Supplementary File 5: Methods

***Manipulation of Data***

*Mortality Rate*

Raw data on mortality rates were provided in units of *total deaths by single age, both sexes combined (thousands).* The single age values were combined into 5-year age groups and multiplied by 1000 to obtain number of deaths per year for each age group and year from 2004 – 2021. Each value was then divided by the total population in that age group for that year to obtain a mortality rate with unit *deaths per person per year*. The *approx.* interpolation function was used to estimate a mortality function for each individual age group to give a value for natural deaths at each time step of the model. These were combined into a vector of length 21 to multiply by each compartment in order to model natural death from that compartment.

*Birth Rate*

Raw birth rate data were given in births per 1000 for each year, with data for 2004-2021 and projections for 2022 – 2040. The values were divided by 1000 to obtain a birth rate with unit *births per person per year* which was used in the model. The *approx.* interpolation function was used to estimate a birth rate value for each time step in the model.

*Sexual Contact*

The small diagonal matrix from the work by Chanthavilay et al., 2016 relating to sexual mixing by age was extrapolated to 21 age groups, with 0.6 on the diagonal and 0.4 elsewhere representing the probability of sexual contact between the same or different age groups. This diagonal matrix was then multiplied by the mean number of annual sexual partners by age group, from which an average of male and female had been taken. Finally, this matrix was multiplied by 0.025 in order to produce transmission parameter (beta) values reflecting those of previous models (Poovorawan et al., 2016).

*Prevalence Data*

Data from a study estimating Thai prevalence by age group in 2004 and 2014 based on a collection of serum samples were used to calibrate the model (Wasitthankasem et al., 2016). The sample size for this study was 5964 which was used to calculated the 95% confidence intervals around each data point, using equation (1), where z is the z score (1.96 for 95% CI), p is the population proportion (the prevalence in this instance), n is the sample size (5964) and N is the population size (number of individuals in the given age group and year).

*Population Data*

Population data by age group were provided in units of total individuals by M/F, age group and year from 2004 to 2021. The Male and Female values were added to obtain an absolute population count for each age group in each year. These values were divided by the total population in the same year to calculate the proportion of the total population held in each age group in any given year. The same source also provided a projection from 2022-2040 of the same values. These data points were used to validate the model and ensure that population structure was preserved.

*Initial Conditions*

The population by age group according to United Nations data was preserved in the initial conditions, with the start year of the model simulation being 2004. Prevalence data by age group from the work of Wasitthankasem et al., 2016 were used to estimate the number of infections in each age group in 2004. Infections were assumed to be evenly distributed across the fibrosis and cirrhosis stages according to three larger age groupings: infections in ages 0-30 were all F0 (early stage fibrosis only); infections in ages 30-59 were all F0-F3 (no cirrhosis or HCC), and prevalence in ages 95+ was zero since individuals likely wouldn’t live to that age with untreated infection.

***Modelling Methods***

*ODE Transmission Model*

The system of equations in Supplementary File 2 was implemented in R and solved using the initial conditions given in Supplementary File 1. The Runge-Kutta 4th order method was used with the General Solver for Ordinary Differential Equations in the deSolve package. An internal time step of h=0.1 was used to minimise divergence and allow efficient solution of the system.

*Aging Matrix*

A 21x21 matrix was initialized representing travel through the aging compartments: travel was only possible from the previous age group and to the next one (an individual cannot travel from 0-4 to 10-14, for example) and all rates were 0.2 years-1 (1/5 years), and as such the matrix is mostly 0s except for the diagonal (-0.2) and the value below each diagonal element (0.2).

*Natural Death Matrix and Mortality/Population Scenarios*

The mortality rates as calculated from data were used for the years 2004-2021. The multiplier for each year (mortality in 2005/mortality in 2004 etc.) and the mean was calculated for each age group to obtain the average factor by which each age group’s mortality had increased (or decreased) between 2004 and 2021. These values were multiplied by a factor (0.975) to fit the overall population data, implying that mortality rates will continue to decrease by 2.5% per year compared to the 2004-2021 average. This multiplication factor was then varied for each of the three other population scenarios, as detailed in the *mortality\_scenarios* tab of Supplementary File 1.

*Age Structure*

Each of the 21 compartments representing the various stages of HCV transmission from a previous model by Poovorawan et al., 2016 were stratified into 21 age groups: 0-4, 5-9, …, 95-99, 100+. An individual moved through the age groups at a set aging rate (1/5 years-1) while simultaneously moving through the transmission cycle at rates determined by transition parameters from the previous model. This was implemented by initialising empty vectors of length 21 for each of the disease compartments and then solving the ODEs with each compartment vector. The flow of new births was injected into the first entry of the susceptible compartment, representing newborn babies entirely susceptible to HCV infection. Individuals moved out of the age group due to multiplication of each compartment by the aging matrix, and died of natural causes due to the multiplication of each compartment by the natural death matrix. They left each disease compartment according to compartment transition rates defined in Supplementary File 4.

*Model Output to Correspond to Data*

The model output was adapted to produce prevalence by the same aggregated age groups given in the 2004 and 2014 data, with 10-year age groups up to 49, and all other cases grouped into 50+. Each prevalence value was calculated by summing the infections over all relevant age groups and dividing by the total number of individuals in those age groups in the given year at baseline screening and population.

***Model Calibration and 95% Confidence Intervals***

*Birth Rate*

The data available for births per person per year were given a multiplication factor in order to ensure a good fit to overall population data. This factor was experimentally found to be a normal distribution with mean 1.1 and standard deviation 0.025, and these were used to calculate upper and lower 95% confidence intervals around the birth rates for each year, to demonstrate a degree of uncertainty.

*Baseline Screening*

The exact screening coverage from 2004 – 2021 is unknown, so the value was experimentally found to be normal distribution with mean 7% and standard deviation 3%. This distribution could be modified if more information on screening coverage was known. The 95% confidence intervals relating to the 2014 model output of prevalence were calculated using the uncertainty in the baseline screening coverage.