Number of years of life lost and multistate models for cancer patients

Antoine Soetewey Catherine Legrand Michel Denuit

antoine.soetewey@uclouvain.be catherine.legrand@uclouvain.be michel.denuit@uclouvain.be

Background & contributions

Over the last decade, the **number of years of life lost** (YLL) became a popular tool in biostatistics and epidemiology to measure differences in life expectancy or mortality, primarily thanks to its ease of interpretation and because information on the cause of death is not required.

On the other hand, **multistate models** (MSMs) are a powerful statistical approach to study the evolution of individuals between different "states", providing a convenient representation for patients diagnosed with cancer.

YLL is usually estimated via the survival curves. In this research, we propose a **new way to estimate** the number of years of life lost—via a 3-state (healthy–cancer–death) multistate model. Usefulness of this approach in the context of the right to be forgotten will also be highlighted.

Data

Data from the **Belgian Cancer Registry**

140,241 cancer patients between 20 and 69 years old:

- Childhood cancers are often considered as different "types" of cancer compared to adults due to their difference in outcome and prognosis
- Non-cancer related deaths increase substantially at the age of 70 and it is not always easy to distinguish the cause of death

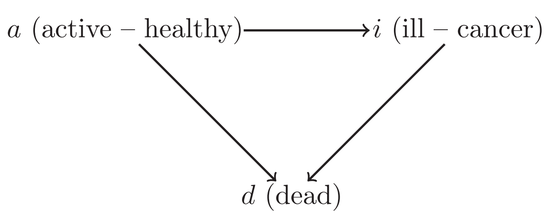
Focus on **breast** (women only), **melanoma** and **thyroid** cancers because:

- High number of incidences, with a significant share occurring before the age of 40 (relevant in the context of the right to be forgotten)
- Cancers with a relatively high survival or cured rate
- "Well-known" to the public

Methods

Multistate models

Multistate models are a powerful statistical approach to **study the evolution of individuals between different "states"** and can be seen as an extension of classical survival analysis. We consider a 3-state MSM, assuming that an individual can either be "healthy", "ill" (diagnosed with cancer), or "dead":



While a MSM represents how an individual moves between a series of states in continuous time, it also provides a convenient representation for patients with diagnosed with cancer.

In this context, it is assumed that only the current state and the sojourn time (i.e., time spent in the current state) influence future transitions. This Semi-Markov assumption ensures that the policyholder's experience is entirely described by the transition probabilities between the three states and the sojourn probabilities.

Similar to the usual force of mortality which is at the heart of life insurance mathematics, transition intensities (derived from transition probabilities) quantify the instantaneous risk of transitioning between two given states in the more general context of multistate models. As mortality depends on attained age, these transition intensities depends on attained age too. However, given that there is also an effect of the duration of stay in the cancer state, transition intensities from this state depend on both attained age (age at entry) and occupation time in the cancer state.

Years of life lost

Over the last decade, the number of years of life lost became a popular tool in biostatistics and epidemiology to measure differences in life expectancy or mortality. The idea behind YLL is to quantify the number of years of life a specific cohort of patients has lost due, for example, to a given disease, compared to the general population.

This measure has the advantage that:

- it is measured on a time metric (usually in years) making its interpretation easy for patients and policy-makers (more meaningful for gauging public health outcomes)
- information on the cause of death is not needed, making it a practical measure for population-based studies in which the cause of death is often unavailable or unreliable
- it can be calculated up to a finite time horizon, in connection with the end of a financial contract or the end of the waiting period defined by the right to be forgotten

YLL is often estimated as the difference between the area under the survival curve of the general population (or some benchmark cohort) and the area under the survival curve of the cohort of interest:

$$YLL_C(\tau) = \int_0^\tau S_P(t)dt - \int_0^\tau S_C(t)dt,$$

where $S_P(\cdot)$ denotes the classical survival function estimated via the population mortality rates, and $S_C(\cdot)$ is the cancer survival curve (in general, estimated via the non-parametric Kaplan-Meier (1958) method but it could be estimated via another method as well).

Methods

Note that if there were no upper time limit, we could work on the whole lifespan and simply compute survival curve of the general population and the one of the cohort (looking at the time from birth to death) to see globally how many years these people would be loosing. Nonetheless, here we look from the time of diagnosis until a certain time τ after diagnosis. Also note that we account for age at diagnosis through the life tables which are stratified by age.

Integrating from 0 to τ, it follows that the number of years of life lost due to cancer before time τ can be written as sums

$$YLL_C(\tau) = \sum_{i=0}^{n-1} S_P(t_i)(t_{i+1} - t_i) - \sum_{i=0}^{n-1} S_C(t_i)(t_{i+1} - t_i).$$

One contribution of the paper is that we can estimate these quantities from our MSM. The idea here is to start from our MSM to compute YLL:

Someone with cancer:

Same if she would have no cancer:

We have thus

$$YLL = \bar{e}_x^{aa} - t$$

with t the number of years spent in state i since diagnosis. But t is a random variable (whose value depends on the time of transition from i to d). To get the expected value of number of years of life lost, we average on this transition time t (where t = time to death for ill patients)

$$YLL_x = \int_0^{\text{upper}} {}_tp_{x;0}^{ii} \mu_{x+t;t}^{id} (\bar{e}_x^{aa} - t)dt$$

where ${}_tp_{x;0}^{ii}$ is the probability for an individual to stay in state i until time t (i.e., the sojourn probability in state i) and $\mu_{x+t;t}^{id}$ is the transition probability from state i to d (i.e., the death at time t of a ill person). Another way to see it is

$$\begin{aligned} YLL &= \text{expected lifetime at age } x \ (\bar{e}_x^{aa}) \text{ if no transition to } i \\ &\quad - \text{expected time spent in } i \text{ if diagnosed at age } x \ (\bar{e}_x^{ii}) \\ &= \bar{e}_x^{aa} - \bar{e}_{x;0}^{ii} \end{aligned} \tag{1}$$

with

$$\bar{e}_x^{aa} = \int_0^{\text{upper}} {}_tp_x^{aa} dt \text{ and}$$
$$\bar{e}_{x;0}^{ii} = \int_0^{\text{upper}} {}_tp_{x;0}^{ii} dt$$

where $p_x^{aa} = P(X_t = a | X_x = a)$ and $p_{x;0}^{ii} = P(X_t = i | X_x = i, Z_t = 0)$. If the upper limit corresponds to some maximum age $(w - x)$ then we have the equivalence between the two expressions with $\bar{e}_x^{aa}$ and $\bar{e}_{x;0}^{ii}$ as defined above. But if the upper limit corresponds to τ, as in our case, then $\bar{e}_x^{aa}$ and $\bar{e}_{x;0}^{ii}$ need to be adapted to have the equivalence between the two expressions.

From Eq. (1):

- Expected lifetime at age x $(\bar{e}_x^{aa})$ if no transition to i can easily be estimated via life tables
- Expected time spent in i if diagnosed at age x $(\bar{e}_x^{ii})$ could be computed via sojourn times using a MSM

$\Rightarrow$ YLL could thus be computed and then compared to the estimation done via survival curves

Further work

- Adapt actuarial notations to a statistical audience
- No information about healthy citizens is available in BCR data
  - $\Rightarrow$ at this stage, impossible to estimate transition rates from (i) healthy to ill and from (ii) healthy to dead
  - $\Rightarrow$ find an optimal way to populate BCR data with information about healthy citizens (e.g., data from life tables, Statbel, etc.)

Acknowledgements

We gratefully acknowledge funding from UCLouvain and the FWO and F.R.S.-FNRS under the Excellence of Science (EOS) programme (project EOS 40007517), and the Belgian Cancer Registry for providing access to the data and for research assistance.