was to assess the accuracy of a claims algorithm to identify incident prostate cancer using data from commercial and Medicare Advantage US health plans from 2007-2015.

METHODS

The HealthCore Integrated Research Database (HIRD®) is a longitudinally integrated database of medical and pharmacy claims and electronic laboratory results from commercial plans in 14 states.

To validate prostate cancer diagnoses in claims, we sought to use cancer registries as a gold standard among the most heavily populated states in the HIRD, California and Georgia. The California cancer registry did not approve this study, so we restricted analyses to the Georgia registry. The Georgia Comprehensive Cancer Registry (GCCR) has collected incident cancer cases among Georgia residents since 1995.[17] All health care providers and medical facilities in the state are mandatory reporters. GCCR also works with neighboring states to capture cancer information for Georgia residents diagnosed or treated out of state.

This validation study was nested within a cohort study examining the relationship between testosterone supplementation and incident prostate cancer. The study included three study groups: men newly exposed to testosterone ("testosterone users"; N=135,454), men newly exposed to phosphodiesterase-5 inhibitors ("PDE5i users"; N=147,620), and testosterone-unexposed men ("testosterone non-users"; N=721,326). Eligible men were at least 30 years old with at least 12 months of enrollment between 1 January 2007 and 31 July 2015 prior to

(and including) the index date (date of medication initiation or matched outpatient visit). The validation was stratified by study group to examine performance in different men with perhaps different levels of healthcare engagement, which would affect detection of prevalent prostate cancer.

We defined prostate cancer cases in claims by a diagnosis for prostate cancer (ICD-9=185) not more than 28 days after a prostate biopsy (HCPCS/CPT=G0416, G0417, G0418, G0419, 55700, 55705, or 55706). The diagnosis date in claims was defined as the biopsy date. To exclude possible prevalent prostate cancer cases and those not at risk for prostate cancer, we identified and excluded men with codes indicating a diagnosis or history of prostate cancer, medical findings strongly correlated with prostate cancer, or procedures associated with prostate cancer prior to their start of follow-up (full exclusions and codes in eTable 2).

We submitted to the GCCR men who had ever resided in Georgia between 2006 and 2015. GCCR first linked the member data to their registry data using exact matches for name, date of birth, and social security number (SSN). No medical information was submitted. Partial matches and patients missing SSN were visually reviewed for errors (e.g. SSN transposition), and other fields, including date of birth and other protected health information (PHI) such as names, county of residence, and sex were employed and visually inspected before ascertaining matching status. The New England and Georgia Department of Public Health Institutional Review Boards approved the registry linkage protocol.

We calculated the following measures of agreement: unweighted kappa, PPV, and sensitivity, as well as their 95% confidence intervals. The PPV was calculated as the number of incident prostate cancer cases identified in claims and confirmed in the registry divided by the total number identified in claims. We calculated sensitivity as the number of cases identified in claims and confirmed in the registry divided by the total number of cases identified in the registry. Characteristics of cases are presented as frequencies and means (SD). All data were analyzed using SAS Enterprise Guide 7.1 (SAS, Cary, NC).

RESULTS

We submitted to the GCCR 74,008 men, including 382 men identified in the claims database as cases of prostate cancer. After registry linkage, GCCR returned 342 prostate cancer cases. Of the registry cases, 91% (95%CI=88, 94) were correctly identified in claims (Table 1). The PPV of the claims algorithm was 82% (78, 85). Stratifying by study group, the highest PPV was in men who had not received testosterone (83%), which accounted for 78% of our overall sample (data in eAppendix 1). The linkage resulted in negative predictive values and specificities approximating 100%.

When the gap between biopsy and diagnosis is lengthened, the sensitivity and PPV values show minor changes (Table 2). Increased gap time is associated with an increase in sensitivity, and decrease in PPV. The group with the largest magnitude of change was in testosterone users where PPV decreased 4% and sensitivity remained unchanged.

Of the 70 false positives, 17 cases (24%) did not have a valid SSN. Thirty-nine men had strong evidence of prostate cancer screening, diagnosis, and/or treatment according to

diagnosis and procedure codes in claims (including 10 men without a valid SSN). Several false positive cases had provider claims for a biopsy and diagnosis from another state.

There was close agreement of diagnosis date, with perfect agreement for 254/312 (81%) of cancer cases and agreement within 2 weeks for 90% (Table 3). It was more common to have a claims cancer date earlier than a registry cancer date (13% vs. 6%). The claims algorithm did not detect 30/343 (9%) GCCR-identified prostate cancer cases. Thirteen (43%) of these 30 cases had an interval between biopsy and prostate cancer diagnosis longer than the stipulated 28 days. The remaining 17 cases lacked a recorded biopsy procedure, recorded prostate cancer diagnosis, or both.

Table 4 provides demographic and clinical characteristics of the prostate cancer cases according to false negative status. Generally, cases not identified in claims were older and had lower tumor grades and Gleason scores.

DISCUSSION

A prostate cancer diagnosis within 28 days following a prostate biopsy accurately identified 91% of incident prostate cancer cases with 9% false positives compared to the GCCR, suggesting this definition may be useful for prostate cancer studies using claims databases. The differences in algorithm performance among our study groups were generally small, but men dispensed PDE5i medications had a PPV about 6% less than the other groups.

The diagnosis dates of prostate cancer in claims and the registry were similar, indicating a low proportion of misclassifying prevalent cases as incident cases. When the diagnosis dates were different, the claims date more frequently preceded the registry date, which is a phenomenon reported in other studies.[18] This could be the result of the claims date being defined as the biopsy date, which technically precedes the diagnosis date.

Study exclusion criteria were broad and used all available baseline time in order to exclude prevalent prostate cancer cases. Using fewer exclusions and limited baseline time would have likely introduced prevalent cases into our population.[19, 20] All men had at least 1 year of baseline data and more than half had at least three years of baseline data. For one study of various cancers (not including prostate cancer), increasing time for cancer-free periods improved PPV and sensitivity when distinguishing incident and recurrent cases.[18] The observation time for the men in our study is similar to other commercial health plans, so similar baseline periods should be available in other claims environments.[21]

Failure to identify prostate cancer cases in claims was primarily due to a time interval between biopsy and prostate cancer diagnosis longer than 28 days. Previous literature reports use of a similar, though not validated, algorithm that required biopsy and ICD-9 diagnosis, but with a time interval of up to 3 months.[10] We originally restricted the time between prostate biopsy and cancer diagnosis to address the concern that a cancer diagnosis more than one month after a biopsy would be less likely associated with a biopsy procedure and more likely attributable to prevalent cancer. Extending the time period generally increased sensitivity at the cost of PPV. When evaluating the false negative cases, missed cases in claims were older than those correctly identified, as had been seen in a previous

study.[5] Furthermore, missed cases had either less aggressive or less developed prostate cancer as evidenced by more unknown tumor grades, lower Gleason scores, and lower tumor stages.

Previous efforts to evaluate prostate cancer algorithms in claims have been conducted in the SEER-Medicare data environment with moderate success.[5, 6] Our study included commercial health plan data, which allowed us to assess younger men (<65). In addition, prior claims-defined prostate cancer validation studies have study periods in the 1980–90s, while this study period was 2007–2015.[5, 6] These newer data are more indicative of current clinical practice because prostate cancer screening guidelines have changed.[14] Additionally, this study is to our knowledge the first to comprehensively evaluate the accuracy of the diagnosis date of prostate cancer ascertainment in claims, and the finding that the dates between claims and the registry are largely similar further supports the validity of this algorithm.

The claims environment has several limitations for identifying incident prostate cancer cases. Our analysis of reasons for algorithm errors demonstrates that claims may not capture a small fraction of incident cancer cases due to lack of recorded biopsy or diagnosis. Another limitation that could affect PPV is the quality of data linkage between claims and the GCCR; an inability to link claims cases to actual registry cases would reduce the PPV of the claims algorithm.

This validation study used one state registry, which may limit generalizability due to data linkage practices, cancer prevalence, case ascertainment, and billing practices. While GCCR has attained the gold standard assigned by the North American Association of Central Cancer Registries, the collection process of prostate cancer cases may differ from other states.[22] GCCR did not capture all of the cases with further evidence of treatment as identified by the algorithm. It is possible that these cases, identified here as false positives according to our GCCR standard, are true cases captured in other state cancer registries. Yet, using even a single state registry has advantages over typical medical record validation studies including its large size and ability to assess sensitivity.

Our prostate cancer algorithm uses ICD-9 codes. Since 2015, ICD-10 codes have been introduced. Because ICD-9 and ICD-10 codes have a one-to-one equivalence for prostate cancer, the diagnostic code C61 should perform similarly when using ICD-10 codes; however, the exclusion criteria must also be updated to ICD-10. The increased granularity of ICD-10 may provide more precise criteria to exclude prevalent prostate cancer cases while retaining incident cases. Thus, future validation studies should test the robustness of the algorithm after updating codes for recent periods (post-2015) and evaluate its performance in other study populations.

In conclusion, this study exhibited that administrative health plan claims can be a valuable and accurate source to identify incident prostate cancer cases. In addition, the case ascertainment algorithm will reduce prostate cancer misclassification compared to a single diagnostic code, thereby increasing internal study validity for future studies performed using claims.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments:

GA Department of Public Health (Rana Bayakly), Bolarinwa Ekezie, Gayathri Sridhar, Philip Cochetti, Anthem (HealthCore, Inc. is a wholly-owned subsidiary of Anthem, Inc.)

Financial support: HealthCore researchers operated under an NCI/NIH contract specific to this study; NCI/NIH researchers performed research under their normal governmental duties.

REFERENCES

1. U.S. Cancer Statistics Working Group. United States Cancer Statistics: 1999–2014 Incidence and Mortality Web-based Report. 2017; Available from: <https://nccd.cdc.gov/USCSDDataViz/rdPage.aspx>.
2. Surveillance, Epidemiology, and End Results (SEER) Program (www.seer.cancer.gov), SEER*Stat Database: Incidence - SEER 9 Regs Research Data, Nov 2016 Sub (1973–2014) National Cancer Institute, DCCPS, Surveillance Research Program, Surveillance Systems Branch.
3. Lanes S, et al., Identifying health outcomes in healthcare databases. *Pharmacoepidemiol Drug Saf*, 2015 24(10): p. 1009–16. [PubMed: 26282185]
4. Czwikla J, Jobski K, and Schink T, The impact of the lookback period and definition of confirmatory events on the identification of incident cancer cases in administrative data. *BMC Med Res Methodol*, 2017 17(1): p. 122. [PubMed: 28806932]
5. McClish DK, et al., Ability of Medicare claims data and cancer registries to identify cancer cases and treatment. *Am J Epidemiol*, 1997 145(3): p. 227–33. [PubMed: 9012595]
6. Cooper GS, et al., The sensitivity of Medicare claims data for case ascertainment of six common cancers. *Med Care*, 1999 37(5): p. 436–44. [PubMed: 10335746]
7. Newschaffer CJ, et al., Causes of death in elderly prostate cancer patients and in a comparison nonprostate cancer cohort. *J Natl Cancer Inst*, 2000 92(8): p. 613–21. [PubMed: 10772678]
8. Etzioni R, et al., Overdiagnosis due to prostate-specific antigen screening: lessons from U.S. prostate cancer incidence trends. *J Natl Cancer Inst*, 2002 94(13): p. 981–90. [PubMed: 12096083]
9. Jang TL, et al., Physician visits prior to treatment for clinically localized prostate cancer. *Arch Intern Med*, 2010 170(5): p. 440–50. [PubMed: 20212180]
10. Perlroth DJ, et al., An economic analysis of conservative management versus active treatment for men with localized prostate cancer. *J Natl Cancer Inst Monogr*, 2012 2012(45): p. 250–7. [PubMed: 23271781]
11. Lafeuille MH, et al., Real-World Corticosteroid Utilization Patterns in Patients with Metastatic Castration-Resistant Prostate Cancer in 2 Large US Administrative Claims Databases. *Am Health Drug Benefits*, 2013 6(6): p. 307–16. [PubMed: 24991366]
12. Crawford ED, et al., A retrospective analysis illustrating the substantial clinical and economic burden of prostate cancer. *Prostate Cancer Prostatic Dis*, 2010 13(2): p. 162–7. [PubMed: 20125121]
13. Smith MR, et al., Gonadotropin-releasing hormone agonists and fracture risk: a claims-based cohort study of men with nonmetastatic prostate cancer. *J Clin Oncol*, 2005 23(31): p. 7897–903. [PubMed: 16258089]
14. The American Cancer Society medical and editorial content team. History of ACS Recommendations for the Early Detection of Cancer in People Without Symptoms. 7 7, 2017 [cited 2018; Available from: <https://www.cancer.org/healthy/find-cancer-early/cancer-screening-guidelines/chronological-history-of-ac-s-recommendations.html>.
15. Fleshner K, Carlsson SV, and Roobol MJ, The effect of the USPSTF PSA screening recommendation on prostate cancer incidence patterns in the USA. *Nat Rev Urol*, 2017 14(1): p. 26–37. [PubMed: 27995937]

16. Maurice MJ and Abouassaly R, Patient opinions on prostate cancer screening are swayed by the United States Preventative Services Task Force recommendations. *Urology*, 2014 84(2): p. 295–9. [PubMed: 24929945]
17. Official Code of Georgia Annotated, Reporting certain diseases and neonatal abstinence syndrome; confidentiality; reporting required of pharmacists; immunity from liability as to information supplied; notification of potential bioterrorism, in 31.
18. Setoguchi S, et al., Agreement of diagnosis and its date for hematologic malignancies and solid tumors between medicare claims and cancer registry data. *Cancer Causes Control*, 2007 18(5): p. 561–9. [PubMed: 17447148]
19. Kent ST, et al., Optimal use of available claims to identify a Medicare population free of coronary heart disease. *Am J Epidemiol*, 2015 182(9): p. 808–19. [PubMed: 26443420]
20. Brunelli SM, et al., Estimation using all available covariate information versus a fixed look-back window for dichotomous covariates. *Pharmacoepidemiol Drug Saf*, 2013 22(5): p. 542–50. [PubMed: 23526818]
21. Barnett ML, et al., Insurance Transitions and Changes in Physician and Emergency Department Utilization: An Observational Study. *J Gen Intern Med*, 2017 32(10): p. 1146–1155. [PubMed: 28523475]
22. North American Association of Central Cancer Registries. Certified Registries. 2017; Available from: <https://www.naaccr.org/certified-registries/>.

Table 1:

Performance of claims algorithm compared to GCCR case ascertainment

	Registry Cases					
Claims Cases	Yes	No	Total	Kappa (95%CI)	Sensitivity (95%CI)	PPV (95%CI)
<i>Entire Cohort</i>						
Yes	312	70	382	0.86 (0.83 – 0.89)	91 (88 – 94)	82 (78 – 85)
No	30	73,596	73,626			
Total	342	73,664	74,008			
<i>Testosterone Non-users</i>						
Yes	218	46	264	0.86 (0.83 – 0.90)	91 (87 – 94)	83 (78 – 86)
No	22	51,669	51,691			
Total	240	51,715	51,955			
<i>Testosterone Users</i>						
Yes	44	10	54	0.86 (0.79 – 0.93)	92 (80 – 98)	82(70 – 89)
No	4	12,438	12,442			
Total	48	12,448	12,496			
<i>PDE5i Users</i>						
Yes	57	20	77	0.81 (0.74 – 0.88)	89 (79 – 96)	74 (65 – 82)
No	7	11,791	11,798			
Total	64	11,811	11,875			

Abbreviations: GCCR – Georgia Comprehensive Cancer Registry; PPV – positive predictive value; CI – confidence interval; PDE5i – phosphodiesterase-5-inhibitor

Definitions: Kappa= (agreement observed – agreement expected) / (1-agreement expected); Sensitivity = true positives / registry cases; PPV = true positives / claims cases

Table 2:

Prostate cancer diagnosis algorithm with different day allowances between biopsy and diagnosis

	28 days (original)		45 days		60 days		90 days	
Group	Sensitivity (95%CI)	PPV (95%CI)	Sensitivity (95%CI)	PPV (95%CI)	Sensitivity (95%CI)	PPV (95%CI)	Sensitivity (95%CI)	PPV (95%CI)
Total	91 (88 - 94)	82 (78 - 86)	92 (90 - 95)	81 (78 - 85)	93 (90 - 96)	81 (77 - 85)	94 (91 - 96)	81 (77 - 85)
Testosterone Users	91 (87 - 95)	83 (78 - 87)	93 (89 - 96)	83 (78 - 87)	93 (89 - 96)	83 (78 - 87)	94 (91 - 97)	83 (79 - 88)
Testosterone Non-users	92 (84 - 100)	82 (71 - 92)	92 (84 - 100)	82 (71 - 92)	92 (84 - 100)	79 (68 - 89)	92 (84 - 100)	77 (66 - 88)
PDE5i Users	89 (81 - 97)	74 (64 - 84)	89 (81 - 97)	72 (62 - 82)	89 (81 - 97)	72 (62 - 82)	89 (81 - 97)	72 (62 - 82)

PDE5i – phosphodiesterase-5-inhibitor; PPV – positive predictive value; CI – confidence interval

Table 3:

Cancer diagnosis date difference between claims and registry (N=312)

Number of days the claims cancer date was from the registry cancer date	N	%
61-91 days before cancer registry date	9	2.9
31-60 days before cancer registry date	5	1.6
15-30 days before cancer registry date	5	1.6
4-14 days before cancer registry date	7	2.2
1-3 days before cancer registry date	14	4.5
0 days (same day as cancer registry date)	254	81
1-14 days after cancer registry date	7	2.2
15-56 days after cancer registry date	5	1.6
57-150 days after cancer registry date	6	1.9

Table 4:

GCCR case characteristics according to identification in claims data, N=342

Characteristics	Total (N=342)	False Negative (FN) ^a (N=30)	True Positive (TP) ^b (N=312)	Difference in proportions (FN- TP) (95% CI)
Age at diagnosis, Mean (SD)	60.9 (7.5)	62.7 (11.8)	60.7 (7.0)	2.0 (-2.4, 6.5)
Age at diagnosis, N(%)				
40-59	148 (44%)	12 (40%)	136 (43%)	-4 (-22, 15)
60-69	156 (46%)	10 (33%)	146 (47%)	-13 (-31, 4)
70+	38 (8%)	8 (27%)	30 (10%)	17 (1,33)
Race				
White	270 (79%)	25 (83%)	245 (79%)	5 (-9, 19)
Black / other	72 (21%)	5 (17%)	67 (21%)	-5 (-19, 9)
Diagnosis Year				
2007-2009	67 (20%)	7 (23%)	60 (19%)	4 (-12, 20)
2010-2012	150 (44%)	11 (37%)	139 (45%)	-8 (-26, 10)
2013-2015	125 (36%)	12 (40%)	113 (36%)	4 (-15, 22)
Deceased	23 (7%)	3 (10%)	20 (6%)	4 (-7, 15)
Tumor grade				
1	27 (8%)	5 (17%)	22 (7%)	10 (-4, 23)
2	145 (45%)	11 (37%)	143 (46%)	-10 (-27, 9)
3	144 (42%)	8 (27%)	136 (44%)	-17 (-34, -0)
Unknown	17 (5%)	6 (20%)	11 (4%)	16 (2, 31)
Treatment status				
Missing / unknown	69 (20%)	7 (23%)	62 (19%)	4 (-8, 22)
Not Missing	273 (80%)	23 (77%)	250 (81%)	-3 (-19, 12)
No treatment	52 (19%)	5 (22%)	47 (19%)	3 (-10, 24)
Treated	208 (76%)	18 (78%)	190 (76%)	2 (-19, 15)
Watchful waiting	13 (5%)	0	13 (5%)	-5 (-8, -2)
Surgical treatment				
Surgery performed	157 (46%)	18 (60%)	139 (45%)	15 (-3, 34)
Surgery not first line treatment	179 (52%)	10 (33%)	169 (54%)	-21 (-39, -3)
Radiation treatment				
Radiation performed	111 (32%)	3 (11%)	108 (35%)	-25 (-37, -11)
Radiation not first line treatment	206 (60%)	23 (77%)	183 (59%)	18 (2, 34)
Other	25 (8%)	4 (13%)	21 (6%)	7 (-6, 19)
PSA levels				
Elevated	288 (84%)	15 (50%)	273 (88%)	-38 (-56, -19)
Normal	13 (4%)	2 (7%)	11 (4%)	3 (-6, 12)
Low	9 (3%)	2 (7%)	7 (2%)	4 (-5, 13)

Characteristics	Total (N=342)	False Negative (FN) ^a (N=30)	True Positive (TP) ^b (N=312)	Difference in proportions (FN- TP) (95% CI)
Unknown	32 (9%)	11 (37%)	21 (7%)	30 (12, 47)
Gleason score on needle core				
Not applicable /not performed/unknown	76 (22%)	13 (43%)	63 (20%)	23 (5, 41)
Performed	266 (78%)	17 (57%)	249 (80%)	-23 (-41, -5)
6	132 (50%)	10 (59%)	122 (49%)	10 (-14, 34)
7	106 (31%)	7 (41%)	99 (40%)	1 (-19, 25)
8	16 (6%)	0	16 (6%)	-6 (-9, -3)
9+	12 (5%)	0	12 (5%)	-5 (-7, -2)
Lymph nodes removed in surgery / biopsy				
None	289 (85%)	25 (83%)	264 (85%)	-1 (-15, 13)
1 - 3	21 (6%)	1 (3%)	20 (6%)	-3 (-10, 4)
4+	31 (9%)	4 (13%)	27 (9%)	5 (-8, 17)
Unknown	1 (0%)	0	1 (0%)	0

Abbreviation: GCCR – Georgia Comprehensive Cancer Registry; PSA – prostate specific antigen

^aFalse negatives (FN) were identified in the registry and not in the claims

^bTrue positives (TP) were identified in both the registry and claims