Changing landscape of liver cancer in California: a glimpse into the future of liver cancer in the United States

Supplementary Methods and Results

Summer S. Han, Ph.D., Scott Kelly, Ph.D., Yuqing Li,Ph.D., Baiyu Yang, Ph.D., Mindie Nguyen, M.D., Samuel So, M.D., Philip S. Rosenberg, Ph.D., and Ann W. Hsing, Ph.D.

APC Forecasting Models and sensitivity analysis

We analyzed 10 subsets of HCC cases in California by sex and ethnicity/race subgroup (White, Black, Hispanic, Asian, and All groups combined). For each subset, incidence rates by single-year of age (ages 35 – 84 years old) and single calendar year were obtained from SEER. Our models included the 14 calendar years from 2000 through 2013 and we made projections for 2014-2030.

In general, the number of cohorts in the APC analysis equals the number of age groups A plus the number of calendar periods P minus 1 , and the sequence of cohorts $c_1, c_2, \ldots, c_{A+P-1}$ is defined by the oldest cases in the earliest period (i.e. 1916) through the youngest cases in the latest period (i.e. 1978), yielding 63 cohorts in all.

We calculated our forecasts of incidence using age-period-cohort (APC) forecasting models. APC model parameters quantify the age-associated natural history, secular perturbations (period effects), and relative increases or decreases from one generation to the next (cohort effects). Projecting ages $a_1, a_2, \ldots, a_A$ (35, …, 84 in our example) forwards through *T* = 17 future periods (2014, …, 2030), the youngest $A-1$ observed cohorts will continue to be followed, and each future year a new cohort will enter follow-up beginning at age a_1.

We developed three approaches to forecast cancer incidence (rate per 100,000 person-years) and cancer burden (absolute numbers of cases) by single-years of age and period for each of the 10 subsets of cases. Please refer to our earlier paper [1] for the definitions of the functions below. The first approach proceeds as follows.

Step 1: Fit a classical APC model [2].

Step 2: Express the expected incidence in longitudinal form, so that the expected rate at age a among individuals born in calendar year c equals $R_{LONG}(a \mid c) = LongAge(a \mid c_0) \times CRR(c \mid c_0) \times e^{PD(c+a)}$. In this equation, the function $LongAge(a \mid c_0)$ describes the expected incidence at age a among individuals born in an arbitrary reference year c_0 and followed longitudinally, $CRR(c \mid c_0)$ is the incidence

rate ratio for cohort c versus c_0, and $PD(c+a)$ are non-linear period effects that are constant for all age groups during calendar period $p = c + a$. Also express the corresponding cross-sectional form $R_{CROSS}(a|p) = CrossAge(a|p_0) \times PRR(p|p_0) \times e^{CD(p-a)}$ described in reference [1].

Step 3: Fit a JoinPoint piecewise linear regression [3] to the logarithm of the $CRR(c \mid c_0)$ values accounting for their estimated variance-covariance matrix. Use the unconditional approach described in that paper to conservatively estimate the variance of the JoinPoint regression accounting for uncertainty about the number and location of the knots. Extrapolate from the last JoinPoint segment to impute a cohort rate ratio value for cohorts that enter follow up after 2013. Also carry out a JoinPoint analysis on the logarithm of the model's period rate ratio curve PRR(p|p_0). See Rosenberg et al. for details [1].

Step 4: Partition the period deviations into an orthogonal quadratic component plus residuals or higher-order period effects. See Chernyavskiy et al. for details [4].

Step 5: [Sensitivity Analysis] Construct forecasts by plugging in age-period-cohort parameter estimates for observed and future ages and periods to the expression for R_{LONG}(a|c) and R_{CROSS}(a|p). We consider three scenarios for sensitivity analysis. Our base age-cohort forecasting model R_{LONG}(a|c) extrapolates an effect for future cohorts using the JoinPoint results for cohorts from Step 3 and assumes that future period deviations are 0. Our first sensitivity model is an age-period forecasting R_{CROSS}(a|p) model that assumes that future cohort deviations are 0, while extrapolating an effect for future periods using the JoinPoint results for periods from Step 3 above. Our second sensitivity model is an age-cohort-period forecasting model. This model adds to our base model, by adding to it an extrapolation from the quadratic component of the period deviations described in Step 4. This additional parameter modulates the period effects for future years to accelerate or decelerate at the average rate of change over the observed years.

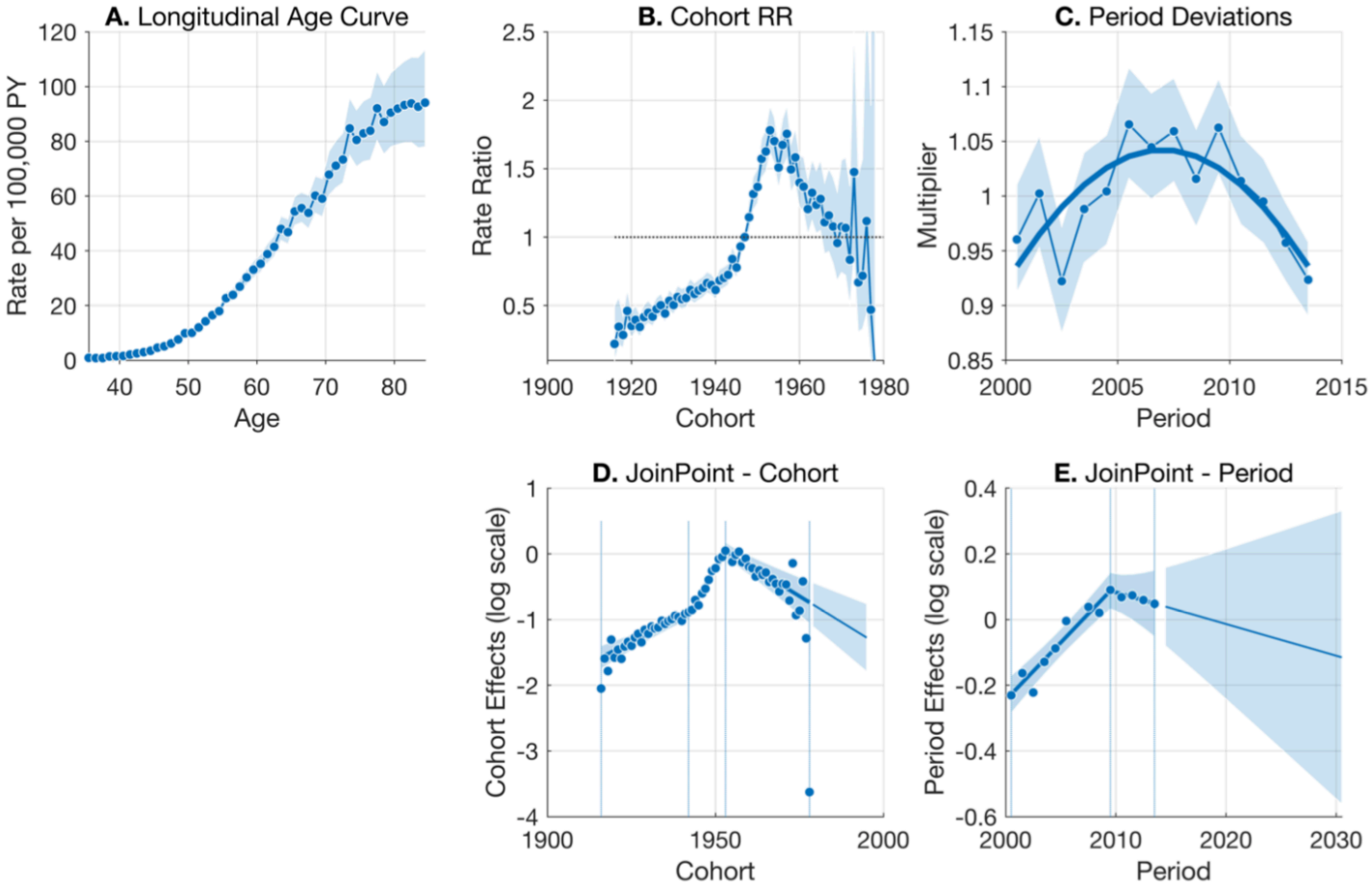
From these results, we calculated plausible ranges centered at the averages of the 3 models described above, ranging from the lowest lower confidence limit to the highest upper confidence limit.

Step 6: Calculate future burden by multiplying the SEER-derived incidence forecasts by single-years of age and period by the corresponding California population projections. For variance calculations assume that the population projections are known quantities.

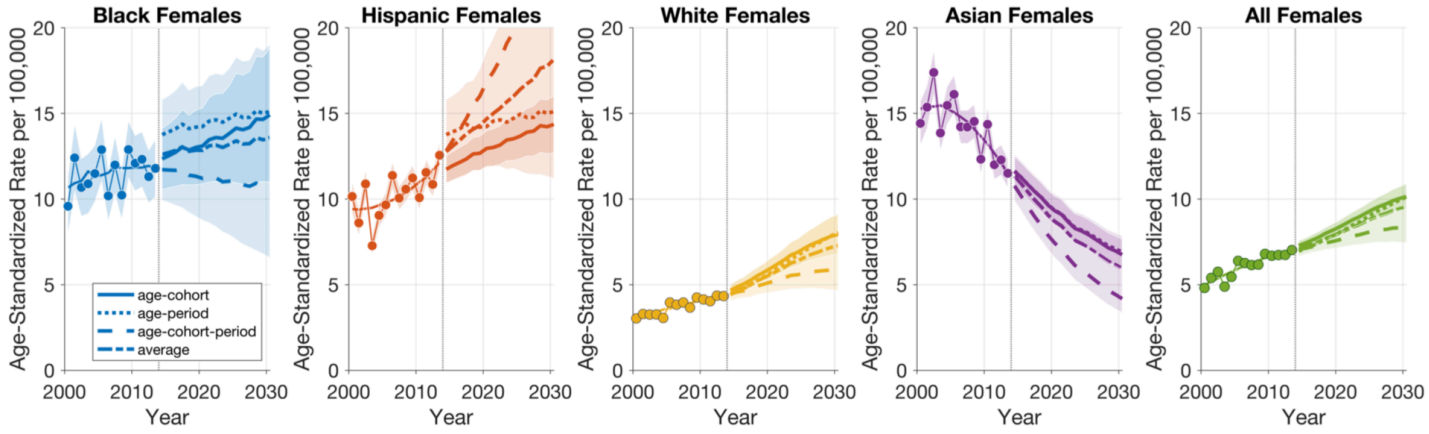
Examples on parameter estimates are shown for illustrative purposes for all males in California in **Supplementary Figure 1**. Age-period-cohort fitted rates are a product of the estimates shown in panels A.-C. Estimates for future rates include estimates in panel D for the age-cohort model, panel E for the age-period model, and panel D and C (smooth curve) for the age-cohort-period model. For all males in California, JoinPoint analysis shows that the risk increased among successive cohorts born from circa 1945 through 1955, and then moderated (panel D). Temporal effects have also moderated slightly in recent years. However, the rapid rise among the oldest baby-boomers will drive increasing incidence for the next 10-15 years, as shown in Figure 1A.

In the main text, we have presented results from the age-cohort model, to be consistent with our previous studies [5]. Quantitative forecasts vary according to the extrapolation model, and the differences between the models increase as we look further forward in time. **Supplementary Figure 2** summarizes quantitative forecasts for Black, Hispanic, White, Asian, and all males by model. **Supplementary Figure 3** summarizes model outputs for females. Taken together, these results suggest that the declines forecast for Asian men and women are supported by each model. Hispanic females may increase at a faster rate than our base model, white males and females may plateau rather than increase, and Black men and women may plateau rather than slowly increase.

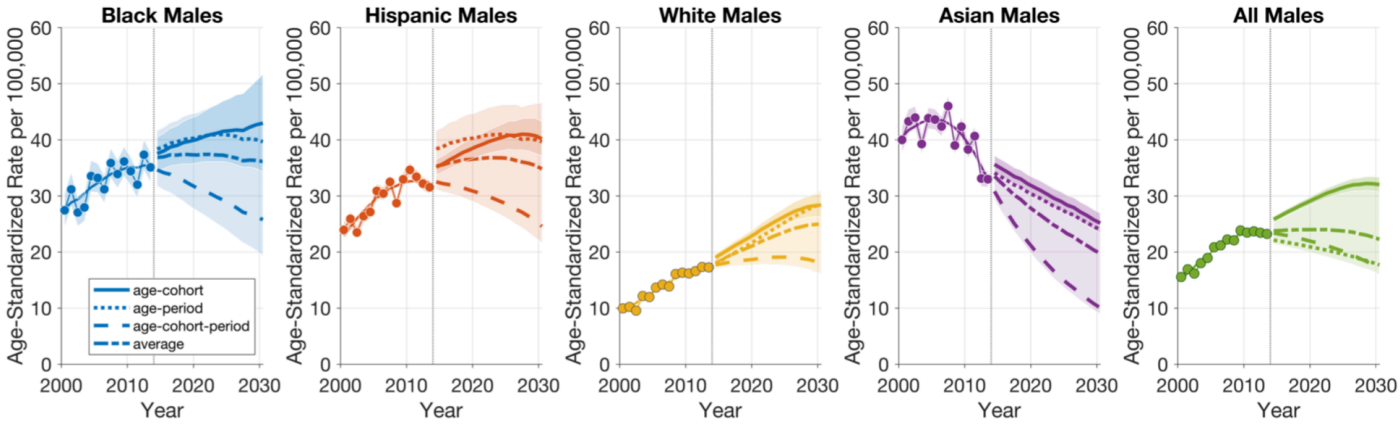
Supplementary Figure 1 Age-Period-Cohort Parameters for HCC incidence among all California males. Age-period-cohort fitted rates are a product of the estimates shown in panels A.-C. Estimates for future rates include estimates in panel D for the age-cohort model, panel E for the age-period model, and panel D and C (smooth curve) for the age-cohort-period model.



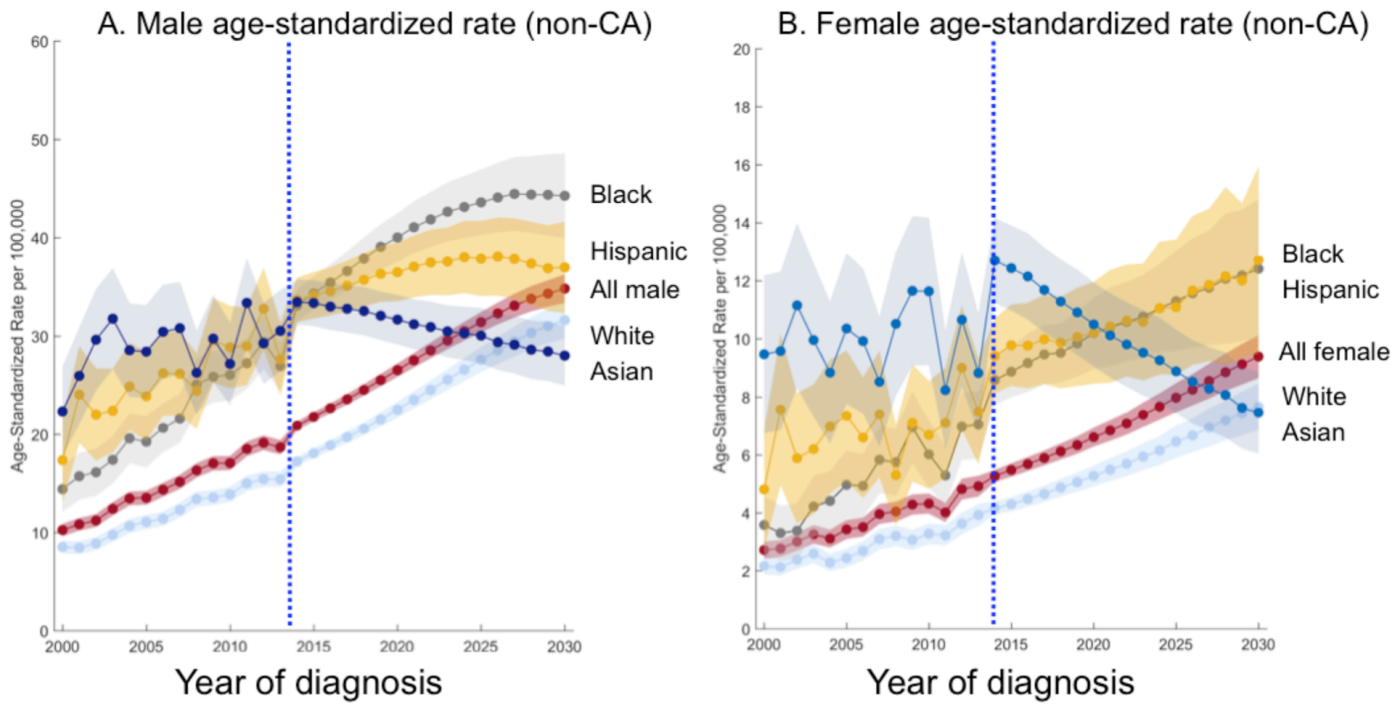
Supplementary Figure 2: Sensitivity analysis for forecasting among females. Each figure summarizes quantitative forecasts for Black, Hispanic, White, Asian, and all males by different forecasting models (base model: age-cohort model)



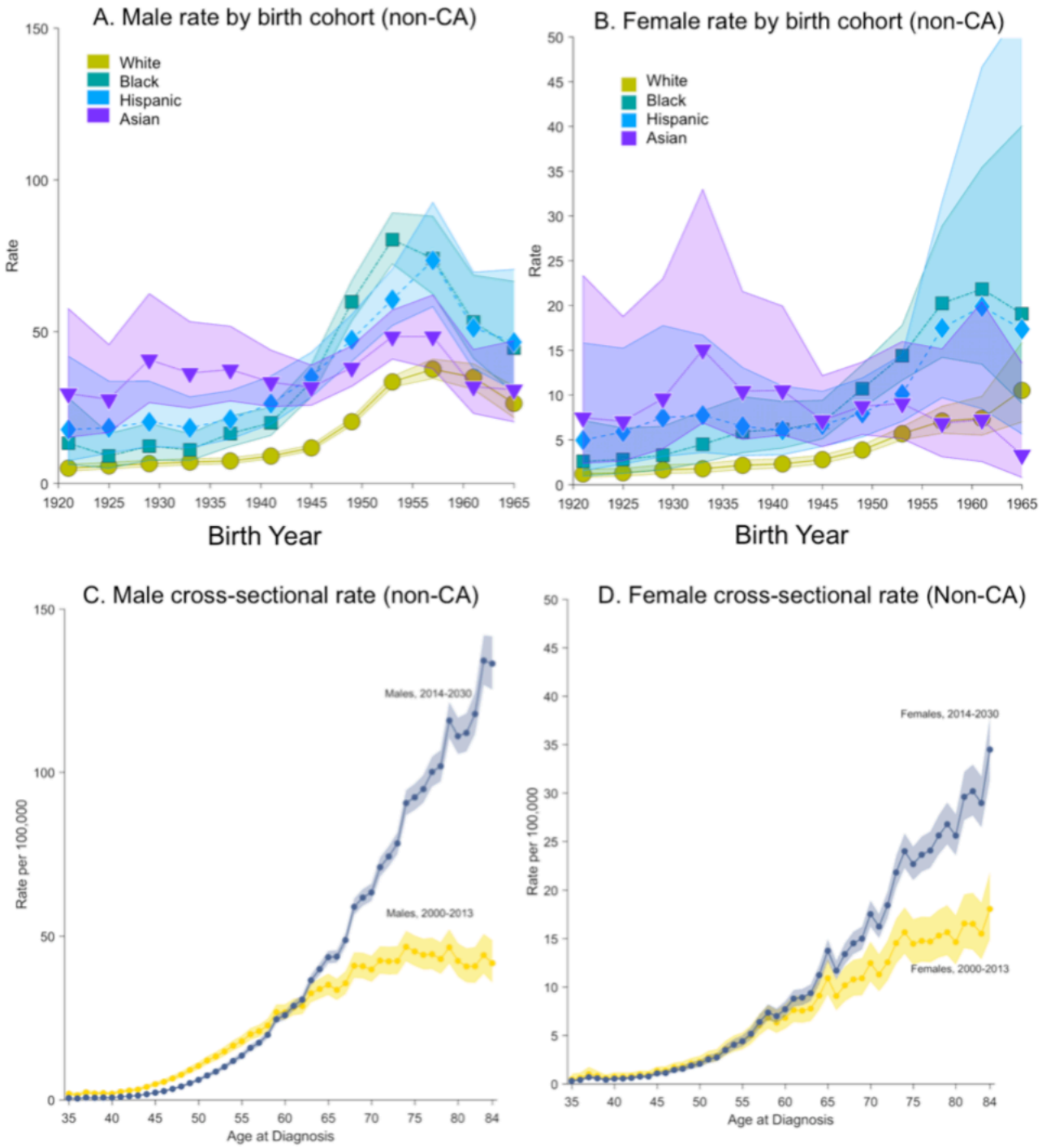
Supplementary Figure 3: Sensitivity analysis for forecasting among males. Each figure summarizes quantitative forecasts for Black, Hispanic, White, Asian, and all males by different forecasting models (base model: age-cohort model)



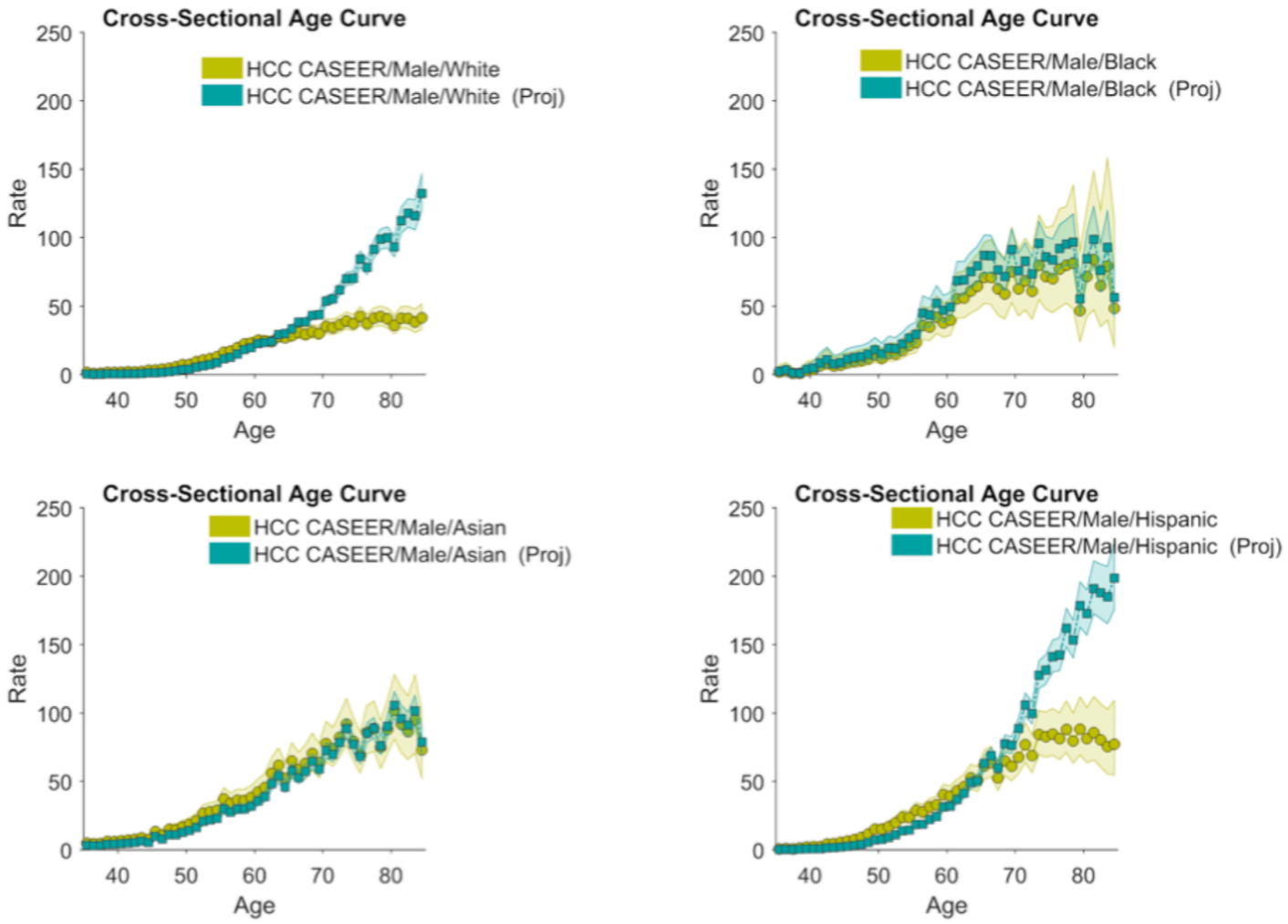
Supplementary Figure 4: Age-standardized HCC incidence rates by period (2000-2013 for the past and 2014-2030 for the future) in the states other than CA (non-CA)



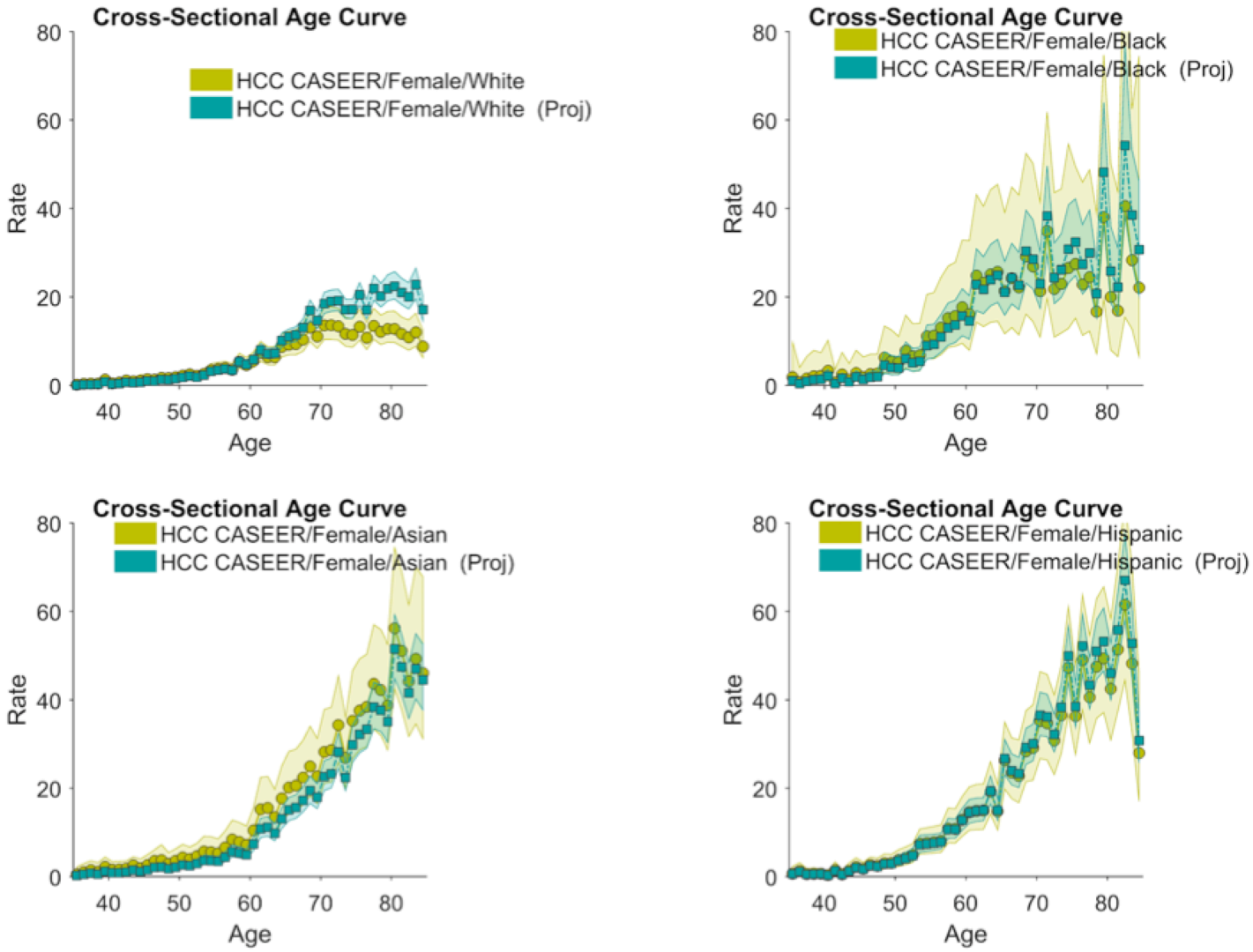
Supplementary Figure 5. HCC incidence rate by birth cohort (A-B); and cross-sectional age curve for the past (2000-2013) versus the future (2014-2030) (**C-D**) in the states other than CA (non-CA).



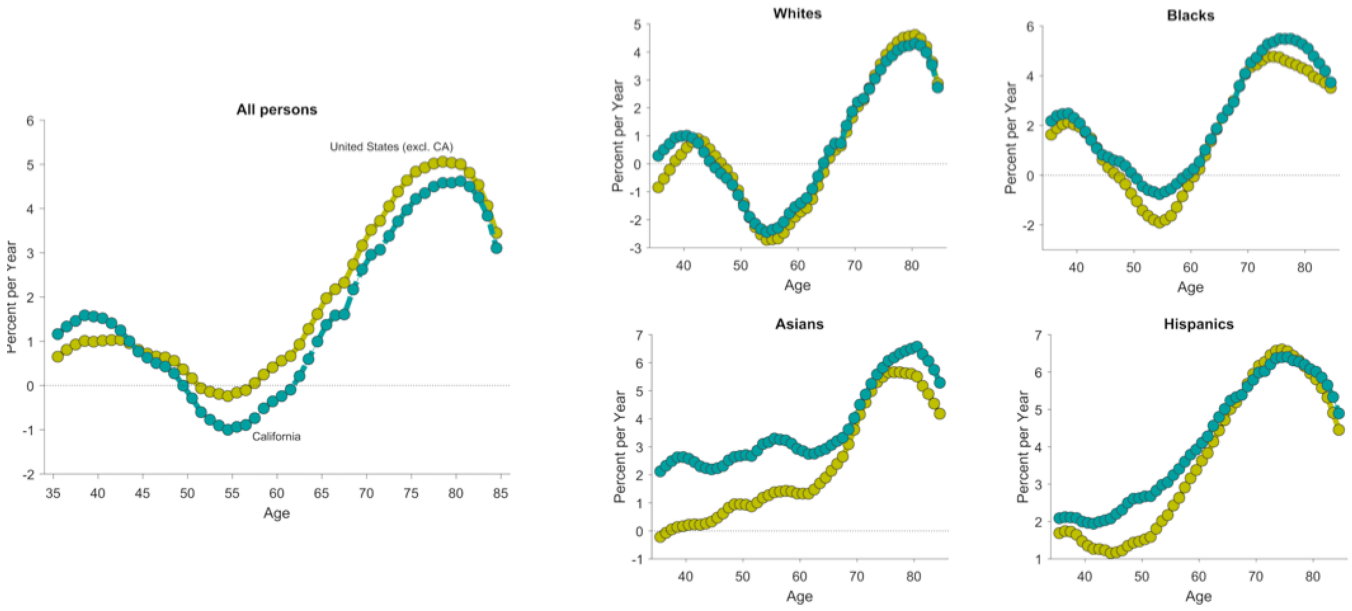
Supplementary Figure 6 Cross-sectional age curve for the past (2000-2013) versus the future (2014-2030) by ethnicity/race in CA (MALE)



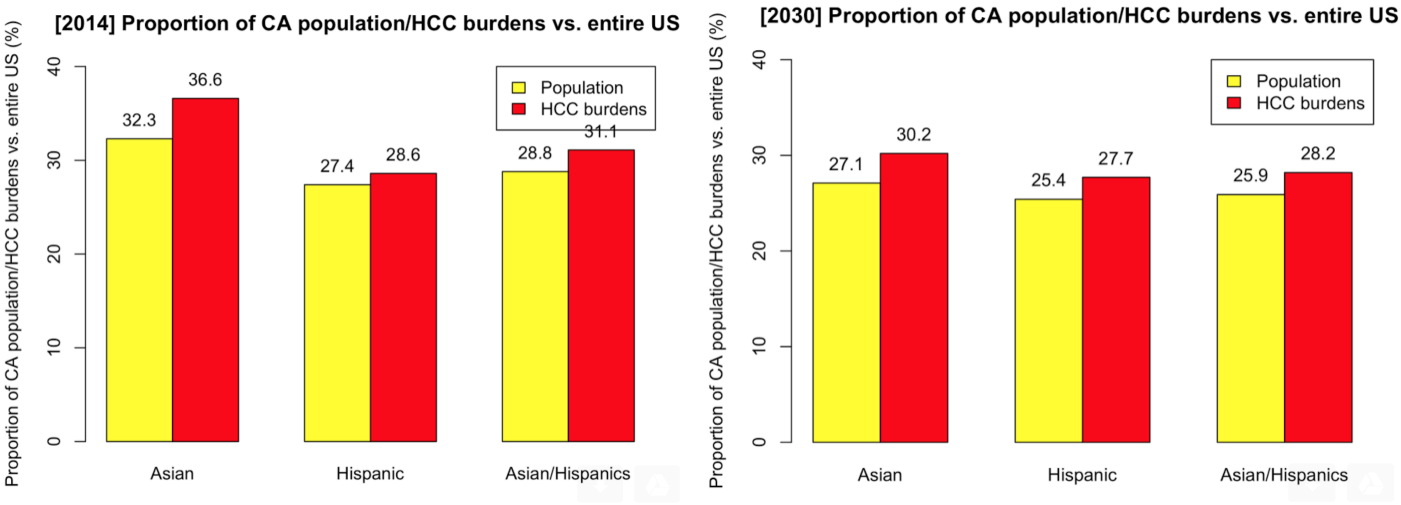
Supplementary Figure 7 Cross-sectional age curve for the past (2000-2013) versus the future (2014-2030) by ethnicity/race in CA (FEMALE)



Supplementary Figure 8 Age-specific annual percent change of the population in CA (dark green) versus non-CA (light green) by race/ethnicity



Supplementary Figure 9 Proportions of HCC cases in CA (red) versus the proportion of individuals who live in CA (yellow) by race/ethnicity. The three groups of bars show the results for Asians, Hispanics, and Asians and Hispanics combined, respectively.



Supplementary Table 1 Predicted versus observed HCC incidence rate and count for 2014 SEER data

2014	Predicted data		Observed data	
	Rate	Count	Rate	Count
All	15.9	3456	15.1	3252
White	11.4	1373	10.4	1248
Black	22.3	292	17.6	240
Asian	21.5	633	21.2	675
Hispanic	22.2	1091	21.3	1042

References

1. Rosenberg PS, Check DP, Anderson WF. A Web Tool for Age–Period–Cohort Analysis of Cancer Incidence and Mortality Rates. Cancer Epidemiology and Prevention Biomarkers 2014;23(11):2296-2302.
2. Holford TR. Understanding the effects of age, period, and cohort on incidence and mortality rates. Annual review of public health 1991;12(1):425-457.
3. Kim HJ, Fay MP, Feuer EJ*, et al.* Permutation tests for joinpoint regression with applications to cancer rates. Statistics in medicine 2000;19(3):335-351.
4. Chernyavskiy P, Little MP, Rosenberg PS. A unified approach for assessing heterogeneity in age–period–cohort model parameters using random effects. Statistical methods in medical research 2017:0962280217713033.
5. Petrick JL, Kelly SP, Altekruse SF*, et al.* Future of hepatocellular carcinoma incidence in the United States forecast through 2030. Journal of Clinical Oncology 2016;34(15):1787-1794.