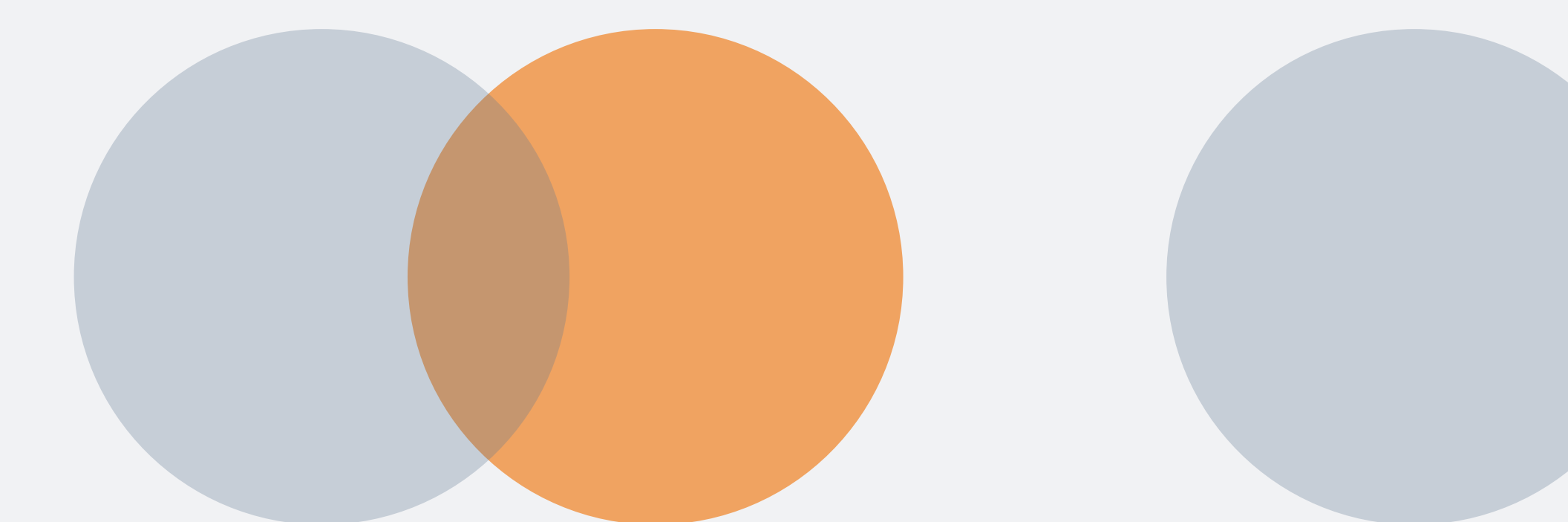


# Combining Data from Medical Registries Reduces The Proportion of Subjects with Missing Demographic Data

Lauren Wiener, MS<sup>1</sup>, Caitlin Rzy, MS<sup>1</sup>, Shrujal S. Baxi, MD, MPH<sup>1</sup>, Aracelis Torres, PhD, MPH<sup>1</sup>, Heather E Moss, MD, PhD<sup>2</sup>

<sup>1</sup>Verana Health, San Francisco, CA  
<sup>2</sup>Stanford University, Stanford, CA



## Purpose

Electronic health record-based medical registries often have missing demographic data, which impedes research. This study aimed to reduce the proportion of missing demographic data for patients with multiple sclerosis (MS) in a neurology medical registry by supplementing with data on the same patients from an ophthalmology registry.

## Methods

- In this cross-sectional study, participants were included if they overlapped temporally in both the American Academy of Neurology Axon Registry® and the American Academy of Ophthalmology IRIS® Registry (Intelligent Research in Sight)
- In the Axon Registry, participants had  $\geq 3$  ICD-9/10 codes for MS on separate dates.
- Age, sex, race, ethnicity and U.S. census region were extracted and classified as conflicting, missing, or not missing in the combined data set.

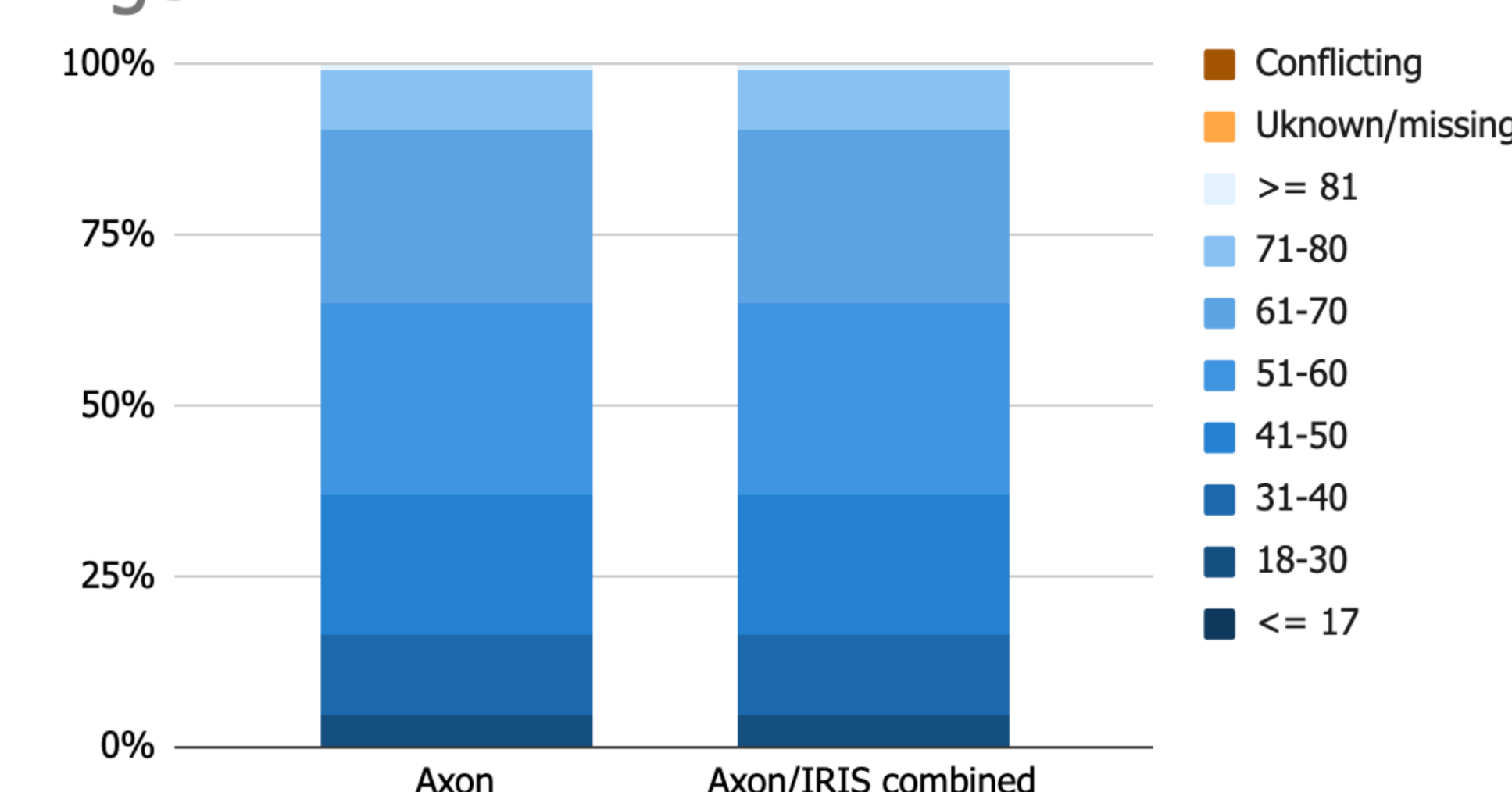
## Results

- Among 60,316 MS patients meeting inclusion criteria in the Axon Registry, 14,085 had temporal overlap in the IRIS Registry.
- Missing demographic data are shown in Table 1.
- When we combined data, there was a reduction in missing data for race, ethnicity, sex and location (all  $p < 0.05$ , McNemar), but not age ( $p = 1.0$ , McNemar).
- When we include the small proportion of conflicting data (Table 1) as “missing,” the missing data for “sex” are no longer significantly reduced.

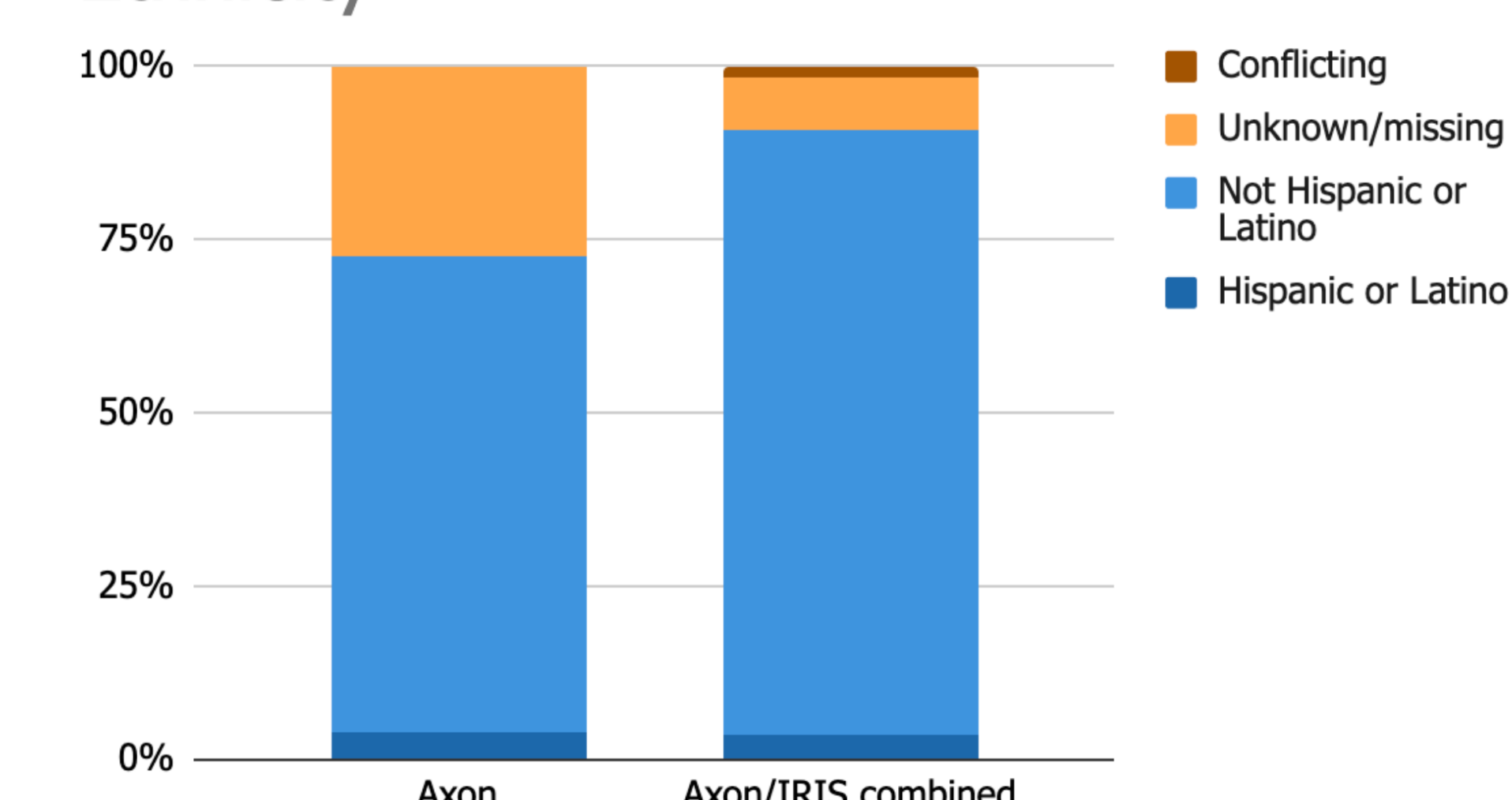
## Conclusion

- Among subjects with MS in the Axon Registry, the proportion of missing demographic data was reduced by supplementing with the IRIS Registry data for race, ethnicity, and location but not for age and sex, for which initial missing proportions were low.
- Sources of conflicting data might include data entry error, data definition differences, and patient travel.

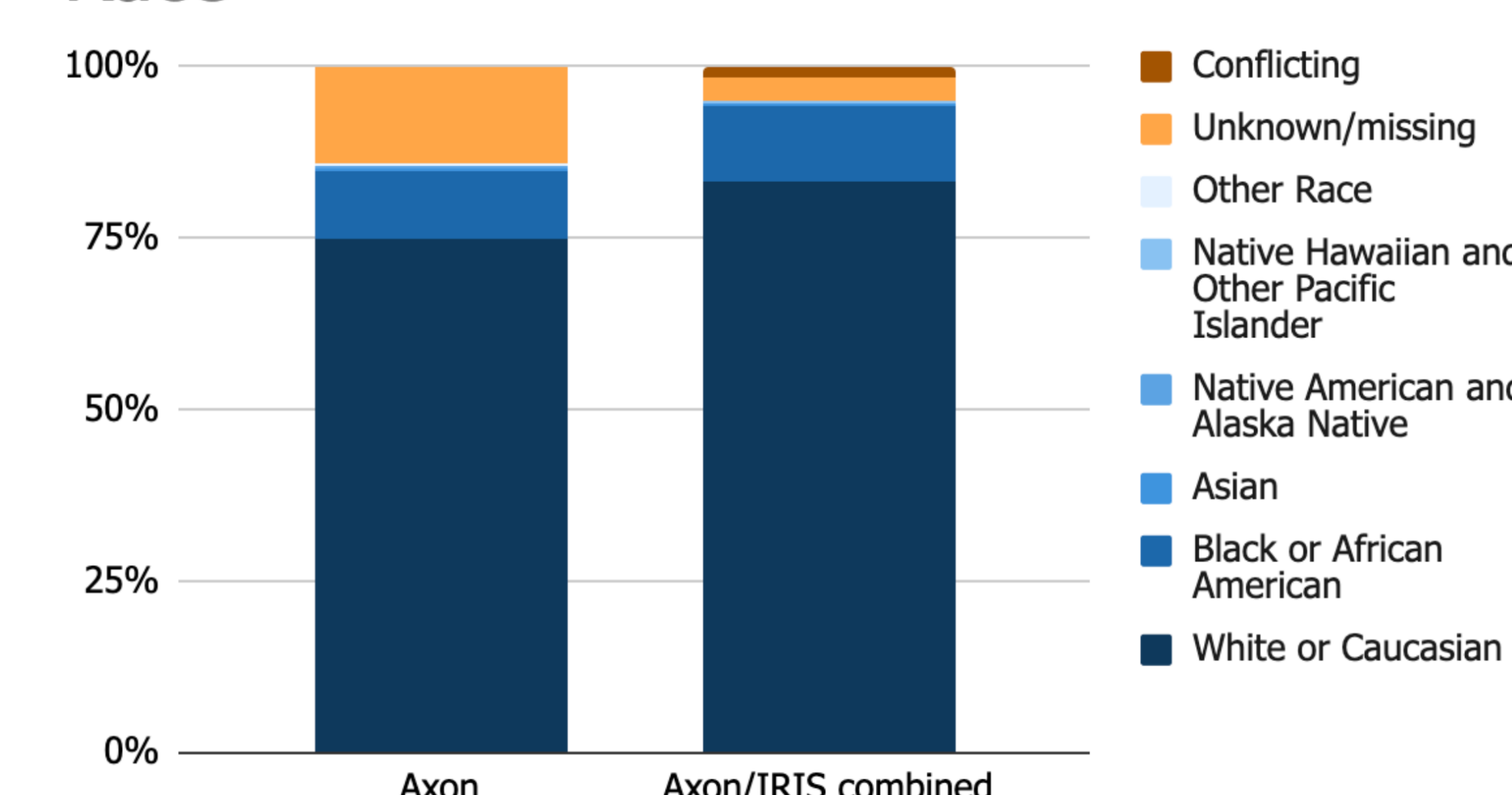
### Age



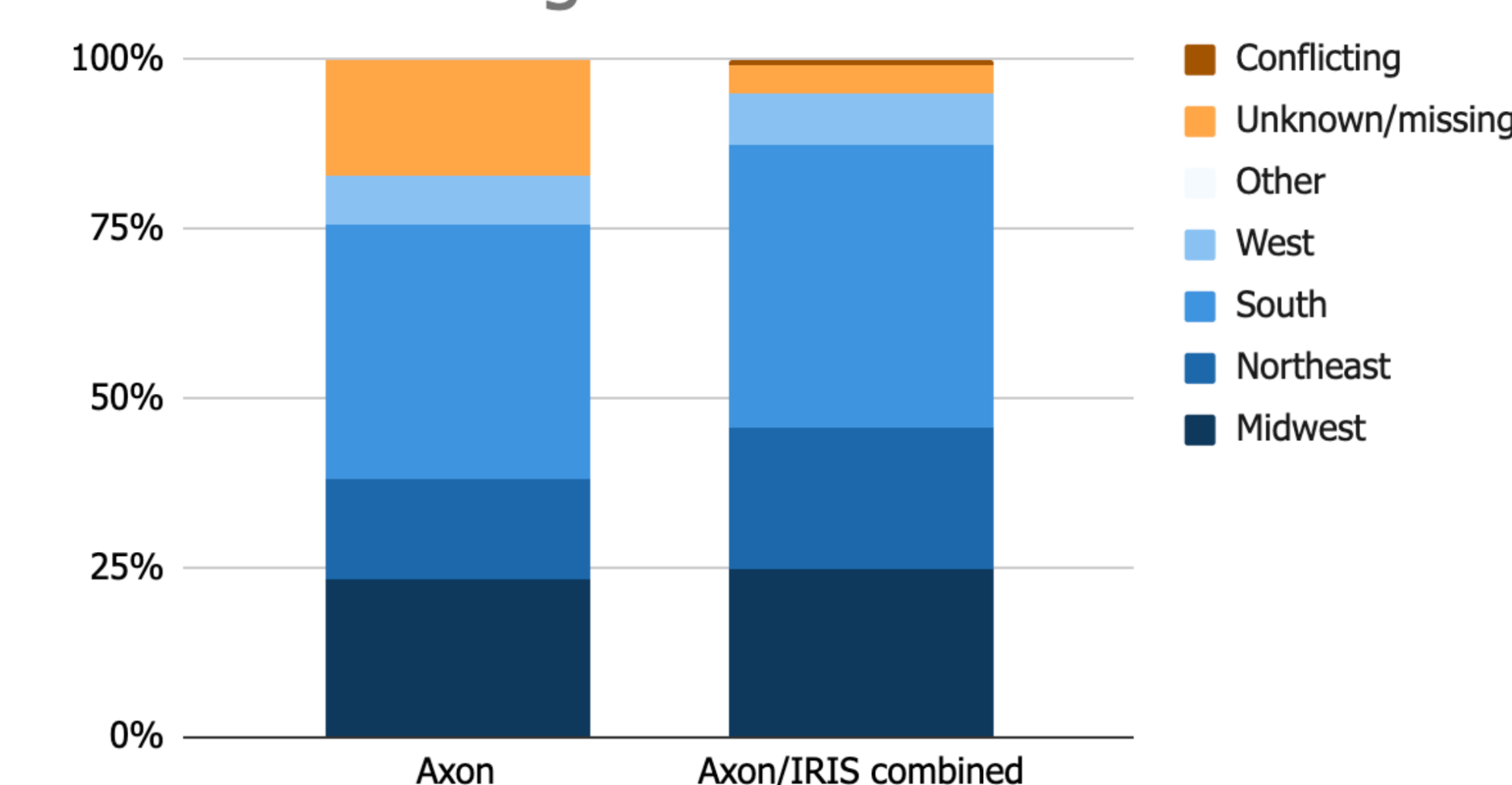
### Ethnicity



### Race



### US Census Region



### Sex

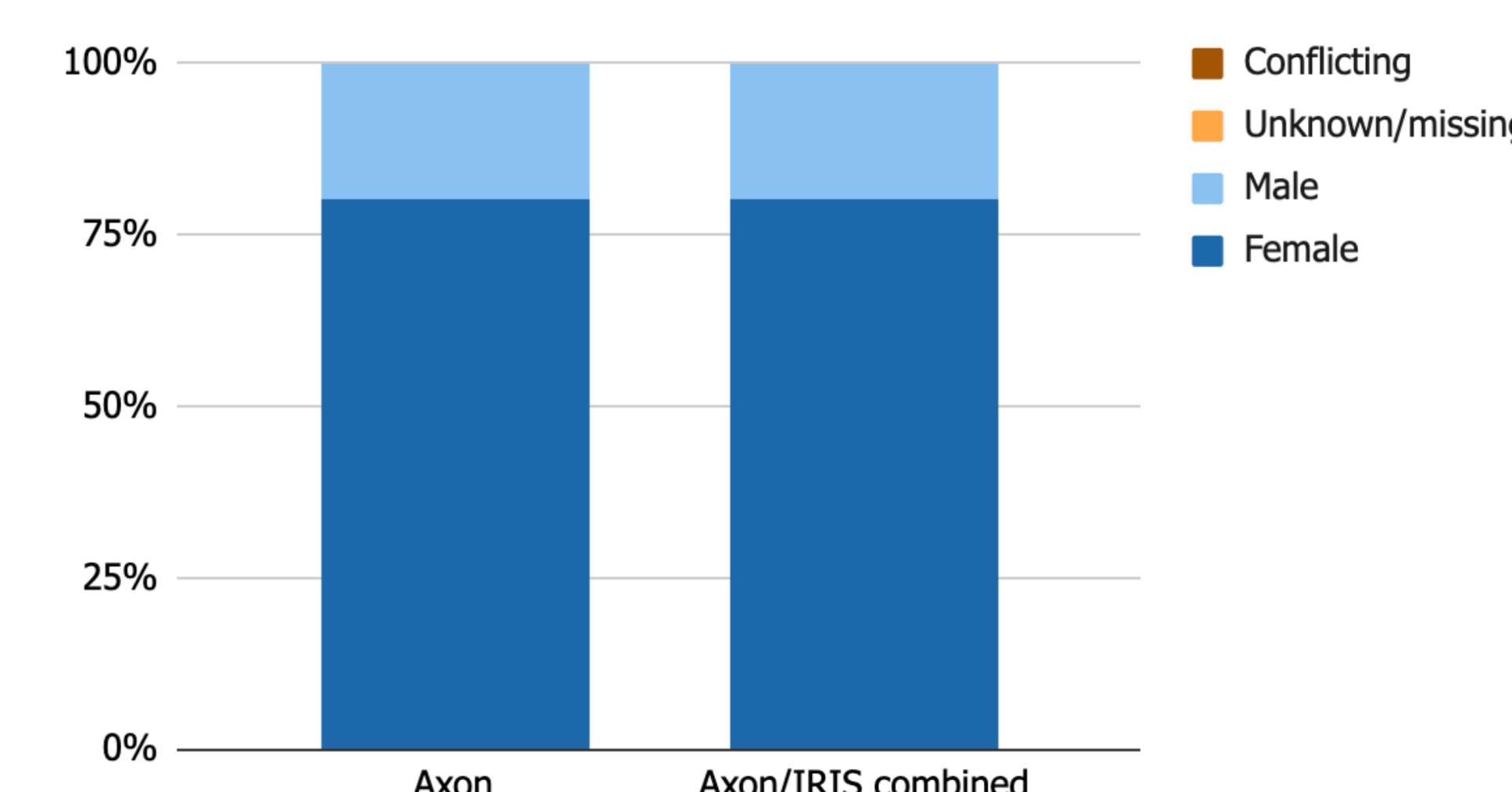


Table 1. Number of participants with missing data

Variable N = 14,085	Axon Registry	IRIS Registry	Combined missing	Combined conflict
Age	0 (0%)	0 (0%)	0 (0%)	1 (0.007%)
Race*	2,005 (14.2%)	2,238 (15.9%)	477 (3.4%)	197 (1.4%)
Ethnicity*	3,876 (27.5%)	3,438 (24.4%)	1,112 (7.9%)	200 (1.4%)
Sex*	17 (0.1%)	45 (0.3%)	0 (0%)	30 (0.2%)
Location*	2,430 (17.3%)	2,415 (17.1%)	574 (4.1%)	117 (0.8%)

\* Reduction in missing data when registries are combined  $p < 0.05$ .

Demographics values for the Axon/IRIS combined cohort were populated as follows:

- If one registry has a null value and the other registry has a non-null value, use the non-null value
- If one registry has an Unknown value, and the other registry is not Unknown, use the not Unknown value
- If both registries are null or Unknown, the final value is Unknown
- If both registries have non-null/non-Unknown values, and the values don't match, the final value is conflicting
- If both registries have values that match, use the matching value