

Comparing Patient Self-Reported Symptoms and SNOMED/ICD-10-CM Codes at Primary Care Visits

Victor M. Castro, MS (1, 2), Danielle M. Crookes, PhD (1), Vivian Gainer, MS (2),
Shawn N. Murphy, MD, PhD (2), Justin Manjourides, PhD (1, 3)

(1) Bouvé College of Health Sciences, Northeastern University, Boston, MA., (2) Research Information Science and Computing, Mass General Brigham, Somerville, MA. (3) OHDSI Center at The Roux Institute, Northeastern University, Portland, ME.

Background

Accuracy of symptomatology is critical to unbiased real-world data (RWD) studies with electronic health records (EHRs). Pharmacoepidemiologic studies evaluating adverse events often define symptom-related information as primary outcomes. These studies commonly define symptoms based on structured ICD-10-CM codes (used for billing) or SNOMED codes (used for encounter and problem list documentation). However, the accuracy and completeness of structured codes for symptoms are often poor [1]. Recently, health systems have begun to routinely collect electronic self-report of symptoms during health care visits [2]. Patients complete symptom screening tools using online patient portals prior to a visit or on tablets provided to patients in the clinic waiting room. While symptom self-report data may improve detection of symptoms in RWD studies, disparities in response may introduce misclassification bias and perpetuate health inequities in observational studies. In this study we evaluated the completeness and accuracy of symptom self-reports collected at primary care annual visits across a large health system in Massachusetts consisting of hospital-based and community care settings. We examine differences between responders and non-responders and evaluate the concordance of self-report data with structured billing and encounter symptom codes.

Methods

The study population includes adults attending a primary care annual physical visit between 2019 and 2021 at Mass General Brigham, an integrated health system in Eastern Massachusetts. Electronic review of symptoms questionnaire and screenings for depression and anxiety are systematically collected by primary care physicians at hospital- and community-based practices. Patients complete the questionnaire either via the patient portal within 5 days prior to the visit or on a tablet in the waiting room.

Self-reported symptoms include cough, shortness of breath, joint pain, anxiety, confusion, depression, dizziness, headaches, sleep disturbances, suicidal ideation, constipation, diarrhea, chest pain, and palpitations. EHR data from MGB Research Patient Data Registry (RPDR) was transformed into the OMOP common data model (CDM version 5.4) and mapped to standard codes (vocabulary version 2022.09.22-v5.0). We manually mapped each self-reported symptom to OMOP standard codes. We then compare the rates of symptom presence based on self-report and existing OMOP mapped EHR visit and billing codes and evaluate the concordance between the two sources.

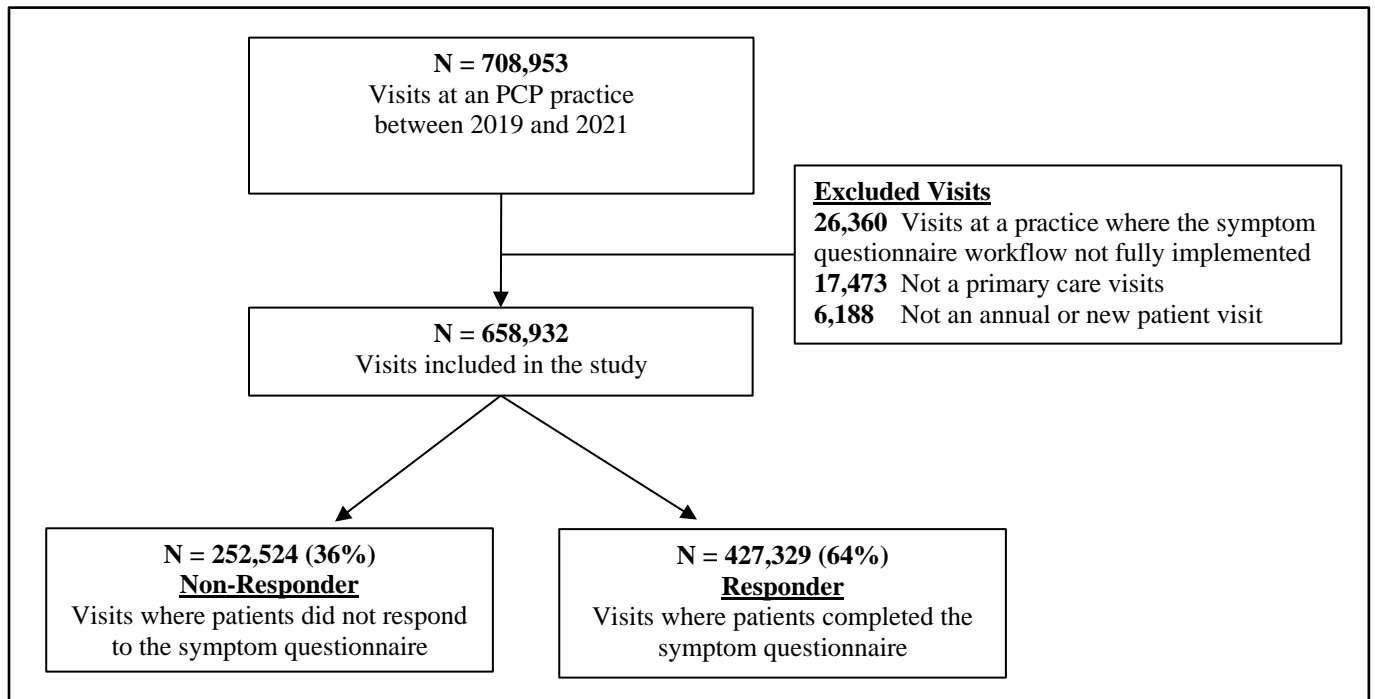
We assessed completeness and disparities in the response of self-report symptom questionnaires by comparing the characteristics of three groups: patients who responded in the online portal, patients who responded using the waiting room tablet, and non-responders. Demographic characteristics examined include age at the visit, sex at birth, self-reported race, self-reported ethnicity, ability to speak

English, and highest education completed. Additionally, we analyzed several visit characteristics, such as insurance (categorized as commercial, public [including Medicaid or Medicare], or self-pay), year of visit, whether the patient was seen via telemedicine (including phone or video visits), and the type of provider seen (medical doctor [MD], nurse practitioner [NP], doctor of osteopathic medicine [DO], or physician assistant [PA]). Lastly, we categorized the type of primary care practice as either located at an academic medical center, a community hospital, a community health center, or an individual community practice not located at a hospital or health center.

Results

Patients completed symptom self-reporting prior to 427,329 of 658,932 (64%) primary care annual physical or new patient visits during the 3-year study period beginning in 2019 (Figure 1).

Figure 1. Study flow diagram



Responders were more likely to be older ($p < 0.001$), White ($p < 0.001$), and non-Hispanic ($p < 0.001$). Responders were also more likely to have greater than a high school education ($p < 0.001$) and speak English as their primary language ($p < 0.001$) (Table 1). Non-response was more likely at visits paid with public insurance ($p < 0.001$), occurring at a community practice ($p < 0.001$), or with a DO credentialed clinician ($p < 0.001$). Non-response was also more common for visits that occurred during the initial phase of the COVID-19 pandemic.

Table 1. Sociodemographic and care site characteristic comparison of response versus non-repose of electronic symptom questionnaires at all PCP annual visits between 2019 and 2021 (N = 665,566).

Characteristic	Overall, N = 665,566 ¹	Patient Portal, N = 201,219 ¹	Clinic Tablet, N = 228,154 ¹	Non-responder, N = 236,193 ¹	p-value ²
Female gender	404,468 (61%)	125,510 (62%)	135,861 (60%)	143,097 (61%)	<0.001
Age at visit (years)	53 (38, 64)	54 (39, 64)	53 (39, 65)	52 (37, 64)	<0.001
Race					<0.001
White	537,023 (83%)	170,850 (86%)	185,698 (84%)	180,475 (79%)	
Black	34,298 (5.3%)	6,369 (3.2%)	11,762 (5.3%)	16,167 (7.1%)	
Asian	38,865 (6.0%)	12,506 (6.3%)	12,922 (5.8%)	13,437 (5.9%)	
Other	31,369 (4.8%)	6,115 (3.1%)	9,832 (4.4%)	15,422 (6.8%)	
Two Or More	6,055 (0.9%)	1,754 (0.9%)	1,946 (0.9%)	2,355 (1.0%)	
Unknown	17,956	3,625	5,994	8,337	
Hispanic ethnicity	46,844 (7.0%)	8,840 (4.4%)	14,645 (6.4%)	23,359 (9.9%)	<0.001
Level of education					<0.001
Did not complete high school	20,649 (3.5%)	3,086 (1.7%)	6,797 (3.4%)	10,766 (5.1%)	
Completed high school or equivalent	105,673 (18%)	23,796 (13%)	38,398 (19%)	43,479 (21%)	
Some college	48,042 (8.2%)	12,820 (7.2%)	15,874 (7.9%)	19,348 (9.2%)	
Bachelor's degree	299,738 (51%)	99,251 (56%)	100,539 (50%)	99,948 (48%)	
Advanced degree	86,369 (15%)	30,686 (17%)	28,956 (14%)	26,727 (13%)	
Other	28,451 (4.8%)	9,190 (5.1%)	9,961 (5.0%)	9,300 (4.4%)	
Unknown	76,644	22,390	27,629	26,625	
Non-English speaker	28,796 (4.3%)	3,421 (1.7%)	8,756 (3.8%)	16,619 (7.0%)	<0.001
Payer					<0.001
Commercial	503,909 (76%)	160,313 (80%)	170,330 (75%)	173,266 (73%)	
Public	158,367 (24%)	40,128 (20%)	56,642 (25%)	61,597 (26%)	
Self-insured	3,290 (0.5%)	778 (0.4%)	1,182 (0.5%)	1,330 (0.6%)	
Provider title					<0.001
MD	558,013 (84%)	171,330 (85%)	193,942 (85%)	192,741 (82%)	
NP	57,489 (8.6%)	16,225 (8.1%)	21,830 (9.6%)	19,434 (8.2%)	
DO	29,565 (4.4%)	7,979 (4.0%)	9,910 (4.3%)	11,676 (4.9%)	
PA	20,499 (3.1%)	5,685 (2.8%)	2,472 (1.1%)	12,342 (5.2%)	
Practice type					<0.001
Academic medical center	150,818 (23%)	50,954 (25%)	54,503 (24%)	45,361 (19%)	
Community health center	44,731 (6.7%)	10,306 (5.1%)	14,591 (6.4%)	19,834 (8.4%)	
Community hospital	108,001 (16%)	35,294 (18%)	42,985 (19%)	29,722 (13%)	
Community practice	362,016 (54%)	104,665 (52%)	116,075 (51%)	141,276 (60%)	
Telemedicine visit	24,480 (3.7%)	15,192 (7.5%)	16 (<0.1%)	9,272 (3.9%)	<0.001
Year of visit					<0.001
2019	251,775 (38%)	22,609 (11%)	148,718 (65%)	80,448 (34%)	
2020	169,145 (25%)	62,091 (31%)	36,147 (16%)	70,907 (30%)	
2021	244,646 (37%)	116,519 (58%)	43,289 (19%)	84,838 (36%)	
Number of symptoms endorsed	2 (0, 3)	1 (0, 3)	2 (0, 3)	-	<0.001
Visit with prior response	176,692 (27%)	126,923 (63%)	49,769 (22%)	0 (0%)	<0.001

¹n (%); Median (IQR); ²Pearson's Chi-squared test; Kruskal-Wallis rank sum test

The most common self-reported symptoms at PCP annual visits were joint pain (21.8%), headaches (15%) and anxiety (11.2%). Agreement between symptom self-report and ICD-10-CM/SNOMED codes was low for most symptoms (κ : 0.04 – 0.31) (Table 2). The rate of structured coding of patient-reported symptoms ranged from 0.46 for anxiety to 0.02 for suicidal ideation (Table 2).

Table 2. Concordance between self-reported symptoms and EHR billing and encounter SNOMED and ICD-10-CM codes

Type	Symptom	Percent of PCP annual physicals		Agreement	Conditional Probabilities	
		Self-report (SR)	SNOMED/ ICD-10-CM (code)	Cohen's Kappa	Pr(SR code)	Pr(code SR)
Respiratory	Cough	5.6%	1.3%	0.22	0.64	0.15
	Shortness of breath	5.8%	1.1%	0.18	0.63	0.12
Orthopedic	Joint pain	28.5%	7.7%	0.19	0.67	0.18
Neuropsychiatric	Anxiety (GAD)	11.2%	17%	0.27	0.30	0.46
	Confusion	1.8%	0.9%	0.12	0.19	0.09
	Depression (PHQ)	9.5%	10.7%	0.27	0.33	0.37
	Dizziness	7.1%	1.6%	0.20	0.61	0.14
	Headaches	15.0%	6.2%	0.28	0.58	0.24
	Sleep disturbance	21.8%	7.1%	0.13	0.45	0.15
	Suicidal ideation (PHQ)	0.9%	0.02%	0.04	0.66	0.02
Gastrointestinal	Constipation	8.9%	0.7%	0.04	0.38	0.03
	Diarrhea	7.3%	0.7%	0.12	0.75	0.07
Cardiometabolic	Chest pain	4.0%	1.8%	0.30	0.51	0.23
	Palpitations	7.0%	2.4%	0.31	0.64	0.22

Conclusion

Patient self-reported symptoms in primary care offer a complementary data source for improving symptom-based outcomes in RWD studies. In this study we found that patients self-report symptoms up to nine times more than clinicians record with structured codes for administrative reasons. Our results suggest that undercoding of ICD-10-CM codes may bias observational studies reliant on symptom incidence. However, we also find significant patient and visit-level disparities in symptom self-report response driven by persistent sociodemographic factors of health inequities. Combining structured symptom codes and patient self-report using concepts sets and/or computable phenotypes should enhance the validity of symptom outcomes in RWD studies.

References

- 1 Koleck TA, Dreisbach C, Bourne PE, *et al.* Natural language processing of symptoms documented in free-text narratives of electronic health records: a systematic review. *J Am Med Inform Assoc* 2019;**26**:364–79.
- 2 Sisodia RC, Dankers C, Orav J, *et al.* Factors associated with increased collection of patient-reported outcomes within a large health care system. *JAMA Netw Open* 2020;**3**:e202764.