Comparing the impact of alternative phenotype definitions: insights from developing cohorts for COVID-19 adverse events of special interest

Rupa Makadia, Azza Shoaibi, Gowtham Rao, Jill Hardin, Joel Swerdel, Stephen Fortin, Xintong Li, Anna Ostropolets, Anthony G Sena, Peter R. Rijnbeek, Marc A Suchard, George Hripcsak, Patrick B. Ryan

Background

COVID-19 vaccines being authorized for emergency use have prompted researchers and regulators to prepare safety surveillance approaches that involve the use of real-world data to study adverse events of the vaccines. The Adverse Events of Special Interest (AESI) represent a collection of phenotypes that can be applied in observational data to study adverse event background rates (1). Developing accurate phenotypes is critical in obtaining reliable and reproducible background rates. Observational data can vary by geographic region and data composition, which can impact phenotype development. Since any given outcome may be represented by multiple alternative phenotype algorithms, we sought to characterize the differences between definitions and assess diagnostic accuracy (sensitivity and specificity errors) and potential index date misclassification errors for a cohort that represents a given phenotype.

Methods

The initial phenotypes used in this analysis were defined by the FDA (23 phenotypes) and ACCESS (37 phenotypes) as the starting definitions (2, 3). Alternative definitions were produced by reviewing prior literature, leveraging the SNOMED vocabulary and PHOEBE. For brevity, the phenotype Guillain-Barre syndrome (GBS) will be presented in the abstract. The three cohort definitions were: (A) GBS codes in an inpatient setting at primary position, (B) GBS codes in an inpatient setting, (C) GBS codes from all places of service. The index date is the date a person has the diagnosis code with the visit qualifier (inpatient primary, inpatient or all) with no GBS codes in the prior 365 days. The phenotypes were developed in ATLAS and evaluated using the CohortDiagnostics (CD) package. Several domains were assessed as part of the phenotype development process: index events/ codes included, incidence rates over time for cohort by age/gender, visit context of cohort before, on and after index, cohort overlap between various cohorts and temporal characterization of all variables in the data at various time intervals. Incidence rates defined time-at-risk as a 365-day period following the index date. People contributed time-at-risk from 1 January to 31 December for each qualifying year in 2017 to 2019 (4). The databases used in this study included Optum© De-Identified Clinformatics® Data Mart Database – Date of Death-(DOD) (Optum DOD) dataset, IBM MarketScan® Databases [Commercial Claims (CCAE), Medicaid (MDCD) and Medicare (MDCR)] and Optum® de-identified Electronic Health Record Dataset (Panther).

Results

Defining GBS based on inpatient diagnosis in the primary position (definition A) resulted in ~30% fewer cases than defining GBS based on inpatient diagnosis in any position (definition B) and ~300% fewer cases than defining GBS based on any diagnosis (definition C), consistently across all US claims databases. The overall annual incidence rate for Definition A ranged from 0.20-0.61 per 1,000 person-years, across age/sex/years. Incidence increased in the age range between 60 to 69 compared to those under 60 and is higher in males; MDCR showed the greatest difference between genders. Approximately, 51%- 58% of GBS cases across databases also concurrently have an emergency room visit under Definition A. Temporal characterization of cohorts identified that a GBS/Acute infective polyneuritis diagnosis occurs in approximately 16.3%-26.6% of GBS cases -30 to -1 days prior to index indicating GBS codes occurring prior to index in another visit. A diagnosis of malaise (symptom of GBS) was recorded for 15.1%-25.8%

across databases on -30 to -1 days prior to index. Gabapentin, a potential treatment for GBS is dispensed in the 1 to 30 days post index ranges from 6.5% to 17.9% of GBS cases across all databases. Comparing Definition A to B (Figure 1), most covariates show no significant standardized differences. Spinal puncture (1 to 30 days after index) has the greatest standardized difference (0.761) between the two definitions. Incidence rates differ amongst the three definitions, by age and sex strata. There were higher rates when including all places of service and marginal rate changes from Definition A to Definition B (Table 1).

Table 1. Incidence rates per 100,000 person-years age-sex stratified (CCAE)

Outcome	Sex	1 - 5	6 - 17	18 - 34	35 - 54	55 - 64	65 - 74
(A) GBS codes in an inpatient setting at primary position	F	1.11	0.67	2.22	2.61	3.41	4.99
	M	1.27	0.45	1.70	3.01	5.17	7.34
(B) GBS codes in an inpatient setting	F	1.11	0.78	2.86	3.57	4.56	5.99
	M	1.48	0.64	2.11	3.66	6.86	9.03
(C) GBS codes all places of service	F	1.33	1.93	5.38	8.89	11.44	16.49
	M	1.80	1.58	3.65	8.25	14.76	22.03

Figure 1. Standardized mean differences plot for covariates in Definition A (y-axis) to Definition B (x axis) (CCAE)
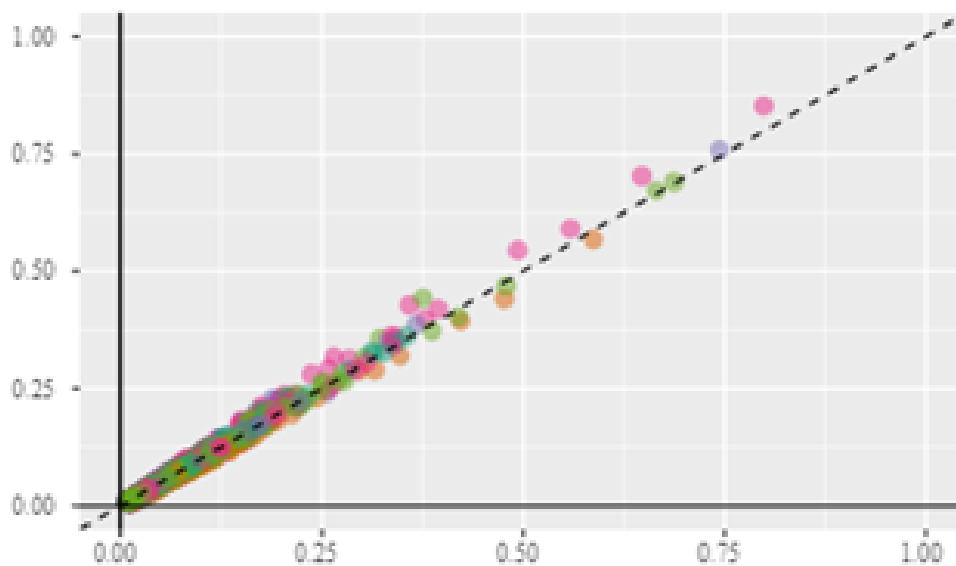
Conclusions

Phenotypes play an important foundational role when conducting any analysis and this work demonstrates an empirical mechanism to evaluate and understand the differences between phenotypes in lieu of a gold standard. The empirical evidence collected by CD represents key metrics to understand for each phenotype the codes being used, stability of incidence rates, understanding symptoms and treatments for any given phenotype. Our GBS case study highlights how alternative definitions can yield substantially different incidence rates without necessarily demonstrating different patient characteristics, which can serve as a useful proxy for evaluating sensitivity/specificity tradeoffs. We observed symptoms prior to the index date indicating a potential for misclassification of the index date in Definition A. These empirical insights can help researchers identify the impact of choices they make for any given phenotype. The incidence rates between phenotype definitions by age-sex strata show significant variation, which can impact the

background rates for these outcomes. Our research plans to evaluate a series of phenotypes through these methodologies and evaluate them on a network of databases.

References

1.   Black S, Eskola J, Siegrist CA, Halsey N, MacDonald N, Law B, et al. Importance of background rates of disease in assessment of vaccine safety during mass immunisation with pandemic H1N1 influenza vaccines. Lancet (London, England). 2009;374(9707):2115-22.
2.   ACCESS: Background rates of Adverse Events of Special Interest for monitoring COVID-19 vaccines. 2020 [Available from: http://www.encepp.eu/encepp/viewResource.htm?id=37274.
3.   Epidemiology. CfBEaROoBa. CBER Surveillance Program Background Rates of Adverse Events of Special Interest for COVID-19 Vaccine Safety Monitoring Protocol 2021 [Available from: https://www.bestinitiative.org/wp-content/uploads/2021/02/C19-Vaccine-Safety-AESI-Background-Rate-Protocol-FINAL-2020.pdf
4.   Li X, Ostropolets A, Makadia R, Shaoibi A, Rao G, Sena AG, et al. Characterizing the incidence of adverse events of special interest for COVID-19 vaccines across eight countries: a multinational network cohort study. medRxiv : the preprint server for health sciences. 2021.