

Blame

Tim Wilson

May 21, 2024

1 Overview

The blame system attributes intervention effects to specific diseases. It answers questions such as “If smoking drops by 10%, how many fewer deaths due to stroke do we see over the next 20 years?”. Previous methods had to create and run a distinct single-disease models to answer such questions, which quickly becomes too expensive in terms of computer time when repeated for many diseases and interventions.

Blame calculates the answers to such questions cheaply, alongside every model run. This allows the PMSLT to routinely generate disease attribution outputs.

2 Conceptualisation

To understand blame we need to keep a firm grip on exactly what it is trying to do. Most importantly, blame *does not* attribute population-level outcomes to disease within any particular scenario. Blame is *always* calculated with respect to two scenarios, an intervention and BAU, and only attributes the *differences* between two scenarios to changes in disease rates. Blame does not purport to know the burden of a particular disease within a single scenario. This matches the underlying structure of the PMSLT, but this constraint can be particularly hard to track when down in the weeds.

Blame attributes deaths, HALYs, health system costs, and income changes to diseases. HALYs are the easiest to calculate because disability rates shift linearly, and do not cause any changes in population size. Deaths are the second-easiest, but are complicated by the fact that deaths are coded as rates rather than risks, and the fact that changes in death rates change the size of the living population. The former is dealt with via $x \approx 1 - e^x$ for $x \approx 0$ followed by rescaling, while the latter is one of the more complicated parts of blame.

Health system costs and income are handled identically, but are each quite complicated due to costs that change by disease and death status. Diseases can have differing first-year, prevalent-year, and last-year of life costs, and people in the main lifetable who live through the year incur different health system costs to those who died.

The first step in the blame algorithm is to attribute deaths to diseases.

3 Deaths

Fix a particular year-of-birth cohort C . Let $p_{\text{INT},y}$, $p_{\text{BAU},y}$, $m_{\text{INT},y}$, and $m_{\text{BAU},y}$, be the population and deaths for C in year y , for the intervention and BAU respectively. By $p_{\text{INT},y}$ we mean population at the end of year y , where year 1 is the first year, and $m_{\text{INT},y}$ is the deaths during year y . Denote the change in deaths between the scenarios, for year y , by

$$\Delta m_y := m_{\text{INT},y} - m_{\text{BAU},y}.$$

Let D be a set of diseases. We wish to find values $m_{d,y}$, for $d \in D$ and each year y , such that

$$\begin{aligned} \Delta m_y &= \sum_{d \in D} m_{d,y} \\ [\Delta m_y]_{d\text{-only model}} &\approx m_{d,y}. \end{aligned}$$

where $[\dots]_{\text{model}}$ denotes an output corresponding to a different model formulation.

Let $a_{\text{INT},y}$ and $a_{\text{BAU},y}$ be the all-cause mortality rate of C , in the intervention and BAU respectively. The PMSLT evolves as follows, for both scenarios $S \in \{\text{INT}, \text{BAU}\}$,

$$\begin{aligned} m_{S,y} &= p_{S,y-1} (1 - e^{-a_{S,y}}) \\ p_{S,y} &= p_{S,y-1} - m_{S,y}. \end{aligned}$$

The value of $a_{\text{BAU},y}$ is sourced from input files, while

$$a_{\text{INT},y} := a_{\text{BAU},y} + \sum_{d \in D} a_{d,y}$$

where $a_{d,y}$ is the all-cause mortality shift calculated for each disease, the details of which do not concern blame. Then, expanding the above,

$$\begin{aligned} \Delta m_y &= p_{\text{INT},y-1} (1 - e^{-a_{\text{INT},y}}) - p_{\text{BAU},y-1} (1 - e^{-a_{\text{BAU},y}}) \\ &= p_{\text{INT},y-1} (1 - e^{-(a_{\text{BAU},y} + \sum_{d \in D} a_{d,y})}) - p_{\text{BAU},y-1} (1 - e^{-a_{\text{BAU},y}}) \end{aligned}$$

Now consider just the first year of the model, year 1. The previous population is the starting point of the model, which is the same in the intervention and BAU, so $p_{\text{INT},0} = p_{\text{BAU},0}$. Therefore,

$$\begin{aligned} \Delta m_1 &= p_{\text{INT},0} (1 - e^{-a_{\text{INT},1}} - (1 - e^{-a_{\text{BAU},1}})) \\ &= p_{\text{INT},0} (1 - e^{-(a_{\text{BAU},1} + \sum_{d \in D} a_{d,1})} - (1 - e^{-a_{\text{BAU},1}})) \\ &= p_{\text{INT},0} (e^{-a_{\text{BAU},1}} - e^{-(a_{\text{BAU},1} + \sum_{d \in D} a_{d,1})}) \\ &= p_{\text{INT},0} e^{-a_{\text{BAU},1}} (1 - e^{-\sum_{d \in D} a_{d,1}}) \\ &\approx p_{\text{INT},0} e^{-a_{\text{BAU},1}} \sum_{d \in D} (1 - e^{-a_{d,1}}) =: \Delta m'_1 \end{aligned}$$

with the final step depending on $x \approx 1 - e^x$ for $x \approx 0$. This approximation is valid because $a_{d,y}$ is the *difference* in population-level fatality rate of disease d between in the BAU and intervention, which will tend to be tiny, even for significant diseases, because it also depends on the effect size of the intervention.

Now, to calculate $m_{d,1}$ we rescale by $\Delta m'_1$, to prevent small errors in approximation adding up over time,

$$m_{d,1} = p_{\text{INT},0} e^{-a_{\text{BAU},1}} (1 - e^{-a_{d,1}}) \frac{\Delta m_1}{\Delta m'_1}$$

so that $\Delta m_1 = \sum_{d \in D} m_{d,1}$.

3.1 Deaths beyond the first year

The problem with deaths beyond the first year is that, in general, $p_{\text{INT},y} \neq p_{\text{BAU},y}$ for years $y \geq 1$. How do we get around this? Well, the difference between the populations is just the sum of the differences in each year, so

$$p_{\text{INT},y} = p_{\text{BAU},y} + \sum_{i \leq y} \Delta m_i,$$

and we have a breakdown of Δm_i by disease. This means changes to the population size itself can be broken down by disease. Denote the change in population caused by disease d , aggregated up to year y , by

$$\Delta p_{d,y} := \sum_{i \leq y} m_{d,i}$$

and note that

$$\sum_{d \in D} \Delta p_{d,y} = \sum_{d \in D} \sum_{i \leq y} m_{d,i} = \sum_{i \leq y} \Delta m_i = p_{\text{INT},y} - p_{\text{BAU},y}.$$

Now we can revisit the general form of the death blame calculation.

$$\begin{aligned} \Delta m_y &= p_{\text{INT},y-1} (1 - e^{-a_{\text{INT},y}}) - p_{\text{BAU},y-1} (1 - e^{-a_{\text{BAU},y}}) \\ &= p_{\text{INT},y-1} (1 - e^{-a_{\text{INT},y}}) - \left(p_{\text{INT},y-1} - \sum_{d \in D} \Delta p_{d,y-1} \right) (1 - e^{-a_{\text{BAU},y}}) \\ &= p_{\text{INT},y-1} (1 - e^{-a_{\text{INT},y}}) - p_{\text{INT},y-1} (1 - e^{-a_{\text{BAU},y}}) + \sum_{d \in D} \Delta p_{d,y-1} (1 - e^{-a_{\text{BAU},y}}) \end{aligned}$$

We know how to handle the first two terms since they are the same as the Δm_1 case, so

$$\begin{aligned} \Delta m_y &= p_{\text{INT},y-1} e^{-a_{\text{BAU},y}} \left(1 - e^{-\sum_{d \in D} a_{d,y}} \right) + \sum_{d \in D} \Delta p_{d,y-1} (1 - e^{-a_{\text{BAU},y}}) \\ &\approx p_{\text{INT},y-1} e^{-a_{\text{BAU},y}} \sum_{d \in D} (1 - e^{-a_{d,y}}) + \sum_{d \in D} \Delta p_{d,y-1} (1 - e^{-a_{\text{BAU},y}}) \\ &= \sum_{d \in D} (p_{\text{INT},y-1} e^{-a_{\text{BAU},y}} (1 - e^{-a_{d,y}}) + \Delta p_{d,y-1} (1 - e^{-a_{\text{BAU},y}})) =: \Delta m'_y. \end{aligned}$$

This gives us $\Delta m'_y$ expressed as the sum of contributions from each disease. Everything can be calculated out, so there is no need to expand the $e^{-a_{\text{INT},y}}$ term. The first term is the change to deaths caused by the disease in the intervention population, while the second term counts the missing deaths due to the population changing size.

Note that interventions will tend to reduce the direct deaths caused by disease, increasing the population, so $\Delta p_{d,y}$ will generally be positive. This makes the second term positive, or in other words, this means that preventing deaths in past years causes those deaths to be attributed to the disease in future years. This is entirely as expected, and in fact, if you run a full closed cohort simulation, then $m_{d,y}$ hits zero in the final year, because there is no change to disease rates that prevents death in the long run.

In any case, we can now calculate $m_{d,y}$ using the same renormalisation trick as before.

$$m_{d,y} = (p_{\text{INT},y-1} e^{-a_{\text{BAU},y}} (1 - e^{-a_{d,y}}) + \Delta p_{d,y-1} (1 - e^{-a_{\text{BAU},y}})) \frac{\Delta m_y}{\Delta m'_y}$$

3.2 Direct deaths

Perhaps you do not want to report that your intervention has no impact on deaths long term. In that case, you might be interested in ‘direct deaths’. This is a form of output produced by the model, I would not recommend using it, but it is produced, and it is used in later steps. So here is how it is calculated.

Let Δc_y be the difference in ‘direct deaths’ for year y , defined as follows

$$\Delta c_y := p_{\text{INT},y-1} (1 - e^{-a_{\text{INT},y}}) - p_{\text{INT},y-1} (1 - e^{-a_{\text{BAU},y}}),$$

and let us find $c_{d,y}$ such that $\Delta c_y = \sum_{d \in D} c_{d,y}$ and $\Delta c_1 = \Delta m_1$, because direct deaths should agree in the first year.

The difference between Δc_y and Δm_y is that the former assumes that the intervention is only coming into effect at year y , i.e. $p_{\text{INT},y-1} = p_{\text{BAU},y-1}$. Appropriately, Δc_y is exactly the single-year case, which we have already solved, so

$$\Delta c_y \approx p_{\text{INT},y-1} e^{-a_{\text{BAU},y}} \sum_{d \in D} (1 - e^{-a_{d,y}}) =: \Delta c'_y$$

Then with renormalisation

$$c_{d,y} = p_{\text{INT},y-2} e^{-a_{\text{BAU},y}} (1 - e^{-a_{d,y}}) \frac{\Delta c_y}{\Delta c'_y}.$$

4 HALYs

With deaths out of the way, HALYs are simple. Keep our fixed cohort C . Let $b_{\text{INT},y}$, $b_{\text{BAU},y}$, $h_{\text{INT},y}$, and $h_{\text{BAU},y}$ be the disability rate of C and the HALYs they accrue in year y , for the intervention and BAU respectively. Denote the change in HALYs caused by an intervention in year y by

$$\Delta h_y := h_{\text{INT},y} - h_{\text{BAU},y}.$$

As with deaths, we wish to find $h_{d,y}$, for $d \in D$ and each year y , such that

$$\begin{aligned} \Delta h_y &= \sum_{d \in D} h_{d,y} \\ [\Delta h_y]_{d\text{-only model}} &\approx h_{d,y} \end{aligned}$$

HALYs are calculated in the PMSLT, for scenarios $S \in \{\text{INT}, \text{BAU}\}$, as follows,

$$h_{S,y} = \left(p_{S,y} + \frac{m_{S,y}}{2}\right) (1 - b_{S,y}).$$

The $m_{S,y}/2$ term is present because the model assumes that deaths are randomly distributed within a year. Therefore, people who are dead by the end of the year have lived, on average, for half a year, and accrued HALYs for that period.

The final piece of the puzzle is the definitions of $b_{\text{BAU},y}$ and $b_{\text{INT},y}$. The former comes from an input file, while the latter is

$$b_{\text{INT},y} := b_{\text{BAU},y} + \sum_{d \in D} b_{d,y}$$

where $b_{d,y}$ is the disability rate shift caused by disease d , the derivation of which, as with mortality rate, does not concern us. To tidy things up, define the ‘ability rate’ $b'_{S,y} := 1 - b_{S,y}$, so

$$b'_{\text{INT},y} = 1 - b_{\text{BAU},y} - \sum_{d \in D} b_{d,y} = b'_{\text{BAU},y} - \sum_{d \in D} b_{d,y}.$$

This all looks linear and easy, so let’s just crank the handle.

$$\begin{aligned} \Delta h_y &= \left(p_{\text{INT},y} + \frac{m_{\text{INT},y}}{2}\right) b'_{\text{INT},y} - \left(p_{\text{BAU},y} + \frac{m_{\text{BAU},y}}{2}\right) b'_{\text{BAU},y} \\ &= \left(p_{\text{INT},y} + \frac{m_{\text{INT},y}}{2}\right) \left(b'_{\text{BAU},y} - \sum_{d \in D} b_{d,y}\right) \dots \\ &\quad - \left(p_{\text{INT},y} + \frac{m_{\text{INT},y}}{2} - \sum_{d \in D} \left(\Delta p_{d,y} + \frac{m_{d,y}}{2}\right)\right) b'_{\text{BAU},y} \\ &= - \sum_{d \in D} b_{d,y} \left(p_{\text{INT},y} + \frac{m_{\text{INT},y}}{2}\right) + \sum_{d \in D} \left(\Delta p_{d,y} + \frac{m_{d,y}}{2}\right) b'_{\text{BAU},y} \\ &= \sum_{d \in D} \left(\left(\Delta p_{d,y} + \frac{m_{d,y}}{2}\right) b'_{\text{BAU},y} - b_{d,y} \left(p_{\text{INT},y} + \frac{m_{\text{INT},y}}{2}\right)\right) \end{aligned}$$

The first term is the change in HALYs relative to the BAU due to the differing population size and number of deaths. The second term is the change in HALYs in this year of the intervention due to the shift in disability rate. That the change in population size due to a disease ended up in Δh_y is entirely expected, since avoiding death is a good way to generate HALYs. Also, note that Δh_y tends to increase over time, because $(\Delta p_{d,y} + \frac{m_{d,y}}{2})$ tends to be positive, and $b_{d,y}$, as the shift in disability rate caused by a disease, tends to be negative when disease burden is reduced.

In any case, we have a decomposition of Δh_y into contributions from each disease, so

$$h_{d,y} = \left(\Delta p_{d,y} + \frac{m_{d,y}}{2}\right) (1 - b_{\text{BAU},y}) - b_{d,y} \left(p_{\text{INT},y} + \frac{m_{\text{INT},y}}{2}\right).$$

5 Income and expenditure

Population income and health system expenditure are handled identically within the model, so without loss of generality, everything that follows will refer just to expenditure. The expenditure blame system is complicated by the fact that expenditure changes in last year of life, as well as

for someone in their first year, death year, or prevalent year with the disease. Counter-intuitive results can arise, such as increased health system due to decreasing the fatality rate from a disease that is relatively cheap to die from.

Let $t_{\text{INT},y}$ and $t_{\text{BAU},y}$ be the total health system expenditure, in year y , for the intervention and BAU respectively. For $S \in \{\text{INT}, \text{BAU}\}$, the PMSLT calculates $t_{S,y}$ as follows

$$t_{S,y} = p_{S,y}r_{S,y} + m_{S,y}\ell_{S,y}$$

where $r_{S,y}$ is the expenditure rate for people who live through the year, and $\ell_{S,y}$ is the expenditure rate for people who die in the year (last year of life). There is a model specification key, **lifetableDeathAsEvent**, that replaces $p_{S,y}$ with $(p_{S,y} + \frac{m_{S,y}}{2})$, which essentially treats $\ell_{S,y}$ as the event cost of death. This document will stick with end-of-year view presented in the equation above.

The blame system calculates the two components of $t_{S,y}$ separately, then adds them together. So denote the living and dead parts of expenditure to be

$$\begin{aligned}\bar{t}_{S,y} &:= p_{S,y}r_{S,y} \\ \hat{t}_{S,y} &:= m_{S,y}\ell_{S,y}\end{aligned}$$

so that $t_{S,y} = \bar{t}_{S,y} + \hat{t}_{S,y}$.

5.1 First, last, and prevalent year for chronic disease

It is worth taking a look at how diseases generate their expenditure rate shifts, as it has implications for what blame is doing. In particular, we will look at chronic disease, because acute diseases simply read per-person living costs (and death cost) directly from their input files.

A chronic disease is a three state Markov model, with states Susceptible, Current, and Dead. The transitions between the states are incidence, fatality, and remission. Denote these states at the end of year y by s_y , c_y , and x_y , and let i , f , and r be the incidence, fatality, and remission rates during that year. Note that $s_y + c_y + x_y = 1$, by the definition of the Markov model.

For chronic disease d , let $\vec{p}_{d,y}$, $\hat{p}_{d,y}$ and $\bar{p}_{d,y}$ denote the proportion of the population in their first year of the disease, the proportion that died from the disease, and the others with the disease that do not fit the first two categories. These are calculated in the PMSLT, after the Markov model is updated, via the following approximations.

$$\begin{aligned}\vec{p}_{d,y} &:= \frac{s_{y-1}}{s_{y-1} + c_{y-1}}(1 - e^{-i}) \\ \hat{p}_{d,y} &:= \frac{x_y - x_{y-1}}{s_{y-1} + c_{y-1}} \\ \bar{p}_{d,y} &:= \max \left\{ \frac{c_y}{s_y + c_y} - \vec{p}_{d,y} - \hat{p}_{d,y}, 0 \right\}\end{aligned}$$

So far we have defined chronic disease within a single scenario. For blame, we need to consider the difference between the intervention and BAU. Let $\vec{p}_{d,y}^S$, $\hat{p}_{d,y}^S$ and $\bar{p}_{d,y}^S$ be the preceding definitions for scenario $S \in \{I, B\}$ (intervention and BAU). Then the shift in expenditure rate for the living population, due to disease d , is

$$r_{d,y} := \left(\vec{p}_{d,y}^I \bar{E}_d^I + \bar{p}_{d,y}^I \bar{E}_d^I \right) - \left(\vec{p}_{d,y}^B \bar{E}_d^B + \bar{p}_{d,y}^B \bar{E}_d^B \right)$$

where \bar{E}_d^S and \bar{E}_d^I are the prevalent and first year expenditure rates of disease d in scenario S , respectively. Note that $\bar{E}_d^I = \bar{E}_d^B$ and $\bar{E}_d^I = \bar{E}_d^B$ in all models at time of writing, it is just something that could be changed.

5.2 Living expenditure

Calculating the expenditure blame for the population that lived through the year is essentially the same as calculating blame for HALYs. We have

$$\Delta \bar{t}_y := \bar{t}_{\text{INT},y} - \bar{t}_{\text{BAU},y}$$

and want to find $\bar{t}_{y,d}$ such that

$$\begin{aligned} \Delta \bar{t}_y &= \sum_{d \in D} \bar{t}_{y,d} \\ [\Delta \bar{t}_y]_{d\text{-only model}} &\approx \bar{t}_{y,d}. \end{aligned}$$

Recall that $\bar{t}_{S,y} = p_{S,y} r_{S,y}$. The PMSLT calculates rates in the intervention by

$$r_{\text{INT},y} := r_{\text{BAU},y} + \sum_{d \in D} r_{d,y}.$$

Now we have enough to crank the handle.

$$\begin{aligned} \Delta \bar{t}_y &= p_{\text{INT},y} r_{\text{INT},y} - p_{\text{BAU},y} r_{\text{BAU},y} \\ &= p_{\text{INT},y} \left(r_{\text{BAU},y} + \sum_{d \in D} r_{d,y} \right) - \left(p_{\text{INT},y} - \sum_{d \in D} \Delta p_{d,y} \right) r_{\text{BAU},y} \\ &= p_{\text{INT},y} r_{\text{BAU},y} + p_{\text{INT},y} \sum_{d \in D} r_{d,y} - p_{\text{INT},y} r_{\text{BAU},y} + r_{\text{BAU},y} \sum_{d \in D} \Delta p_{d,y} \\ &= \sum_{d \in D} (p_{\text{INT},y} r_{d,y} + r_{\text{BAU},y} \Delta p_{d,y}) \end{aligned}$$

So $\bar{t}_{y,d} = p_{\text{INT},y} r_{d,y} + r_{\text{BAU},y} \Delta p_{d,y}$.

5.3 Death expenditure

This is where it gets tricky. Rather than calculate a general rate of expenditure for people that lived through the year, we are trying to calculate the rate for people who died. This implies a few properties that our solution should satisfy.

- If a disease causes no changes in mortality rate during a year, then it should also cause no change to the average cost of dying in that year.
- A disease that is relatively cheap to die from will *increase* the average cost of dying, when we *decrease* mortality from that disease.

This is weird when compared to living expenditure, as reducing the prevalence of any disease with positive health expenditure costs will always reduce health expenditure costs, from a single-year

perspective. To deal with this, the PMSLT uses the blame system to calculate death expenditure, then uses the result to derive a rate.

The first step is to attribute direct deaths, $c_{d,y}$, to each disease d , as per Section 3.2. Recall that direct deaths are the deaths caused by the intervention on each disease, but with the assumption that it is always the first year of the intervention. Changes in death rate due to any difference between the BAU and intervention populations are ignored.

With direct deaths, we can calculate the difference in direct death expenditure, $\Delta \hat{t}_y$, as follows

$$\Delta \hat{t}_y = \sum_{d \in D} c_{d,y} \hat{E}_d$$

where \hat{E}_d is the expenditure rate per death by disease d . This yields the change in deaths, but there are many deaths in the intervention that are not directly attributable to a change in disease rates. For these deaths, of which there are $(m_{\text{INT},y} - \Delta c_y)$, we use the population-level death rate, \hat{E}_{INT} . So then the total death expenditure of the intervention is

$$\hat{t}_{\text{INT},y} := (m_{\text{INT},y} - \Delta c_y) \hat{E}_{\text{INT}} + \Delta \hat{t}_y.$$

Now to calculate the new death expenditure rate, $\ell_{\text{INT},y}$, we just divide by the number of deaths

$$\ell_{\text{INT},y} = \frac{\hat{t}_{\text{INT},y}}{m_{\text{INT},y}}.$$