

Validated case-identifying algorithms using Canadian administrative health data: a targeted literature review

Bowie AC,¹ Tinajero MG,¹ Qian C¹
¹Broadstreet HEOR, Vancouver BC Canada

Background

- **Case-identifying algorithms** are essential for identifying patients and events in real-world evidence studies
- Few centralized resources exist to support the implementation of case-identifying algorithms
- Single-payer health care systems in Canada provide comprehensive, population-level data to support the development and validation of algorithms

Objective

To identify and summarize published, validated case-identifying algorithms developed using Canadian administrative health data

Methods

- A targeted literature review (TLR) was conducted in Ovid MEDLINE, PubMed, and publications lists of Canadian administrative health databases
- English-language studies were included if they used Canadian data and reported ≥1 of the following performance measures: sensitivity (SE), specificity (SP), positive predictive value (PPV), or negative predictive value (NPV)
- No date restrictions were applied
- Three independent reviewers screened all identified abstracts and extracted study and algorithm details from eligible studies
- Studies were categorized by individual disease and disease areas based on the International Classification of Diseases (ICD)-10 classification system

Abbreviations: CKD, chronic kidney disease; eGFR, estimated glomerular filtration rate; ICD-10, International Classification of Diseases – 10th Revision; NPV, negative predictive value; PPV, positive predictive value; RWE, real world evidence; SE, sensitivity; SP, specificity; TLR, targeted literature review

Results

- **270 studies** were identified, and **82 eligible studies** reporting **1453 algorithms** were summarized
- Of the 82 studies, 48 (58.5%) used data from the most populous province of **Ontario**
- **Chronic kidney disease, diabetes, hypertension, and rheumatoid arthritis** were the most frequent areas of focus (each representing ~4% of included studies)

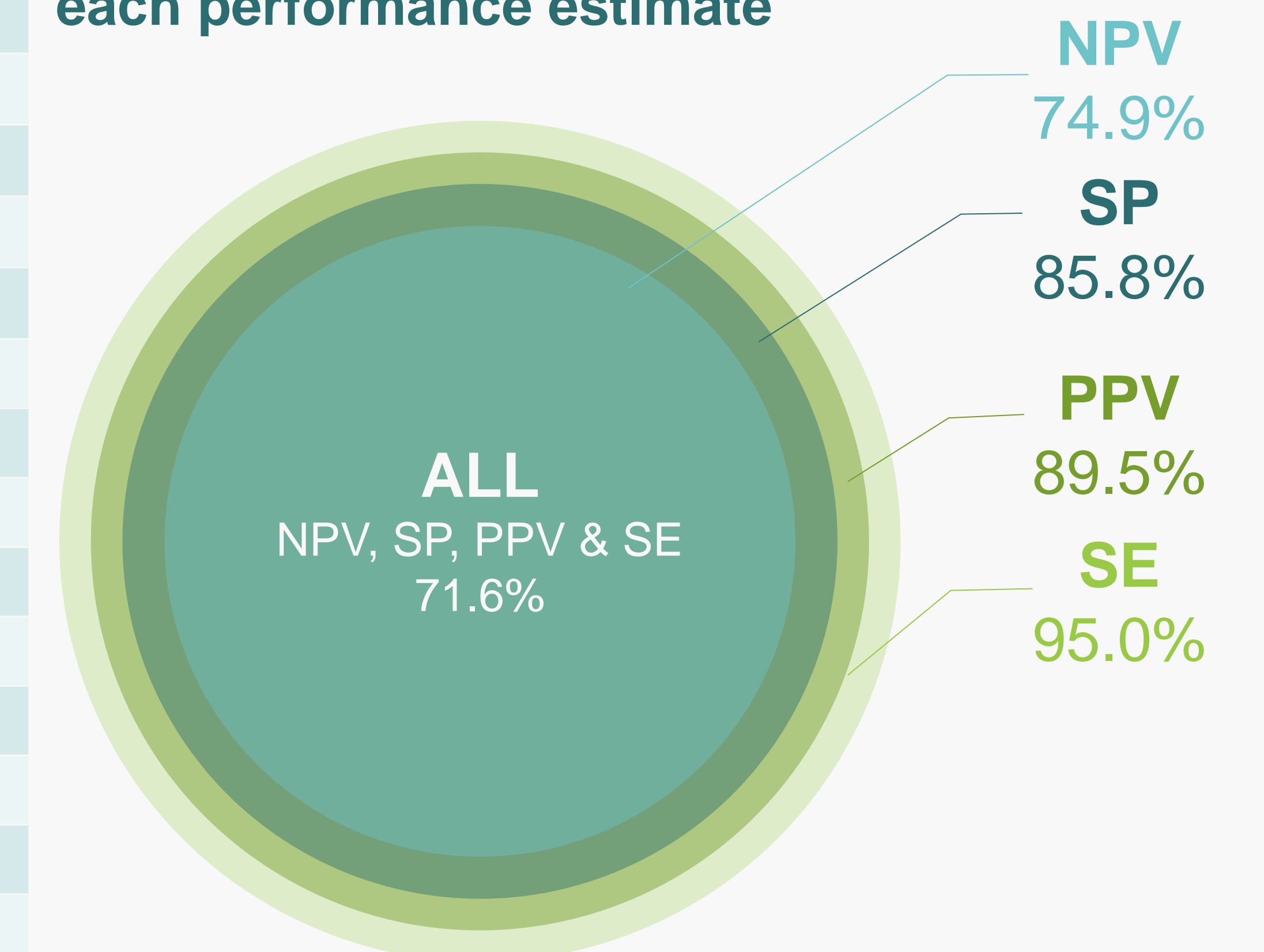
	n (%)
Number of diseases with ≥ 1 algorithm (n = 82)	71 (86.6%)
Number of ICD-10 chapters with ≥ 1 algorithm available (n = 20)	15 (75%)
Algorithm components (n = 1453)	
Diagnosis codes	1341 (92.3%)
Diagnosis codes ONLY	934 (64.3%)
Lab findings	63 (4.3%)
Procedure codes	221 (15.2%)
Medication	225 (15.5%)

- No algorithms were identified for:**
- diseases of the eye or ear
 - conditions originating in the perinatal period
 - external causes of morbidity and mortality
 - symptoms, signs, and abnormal clinical and laboratory findings
- highlighting opportunities for future algorithms*

Relied solely on ICD-9/ICD-10 diagnosis codes - highlighting opportunities for direct application beyond the Canadian context

Availability of performance estimates	Algorithms n = 1453
Any performance measure, n (%)	
Sensitivity (SE)	1380 (95.0%)
Specificity (SP)	1246 (85.8%)
Positive predictive value (PPV)	1300 (89.5%)
Negative predictive value (NPV)	1088 (74.9%)
Specific categories, n (%)	
SE SP PPV NPV	1041 (71.6%)
SE & PPV	131 (9.0%)
SE & SP	125 (8.6%)
SE, SP & PPV	44 (3.0%)
PPV only	43 (3.0%)
SP, PPV & NPV	30 (2.1%)
SE only	22 (1.5%)
SE, PPV & NPV	11 (0.8%)
SE, SP & NPV	6 (0.4%)

Figure 1. Proportion of algorithms reporting each performance estimate

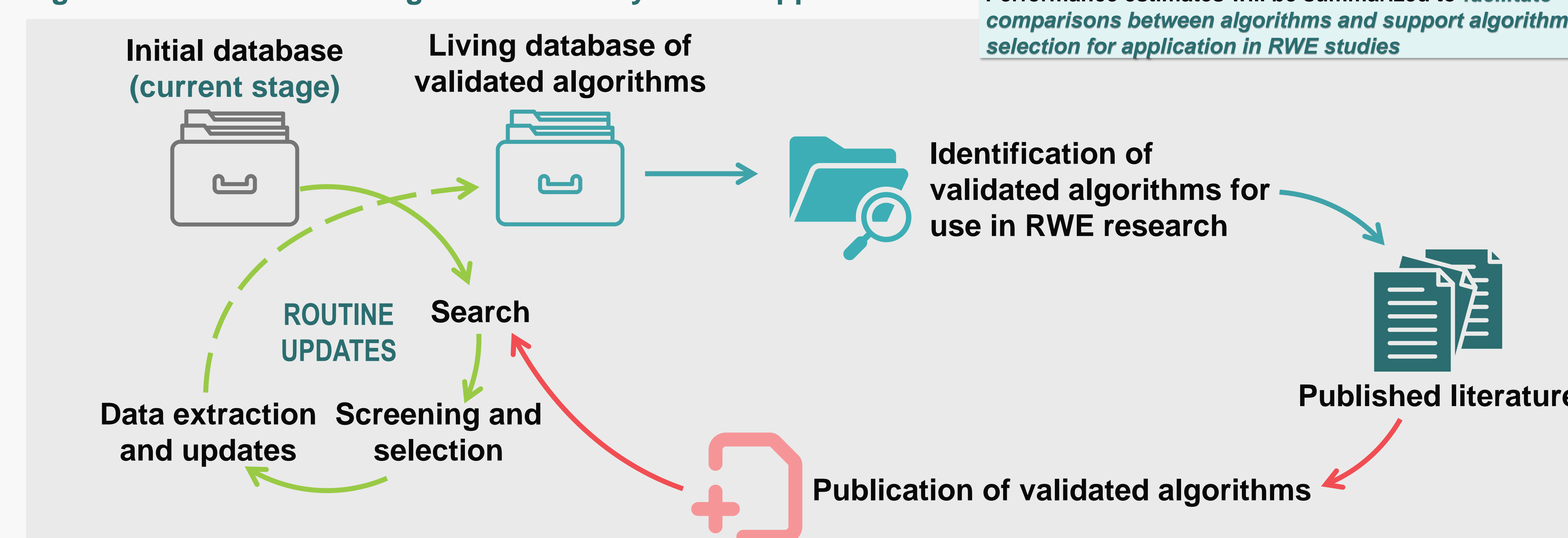


Future applications

Figure 2. Example of database layout

Author (Year)	Disease	Diagnosis	Algorithm components				Description	Performance				Gold standard
			Findings	Procedures	Medication	SE		SP	PPV	NPV		
Fleet (2013)	CKD	ICD-10: N18	No	No	No	≥ 1 hospital discharge OR ≥ 1 physician claim in 5-years	12.24	99.45	80.08	86.17	eGFR <45 mL/min per 1.73 m2	

Figure 3. Flowchart of living database lifecycle and application



Limitations

- Generalizability of the summarized algorithms beyond the Canadian context may require additional validation due to region-specific variations in coding

Conclusions

- This study showed good coverage of Canadian-specific case identifying algorithms, with at least one algorithm identified for most ICD-10 chapters
- Limited resources to support the identification and application of case-identifying algorithms exist
- To support validity and efficiency in RWE evidence generation, algorithms identified through this TLR will be transformed into a publicly available resource

REFERENCES

Fleet JL et al. Detecting chronic kidney disease in population-based administrative databases using an algorithm of hospital encounter and physician claim codes. BMC Nephrol. 2013 Apr 5;14:81

For any questions please email Christina Qian at cqian@broadstreetheor.com