Name:	Levon Utidjian, MD MBI
Affiliation:	The Children's Hospital of Philadelphia
Email:	utidjianl@email.chop.edu
Presentation type (select one):	Poster

Experience of Developing Computable Phenotypes for Pediatric Chronic Conditions in PEDSnet

Levon H. Utidjian, MD, MBI[1], Ritu Khare, PhD[2], Janet Zahner, BS[2], Nandan Patibandla, MS[3], L. Charles Bailey, MD, PhD[1]
[1]The Children's Hospital of Philadelphia, Philadelphia, PA; [2]Cincinnati Children's Hospital Medical Center, Cincinnati, OH; [3]Boston Children's Hospital, Boston, MA

Abstract

Computable phenotypes provide researchers with a way to assemble complex and highly specific cohorts of patients by leveraging multiple clinical variables present in electronic health record (EHR) data. Here, we describe our experience developing computable phenotypes in response to two study requests for pediatric Crohn's disease and type 2 diabetes from the PEDSnet clinical data research network (CDRN). Our examples show how simple changes to computable phenotype criteria can result in large changes in cohort size and improve cohort specificity.

Introduction

The ability to accurately and reliably identify patients with specific medical conditions is essential for constructing study cohorts. Traditional cohort construction typically relies upon administrative datasets that are limited to billing data like diagnosis codes and procedure codes. EHR data provide a larger number of clinical variables to allow for more complex cohort construction. Computable phenotypes utilize multiple EHR data elements to define cohorts of patients with specific conditions by augmenting billing data with additional data like laboratory results, medications, vital signs and clinical notes.[1]

In pediatrics, achieving adequate cohort sizes is a challenge given that most chronic diseases of childhood are rare. PEDSnet is a CDRN that aggregates EHR data from eight of the nation's largest children's hospitals, and uses a version of the OMOP CDM v5 data model. Here we describe our initial results developing two computable phenotypes for pediatric Crohn's Disease and Type 2 Diabetes based on research study requests.

Methods

Study requests received by the PEDSnet data science group contained inclusion and exclusion criteria, which consisted of lists of patient characteristics, diagnosis codes and other features like medications or abnormal lab values meant to identify patients with the condition of interest. These criteria were translated into standardized concepts within the OMOP CDM v5, and SQL code was then generated for each computable phenotype. These SQL queries could then be run across our CDRN.

The Crohn's disease phenotype was based on either having at least one encounter with any ICD9 diagnosis codes for Crohn's disease, while excluding patients with a ratio of ulcerative colitis:Crohn's disease diagnosis codes of >1, or the combination of diagnosis codes plus the presence of any medications for the treatment of Crohn's. The type 2 diabetes phenotype in the study request was for patients with any ICD9 diagnosis codes for type 2 diabetes diagnosed at ≥24 months of age or the presence of abnormal lab values for either fasting glucose, hemoglobin A1C (HbA1C) or 2-hour oral glucose tolerance tests (OGTT). Phenotypes were then refined by adjusting these criteria with the goal of improving the specificity of the computable phenotypes.

Results

The results for the Crohn's disease phenotype development are summarized in Table 1. By requiring that a patient have at least 3 occurrences of a diagnosis of Crohn's disease instead of a single occurrence, the cohort size decreased by 17-25% depending on the criteria used.

Table 1. Crohn's Disease Computable Phenotype Results at 8 Sites

Condition	Criteria	# of Patients (% of total)	Difference in # of patients (% change)
1a	At least 1 Crohn's diagnosis occurrence	9,407 (0.18%)	--
1b	At least 3 Crohn's diagnosis occurrences	7,078 (0.14%)	-2,399 (25%)
2a	At least 1 Crohn's diagnosis occurrence + Crohn's medication	7,713 (0.15%)	--
2b	At least 3 Crohn's diagnosis occurrences + Crohn's medication	6,402 (0.12%)	-1,311 (17%)

A different approach was taken to refine the type 2 diabetes computable phenotype as it was realized that it would be important to exclude patients with any history of type 1 diabetes diagnoses as it is the more common form of diabetes in children and they would also likely have similar abnormal lab values (Table 2). Excluding patients with a diagnosis of type 1 diabetes led to large decreases in cohort size as 57% of patients with a type 2 diabetes diagnosis code and 87% of patients with abnormal HbA1C values had at least one type 1 diabetes code. This exclusion led to an overall decrease in cohort size of 73% compared to the original study request's criteria.

Table 2. Type 2 Diabetes Computable Phenotype Results at 6 Sites

Criteria	Type 2 diabetes ICD9 code alone	Fasting glucose ≥126 alone	HbA1C ≥6.5% alone	2-hour OGTT ≥200 alone	Any 1 of the criteria
# of Patients (% of total)	11,124 (0.309%)	141 (0.004%)	15,529 (0.432%)	145 (0.004%)	21,350 (0.594%)
# of Patients after excluding those with any type 1 diabetes ICD9 codes	4,797 (0.18%)	100 (0.001%)	2,045 0.38%	121 (0.001%)	5,665 (0.44%)
Difference in # of Patients (% change)	-6,327 (57%)	-41 (29%)	-13,484 (87%)	-24 (17%)	-15,685 (73%)

Conclusion

Our initial experience developing computable phenotypes for Crohn's disease and type 2 diabetes shows how adjustments to exclusion criteria to improve specificity can result in large changes in cohort size. It is particularly important to be aware of criteria that could be shared between similar conditions, like abnormal lab values in type 1 & 2 diabetes, and consider adding other criteria to ensure only patients with the condition of interest are obtained. As such, careful attention should be paid to the selection of computable phenotype criteria that strike a balance between the competing goals of having large cohorts while simultaneously maintaining high specificity.

References

1. Richesson RL, Hammond WE, Nahm M, Wixted D, Simon GE, Robinson JG, et al. Electronic health records based phenotyping in next-generation clinical trials: a perspective from the NIH Health Care Systems Collaboratory. Journal of the American Medical Informatics Association. The Oxford University Press; 2013 Dec;20(e2):e226–31.