Assessing Missing Data Mechanisms for Unspecified Diabetic Retinopathy Disease Severity Encounters in the Electronic Health Record: An IRIS® Registry Analysis

Meghan Hatfield, MPH[1], Durga S. Borkar, MD, MMCi[1,2], Thai Hien Nguyen, MPH[1], Richard Chapman, MS[1], Theodore Leng, MD, MS[1,3], Michael Mbagwu, MD[1,3], Matthew T. Roe, MD, MS[1], Shrujal Baxi, MD, MPH[1], Aracelis Z. Torres, PhD, MPH[1]

[1]Verana Health, San Francisco, CA

[2]Duke Eye Center, Duke University School of Medicine, Durham, NC

[3]Byers Eye Institute, Stanford University School of Medicine, Palo Alto, CA

Background:
International Classification of Diseases (ICD) codes can be used in observational electronic health record (EHR) studies. For diabetic retinopathy (DR), a chronic, progressive disease and a leading cause of blindness, disease severity may be missing from the EHR when unspecified ICD codes are used. Unspecified clinical encounters are often excluded in research studies, potentially introducing selection bias. We sought to understand the mechanism of missingness for DR disease severity.

Methods:
Using data from the American Academy of Ophthalmology IRIS® Registry (Intelligent Research in Sight), we identified all encounters with a DR ICD code from January 1, 2014 to June 30, 2021. We compared clinical and demographic characteristics between encounters with specified and unspecified disease severity.

Results:
Of the 11,215,870 ICD-9 and 25,375,003 ICD-10 clinical encounters, 2,171,971 (19.37%) ICD-9 and 1,168,827 (4.61%) ICD-10 clinical encounters had unspecified DR disease severity. We found significant differences between specified and unspecified encounters: unspecified encounters were found in patients who were younger at the time of encounter, had better visual acuity, fewer ophthalmic procedures during the visit, or who were at their first encounter. Specified encounters occurred more commonly when retinal specialists were the provider for the visit.

Conclusions:
Our findings suggest that DR disease severity is not missing completely at random (MCAR) and is most likely missing not at random (MNAR). The differences observed between the unspecified and specified clinical encounters suggest that unspecified clinical encounters are likely associated with less severe DR. Disease severity may also be missing at random (MAR) due to unobserved practice level differences. Understanding the mechanism of missingness can help inform how to impute data points in the future to understand disease progression and limit biases in observational studies.