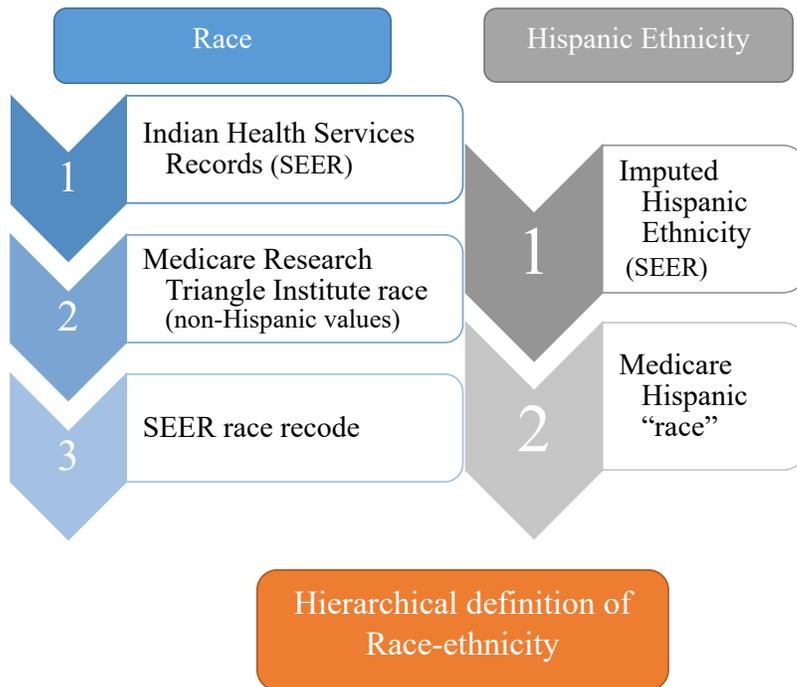


**Supplementary Figure 1. Hierarchy of Data Sources used to Define Race and Ethnicity.**



Race was defined using data from both SEER and Medicare in the following order. Patients with a matching record in the Indian Health Services system were considered AIAN regardless of any other race or ethnicity values. Next, we obtained Medicare Research Triangle Institute race from enrollment files, which has optimal sensitivity compared to patient self-report as the gold standard.<sup>36</sup> For patients categorized as Hispanic or unknown Medicare race and those with more than one value reported, we obtained additional information from SEER. We manually compared data from all sources to identify an informative category where possible. For patients with both a known and unknown race value, the known value was used. Patients with two distinct racial groups were considered “other” race.

The SEER derived Hispanic origin variable was the primary source of ethnicity data; this relies upon the validated North American Association of Central Cancer Registries Hispanic Identification Algorithm when Hispanic ethnicity is not otherwise recorded.<sup>73-75</sup> For patients classified as non-Hispanic, we evaluated Medicare race to capture any additional patients who classified themselves as Hispanic.