Comparing Performance Characteristics of Phenotype Algorithms for Dermatological and Renal Diseases.

Joel N. Swerdel, PhD MS MPH[1,3]; Patrick B. Ryan, PhD[1,2,3]

[1] Janssen Research and Development, Titusville, NJ, USA

[2] Columbia University, New York, NY, USA

[3] Observational Health Data Sciences and Informatics (OHDSI), New York, NY

Abstract - ***Background****: Phenotype algorithms (PAs) are commonly used to determine subjects with specific health conditions in observational research. There has been few research studies performed to examine the performance characteristics of PAs for diseases within dermatology. The objective of this research was to use PheValuator, a package within the OHDSI toolset, to determine the performance characteristics, e.g., sensitivity and positive predictive value (PPV), of PAs from dermatological diseases and to compare those results to results from renal diseases.* ***Method:*** *Data for this study were collected between January 1, 2010 and December 31, 2018 from 4 administrative claims data sets: IBM MarketScan's Commercial Claims and Encounters, Medicare Supplemental Beneficiaries, and Multi-State Medicaid; OptumInsight's de-identified Clinformatics™ Datamart (Eden Prairie,MN); and Optum© de-identified Electronic Health Record Dataset (OptumInsight, Eden Prairie,MN). For this research, we examined the performance characteristics of PAs for 8 health conditions: atopic dermatitis, plaque psoriasis, candidiasis, malignant skin cancer, chronic kidney disease, acute renal failure, kidney stones, and renal cell carcinoma. We used two PAs in this study: 1) a PA where subjects were included if a diagnosis code appeared 1 or more times within the health record ("≥1 X Outcome") and 2) a PA where subjects were included in the cohort if a diagnosis code appeared 3 or more times in the patient record or if there was 1 or more diagnosis code from a hospital in-patient setting ("≥3 X Outcome/1 X IP").* ***Results:*** *PAs involving dermatological diseases consistently demonstrated low values for sensitivity and PPV. PAs involving renal diseases consistently demonstrated higher values for sensitivity and PPV compared to those from dermatological diseases.* ***Conclusion:*** *We found consistently low values for performance characteristics from PAs for several dermatological diseases. Results from research involving these PAs may be prone to significant misclassification bias.*

Background: The primary approach for defining disease in observational healthcare databases is to construct phenotype algorithms (PAs), rule-based heuristics predicated on the presence, absence, and temporal logic of clinical observations. There has been few research studies performed to examine the performance characteristics of PAs for diseases within dermatology. The objective of this research was to use PheValuator, a package within the OHDSI toolset, to determine the performance characteristics, e.g., sensitivity and positive predictive value (PPV), of PAs from dermatological diseases and to compare those results to results from renal diseases.

Methods: Data for this study were collected between January 1, 2010 and December 31, 2018 from four US administrative claims data sets: IBM MarketScan's Commercial Claims and Encounters, Medicare Supplemental Beneficiaries, and Multi-State Medicaid; and OptumInsight's de-identified Clinformatics™ Datamart (Eden Prairie,MN).

For this research, we examined the performance characteristics of PAs for 8 health conditions, 4 from dermatology (atopic dermatitis, plaque psoriasis, candidiasis, malignant skin cancer) and 4 involving renal disease (chronic kidney disease, acute renal failure, kidney stones, and renal cell carcinoma). We used two PAs in this study: 1) a PA where subjects were included if a diagnosis code appeared 1 or more times within the health record ("≥1 X Outcome") and 2) a PA where subjects were included in the cohort if a diagnosis code appeared 3 or more times in the patient record or if there was 1 or more diagnosis code from a hospital in-patient setting ("≥3 X Outcome/1 X IP"). We chose to test the same PAs for each disease to allow inter-disease performance characteristic comparisons.

PheValuator is an R package within the OHDSI toolset that allows for evaluating the performance characteristics of PAs[1]. The tool uses diagnostic predictive modeling to determine the probability of a health condition in a large set of subjects, the "evaluation cohort". The predictive model is developed using a set of labeled data where the label represents the presence or absence of the phenotype for each subject in the dataset. For the subjects labeled as having the phenotype, we used an extremely specific ("xSpec") PA, ensuring that these subjects would have the phenotype with a very high likelihood. For the diseases that often require a hospital in-patient visit, we used an xSpec PA requiring at least 5 occurrences of a diagnostic condition code for the disease with at least one occurrence from a hospital in-patient visit. In this study we used this PA for malignant skin cancer, acute renal failure, kidney stones, and renal cell carcinoma. For the remainder of the diseases the xSpec PA required at least 5 occurrences of a diagnostic condition code for the disease from any type of clinical visit.

The evaluation cohort of subjects may be used to test PAs for sensitivity, specificity, and positive and negative predictive value. In this study we also calculated the F1 score which is defined as the harmonic mean of sensitivity and PPV.

Results: We found the performance characteristics from dermatological diseases to be low with the exception of malignant skin cancer (Table 1). The highest value for sensitivity in skin diseases other than cancer was found in ≥ 1 X candidiasis where the mean across the four datasets was 41%; the highest PPV was found in ≥ 3 X psoriasis at about 35%. By contrast, in diseases of the kidney, the highest mean sensitivity was found in ≥ 1 X chronic kidney disease at about 79% and the highest PPV was found in ⩾3 X chronic kidney disease at about 66%. The highest F1 score for non-cancer skin diseases was about 28% (≥ 1 X candidiasis) while the lowest F1 score for kidney diseases was 42% (≥ 1 X kidney stone).

Conclusion: We found consistently low values for the performance characteristics from PAs for the non-cancer associated dermatological diseases tested. As sensitivity, in particular, is rarely determined in traditional validation studies, these results represent an important addition to the body of knowledge for these diseases. There may be several reasons for these findings. The low values for sensitivity and PPV may be due to alternate diagnoses for diseases with similar characteristics. For example, Silverberg et al found many alternate diagnosis designations for atopic dermatitis including eczema, dermatitis, and atopic eczema[2]. The lack of definitive diagnostic tests for these diseases is also an issue[3]. Definitive diagnostic tests are important for informing the model and allowing better discrimination between disease states. However, improving the performance characteristics of PAs for these diseases is possible. Researchers, e.g., Dobson-Belaire and colleagues studying psoriasis, have found higher PPVs using PAs that include prescribed medications specific for the disease[4]. Anti-psoriasis and anti-fungal medications were important predictors in the models for psoriasis and candidiasis, respectively, and the inclusion of these elements into the PAs would likely improve the performance characteristcs. It appears that caution should be used when developing PAs to study common dermatological conditions. Results from research involving these PAs may be prone to significant misclassification bias.

Table 1: Mean Performance Characteristics of Two Phenotype Algorithms for Eight Health Conditions from Four Datasets using PheValuator

Phenotype Algorithm	Sens	PPV	Spec	NPV	F1	Phenotype Algorithm	Sens	PPV	Spec	NPV	F1
≥1 x Atopic Dermatitis	0.280	0.161	0.979	0.989	0.203	≥1 x Acute Renal Failure	0.777	0.447	0.963	0.992	0.567
≥3 x Atopic Dermatitis/≥1 x IP	0.070	0.302	0.998	0.986	0.114	≥3 x Acute Renal Failure/1 x IP	0.727	0.484	0.969	0.990	0.580
≥1 x Psoriasis	0.388	0.199	0.983	0.993	0.262	≥1 x CKD	0.787	0.540	0.953	0.983	0.638
≥3 x Psoriasis/≥1 x IP	0.235	0.366	0.995	0.991	0.285	≥3 x CKD/≥1 x IP	0.642	0.658	0.977	0.971	0.649
≥1 x Candidiasis	0.410	0.223	0.946	0.977	0.288	≥1 x Kidney Stone	0.756	0.294	0.961	0.994	0.422
≥3 x Candidiasis/≥1 x IP	0.141	0.351	0.989	0.968	0.198	≥3 x Kidney Stone/≥1 x IP	0.573	0.463	0.986	0.991	0.510
≥1 x Skin Malignancy	0.688	0.340	0.959	0.990	0.448	≥1 x RenalCarc	0.694	0.378	0.997	0.999	0.488
≥3 x Skin Malignancy/≥1 x IP	0.496	0.411	0.979	0.981	0.444	≥3 x RenalCarc/≥1 x IP	0.635	0.516	0.998	0.999	0.568

Sens - Sensitivity; PPV - Positive Predictive Value; Spec - Specificity; NPV - Negative Predictive Value; ≥1 X - ≥1 X Health Condition; ≥3 X - ≥3 X Health Condition; ≥1 X IP - ≥1 X Health Condition from a hospital in-patient setting; CKD - Chronic Kidney Disease; RenalCarc – Renal Cell Carcinoma; F1 – F1 Score
Values represent the mean across 4 datasets, IBM MarketScan's Commercial Claims and Encounters, Medicare Supplemental Beneficiaries, and Multi-State Medicaid; and OptumInsight's de-identified Clinformatics™ Datamart (Eden Prairie,MN). The continuous 3-color heat map for the data in the table was defined as Red (value = 0), Yellow (value = 0.5), and Green (value = 1).

1. PheValuator. https://githubcom/OHDSI/PheValuator 2019 [cited 2019; Available from: https://github.com/OHDSI/PheValuator
2. Silverberg JI, Thyssen JP, Paller AS, Drucker AM, Wollenberg A, Lee KH, et al. What's in a name? Atopic dermatitis or atopic eczema, but not eczema alone. Allergy. 2017 2017/12/01;72(12):2026-30.
3. Griffiths CE. Psoriasis: future research needs and goals for the twenty-first century. Dermatol Clin. 2004 Oct;22(4):493-9, x.
4. Dobson-Belaire W, Goodfield J, Borrelli R, Liu FF, Khan ZM. Identifying Psoriasis and Psoriatic Arthritis Patients in Retrospective Databases When Diagnosis Codes Are Not Available: A Validation Study Comparing Medication/Prescriber Visit–Based Algorithms with Diagnosis Codes. Value Health. 2018 2018/01/01/;21(1):110-6.