

Calculating lifetime cancer risk in a population

Karl Pettersson

Introduction

It is common to hear statements such as

One in three One in three persons will develop cancer during their lifetime.

One in nine One in nine women will develop breast cancer.

Most often, such statements are based on a simple calculation of cumulative risk, i.e. age-specific incidence rates for a given year and cancer diagnosis are summed up to a chosen maximum age, e.g. 75 years, and the resulting cumulative incidence rate r is then converted into a probability using the formula $1 - \exp(-r)$. However, if “lifetime cancer risk” is interpreted as the proportion of the population which will be diagnosed with cancer during their lifetime, this method gives incorrect results, because it does not take the following into account:

1. Future changes in cancer rates.
2. People who die before they reach the maximum age, due to causes unrelated to cancer.
3. People who develop cancer at ages above the maximum age.
4. People who are diagnosed with multiple primary cancers during their lifetime.

The AMP method

The first problem will not be further discussed in this post, as dealing with it obviously would require projections into the future. The other problems can be assessed with a method described by Sasieni et al. (2011), which they call AMP (“adjusted for multiple primaries”), and which only requires routinely available data. Their idea is to build a life table where it is possible to be eliminated from the population either by being diagnosed with cancer or by dying from something other than cancer. It is then possible to calculate the proportions eliminated in these different ways. The AMP method hinges on the independence assumption that primary cancer incidence and mortality from causes other than cancer are the same among people who have had cancer as in

the general population, because these groups cannot normally be differentiated in official statistics. Only the following data are required:

1. Age-specific population size, in order to calculate incidence and mortality rates.
2. Age-specific number of cancer cases.
3. Age-specific number of deaths due to all causes.
4. Age-specific number of deaths due cancer. Note that official statistics normally reports so-called underlying causes of deaths, which means that this should include complications of cancer and cancer treatment (otherwise, the independence assumption given above would be violated).

Using my LifeTable package, <https://github.com/klpn/LifeTable.jl>, the AMP method can be easily implemented in Julia. I will give examples with calculations for Sweden 2014, using data from Statistics Sweden (2016) for population size, National Board of Health and Welfare (2019a) for cancer cases and National Board of Health and Welfare (2019b) for deaths. The data are given in 5-year age intervals from 0–4 to 80–84 years, with an open interval for ages above 85 years. The files used in the example are available via a gist, <https://gist.github.com/klpn/3ab1feff67d6e938fecc61c8307ce394>. The Julia file, <https://gist.github.com/klpn/3ab1feff67d6e938fecc61c8307ce394#file-amplt-jl> contains the following code:

```
1 using LifeTable, DataFrames
2 function AmpLt(inframe, sex, rate = "inc")
3     age = inframe[1]
4     pop = inframe[2]
5     acd = inframe[3]
6     cd = inframe[4]
7     cc = inframe[5]
8     if rate == "inc"
9         ncol = cc
10        dcol = acd .- cd .+ cc
11    elseif rate == "mort"
12        ncol = cd
13        dcol = acd
14    end
15    df = DataFrame(age = age, pop = pop, dcol = dcol)
16    cprop = ncol ./ dcol
17    lt = PeriodLifeTable(df, sex)
18    return CauseLife(lt, cprop)
19 end
```

Assuming the LifeTable package is installed and the files have been downloaded, you can calculate tables with lifetime cancer risk for Swedish females and males, at a given age:

```

1 include("amplt.jl")
2 fse14 = readtable("fse14.csv")
3 mse14 = readtable("mse14.csv")
4 ampfse14 = AmpLt(fse14, 2)
5 ampmse14 = AmpLt(mse14, 1)

```

The first row in the `f` column in a frame returned by `AmpLt` gives the lifetime cancer risk at birth, which should be about 45.7 percent for females and 49.3 percent for males. It is also possible to calculate lifetime risk for cancer mortality, rather than incidence:

```

1 mampfse14 = AmpLt(fse14, 2, "mort")
2 mampmse14 = AmpLt(mse14, 1, "mort")

```

The first row in these frames should be about 22.3 and 26.2 percent for females and males. With `PyPlot`, the frames can be plotted:

```

1 plot(ampfse14[:age], ampfse14[:f], label = "incidence, females")
2 plot(ampfse14[:age], ampmse14[:f], label = "incidence, males")
3 plot(mampfse14[:age], mampmse14[:f], label = "mortality, females")
4 plot(mampfse14[:age], mampmse14[:f], label = "mortality, males")
5 title("Lifetime cancer risk Sweden 2014")
6 xlim(0, 85)
7 ylim(0, 0.5)
8 legend(loc=3)
9 grid(1)

```

As the chart shows, the probabilities tend to decrease with age, especially after age 60, which is due to increasing competition from other causes of death, e.g. circulatory disorders.

Table 1: Lifetime probability of cancer incidence and mortality for Swedish females and males 2014.

Measure	Female	Male
Incidence	45.7	49.3
Mortality	22.3	26.2

If cancer incidence and mortality are changed, this might also influence mortality from some non-cancer causes. For example, decreased smoking tends to decrease lung cancer incidence and mortality, as well as mortality from nonmalignant respiratory diseases and atherosclerotic diseases.¹ One might ask how such risk

¹This shared risk factor can be expected to violate the independence assumption in the AMP method to some extent. However, as noted by Sasieni et al. (2011), these effects should not be serious when all cancers are studied, because there are few lung cancer survivors in the population.

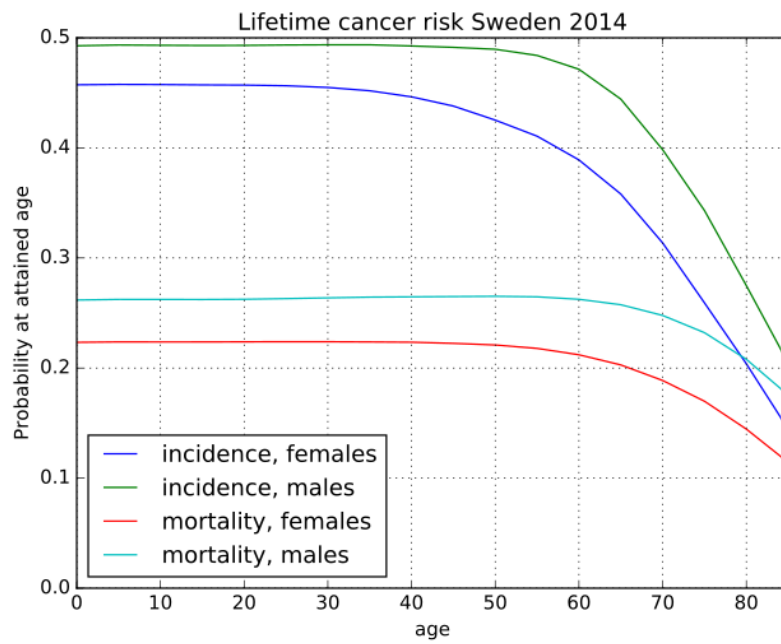


Figure 1: Lifetime probability of cancer incidence and mortality for Swedish females and males 2014

factor changes would influence lifetime cancer risk, which might be decreased, as well as unchanged, or even increased, due to diminished competition. The following function recalculate a frame with cancer cases, as well as cancer deaths and non-cancer deaths, changed by the same factor for all age groups:

```
1 function RateChange(inframe, changefac)
2     ncd = inframe[4] .* changefac
3     ncc = inframe[5] .* changefac
4     nacd = (inframe[3].-inframe[4]).*changefac .+ ncd
5     return DataFrame(age = inframe[1], pop = inframe[2],
6         acd = nacd, cd = ncd, cc = ncc)
7 end
```

To calculate lifetime cancer risk for Swedish females with all three rates reduced by one third, give `AmpLt(RateChange(fse14, 2/3), 2)`. The frame returned by this call gives a lifetime risk at birth of 37.5 percent. For males, the corresponding risk is 42.6 percent. For mortality, the risks would be 19.2 and 24.0 percent for females and males respectively. If age-specific cancer incidence and mortality and non-cancer mortality are reduced by the same factor, in a society such as Sweden, which already has high life expectancy, this tends to reduce the lifetime risk of getting cancer or dying from it, because more people will survive to higher ages where the probability of getting cancer before succumbing to something else is lower (with greater reductions, the risks at lower ages asymptotically approach the corresponding risks at the highest age, i.e. 85 years in this example).

References

- National Board of Health and Welfare. 2019a. “Cancer.” 2019. https://sdb.socialstyrelsen.se/if_can/val_eng.aspx.
- . 2019b. “Cause of Death.” 2019. https://sdb.socialstyrelsen.se/if_dor/val_eng.aspx.
- Sasieni, P. D., J. Shelton, N. Ormiston-Smith, C. S. Thomson, and P. B. Silcocks. 2011. “What Is the Lifetime Risk of Developing Cancer?: The Effect of Adjusting for Multiple Primaries.” *British Journal of Cancer* 105. <https://doi.org/10.1038/bjc.2011.250>.
- Statistics Sweden. 2016. “Mean Population by Region, Marital Status, Age and Sex.” 2016. <http://www.statistikdatabasen.scb.se/goto/en/ssd/MedelfolkHandelse>.