Contents

eMethods. 2

eMethods.

Covariates

We identified the following factors from the SEER database: sex, race/ethnicity, age at diagnosis, year of diagnosis, tumor factors (histological classification, stage at diagnosis, and tumor grade), treatment factors (radiation therapy, chemotherapy, surgical recommendations, surgery status, surgery of primary site and scope of regional lymph node surgery), and socioeconomic factors (marital status, median household income, and rural-urban classification). We utilized the SEER race/ethnicity classification framework, which was operationalized by combining Race recode and Race and origin recode into five categories: Non-Hispanic White, Non-Hispanic Black, Non-Hispanic API, Non-Hispanic AI/AN, and Hispanic. Due to limited sample size, AI/AN patients were excluded in subsequent analyses. Tumor stage was consolidated from multiple systems (including SEER historic stage, AJCC6th and7th editions, and SEER Combined Stage) into stages I–IV, with non-applicable cases labeled as such. Surgical recommendations and surgery status were derived from raw SEER variable 'Reason no cancer-directed surgery', which captures the reason for the presence or absence of cancer-directed surgery.

Outcomes

The follow-up of the patients was from the date of Pulmonary Neuroendocrine Neoplasms diagnosis until the date of the outcome (that is, death), censored or the end of the study (that is, 31 December 2019), whichever occurred first. The outcomes of interest were all-cause and cause-specific death. All-cause death means death due to any cause. Cause-specific death was classified into dead attributable to this cancer, other-cancer-cause and non-cancer-cause death. Patients with confirmed cancer-attributable mortality in the SEER cause-specific death classification were directly mapped to the outcome. The remaining death samples were classified as other-cancer-cause and non-cancer-cause death based on the values of COD to site recode (eTable 2).

Statistical analysis

To comprehensively assess temporal trends in racial/ethnic disparities, we evaluated both absolute and relative measures. Cumulative incidence of death was estimated using the Fine and Gray model (via the cuminc function in R's tidycmprsk package). Analyses were conducted for the overall cohort and stratified by diagnostic period (categorized as 2000–2004, 2005–2009, 2010–2014, and 2015–2019) and race/ethnicity. Absolute disparities were quantified as rate differences in 5-year cumulative mortality incidence between the earliest (2000–2004) and most recent period available (2015–2019). To ensure adequate follow-up for 5-year estimates, the analytical cohort was restricted to patients diagnosed through 2015.

We evaluated both absolute and relative measures. Absolute disparities were quantified by the differences in 5-year cumulative mortality incidence. We employed Fine and Gray model (implemented via the cuminc function in the R package tidycmprsk) to estimate cumulative incidence of death. Analyses were performed for the overall cohort and stratified by race/ethnicity and diagnostic era. Diagnostic years were categorized into four temporal cohorts: 2000–2004, 2005–2009, 2010–2014, and 2015–2019. For the 5-year cumulative incidence estimates, we restricted the analytical sample to patients diagnosed through 2015 to ensure adequate follow-up duration. We quantified absolute racial/ ethnic disparities using absolute rate difference for the 5-year cumulative incidence of death between the period 1995–1999 and the period 2010–2014.

To quantify relative disparities, we employed the Accelerated Failure Time (AFT) model (R package survival) to estimatetime ratios (TRs) with 95% confidence intervals (CIs), using non-Hispanic White patients as the reference population. Initial analyses examined all-cause death, followed by cause-specific death in separate models. Temporal trends in relative disparities were assessed by incorporating multiplicative interaction terms between race/ethnicity and diagnostic period. Statistical significance of interaction terms ($P < 0.05$) was evaluated using Wald χ^2 tests. The use of the AFT model is due to the violation of the proportional hazards assumption for key covariates such as race and diagnostic period (eFigure 2).

To assess the cumulative impact of socioeconomic, tumor, and treatment factors on expected survival time, a series of stepwise AFT models were used to calculate the time ratios and their 95% confidence intervals for race with respect to all-cause and cause-specific mortality in different periods and the overall cohort. Socioeconomic, tumor, and treatment factors were adjusted in sequence. In model 1, we adjusted for age, sex, and race/ethnicity. Model 2 additionally adjusted for socioeconomic factors based on model 1. Model 3 additionally adjusted for tumor factors based on model 2. Model 4 additionally adjusted for treatment factors based on model 3.

To further evaluate mediating factors underlying racial/ethnic disparities, we implemented mediation analyses using AFT models. Based on prior knowledge and data availability, we selected candidate mediators associated with disparities, including socioeconomic, tumor and treatment factors. Variables significantly associated with outcomes in univariate analyses ($P<0.05$) were retained for mediation testing. We constructed a causal inference diagram (DAG) to describe the causal problem of this study. Factors such as age and sex act as confounding factors. Based on this DAG, we conducted a series of causal mediation analyses to assess the individual mediation effect of each mediator and the cumulative mediation effect of each category of risk factors (socioeconomic, tumor, and treatment factors).

Benjamini-Hochberg (BH) correction was used for false discovery rate (FDR) control at $\alpha = 0.05$. All statistical analyses and data visualizations were executed using R statistical software (version 4.4.1).