

Assessing the representativeness of survey-eligible individuals identified from a large US national claims database

Judith J. Stephenson*, SM; Lauren E. Parlett, PhD; Amita Ketkar, SM, BDS; Michael Grabner, PhD; Katherine M. Harris, PhD; Vincent J. Willey, PharmD

*Presenting author
Carelon Research, Wilmington, DE

RWD170

Background

- Administrative claims are an efficient and accurate method of identifying members with specific characteristics for direct-to-patient studies.
- The Healthcare Integrated Research Database (HIRD[®]) has served as a sampling frame for numerous survey and qualitative studies.
- However, a better understanding of the characteristics of survey-eligible members (i.e., fully-insured members enrolled in commercial or Medicare Advantage health plans who can be included in direct-to-patient studies) is required.

Objective

To evaluate similarities and differences between HIRD Survey-Eligible and Not Survey-Eligible populations across various demographic and clinical measures.

Methods

Data Sources

- The HIRD is a large, US database curated for health-related research.
- The HIRD includes claims data from health plans in the Northeast, Midwest, South, and West geographic regions of the US and is updated monthly.
- Details about the HIRD's covered population, data structure, data provenance and quality, and example applications have been previously published (Barron 2025).

Population

- HIRD members were required to be continuously enrolled from January 1, 2024, through December 31, 2024, with both medical and pharmacy benefits.
- 'Survey-Eligible' members consisted of those fully-insured members with commercial or Medicare Advantage health insurance who could be included in direct-to-patient studies (e.g., currently active, not enrolled in an administrative services (ASO) health plan, not on any do-not-call lists); otherwise, members were considered 'Not Survey-Eligible'.

Variables

- Compared variables were:

- Insurance Type** (Commercial; Medicare Advantage)
- Sex** (Male, Female)
- Age-Group** (0-17; 18-24; 25-44; 45-64; 65-75; >75)
- Region** (Northeast; South; Midwest; West; Missing)
- Race and Ethnicity** (American Indian or Alaska Native, Non-Hispanic (NH); Asian, NH; Black or African American, NH; Hispanic or Latino of any race; Native Hawaiian or Other Pacific Islander, NH; White, NH; Other race, NH; Unknown or Undisclosed)
- Socioeconomic Status (SES) Index Quartiles** (1 [lowest quartile], 2, 3, 4 [highest quartile]) based on 2022 5-year estimates from the American Community Survey
- Quan-Charlson Comorbidity Index (QCI)** (QCI=0; QCI=1; QCI=2; QCI≥3) (Quan 2005)

Analysis

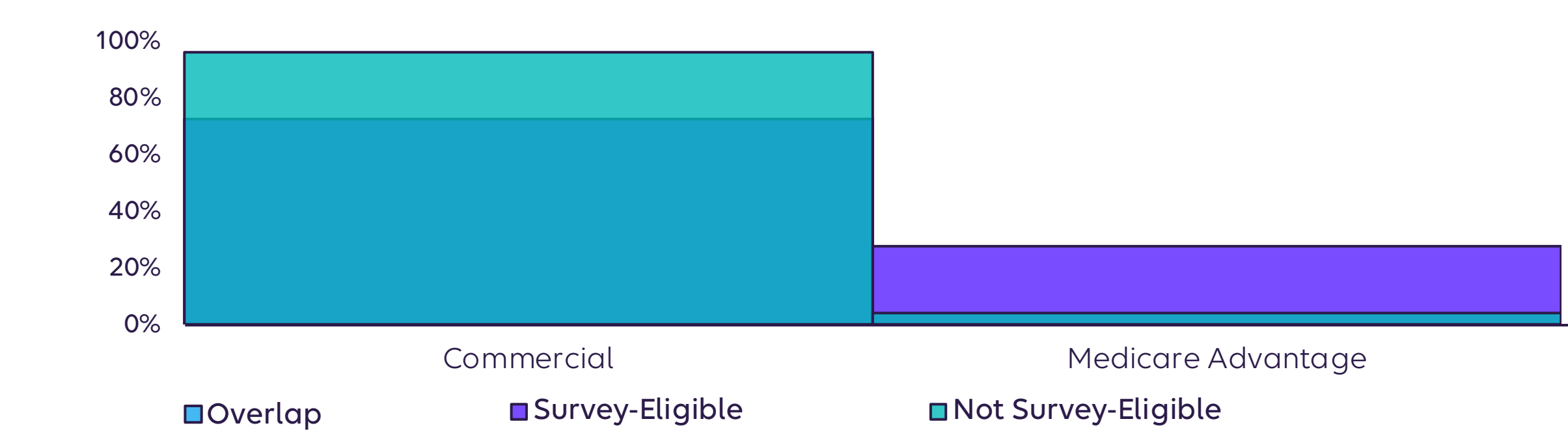
- For categorical variables, frequencies and proportions are reported; for continuous variables, mean, median and standard deviation are reported.
- The overlap index (η) was calculated to quantify the percentage overlap between the probability distributions of categorical variables common to both populations (Pastore 2019).
 - $\eta = 0\%$ means the two distributions are completely separated whereas $\eta = 100\%$ means the two distributions are completely the same.
- The standardized mean difference (SMD) was computed to assess the magnitude of the difference or effect size between the probability distributions of continuous characteristics common to both populations (Yang 2012).
 - Cohen's interpretation of SMD magnitude was used where 0.2 represents a small effect size; 0.5 represents a medium effect size, and 0.8 represents a large effect size.
 - Statements about the representativeness of the Survey-Eligible population compared to the Not Survey-Eligible population were made based on the results of the comparisons.

Results

Comparison of 2024 HIRD Researchable Survey-Eligible and Not Survey-Eligible Populations; 2024 HIRD Researchable Population = 13,970,339

DEMOGRAPHIC CHARACTERISTICS

Figure 1: Insurance Type; Overlap (η) = 76.5%

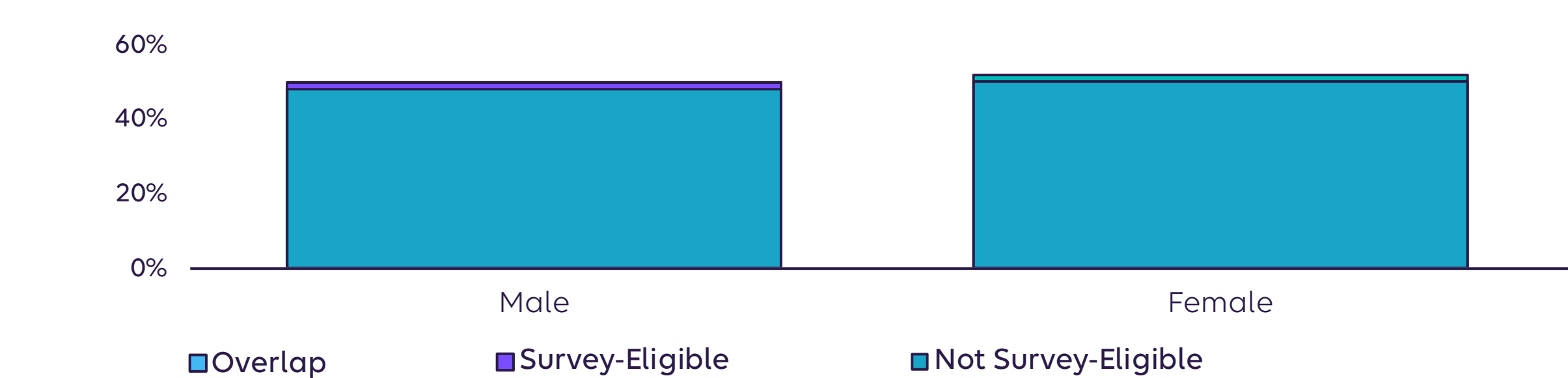


	2024 Survey-Eligible	2024 Not Survey-Eligible
Total Population, N	2,840,653	11,129,686
Insurance Type, %		
Commercial	72.5%	96.0%
Medicare Advantage	27.5%	4.0%

Note: Black font comparisons mean the Not Survey-Eligible percentage is greater than the Survey-Eligible percentage; blue font comparisons mean the opposite.

- Approximately 20% of the 2024 HIRD Researchable Population was survey eligible.
- The Survey-Eligible and Not Survey-Eligible populations are less similar due to a significantly greater percentage of Medicare Advantage members in the Survey-Eligible population, resulting in an $\eta < 80.0\%$.

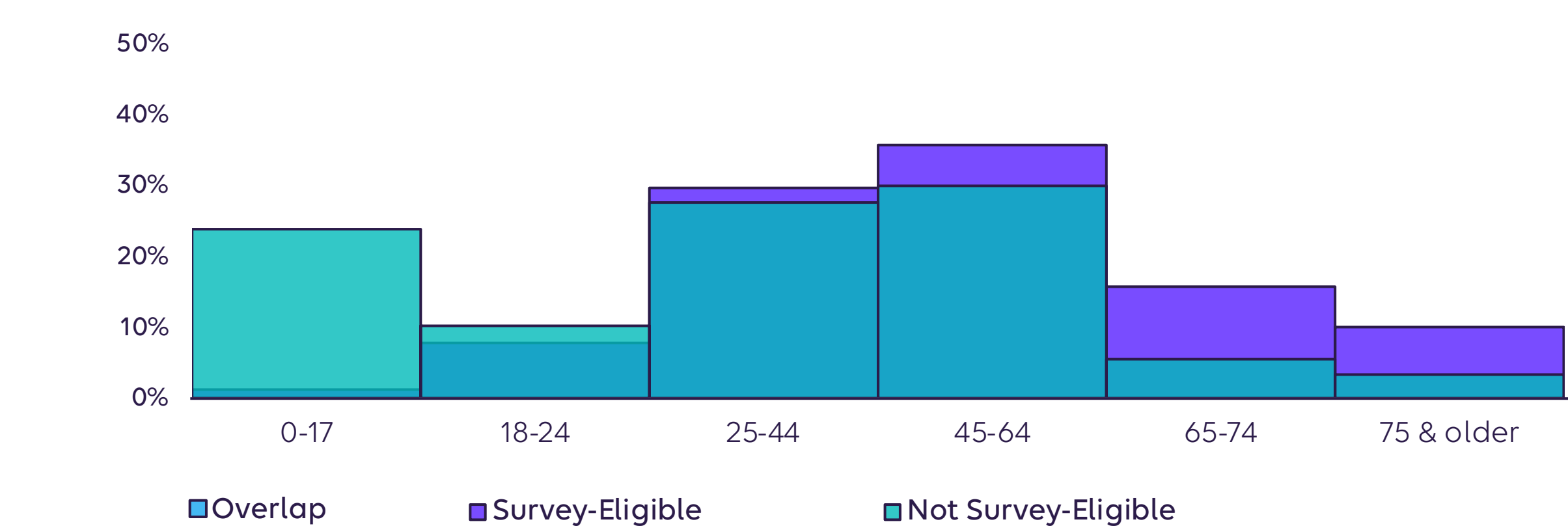
Figure 2: Sex; Overlap (η) = 98.3%



	2024 Survey-Eligible	2024 Not Survey-Eligible
Total Population, N	2,840,653	11,129,686
Sex, %		
Male	49.1%	49.9%
Female	50.9%	50.1%

- The sex distributions of the Survey-Eligible and Not Survey-Eligible populations are very similar with $\eta > 95.0\%$.

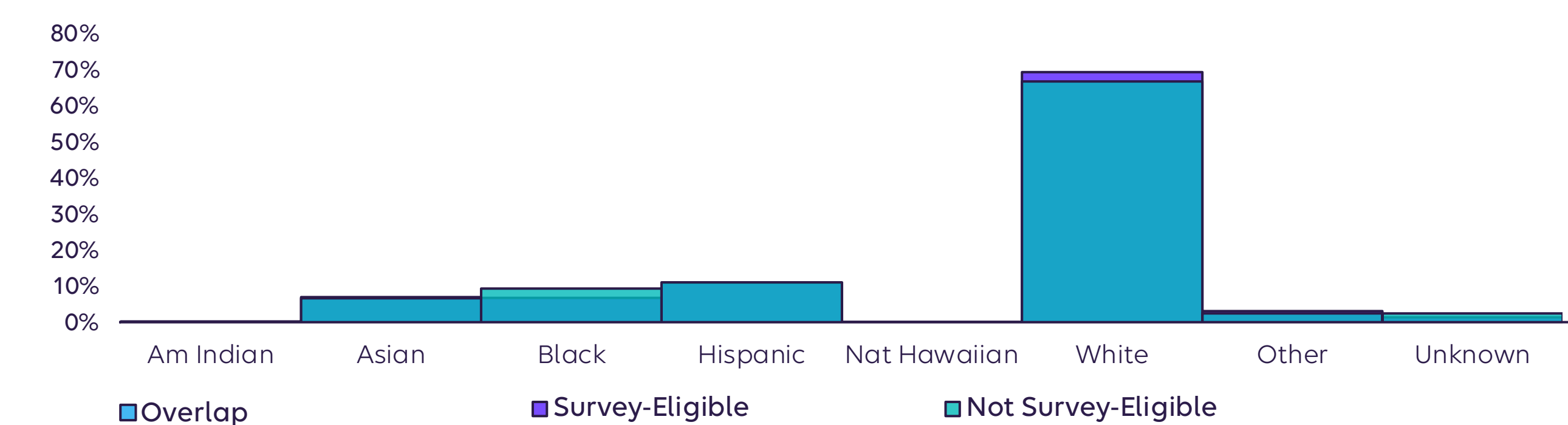
Figure 3: Age Group; Overlap (η) = 75.0%



	2024 Survey-Eligible	2024 Not Survey-Eligible
Total Population, N	2,840,653	11,129,686
Age Group, %		
0 to 17 years	1.2%	23.8%
18 to 24 years	7.8%	10.2%
25 to 44 years	29.6%	27.6%
45 to 64 years	35.7%	29.9%
65 to 74 years	15.7%	5.5%
75 & older years	10.0%	3.3%

- The Survey-Eligible Age Group population is older than the Not Survey-Eligible population due to the greater percentage of Medicare Advantage members, as shown by $\eta < 80.0\%$

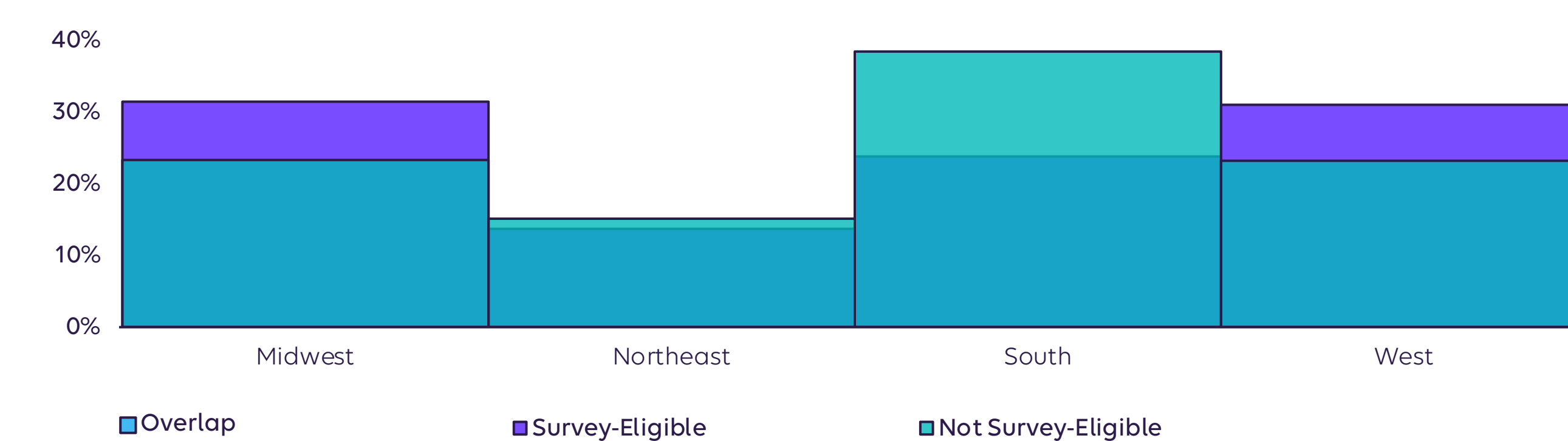
Figure 4: Race and Ethnicity; Overlap (η) = 96.3%



	2024 Survey-Eligible	2024 Not Survey-Eligible
Total Population, N	2,840,653	11,129,686
Race and Ethnicity, %		
American Indian or Alaska Native, Not Hispanic or Latino	0.4%	0.3%
Asian, Not Hispanic or Latino	7.1%	6.8%
Black or African American, Not Hispanic or Latino	6.9%	9.5%
Hispanic or Latino of any race	11.1%	11.2%
Native Hawaiian or Other Pacific Islander, Not Hispanic or Latino	0.2%	0.2%
White, Not Hispanic or Latino	69.5%	66.9%
Other race, Not Hispanic or Latino	3.3%	2.6%
Unknown or Undisclosed	1.5%	1.5%

- The Race and Ethnicity distributions of the Survey-Eligible and Not Survey-Eligible populations are very similar, as shown by $\eta > 95.0\%$

Figure 5: Geographic Region of Residence; Overlap (η) = 84.1%

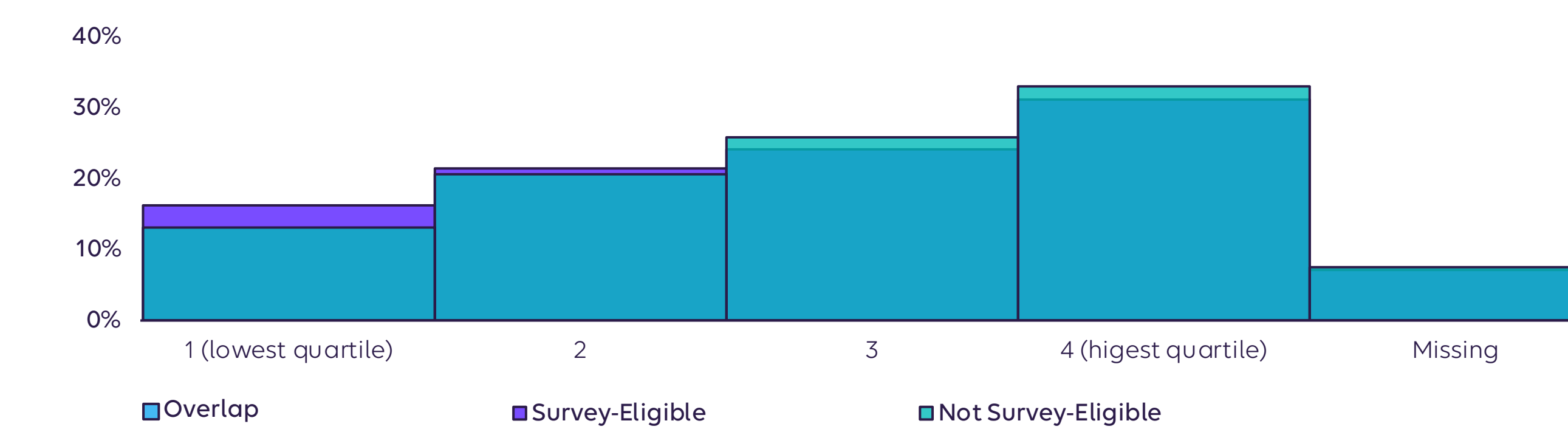


	2024 Survey-Eligible	2024 Not Survey-Eligible
Total Population, N	2,840,653	11,129,686
Geographic Region of Residence, %		
Midwest	31.4%	23.3%
Northeast	13.7%	15.1%
South	23.8%	38.4%
West	31.0%	23.2%

- The Survey-Eligible and Not Survey-Eligible populations are less similar by Geographic Region, as shown by $\eta < 90.0\%$
- A greater percentage of the Survey-Eligible population is found in the Midwest and West regions whereas a greater percentage of the Not Survey-Eligible population is found in the South and Northeast regions

SOCIAL AND HEALTH EQUITY CHARACTERISTICS

Figure 6: SES Index Category; Overlap (η) = 96.1%

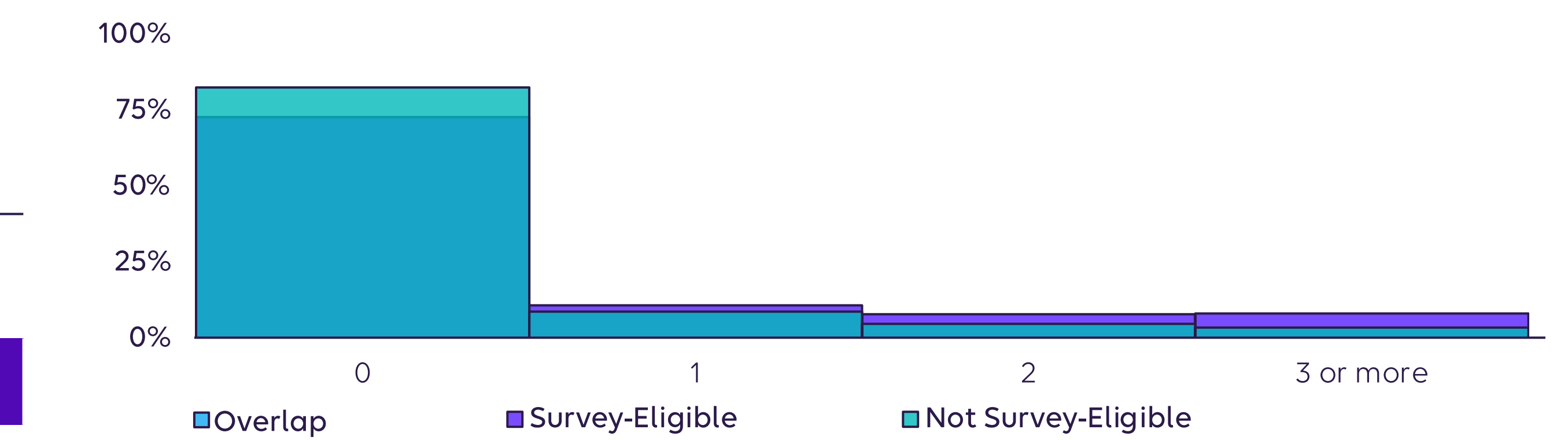


	2024 Survey-Eligible	2024 Not Survey-Eligible
Total Population, N	2,840,653	11,129,686
SES Index Category, %		
1 (lowest quartile)	16.2%	13.3%
2	21.4%	20.6%
3	24.1%	25.8%
4 (highest quartile)	31.1%	33.0%
Missing	7.1%	7.5%

- Although the SES distributions of the Survey-Eligible and Not Survey-Eligible populations are very similar, as shown by $\eta > 95.0\%$, a greater proportion of the Survey-Eligible population resides in the lower 2 SES index category areas (37.6% vs 33.9%).

SOCIAL AND HEALTH EQUITY CHARACTERISTICS

Figure 7: Quan Charlson Comorbidity Index (QCI); Overlap (η) = 90.3%; SMD = 0.28



	2024 Survey-Eligible	2024 Not Survey-Eligible
Total Population, N	2,840,653	11,129,686
QCI, mean (SD)	0.6 (1.3)	0.3 (1.0)
QCI Score Category, %		
0	72.9%	82.6%
1	10.9%	8.9%
2	8.0%	4.9%
3 or more	8.3%	3.6%

- The Survey-Eligible and Not Survey-Eligible populations are similar in terms of their QCI Score probability distributions with η slightly greater than 90%.
- The SMD for the mean QCI score is also low, < 0.3 .

Limitations

- Survey eligibility for a given member of the HIRD is subject to several selection processes; the survey-eligible subgroup is not randomly determined. These selection processes may drive some of the population characteristics' differences.
- Results are derived for 2024; population characteristics may differ for prior and/or subsequent years.
- These results are at the population level and do not reflect any differences based on the population of interest for a given study.

Conclusion

The Survey-Eligible and Not Survey-Eligible populations are highly similar ($\eta > 90\%$) for sex, race and ethnicity, SES Index, Urbanicity and QCI. They are less similar ($\eta < 90\%$) for insurance type, age group, and geographic region due to the greater percentage of Medicare Advantage patients in the Survey Eligible population.

Overall, the Survey-Eligible HIRD population resembles the Not Survey-Eligible population and strengthens confidence in generalizing survey and qualitative research findings to the broader source population.

References

- Barron JJ, Willey VJ, Doherty BT, et al. The Healthcare Integrated Research Database (HIRD) as a Real-World Data Source for Pharmacoepidemiologic Research. *Pharmacoeconomic Drug Saf* 2025 Feb; 34(2): e70110. <https://doi.org/10.1002/pds.70110>
- Pastore M, Calcagni A. Measuring Distribution Similarities Between Samples: A Distribution-Free Overlapping Index. *Front. Psychol*. 2019; 10:1089. <https://doi.org/10.3389/fpsyg.2019.01089>
- Quan H, Sundararajan V, Halfon P, et al. Coding algorithms for defining comorbidities in ICD-9-CM and ICD-10 administrative data. *Med Care* 2005 Nov; 43(11):1350-9. <https://doi.org/10.1097/01.mir.0000182534.19832.88>
- Yang D, Dalton J. A unified approach to measuring the effect size between two groups using SAS. *SAS Global Forum* 2012, Paper 335-2012. <https://support.sas.com/resources/papers/proceedings12/335-2012.pdf> [Accessed 19 March 2026]

Funding & disclosures

This work was supported by Carelon Research. All co-authors are employees of Carelon Research.

Judith Stephenson, Principal Scientist
judith.stephenson@carelon.com

Poster presented at ISPOR 2026, May 17-20, Philadelphia, PA

