Diagnostic Accuracy of Code-Based Algorithms to Identify Urinary Tract Infection in U.S. Administrative Claims Databases

Stephen P Fortin, Jeroen Geurtsen, Michal Sarnecki, Joachim Doua, Jamie Colasurdo, Joel Swerdel

Background

Occurring in nearly half of women over the course of their lifetime, urinary tract infections (UTI) are among the most common infections in the United States[1,2,3]. Administrative claims databases are widely used to perform observational research on disease burden, epidemiology, treatment pathways, and healthcare resource utilization and costs. Nevertheless, limited prior literature exists assessing the performance characteristics of code-based algorithms to identify UTIs, and prior publications are mainly limited to single center studies or studies of pediatric patients[4,5]. The current study evaluates the performance characteristics of 10 code-based algorithms to identify UTI, derived from diagnosis, medication, and clinical procedure codes, among adult patients in three large U.S. administrative claims databases.

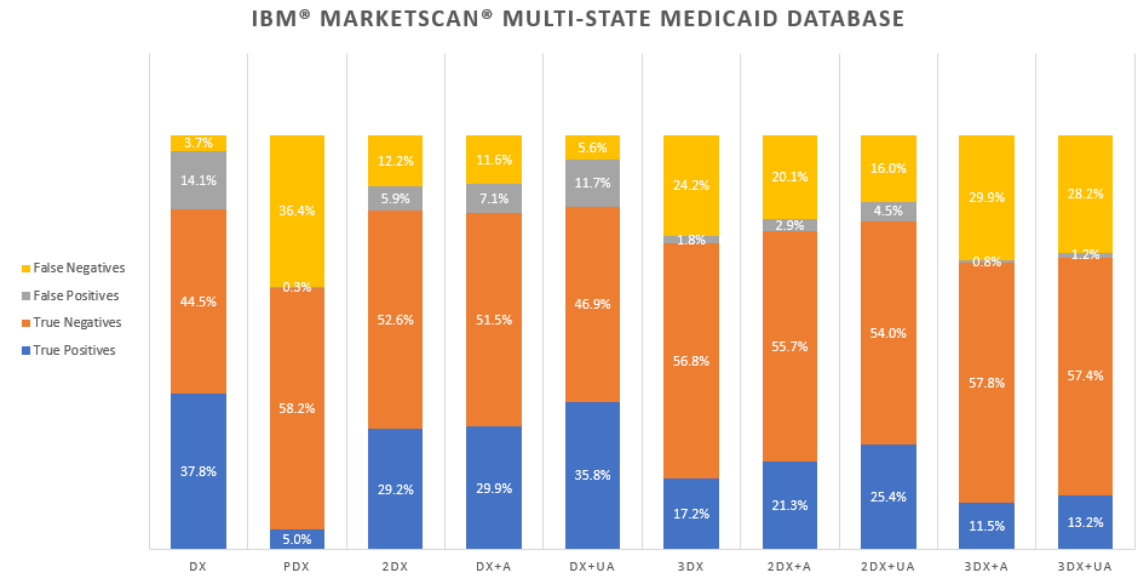
Methods

We identified all patients observed within IBM® MarketScan® Multi-State Medicaid Database (MDCD); IBM® MarketScan® Medicare Supplemental and Coordination of Benefits Database (MDCR); and IBM® MarketScan® Commercial Claims and Encounters Database (CCAE) between 01-01-2010 to 12-31-2019 (MCD) and 10-31-2020 (MDCR and CCAE). The study included patients with at least 365 days of prior observation in their claims record aged ≥18 years in MDCD and CCAE, and ≥66 years in MDCR. A systematic literature review was performed to inform the development of 10 code-based algorithms to identify UTI alongside clinical subject matter expert input. Algorithms were required to possess at least a single UTI diagnosis code and included combinations of a primary UTI diagnosis code, multiple UTI diagnosis codes, exposure to antibiotics used in the treatment of UTI, and/or a urinalysis or urine culture (regardless of outcome). For each database, a probabilistic gold standard based on diagnostic predictive models for UTI was developed using the PheValuator tool[6]. The probabilistic gold standards were used to assess the performance characteristics of code-based algorithms; specifically, the sensitivity, positive predictive value (PPV), specificity, and negative predictive value (NPV) of each algorithm were measured. Furthermore, the F_1 score, defined as 2*PPV*sensitivity / (PPV + sensitivity), was measured.

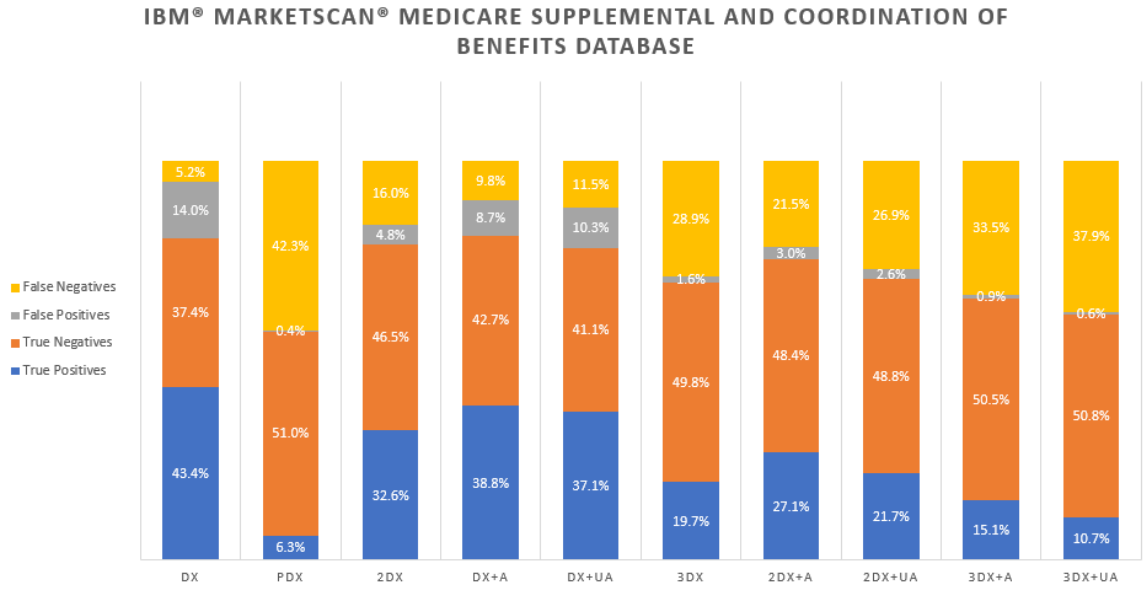
Results

A total of 2,950,641, 1,831,405 and 2,294,929 patients meeting the study criteria were identified in MDCD, MDCR and CCAE, respectively. Furthermore, 14,230, 11,613 and 15,274 covariates were included in the diagnostic predictive models for each respective database. The performance characteristics of code-based algorithms is summarized in **Figures 1-3**.

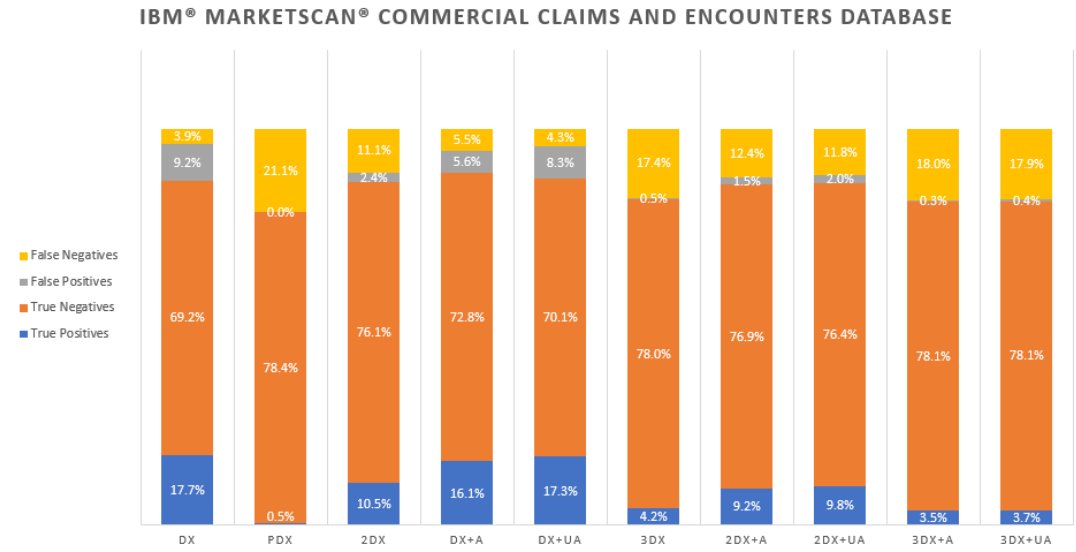
Figures 1-3. Diagnostic accuracy of code-based algorithms in MDCD (top), MDCR (middle), and CCAE (bottom)



	DX	PDX	2DX	DX+A	DX+UA	3DX	2DX+A	2DX+UA	3DX+A	3DX+UA
Sensitivity	91.1%	12.0%	70.6%	72.1%	86.5%	41.6%	51.5%	61.3%	27.7%	31.9%
PPV	72.8%	93.8%	83.1%	80.9%	75.4%	90.5%	88.1%	84.8%	93.6%	91.9%
Specificity	76.0%	99.4%	89.8%	87.9%	80.1%	96.9%	95.1%	92.2%	98.7%	98.0%
NPV	92.4%	61.5%	81.2%	81.7%	89.3%	70.1%	73.5%	77.1%	65.9%	67.1%
F1 score	81.0%	21.3%	76.3%	76.2%	80.6%	57.0%	65.0%	71.2%	42.8%	47.4%



	DX	PDX	2DX	DX+A	DX+UA	3DX	2DX+A	2DX+UA	3DX+A	3DX+UA
Sensitivity	89.2%	12.9%	67.1%	79.8%	76.4%	40.5%	55.7%	44.7%	31.1%	22.1%
PPV	75.6%	94.4%	87.1%	81.7%	78.3%	92.5%	90.0%	89.2%	94.2%	94.4%
Specificity	72.8%	99.3%	90.6%	83.1%	79.9%	96.9%	94.1%	94.9%	98.2%	98.8%
NPV	87.7%	54.7%	74.4%	81.3%	78.1%	63.3%	69.2%	64.5%	60.1%	57.3%
F1 score	81.9%	22.7%	75.8%	80.7%	77.3%	56.4%	68.8%	59.5%	46.8%	35.8%



	DX	PDX	2DX	DX+A	DX+UA	3DX	2DX+A	2DX+UA	3DX+A	3DX+UA
Sensitivity	81.9%	2.3%	48.7%	74.5%	80.0%	19.3%	42.6%	45.4%	16.3%	17.0%
PPV	65.8%	93.1%	81.6%	74.2%	67.5%	89.9%	85.6%	82.8%	92.1%	90.9%
Specificity	88.3%	100.0%	97.0%	92.9%	89.4%	99.4%	98.0%	97.4%	99.6%	99.5%
NPV	94.7%	78.8%	87.3%	93.0%	94.2%	81.8%	86.1%	86.7%	81.2%	81.4%
F1 score	73.0%	4.5%	61.0%	74.3%	73.2%	31.8%	56.9%	58.7%	27.8%	28.7%

DX: single UTI diagnosis code; PDX: single primary UTI diagnosis code; 2DX: 2 UTI diagnosis codes; DX+A: single UTI diagnosis code and antibiotic exposure; DX+UA: single UTI diagnosis code and urinalysis or urine culture; 3DX: 3 UTI diagnosis codes; 2DX+A: 2 UTI diagnosis codes and antibiotic exposure; 2DX+UA: 2 UTI diagnosis codes and urinalysis or urine culture; 3DX+A: 3 UTI diagnosis codes and antibiotic exposure; 3DX+UA: 3 UTI diagnosis codes and urinalysis or urine culture

Although the prevalence of UTI was lower in CCAE, similar trends in the performance characteristics of code-based algorithms were observed across all databases. A single primary UTI diagnosis code achieved both the highest PPV (>93.1%) and lowest sensitivity (<12.9%) translating to a low F_1 score (<22.7%) across all databases. In comparison, algorithms requiring three UTI diagnosis codes achieved high PPV (>89.9%) and improved, albeit low, sensitivity (<41.6%). The algorithm requiring a single UTI diagnosis code achieved sensitivities and PPVs greater than 81.9% and 65.8%, respectively, translating to higher F_1 scores (>73.0%). The requirement for an antibiotic exposure alongside a single diagnosis code led to improvements in PPV (>74.2%) alongside a small reduction in sensitivity (>72.1%) as compared to a single UTI diagnosis code.

Conclusion

We compared the performance characteristics of 10 code-based algorithms to identify UTI in 3 U.S. administrative claims databases. Each algorithm represented a unique trade-off in sensitivity and PPV. The selection of the optimal code-based algorithm should consider the intended application of the algorithm and characteristics of the data source. Based on this study, we recommend algorithms requiring a single UTI diagnosis code, which are associated with improved balance in PPV and sensitivity for studies where high sensitivity is critical. While a single UTI diagnosis code is associated with excellent sensitivity and fair PPV, the additional requirement of antibiotic exposure is associated with improved PPV and fair sensitivity, which may represent an alternative for studies where PPV is critical.

References

1. Foxman B. Epidemiology of urinary tract infections: incidence, morbidity, and economic costs. *Dis Mon*. 2003;49(2):53–70.
2. American College of Obstetricians and Gynecologists ACOG Practice Bulletin No. 91: Treatment of urinary tract infections in nonpregnant women. *Obstet Gynecol*. 2008;111:785–94.
3. Foster RT., Sr Uncomplicated urinary tract infections in women. *Obstet Gynecol Clin North Am*. 2008;35:235–48, viii.Tieder JS, Hall M, Auger KA, et al. Accuracy of administrative billing codes to detect urinary tract infection hospitalizations. *Pediatrics*. 2011;128(2):323-330. doi:10.1542/peds.2010-2064.
4. Tieder JS, Hall M, Auger KA, et al. Accuracy of administrative billing codes to detect urinary tract infection hospitalizations. *Pediatrics*. 2011;128(2):323-330. doi:10.1542/peds.2010-2064.
5. Landers T, Apte M, Hyman S, Furuya Y, Glied S, Larson E. A comparison of methods to detect urinary tract infections using electronic data. *Jt Comm J Qual Patient Saf*. 2010;36(9):411-417. doi:10.1016/s1553-7250(10)36060-0.
6. Swerdel JN, Hripcsak G, Ryan PB. Phevaluator: Development and evaluation of a phenotype algorithm evaluator. *Journal of Biomedical Informatics*. 2019;97:1