PheValuator 2.0: Methodological improvements for the PheValuator approach to semi-automated phenotype algorithm evaluation

Joel N. Swerdel, Martijn Schuemie, Gayle Murray, and Patrick B. Ryan

Journal of Biomedical Informatics 2022 Nov;135:104177.
doi: 10.1016/j.jbi.2022.104177.

Background

- PheValuator 1.0 introduced in 2019
- Developed to provide a method for determining the performance characteristics of phenotype algorithms
- Results in first paper suggested that PheValuator 1.0 underestimated the positive predictive value of algorithms

Objectives

1. Develop a comparative benchmark using prior validation studies
2. Test improvements in PheValuator 2.0 against benchmark

Methods

- Changes to PheValuator method
  - Re-define xSpec and xSens algorithms to potentially use the phenotype diagnosis codes in the predictive model
  - Use multiple time windows to add a dimension of temporality to the model
- Develop phenotype benchmark
  - Found 825 prior validation studies
  - 34 studies provided validation statistics for 17 phenotypes
  - Included 2 more phenotypes (MI, stroke) from first paper
  - Final comparison: 19 phenotypes and 90 algorithms

Results

Table 1
Studies with Phenotype Algorithm Validation Results for Selected Phenotypes.

Disease Area	Phenotype	Author (Year; Country)Reference	Number of Algorithms	Sensitivity	Specificity	PPV
Cardiovascular	Atrial Fibrillation	Navar-Boggan et al. (2015; US) [14]	1			●
	Pulmonary Embolism	Alotaibi et al. (2015; Canada) [11]	2	●	●	●
		White et al. (2010;) [15]	2			●
	Venous Thromboembolism	Alotaibi et al. (2015; Canada) [11]	2	●	●	●
		McPeek Hinz et al. (2013; US) [12]	1			●
		White et al. (2010;) [15]	2			●
Immunology	Ankylosing spondylitis	Curtis et al. (2016; US) [16]	4			●
		Dubreuil et al. (2017; UK) [17]	6			●
	Atopic dermatitis	Hsu et al. (2017; US [)9]	8			●
	Crohns Disease	Ananthakrishnan et al. (2013; US) [18]	4			●
		Stepaniuk et al. (2015; Canada) [19]	1			●
		Thirumurthi et al. (2010; US) [20]	1			●
	Psoriasis	Eder et al. (2020; Canada) [21]	3	●	●	●
		Icen et al. (2008; US) [22]	1			●
		Seminara et al. (2011; UK) [23]	3			●
	Rheumatoid Arthritis	Hanly et al. (2015; Canada) [24]	3			●
		Ng et al. (2012; US) [25]	6			●
		Widdifield et al. (2014; Canada) [10]	4	●		●
	Systemic Lupus Erythematosus	Hanly et al. (2014; Canada) [26]	3			●
	Ulcerative Colitis	Ananthakrishnan et al. (2013; US) [18]	4			●
		Stepaniuk et al. (2015; Canada) [19]	1			●
		Thirumurthi et al. (2010; US) [20]	1			●
Infectious Disease	Viral Hepatitis B	Niu et al. (2016; US) [27]	3			●
	Viral Hepatitis C	Niu et al. (2016; US) [27]	3			●
		Lo Re et al. (2009; UK) [28]	1			●
Neurology	Autism	Burke et al. (2014; US) [29]	2			●
	Bipolar	Castro et al. (2015; US) [30]	2			●
	Epilepsy	Christensen et al. (2007; Denmark) [31]	1			●
		Jette et al. (2010; Canada) [32]	1			●
		Moura et al. (2017; US) [33]	1			●
		Pugh et al. (2008; US) [34]	1			●
		Tan et al. (2015; Australia) [35]	1	●	●	●
Oncology	Multiple Myeloma	Brandenburg et al. (2019; US) [36]	4	●		●
	Prostate Cancer	Parlett et al. (2019; US) [13]	4	●		●

PPV – Positive Predictive Value.

Results

Table 2
Differences in estimates for Positive Predictive Value from PheValuator Version 1.0 and 2.0 and the gold standard estimates from prior validation studies for Acute Myocardial Infarction.

Algorithm	Publication PPV	CDM	Version 1.0 PPV	Version 1.0 Difference	Version 2.0 PPV	Version 2.0 Difference
Choma[1]	93	CCAE	80	−13	86	−7
		DOD	50	−43	86	−7
		MDCD	60	−33	93	0
		MDCR	44	−49	86	−7
		PanTher	73	−20	90	−3
Cutrona[2]	86	CCAE	80	−6	89	3
		DOD	50	−36	85	−1
		MDCD	60	−26	92	6
		MDCR	44	−42	85	−1
		PanTher	66	−20	89	3
Wahl[3]	88	CCAE	78	−10	81	−7
		DOD	57	−31	83	−5
		MDCD	54	−34	89	1
		MDCR	42	−46	82	−6
		PanTher	66	−22	88	0

Median: −31 (IQR −39, −20) | −1 (IQR −6.5, 0.5)

Results (cont.)

Therapeutic Area	Condition	Positive Predictive Value Difference	
		Version 1.0	Version 2.0
Overall	Overall (all therapeutic areas)	Median (IQR) −21 (-34, −3)	Median (IQR) 4 (-3, 15)

Therapeutic Area	Condition	Specificity Difference	
		Version 1	Version 2
Overall	Overall	Median (IQR) 3 (1, 4.25)	Median (IQR) 3 (1, 4)

Therapeutic Area	Condition	Specificity Difference	
		Version 1	Version 2
Overall	Overall	Median (IQR) 3 (1, 4.25)	Median (IQR) 3 (1, 4)
Cardiovascular	Pulmonary Embolism	5 (3, 7)	5.5 (3, 8)
	Venous Thromboembolism	4 (3, 6)	3.5 (3, 5)
Immunology	Psoriasis	1 (1, 2.5)	1 (1, 4)
Neurology	Epilepsy	1 (1, 1)	1 (1, 1)

	Thromboembolism		
	Ischemic Stroke	−22.5 (-28, −19.75)	4 (2, 5)
	Myocardial Infarction	−31 (-39, −20)	−1 (-6.5, 0.5)
Immunology	Overall (Immunology)	−27 (-39, −4)	7 (-2, 30.25)
	Ankylosing spondylitis	−51 (-59, −41.5)	−2.5 (-10, 5)
	Atopic dermatitis	−8 (-16.5, −2)	39 (26.5, 42.5)
	Ulcerative Colitis	−31 (-37.5, −21.25)	−1 (-5.75, 4)
	Crohns Disease	−32.5 (-39, −26.25)	−2 (-6, 2.75)
	Rheumatoid Arthritis	6 (-11, 16)	38 (23, 48)
	Psoriasis	−23 (-34, −6.5)	9 (-3, 26)
	Systemic Lupus Erythematosus	−34 (-42.5, −30)	8 (0, 9)
Infectious Disease	Overall (Infectious disease)	−1 (-12.5, 3)	2 (-3, 6)
	Viral Hepatitis C	−2.5 (-11.5, 3)	0.5 (-3, 3.75)
	Viral Hepatitis B	1 (-14.5, 1)	5 (-5.5, 6)
Neurology	Overall (Neurology)	−24 (-31.75, −17)	7 (-8, 12.75)
	Autism	−31.5 (-36, −26.5)	2.5 (-5.25, 7.5)
	Bipolar	−28 (-31.5, −26.25)	13 (11.5, 14.75)
	Epilepsy	−20 (-24, −10.75)	−7 (-18.75, 9.5)
Oncology	Overall (Oncology)	−1 (-4, 4.5)	14 (10, 18)
	Multiple Myeloma	0 (-3.25, 6.5)	13.5 (10, 22)
	Prostate Cancer	−2 (-4, 1.25)	14 (9.5, 16.5)

Conclusions

- PheValuator 2.0 provides better agreement with PPV than previous version using a comparison to chart review validation studies.
- PheValuator 2.0 produces estimates that are less than 5% different than estimates from prior studies for positive predictive value and specificity.
- Considering low cost and effort, PheValuator should be considered when algorithm validation is needed.
- Using PheValuator would allow quantitative bias analysis to improve the reliability and reproducibility of research studies using observational data.
- …and now in HADES!