Supplementary Materials Racial inequality in the annual risk of Tuberculosis infection, 1910-1933.

# Supplementary Data

## Panel Cities:

**North** (N = 11): Baltimore,MD; Brooklyn,NY; Chicago,IL; Cincinnati,OH; Indianapolis,IN; Kansas City,MO; Manhattan,NY; Philadelphia,PA; Pittsburgh,PA; St. Louis,MO; Washington,DC

**South** (N = 5): Atlanta,GA; Birmingham,AL; Louisville,KY; New Orleans,LA; Richmond,VA

# Supplementary Methods

## City-level variation in per-case risk

Because the population of small children in some cities in our panel is relatively small, particularly among African-Americans in the early years of the observation period, we cannot reliably interpolate population sizes for children under 5 between census years. We can, however, investigate the factors associated with variation in per-case risk at the city level during the 1910, 1920 and 1930 census years. We can model the per-case infection risk as a function of city size, including yearly fixed effects by race. We can estimate the factors impacting city-level variation in using a modified Poisson regression model, similar to the time series SIR model commonly applied to measles1. To do this, we estimate a Poisson regression model with the following underlying log-linear model, where is a vector of city-specific covariates and is a vector of city/race regression coefficients:

**Equation S1**

**Equation S1**

The inclusion of the offset term for pulmonary TB prevalence () ensures that the coefficients in reflect the impact of covariates on per-pulmonary-case risk of infection rather than variation in ARTI. We do not include city-level random effects in this model to understand the relationship between city characteristics and variation in ARTI. This decision is justified by the relatively small number of cities in our sample. With more cities, we would be better able to differentiate between city characteristics and the impact of specific characteristics. However, our city-level results should obviously be interpreted in light of this small sample size.

In this analysis, we do not adjust for the differential reporting of TB meningitis mortality for blacks vs. whites and instead use the total count of extra-pulmonary TB cases, since we are primarily interested in factors impacting the relative risk of infection, such as city size, and these relationships would not be impacted by this adjustment.

# Supplementary Results

## Age-adjusted pulmonary TB mortality

We standardized all pulmonary TB mortality rates using the distribution of ages in 1910 for people of all racial groups in the entire panel of cities. We linearly interpolated the population size at each age for the the entire panel of cities and calculated age-group specific pulmonary TB mortality rates for the age groups in the mortality data.

We used individual ages because, across years and cities, the bins used in the original data changed, from 5-year bins in some instances, to 10-year bins in others. For children under 5, mortality for exact ages were reported. Using the interpolated population data, we aggregated the population for the age groups included in each data bin (i.e. 5-9 years, 25-34 years, etc.) to estimate age-specific rates.

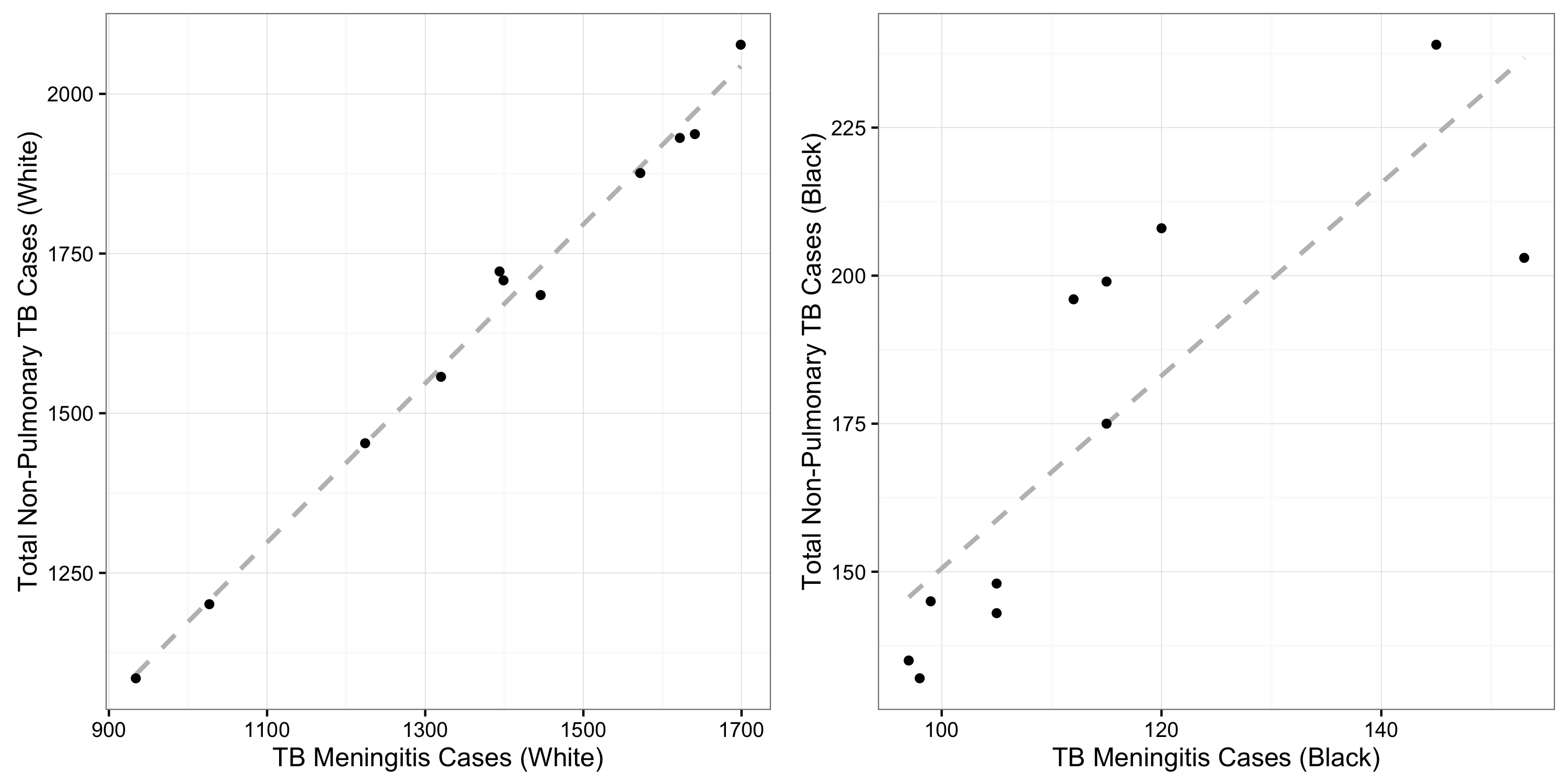
We estimated age-specific rates for each bin in the data using a Poisson model fit via MCMC. On each step of the MCMC estimation we calculated an age-adjusted pulmonary TB mortality rate by taking the weighted sum of these rates using population weights corresponding to the distribution of ages across all races in the 1910 census data for all cities in our data.

## City-level variation in ARTI

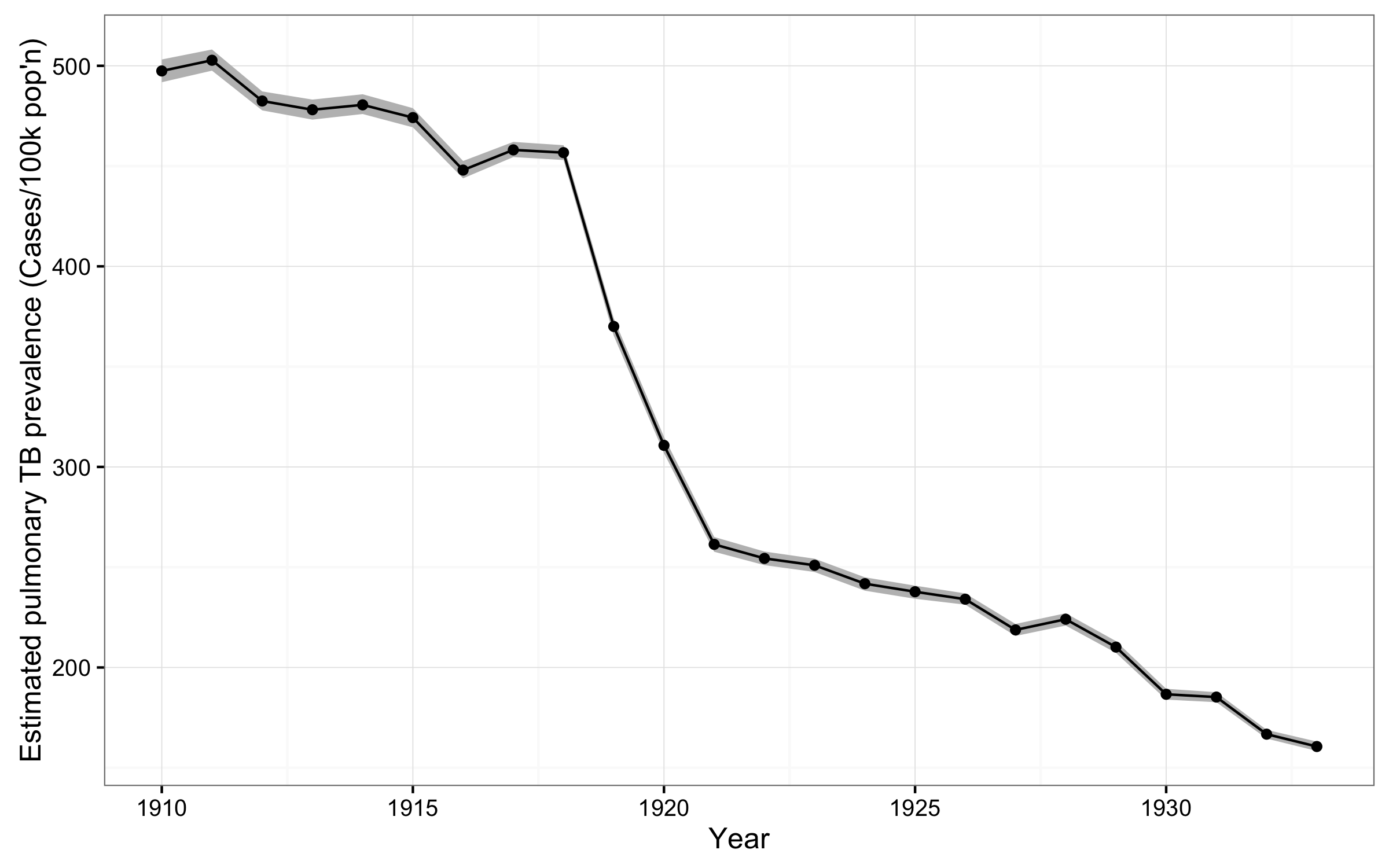
Table 1 shows relative risks (RR) associated with city-level factors during the 1910, 1920 & 1930 census years. The model contains fixed-effects by race for each of the census years, to account for unmodeled change over time by race. It is important to note that because the estimates in this table reflect changes in the per-pulmonary-case risk of extrapulmonary TB in children under 5, the coefficient for race does not reflect the differential reporting of TB meningitis by race reported in the main text.

## TBM cases as a fraction of all extra-pulmonary TB cases: 1910-1920

Figure S1 illustrates the relationship between the count of extrapulmonary TB cases in our data and the number of TB meningitis cases in children under 5 among whites and African-Americans, respectively.



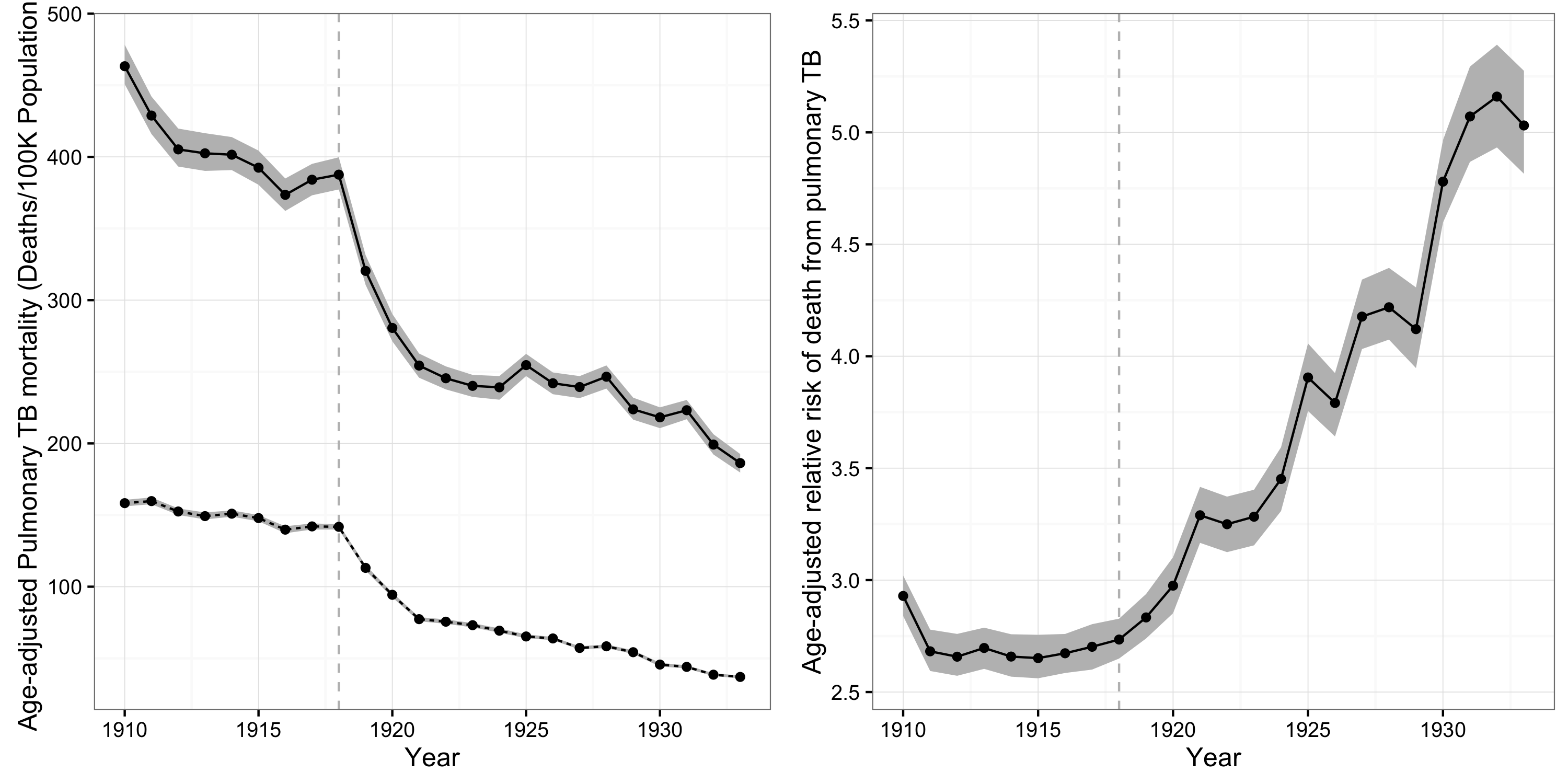
**Figure S1. TBM as a fraction of all non-pulmonary TB cases by race, 1910-1920.** The figure shows the ratio of aggregated non-pulmonary TB deaths to TB meningitis deaths for whites (left-hand panel) and blacks (right-hand panel). Each dot represents the total number of TBM and aggregated non-pulmonary TB deaths for all cities in the panel data.



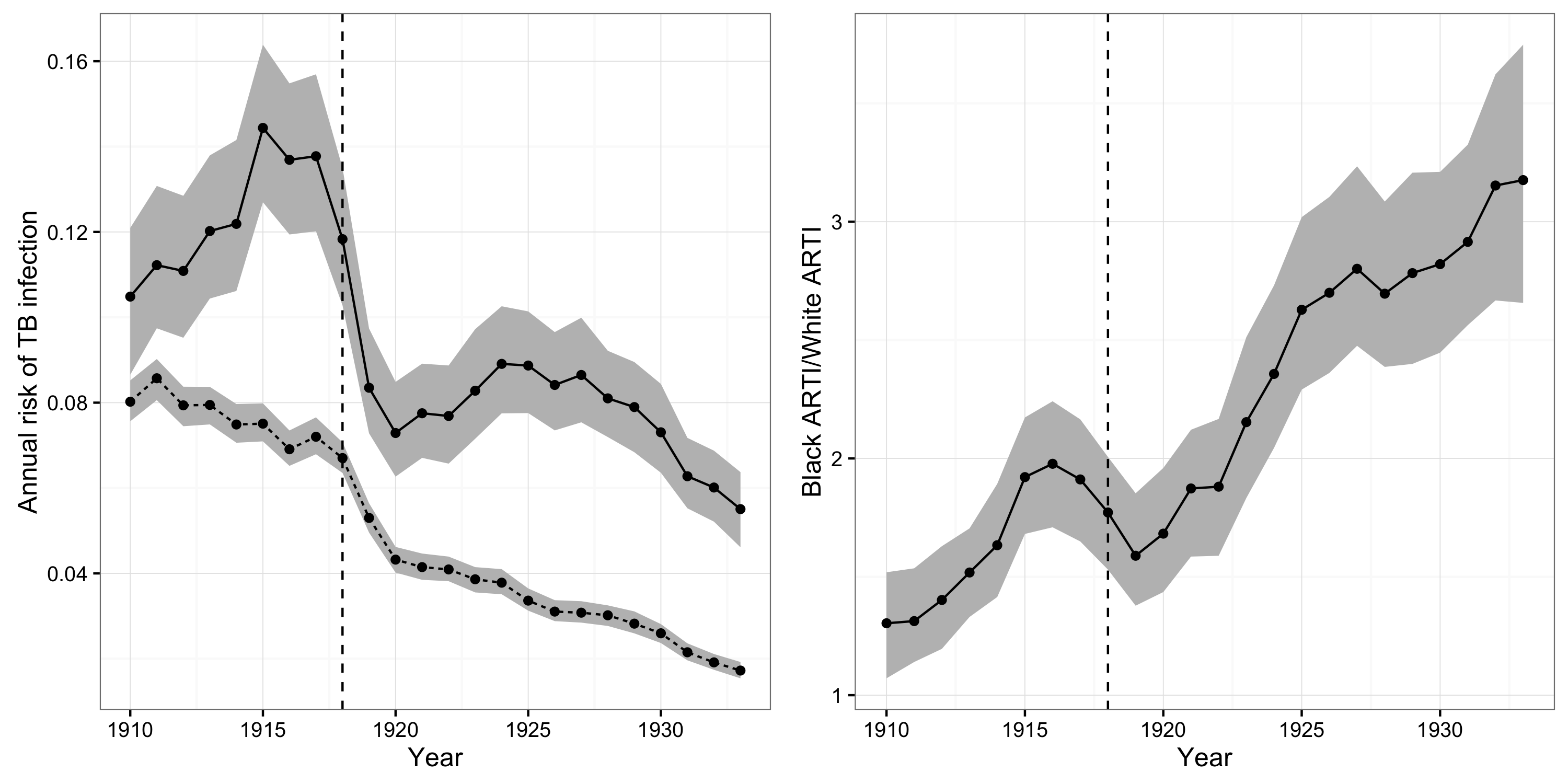
**Figure S2. Crude TB prevalence, 1910-1933.** Estimated population-level prevalence of infectious (i.e. smear-positive) TB estimated using aggregated crude TB mortality data across all Northern panel cities.

## Combined Northern and Southern Results

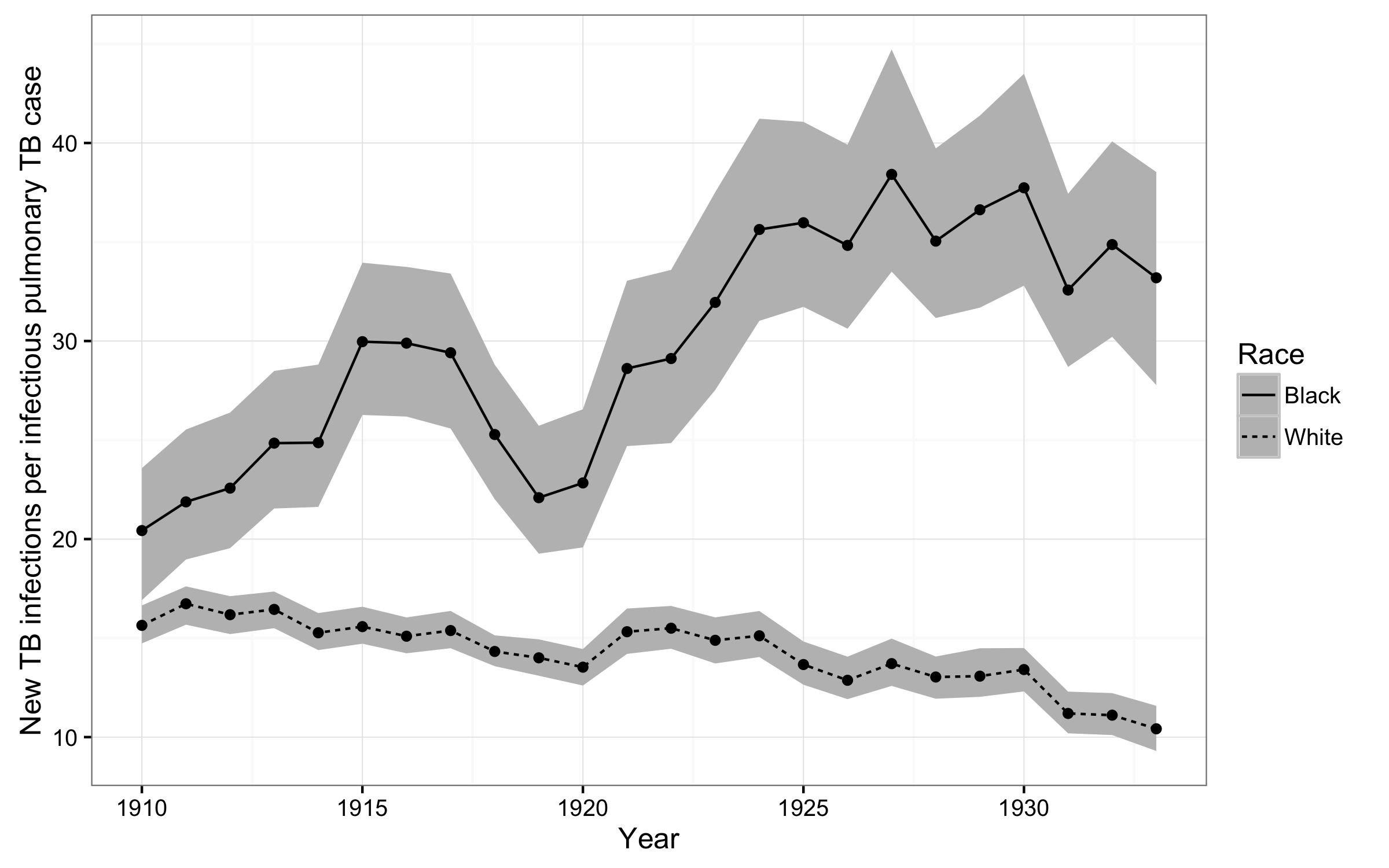
In this section, we present estimates of age-adjusted TB mortality, the annual risk of TB infection by race, and the per-case risk of infection among individuals living in both the northern and southern cities in our dataset. These results are presented to demonstrate that our qualitative findings are not sensistive to the exclusion of these cities.



**Figure S3. Age-standardized TB mortality rates, by race.** The left-hand panel shows age-standardized TB mortality rates in the panel of cities from 1910 to 1933 in deaths per 100K population. Black mortality rates are indicated by the solid line and white TB mortality rates by the dashed line. The right-hand panel shows the adjusted relative risk (ARR) of death from pulmonary TB for African-Americans vs. whites from 1910 to 1933. The vertical dashed line indicates the timing of the 1918 influenza pandemic.



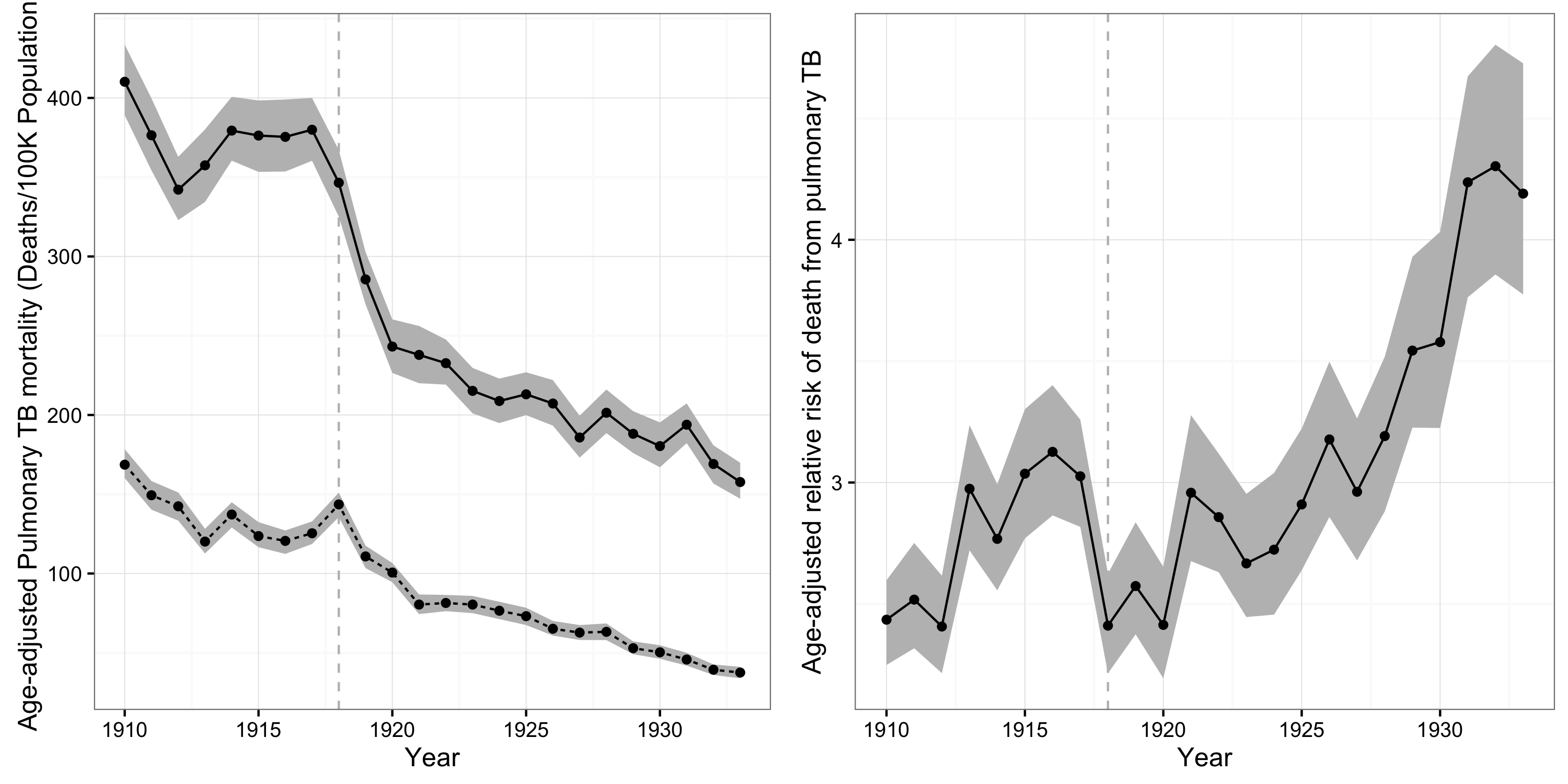
**Figure S4. Annual risk of TB infection (ARTI) by race, 1910-1933.** The left-hand panel illustrates the ARTI for African-Americans (solid line) and whites (dashed line) from 1910-1933. The right-hand panel illustrates the ratio of the ARTI for African-Americans vs. whites during this period. The gray shaded area in both panels illustrates the 95% posterior credible intervals (CIs) for these quantities.



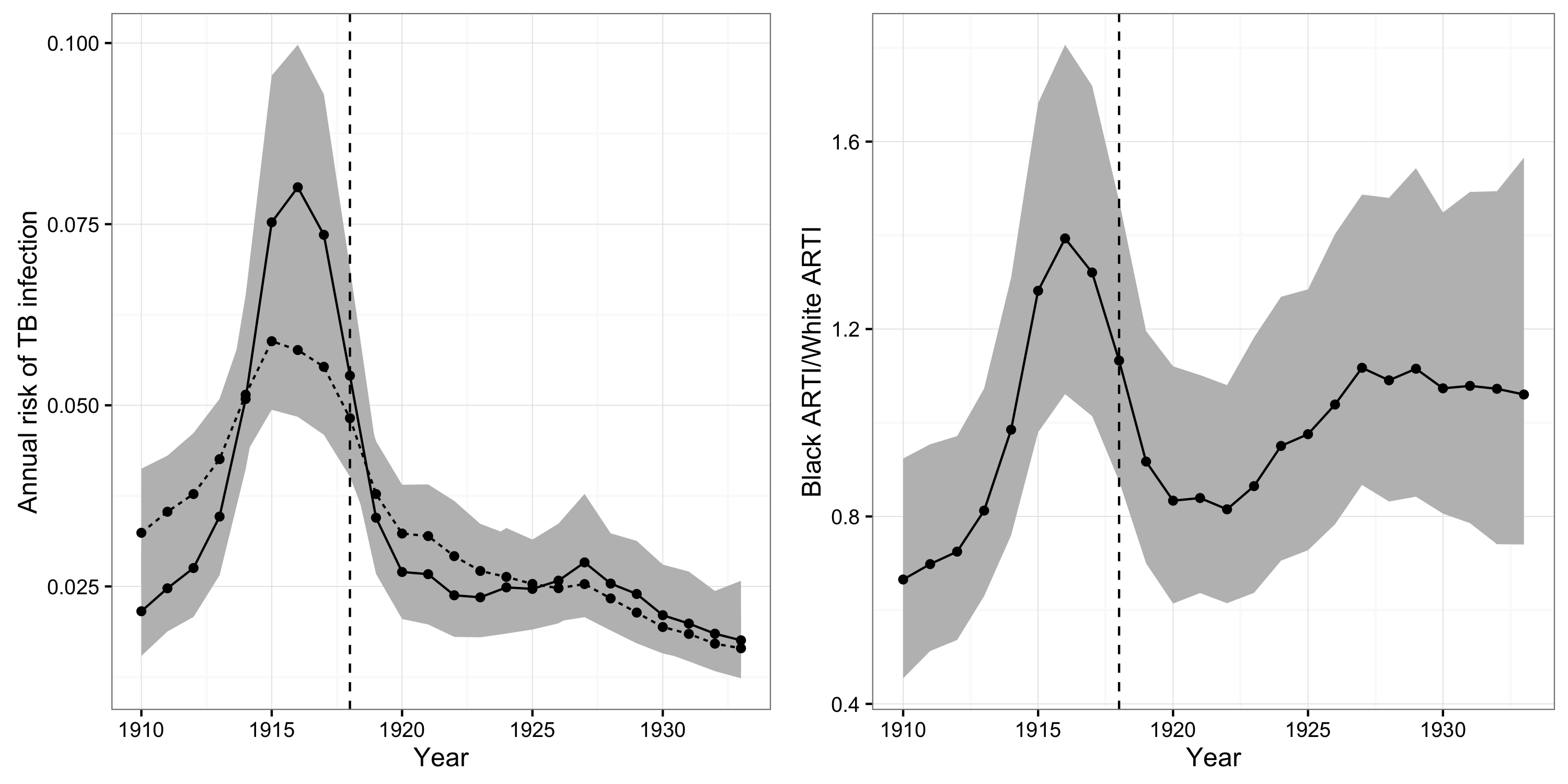
**Figure S5. New TB infections per prevalent pulmonary TB cases, 1910-1933.** The figure illustrates trends in the number of new TB infections among African-Americans (solid line) and whites (dashed line) for every prevalent pulmonary TB case during the period from 1910-1933.

## Southern Results

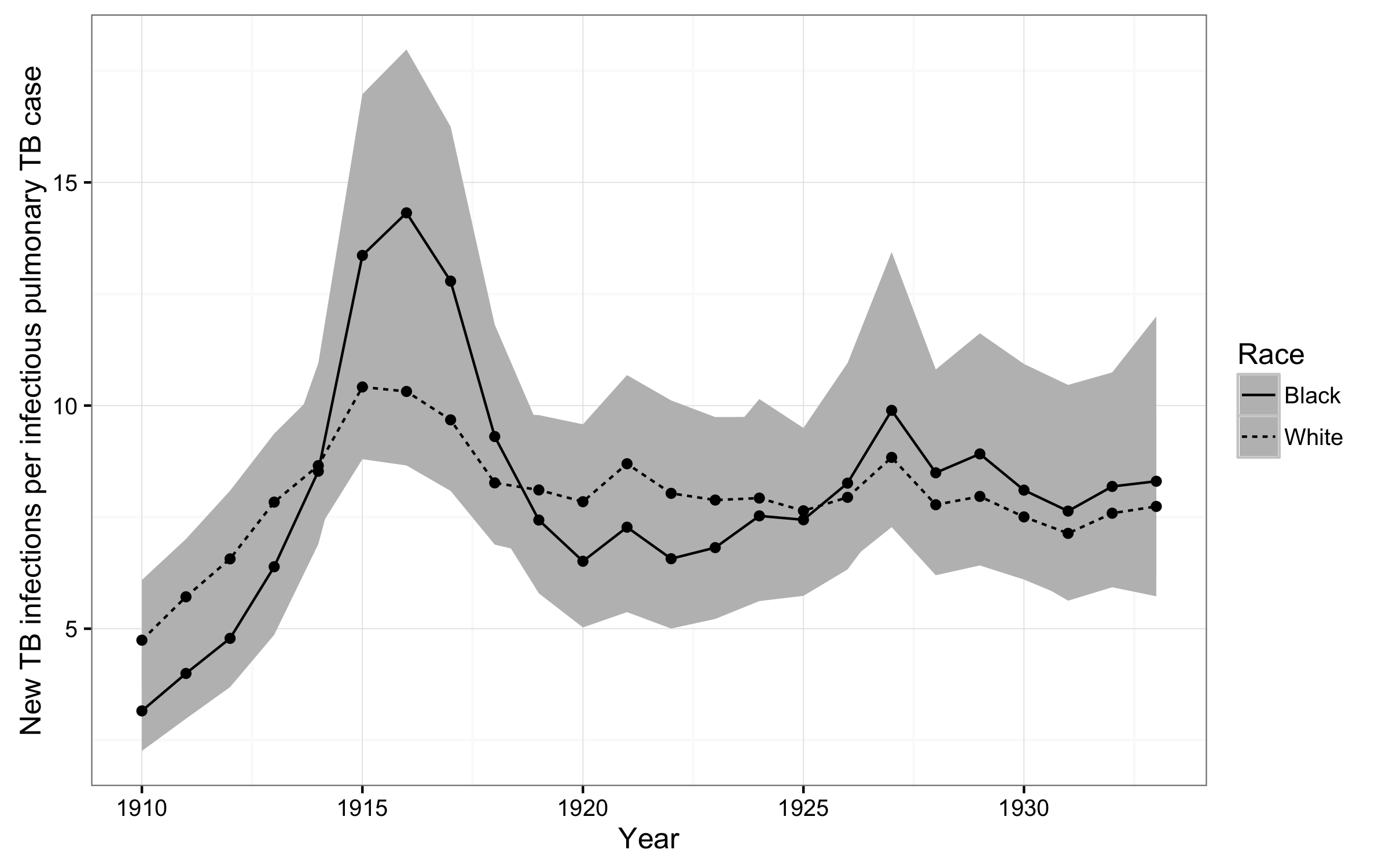
In this section, we present estimates of age-adjusted TB mortality, the annual risk of TB infection by race, and the per-case risk of infection among individuals living in the southern cities in our dataset.



**Figure S6. Age-standardized TB mortality rates, by race.** The left-hand panel shows age-standardized TB mortality rates in the panel of cities from 1910 to 1933 in deaths per 100K population. Black mortality rates are indicated by the solid line and white TB mortality rates by the dashed line. The right-hand panel shows the adjusted relative risk (ARR) of death from pulmonary TB for African-Americans vs. whites from 1910 to 1933. The vertical dashed line indicates the timing of the 1918 influenza pandemic.



**Figure S7. Annual risk of TB infection (ARTI) by race, 1910-1933.** The left-hand panel illustrates the ARTI for African-Americans (solid line) and whites (dashed line) from 1910-1933. The right-hand panel illustrates the ratio of the ARTI for African-Americans vs. whites during this period. The gray shaded area in both panels illustrates the 95% posterior credible intervals (CIs) for these quantities.



**Figure S8. New TB infections per prevalent pulmonary TB cases, 1910-1933.** The figure illustrates trends in the number of new TB infections among African-Americans (solid line) and whites (dashed line) for every prevalent pulmonary TB case during the period from 1910-1933.

# Model code

data {  
  
 int TBM\_N; //Number of observations with disaggregated TBM and non\_pulmonary\_tb in data  
 int<lower=TBM\_N+1> N; //Total number of observations  
 int tbm\_deaths[TBM\_N]; //Count of TBM deaths by year by race  
 int non\_pulmonary\_deaths[N]; //Count of aggregate non-pulmonary deaths  
 int<lower=0,upper=1> black[N];  
 real<lower=0> pop[N];  
 int T; //Number of years in aggregated observations  
 int pulmonary\_deaths[T]; //Aggregated pulmonary deaths  
 real<lower=0> total\_pop[T]; //Total population  
  
 int year[N]; //Year indicator for indexing year terms  
}  
  
  
parameters {  
  
 ## White and black rates of TBM:non-pulmonary-tb  
 vector[2] tbm\_beta;  
  
 ## Non-pulmonary TB mortality rate for whites  
 real np\_tb\_rate\_w\_intercept;  
 vector[T-1] np\_tb\_rate\_w\_incr;  
  
  
 ## Relative risk of non-pulmonary TB for blacks  
 real np\_tb\_rate\_rr\_b\_intercept;  
 vector[T-1] np\_tb\_rate\_rr\_b\_incr;  
  
 vector[T] pulmonary\_rate;  
  
}  
  
transformed parameters {  
 vector[T] black\_ari;  
 vector[T] white\_ari;  
  
 vector[T] black\_beta;  
 vector[T] white\_beta;  
  
 vector[T] np\_tb\_rate\_w;  
 vector[T] np\_tb\_rate\_rr\_b;  
 vector[T] prevalence;  
  
 real white\_beta\_change;  
 real black\_beta\_change;  
  
 ## Add up increments to smooth term for white rate  
 np\_tb\_rate\_w[1] = np\_tb\_rate\_w\_intercept;  
 for (t in 2:T) {  
 np\_tb\_rate\_w[t] = np\_tb\_rate\_w[t-1] + np\_tb\_rate\_w\_incr[t-1];  
 }  
  
 np\_tb\_rate\_rr\_b[1] = np\_tb\_rate\_rr\_b\_intercept;  
 for (t in 2:T) {  
 np\_tb\_rate\_rr\_b[t] = np\_tb\_rate\_rr\_b[t-1] + np\_tb\_rate\_rr\_b\_incr[t-1];  
 }  
  
 for (t in 1:T) {  
 prevalence[t] = 4\*exp(pulmonary\_rate[t]);  
 }  
  
 for (t in 1:T) {  
 white\_ari[t] = 100\*exp(tbm\_beta[1] + np\_tb\_rate\_w[t]);  
 black\_ari[t] = 100\*exp(sum(tbm\_beta) + np\_tb\_rate\_w[t] + np\_tb\_rate\_rr\_b[t]);  
  
 black\_beta[t] = black\_ari[t]/prevalence[t];  
 white\_beta[t] = white\_ari[t]/prevalence[t];  
  
 }  
}  
  
model {  
  
 tbm\_beta ~ normal(0, 1);  
 np\_tb\_rate\_w\_incr ~ normal(0, .1);  
 np\_tb\_rate\_rr\_b\_incr ~ normal(0, .1);  
 pulmonary\_rate ~ normal(0, 1);  
  
 for (i in 1:TBM\_N) {  
 tbm\_deaths[i] ~ poisson\_log(log(non\_pulmonary\_deaths[i]) + tbm\_beta[1] + tbm\_beta[2]\*black[i]);  
 }  
  
 for (i in 1:N) {  
 int t;  
 t = year[i];  
 non\_pulmonary\_deaths[i] ~ poisson\_log(log(pop[i]) + np\_tb\_rate\_w[t] + np\_tb\_rate\_rr\_b[t]\*black[i]);  
 }  
  
 for (i in 1:T) {  
 pulmonary\_deaths[i] ~ poisson\_log(log(total\_pop[i]) + pulmonary\_rate[i]);  
 }  
}

# References

1. Bjørnstad ON, Finkenstädt B, Grenfell BT. Dynamics of measles epidemics: estimating scaling of transmission rates using a time series SIR model. *Ecological Monographs*. 2002;72:169-184.