Supplementary Appendix

Table of Contents

Supplement 1: Data collection , and the definition of never-smoker, former smoker, current smoker, pack-years, and quit-years

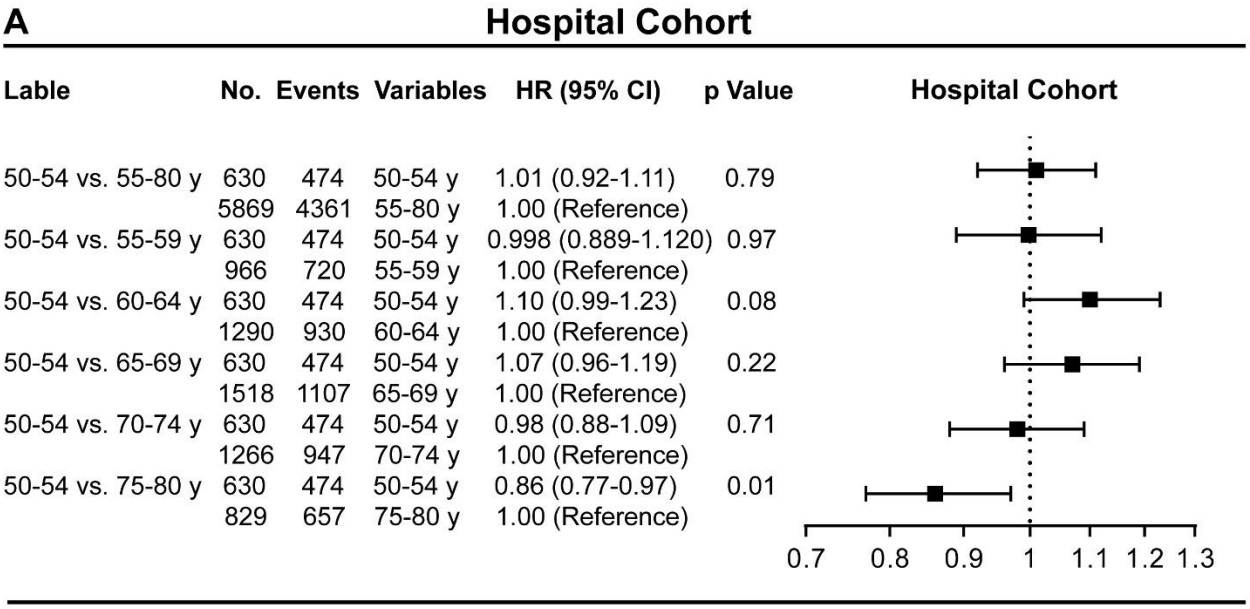
For each new patient, medical records were reviewed for eligibility and the following information was collected: demographic characteristics (age at diagnosis, sex, and race), cigarette smoking history (smoking status, history of cigarette smoking in pack-years, duration of smoking and, if quit, the number of years since the person quit smoking), and clinical characteristics (histologic diagnosis of lung cancer, stage, and treatment modality). The detailed information above was supplemented with additional information from structured subject interviews, follow-up questionnaires, or both. For patients who have received medical care outside of Mayo Clinic, copies of the relevant medical records were requested. All enrolled patients were actively followed up. Follow-up data were collected through comprehensive medical record abstraction and self-administered questionnaires, including current health information and treatment updates starting within 6 months after diagnosis and annually thereafter. Annual verification of patients' status was performed through the Mayo Clinic's electronic medical records and registration database, death certificates, next-of-kin reports, obituary documents filed in the patients' medical records through the Mayo Clinic Tumor Registry and the Social Security Death Index website. Our manually abstracted data were used as a gold standard to train a Natural Language Processing (NLP) tool for accurately retrieving medical records such as smoking status. (Reference 18 in the manuscript) While training the algorithms of the NLP tool, we discovered human errors (approximately 1/1,000) in our original data, for which we made the needed correction.

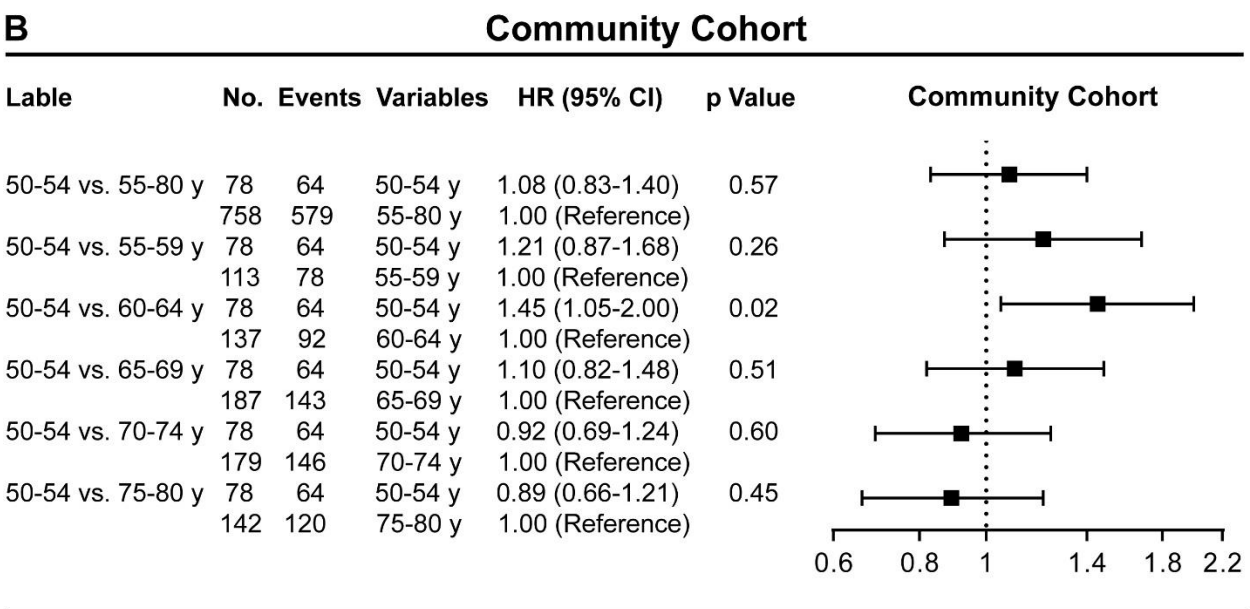
Pack-years were calculated by multiplying the number of cigarette packs smoked daily by the number of years smoked. Quit-years for former smokers were calculated by subtracting the age at which the patient last stopped smoking regularly from their current age.

Supplement 2: Age-group stratified analysis

As shown in **Supplement Figure 1A**, 5-year survival analyses of the YAG (patients aged 50–54·9) using unadjusted Cox proportional hazards regression showed similar survival when compared with the USPSTF subgroup aged 55–59 years (HR = 0·998, p=0·97), and a trend toward worse survival than two subgroups, including patients aged 60–64 years (HR = 1·10, p=0·08) and those aged 65–69 years (HR=1·07, p=0·22), in the hospital cohort. The YAG tended to have better survival than the other two subgroups, including patients aged 70–74 years (HR = 0·98, p=0·71) and those aged 75–80 years (HR = 0·86, p= 0·01). The trend for patient survival of USPSTF subgroups in the community cohort was relatively consistent with that in the hospital cohort (**Supplement Figure 1B**). The 5-year survival of YAG demonstrated a trend toward worse survival than three subgroups, including patients aged 55–59 years (HR = 1·21, p=0·26), those aged 60–64 years (HR = 1·45, p=0·02), and those aged 65– 69 years (HR = 1·10, p=0·51), in the community cohort. The YAG showed a trend toward better survival than another two subgroups, including patients aged 70–74 years (HR = 0·92, p=0·60) and those aged 75–80 years (HR = 0·89, p=0·45).

Supplement Figure 1: Survival analyses of the younger age group (YAG) vs. United States Preventive Services Task Force (USPSTF) subgroups





The 5-year survival analyses of the YAG vs. USPSTF subgroups were analyzed by unadjusted Cox proportional hazards regression and presented in (A) the hospital cohort and (B) the community cohort.