



Max-Planck-Institut für demografische Forschung  
**Max Planck Institute for Demographic Research**

---

Konrad-Zuse-Strasse 1 • D-18057 Rostock • Germany • Tel +49 (0) 3 81 20 81 - 0 • Fax +49 (0) 3 81 20 81 - 202 • [www.demogr.mpg.de](http://www.demogr.mpg.de)

MPIDR Technical Report TR 2018-004 | March 2018

## **Two demographic models for estimation of population-level burden of cancer: an Excel spreadsheet**

**E. M. Andreev**

**V. Shkolnikov** | [shkolnikov@demogr.mpg.de](mailto:shkolnikov@demogr.mpg.de)

**D. Jasilionis** | [jasilionis@demogr.mpg.de](mailto:jasilionis@demogr.mpg.de)

For additional material see [www.demogr.mpg.de/tr/](http://www.demogr.mpg.de/tr/)

This technical report has been approved for release by: Vladimir Shkolnikov ([shkolnikov@demogr.mpg.de](mailto:shkolnikov@demogr.mpg.de)),  
Editor of the MPIDR Technical Reports.

© **Copyright is held by the authors.**

Technical reports of the Max Planck Institute for Demographic Research receive only limited review. Views or opinions expressed in technical reports are attributable to the authors and do not necessarily reflect those of the Institute.

## Two demographic models for estimation of population-level burden of cancer: an Excel spreadsheet

MPIDR Technical Report

### *Background*

Demographers have developed models of the population-level burden of disease which typically include: a) intensity of incurring the disease; b) survival in sub-groups living with and without the disease; and c) intensities of dying from the disease and from other causes in these sub-groups. This approach allows assessing the importance of cancer as a part of the general population dynamics.. In particular, one can assess the amounts of time spent with and without cancer in the overall life expectancy (the total lifetime lived by the life table cohort).

Demographic models of cancer may be based either on data on incidence, mortality, and lethality (IML models) or only on incidence and mortality. IM models). Technically, IML models are period increment-decrement life tables, which construct a stationary population consisting of the cancer and non-cancer sub-groups. These groups are being shaped by: 1) age-specific incidence of cancer and 2) age-specific mortality by cancer in the cancer sub-group (lethality) and age-specific mortality by other causes. The IML models are derived using conventional life table estimation methods that belong to the classic toolkit of demography (Chiang 1984, 245-273; Egidi and Frova 2006). IM models have a similar structure, but lethality in the cancer sub-group is estimated by balancing between numbers of individuals incurring cancer and death from cancer and from causes other than cancer over the life course (Pollard, 1980; Duchêne, 2002).

A conventional multi-state model of cancer mortality process in continuous notation can be described by six functions:

$l^H(x)$  - survival in the cancer-free ("healthy") sub-group as a function of age  $x$ ;  $l^H(0) = 1$ ;

$l^C(x)$  - survival function in the cancer sub-group,  $l^C(0) = 0$ ;

$I(x)$  - the density of new cancer cases;

$\delta^H(x)$  - the density of deaths in the cancer-free sub-group;

$\delta^C(x)$  - the density of deaths from cancer in the cancer sub-group. It is assumed that deaths from cancer occur in the cancer sub-group only;

$\delta^{O^C}(x)$  - the density of deaths from causes other than cancer in the cancer sub-group.

### *The Incidence-Mortality-Lethality model*

The general model of morbidity-mortality from a disease (Egidi and Frova 2006) does not consider one important peculiarity of cancer - strong dependence of the risk of death on the time elapsed after the first diagnosis (duration of the disease). The IML model addresses this

peculiarity by defining the lethality as a function of the two variables  $a, t$  where the first one is the age at incurring cancer and the second variable is the duration of the disease  $l^C(a, t)$ ,  $\delta^C(a, t)$ , and  $\delta^{O^C}(a, t)$ .

Four exogenous variables express intensities of cancer morbidity and mortality in the two sub-groups:

$\gamma^H(x)$  - the force of incurring cancer in the cancer-free sub-group;

$\eta^H(x)$  - the force of mortality (from all causes but cancer) in the cancer-free sub-group;

$\chi^C(a, t)$  - the force of mortality from cancer in the cancer sub-group ;

$\nu^C(a, t)$  - the force of mortality from causes other than cancer in the cancer sub-group.

The cancer-free survival to age  $x$  in the cancer-free sub-group is

$$(1) \quad l^H(x) = \exp\left(-\int_0^x (\eta^H(u) + \gamma^H(u)) du\right).$$

The density distribution of new cancer cases by age is

$$(2) \quad I(x) = l^H(x) \cdot \gamma^H(x),$$

and the number of deaths at an age  $x$  in the cancer-free sub-group is

$$(3) \quad \delta^H(x) = l^H(x) \cdot \eta^H(x).$$

Individuals entering the cancer sub-group are exposed to hazards of death from cancer and from causes other than cancer. Thus, survival of those who contracted cancer at age  $a$  to the duration of disease  $t$  is

$$(4) \quad l^C(a, t) = I(a) \cdot \exp\left(-\int_0^t (\chi^C(a, \tau) + \nu^C(a, \tau)) d\tau\right).$$

The total size of the cancer sub-group by age  $x$  is

$$(5) \quad l^C(x) = \int_{a+t=x} l^C(a, t) da dt.$$

The size of the entire cohort is  $l(x) = l^H(x) + l^C(x)$ .

The number of deaths from cancer among individuals, who have contracted cancer

$$(6) \quad \delta^C(a, t) = l^C(a, t) \cdot \chi^C(a, t)$$

The corresponding number of deaths from causes other than cancer is

$$(7) \quad \delta^{O^C}(a, t) = l^C(a, t) \cdot \nu^C(a, t).$$

Finally, the density of deaths at the age  $x$  from cancer and from causes other than cancer in the cancer sub-group are:

$$(8) \quad \delta^C(x) = \int_{a+t=x} \delta^C(a, t) da dt$$

and

$$(9) \quad \delta^{O^C}(x) = \int_{a+t=x} \delta^{O^C}(a, t) da dt$$

respectively.

Equations (1)-(3), (5), (8)-(9) allow us to calculate six functions:  $l^H(x)$ ,  $l^C(x)$ ,  $I(x)$ ,  $\delta^H(x)$ ,  $\delta^C(x)$ ,  $\delta^{O^C}(x)$ . These functions fully determine the IML model.

### The Incidence-Mortality model

The Incidence-Mortality model is based on three exogenous variables:

$\mu(x)$  - the force of all-cause mortality in the entire population;

$\mu^C(x)$  - the force of mortality from cancer in the entire population;

$\zeta(x)$  - the force of contracting cancer in the entire population. In contrast to the similar measure for the cancer-free sub-group, this indicator may be named *unconditional<sup>1</sup> force of contracting cancer*.

The number of survivors in the whole cohort is

$$(10) \quad l(x) = \exp\left(-\int_0^x \mu(u) du\right).$$

It is assumed that the force of mortality from causes other than cancer is the same as the force of mortality in the non-cancer sub-group and is equal to  $\mu^O(x) = \mu(x) - \mu^C(x)$ .

The total size of the cancer-free sub-group by age  $x$  is

$$(11) \quad l^H(x) = e^{-\int_0^x \mu^O(z) dz} \cdot \left(1 - \int_0^x l(z) \cdot e^{-\int_0^z \mu^O(y) dy} \cdot \zeta(z) dz\right)$$

The total size of the cancer sub-group by age  $x$  is

$$(12) \quad l^C(x) = e^{-\int_0^x \mu^O(z) dz} \cdot \left(\int_0^x l(z) \cdot e^{-\int_0^z \mu^O(y) dy} \cdot (\zeta(z) - \mu^C(z)) dz\right)$$

The density function of new cancer cases at age  $x$  is

$$(13) \quad I(x) = \zeta(x) \cdot l(x).$$

The density of deaths in the non-cancer sub-group is

$$(14) \quad \delta^H(x) = \mu^O(x) \cdot l^H(x).$$

The density of deaths from cancer is

$$(15) \quad \delta^C(x) = \mu^C(x) \cdot l(x).$$

The density of deaths from causes other than cancer in the cancer sub-group is

$$(16) \quad \delta^{O^C}(x) = \mu^O(x) \cdot l^C(x).$$

---

<sup>1</sup> *Unconditional* forces or rates relate numbers of events in question (e.g. new disease cases, deaths from cancer or from other causes) to the overall population. *Conditional* forces or rates relate numbers of events in question to specific-population groups exposed to risks of these events. While the unconditional force of contracting cancer is offset the whole population, the conditional force of contracting cancer is offset the non-cancer sub-group. Similarly, the unconditional force of dying from cancer is offset the whole population and the respective conditional force is offset the cancer sub-group.

Equations (11)-(16) allow us to obtain six functions  $l^H(x)$ ,  $l^C(x)$ ,  $I(x)$ ,  $\delta^H(x)$ ,  $\delta^C(x)$ ,  $\delta^{O/C}(x)$ . They fully determine the IM model.

Unlike the IML model, it is possible that set of functions  $\mu(x)$ ,  $\mu^C(x)$ ,  $\zeta(x)$  is self-repugnant and one of the six functions calculated according to (11)-(16) may have negative values or values exceeding 1.

### Summary indicators

These indicators present cancer as a population phenomenon. The summary indicators are calculated from the six functions  $l^H(x)$ ,  $l^C(x)$ ,  $I(x)$ ,  $\delta^H(x)$ ,  $\delta^C(x)$ ,  $\delta^{O/C}(x)$  (Table 1).

**Table 1. Length of life and morbidity indicators in the stationary population of the life table**

Name	Formula
Life expectancy at age 0	$e_0 = \int_0^{\infty} (l^H(x) + l^C(x)) dx$
The expectation of cancer-free life at age 0	$e_0^H = \int_0^{\infty} l^H(x) dx$
The expectation of life with cancer at age 0	$e_0^C = \int_0^{\infty} l^C(x) dx$
Number of cancer cases	$I_0^C = \int_0^{\infty} I(x) dx$
Average age at contracting cancer	$XI_0^C = \frac{\int_0^{\infty} x \cdot I(x) dx}{I_0^C}$
Number of deaths from cancer in the stationary model population	$D_0^C = \int_0^{\infty} \delta^C(x) dx$
Average age at death from cancer	$XD_0^C = \frac{\int_0^{\infty} x \cdot \delta^C(x) dx}{D_0^C}$
Number of deaths from causes other than cancer in the cancer sub-group	$D_0^{O/C} = \int_0^{\infty} \delta^{O/C}(x) dx$
Expected age at death from causes other than cancer in the non-cancer sub-group	$XD_0^{O/C} = \frac{\int_0^{\infty} x \cdot \delta^{O/C}(x) dx}{D_0^{O/C}}$

Average age at death in the cancer sub-group	$XD_0^{C\&O/C} = \frac{\int_0^{\infty} x \cdot (\delta^C(x) + \delta^{O/C}(x)) dx}{I_0^C}$
--	--

*Calculations using discrete data*

The Excel spreadsheet enclosed to this report can be used for calculations of IML and IM models and the summary indicators. The rules of transition from continuous notation given in the text above and the discrete notation are simple.

Forces of incurring or mortality correspond to rates of incurring or mortality rates. In the spreadsheet, the incurring or mortality rates are calculated by the standard way as ratios of numbers of events to corresponding population exposure values (mid-year populations). Survival and density functions (functions of age or duration of disease) correspond to numbers of survivors to respective ages or duration values and numbers of deaths at respective ages or duration values, respectively. At the initial age 0  $l^H(0) = l(0) = 100000$ .

Integrals should be replaced by sums of respective indicators over age or duration intervals. The formulas (11) - (12) are solutions of the linear differential equations of the first order:

$$\frac{dl^H(x)}{dx} + \mu^O(x) \cdot l^H(x) = -\zeta(x) \cdot l(x),$$

$$\frac{dl^C(x)}{dx} + \mu^O(x) \cdot l^C(x) = (\zeta(x) - \mu^C(x)) \cdot l(x).$$

In discrete form, the equations transform to finite-difference equations. We use these equations for a recurrent relation that enables calculation of  $l^H(x+1)$  and  $l^C(x+1)$  from  $l^H(x)$  and  $l^C(x)$ .

*How to use the Excel spreadsheet*

For calculating the cancer models from an empirical data set, it is necessary to place the available data to the appropriate ranges of cells. All other cells of the workbook are protected from changes.

The IML model is given in the worksheet "IML Model".

The *input data* for the IML model.

For the cancer-free ("Healthy") sub-group:

Population exposure by age	B4:B105
Number of deaths from all causes but cancer by age	C4:C105
Number of new cancer cases by age	D4:D105

Cancer sub-group

Population exposure-to-risk by age	G4:G105
Number of deaths from cancer	H4:H105

Number of deaths from other causes	I4:I105
Population exposure by age at diagnosis $a$ and by time after diagnosis $t$ for $t=0, \dots, 10+$	L4:V105
Number of deaths from Ca by age at diagnosis $a$ and by time after diagnosis $t=0, \dots, 10+$	I4:A1105

These input data are used in calculations of mortality rate at age  $x = 0, \dots, 100+$  in the non-cancer sub-group in the range B111: B211, calculation of incidence rate by age in the non-cancer sub-group in the range B111: B211, calculation of probability dying from cancer among individuals, who contracted cancer at age  $a = 0, \dots, 100+$  at the duration of disease  $t = 0, \dots, 10+$  in the range U111: AE222, and mortality rate from causes other than cancer in the cancer sub-group at age  $x$  in the range Q111: Q211.

Formulas for calculation of characteristics of the cancer-free sub-group of the IML model are placed in the range A111:L2011. For characteristics of the cancer sub-group, the calculation formulas are given in the range N111:AZ211. For the entire population, the calculations' range BC111:BF211.

The summary indicators for the IML model are presented in the range BH4: B113.

#### The IM model

The input data for the IM model include (All indicators for the whole population)

Population exposure-to-risk by age	B4:B105
Number of deaths from all causes by age	C4:C105
Number of new incidents of cancer by age	D4:D105
Number of deaths from cancer	E4:E105

The three exogenous discrete variables of the IM model are unconditional rates of mortality (the first two) and morbidity (the last one). They are computed as conventional central event rates with the mid-year population being the denominator.

The formulas for calculation of characteristics of the cancer morbidity and mortality in the entire population are given in the range A111:N211 and for the cancer sub-group in the range O111:T211, and for the summary indicators for the IM model in the range X4: Y13.