

# The Validity of Electronic Health Databases for Measuring Smoking Status

## A Scoping Review

Md Ashiqul Haque; Lisa M Lix

Department Of Community Health Sciences, University of Manitoba, Winnipeg.

### INTRODUCTION & AIM

- Smoking status is an important covariate in many disease risk prediction models
- Many electronic health databases do not contain direct measures of smoking status
- Algorithms can be developed to derive proxy measures of smoking status from diagnosis codes, text information, and prescription medication records

**AIM:** To identify existing algorithms to ascertain smoking status from electronic health databases.

### METHODS

- The five-step Arksey and O'Malley framework for a scoping review was adopted
- PubMed, Scopus, and Web of Science were searched
- The search terms included: validity, administrative data, electronic medical records (EMRs), smoking, and tobacco use
- Articles published from 1990 to 2021 were screened
- Algorithm validity measures including sensitivity and specificity were analyzed descriptively (e.g., means and 95% confidence intervals).

### RESULTS

Figure 1: Flow diagram for scoping review

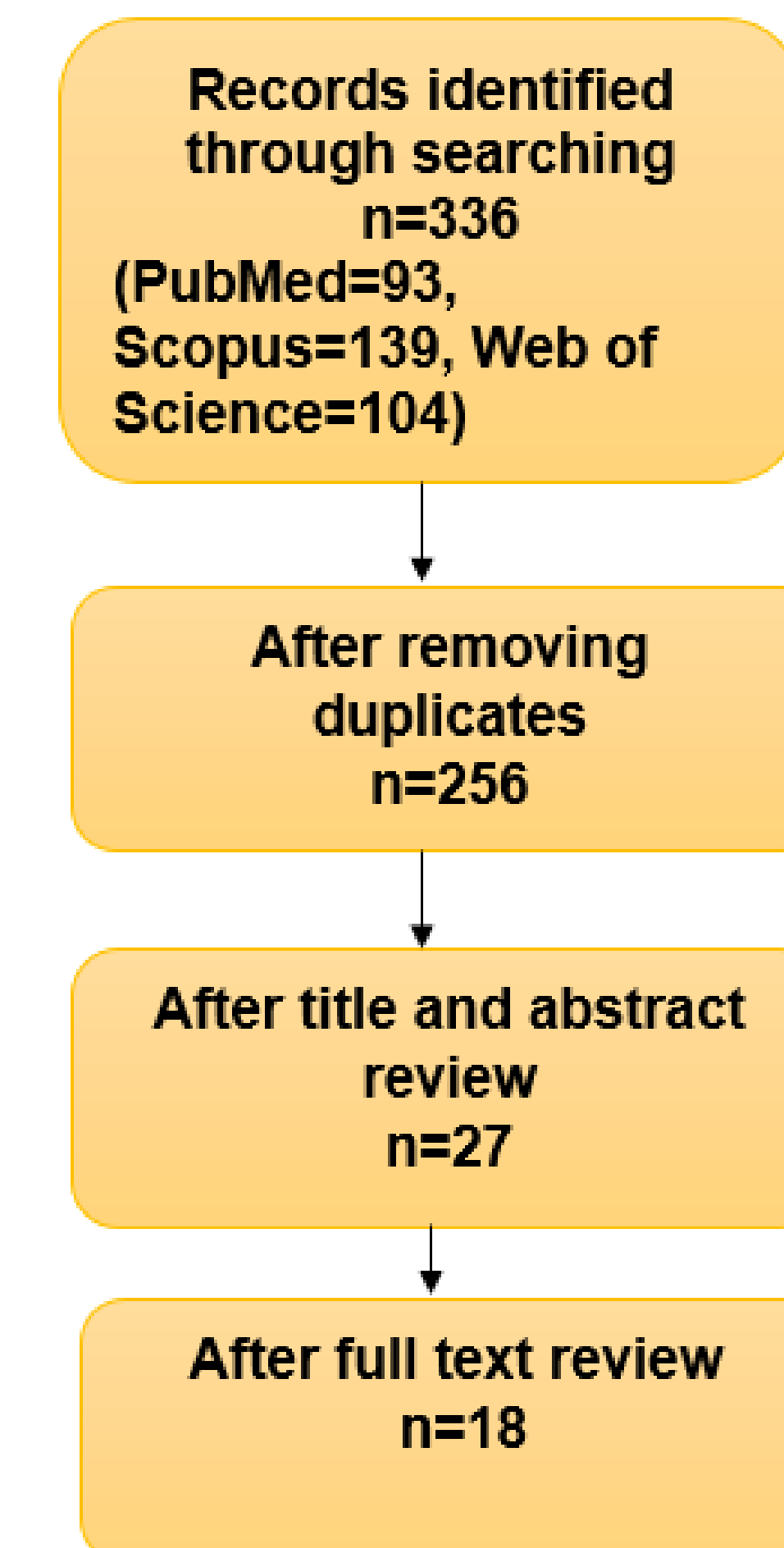


Figure 2: Sensitivity and specificity of smoking status algorithms from reviewed articles

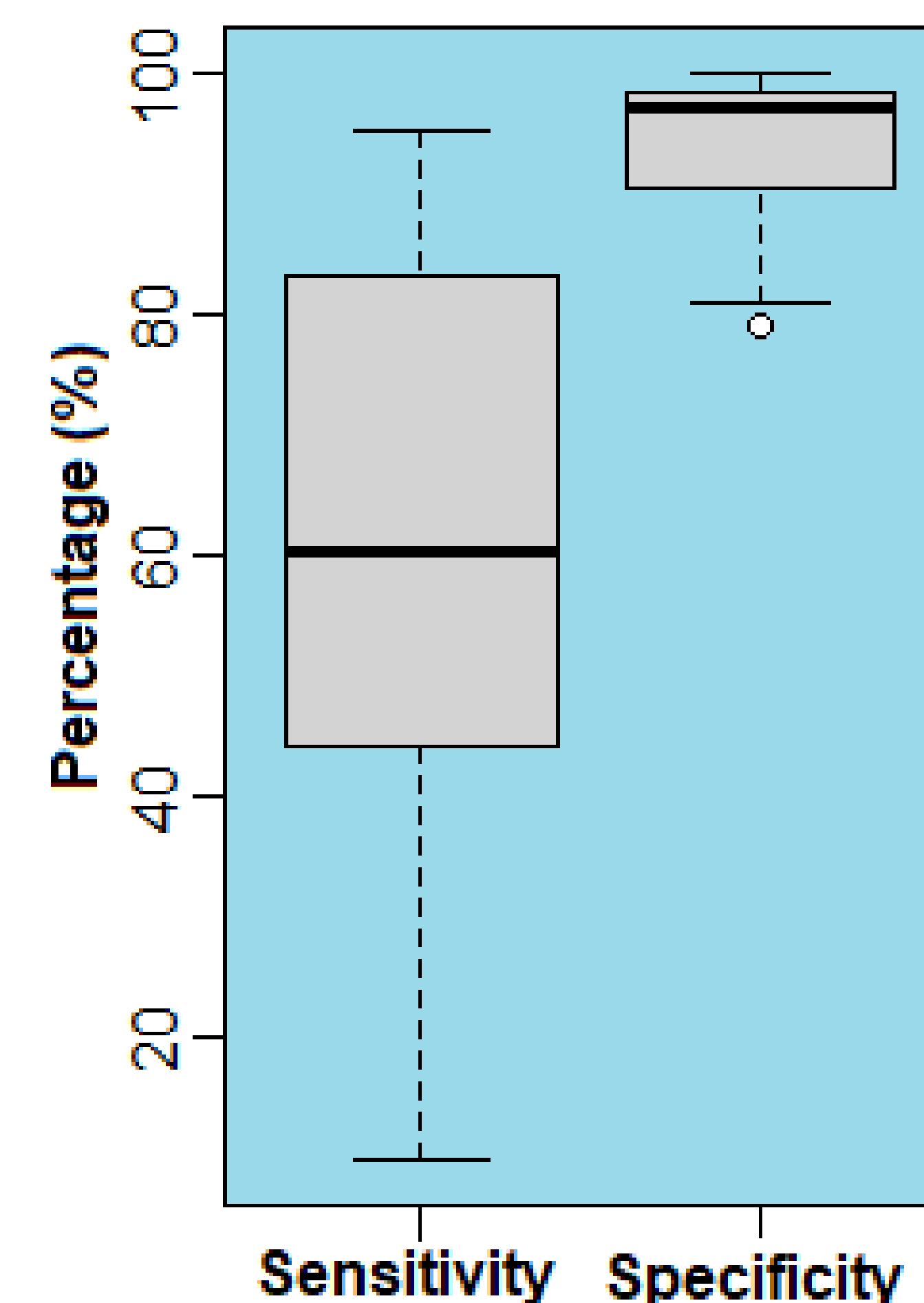


Table 1: Characteristics of validated algorithms

Data source	Source of algorithm codes	Validation data
Prescription drug dispensation information about type of smoking therapy, timing and quantity of supply and Hospital data	ICD-10 & ATC	Face validity was assessed
Smoking status was derived from Medicaid physician claims or as an encounter from a managed care plan	ICD-9, ICD-10, NDC, CPTT, HCPCS	Survey data
Veterans Health electronic medical records (EMRs)	N/A	Survey data
Medicare claims in Part-A (inpatient/hospital coverage), B (outpatient/medical coverage), and D (prescription drug coverage).	ICD-9 & CPT	Survey data
Hospital records	ICD-10	Survey data
A team of clinicians and researchers reviewed text entries (EMRs) and created an algorithm to assign each text value entry to a smoking status	N/A	Survey data
Clinical notes (EMRs) and records in social history	ICD-9 and CPT	Survey data
Clinical notes (EMRs) and semi-structured data	N/A	Survey data
Social history documentation and clinical encounter data	Not reported	Survey data

ICD = International Classification of Diseases; ATC = Anatomic, Therapeutic, Chemical Classification System; NDC = National Drug Code; CPT = Current Procedural Terminology; HCPCS = Healthcare Common Procedure Coding System; N/A= No codes were used to develop the algorithm

### CONCLUSIONS

- Several algorithms have been proposed to ascertain smoking status in electronic health databases.
- Most of the existing algorithms to identify smoking status have high specificity and low sensitivity.
- Future research could compare their accuracy and develop new algorithms by linking multiple data sources that contain smoking-related information.

### ACKNOWLEDGEMENTS



Contact: Ashiq Haque haquema4@myumanitoba.ca

