Blame

Tim Wilson

May 16, 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?". The previous method calculates the answer by running an entire stroke-only model. This quickly becomes expensive when applied generally, as attribution for 20 diseases requires 20 times the computer time.

Blame calculate the answers to such questions cheaply, within calculations running alongside every model run. This allows for routine disease attribution outputs, rather than limiting disease attribution to special cases.

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 difference in outcomes between those scenarios to diseases. At no point that blame 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 $x \approx 0$ and rescaling, 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. Population is population at the end of the year, and deaths is deaths during the year. Denote the difference in deaths in one year 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

$$\Delta m_y = \sum_{d \in D} m_{d,y}$$

$$[\Delta m_y]_{d\text{-only model}} \approx m_{d,y}$$

where $[\ldots]_{\text{model}}$ denotes a value corresponding to a differently formulated model.

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

$$m_{S,y} = p_{S,y-1}\left(1 - e^{-a_{S,y}}\right)$$

$$p_{S,y} = p_{S,y-1} - m_{S,y}$$

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,

$$\Delta m_y = p_{\text{INT},y-1}\left(1 - e^{-a_{\text{INT},y}}\right) - p_{\text{BAU},y-1}\left(1 - e^{-a_{\text{BAU},y}}\right)$$

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

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,

$$\Delta m_1 = p_{\text{INT},0}\left(1 - e^{-a_{\text{INT},1}} - \left(1 - e^{-a_{\text{BAU},1}}\right)\right)$$

$$= p_{\text{INT},0}\left(1 - e^{-(a_{\text{BAU},1} + \sum_{d \in D} a_{d,1})} - \left(1 - e^{-a_{\text{BAU},1}}\right)\right)$$

$$= p_{\text{INT},0}\left(e^{-a_{\text{BAU},1}} - e^{-(a_{\text{BAU},1} + \sum_{d \in D} a_{d,1})}\right)$$

$$= p_{\text{INT},0}e^{-a_{\text{BAU},1}}\left(1 - e^{-\sum_{d \in D} a_{d,1}}\right)$$

$$\approx p_{\text{INT},0}e^{-a_{\text{BAU},1}} \sum_{d \in D}\left(1 - e^{-a_{d,1}}\right) =: \Delta m_1'$$

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. Performing the blame calculation on the deltas allows for such approximations.

Now, to calculate $m_{d,1}$ we rescale by the approximation $\Delta m_1'$ as follows,

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

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 can attribute changes in population size to each diseases. Denote the change in population caused by disease d by 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.

$$\Delta m_y = p_{\text{INT},y-1}\left(1 - e^{-a_{\text{INT},y}}\right) - p_{\text{BAU},y-1}\left(1 - e^{-a_{\text{BAU},y}}\right)$$

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

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

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

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

$$\approx p_{\text{INT},y-1} e^{-a_{\text{BAU},y}} \sum_{d \in D}\left(1 - e^{-a_{d,y}}\right) + \sum_{d \in D} \Delta p_{d,y}\left(1 - e^{-a_{\text{BAU},y}}\right)$$

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

TODO CHECK THE THREE LINES ABOVE

Stopping here is fine (is it?), we have $\Delta m_y'$ expressed as a sum of contributions from diseases. Everything here can be calculated, so we do not need to expand the $e^{-a_{\text{INT},y}}$. The left term captures the change in the number of deaths due to the change in the size of the population. Note that $\Delta p_{d,y}$ will often be negative, representing an increase in population in the intervention, so more deaths will be attributed to diseases that prevented deaths in the past.