Annex

Data sources

The data on the Hungarian population and mortality between 2000 and 2014 come from the population and mortality database of the Hungarian Central Statistical Office. The annual population numbers are based on the censuses with the reference dates of 1 February 2001 and 1 October, 2011 as well as on their further and back calculations. Full-scale mortality data by gender, age and cause of death are recorded on the certificates on the post mortem examination and on death records.

Categories of used and examined data:

– gender: men, women, men and women together;

– age group (completed age): 0, 1–4, 5–9, 10–14, 15–19, 20–24, 25–29, 30–34, 35–39, 40–44, 45–49, 50–54, 55–59, 60–64, 70–74, 75–79, 80–84, 85–;

– cause of death: the disease recorded in the box “underlying disease” on the certificate on the post mortem examination. During the reference period of the research, the classification was performed according to the International Classification of Diseases, 10th Revision. The method of causes-of-death data processing was changed in 2005. Manual coding has been replaced by automated processing in which the coding of diagnose texts and the selection of the underlying cause are carried out by a software. Therefore, there are breaks in the time series of causes of deaths data between 2004 and 2005.

The international data (on the selected smoking-related causes calculated for hundred thousand population) are included in the WHO-HFA database (http://data.euro.who.int/hfadb/):

Methods

Death rates (*Józan–Radnóti* ([2002] p. 40.):

The quotient of the absolute number of deaths in the observed period and the mid-year population exposed to the risk of death is:

$$m = \frac{D}{P},$$

where

> D – number of death,
> P – mid-year population,
> m – death rate.

Standardisation (*Józan–Radnóti* [2002] p. 40.):

The method of direct standardisation was applied as follows: the indicator was generated for one hundred thousand population and standardised for the age structure of the European population in 2012 as recommended by Eurostat:

$$m^s = \frac{\sum P_i^s m_i}{\sum P_i^s},$$

where

> m^s – standardised death rate,
> P_i^s – standard population,
> m_i – age-specific death rate.

Smoking-attributable excess mortality by cause of death

The concept of the method was elaborated by *Levin* [1953] and used later by *Shultz–Novotny–Rice* [1991]; in the 1990s, it was built into the internationally recommended SAMMEC (smoking-attributable mortality, morbidity and economic costs) methodology of the Centers for Disease Control and Prevention This was applied to domestic conditions by *Józan–Radnóti* [2002], and its updated version was used by *Vitrai et al.* [2012] and by *Wéber–Faragó* [2014]. This research uses the latest available version of the method, which was developed in 2013 and its description was published in 2014 in the summary volume titled "The Health Consequences of Smoking – 50 Years of Progress: A Report of the Surgeon General"[1]. The renewed methodology includes, among others, the quantification of deaths caused by passive smoking as well.

The methodology of calculating smoking-attributable excess mortality relies on the following three data sources: *1*. relative risks, *2*. smoking prevalence and *3*. number of deaths due to certain causes of death. Among these, the relative risks are the bottleneck as these values originate from a follow-up survey conducted in the

[1] See http://www.cdc.gov/tobacco/data_statistics/sgr/.

United States. In 2013, *Thun et al.* [2013] re-evaluated the data of the CPS-II (cancer prevention study) survey. Based on this, the results included in the Excel annex of the present study (see www.ksh.hu/statszemle) were obtained by gender, smoking status (smokers, ex-smokers), age and causes of death.

Relative risks quantify how higher the exposure of smokers (*RR*1) and ex-smokers (*RR*2) is than that of non-smokers; the relative risk for non-smokers is always 1. These values are known by different groups of diseases. Due to the slow progression of smoking-attributable chronic diseases and the fact that the addiction generally begins when people are in their teens, the methodology assumes that no one dies of diseases caused by nicotine addiction before the age of 35. (Smoking may cause infrequently fatal accidents as well, but their number is insignificant.) The few exceptions are those – fortunately very rare – cases when an infant death occurs due to the mother's smoking during pregnancy.

People who currently smoke (p_1) and those who have quit smoking (p_2) form the basis of smoking-attributable excess mortality. Its prevalence broken down by smoking status (smokers, ex-smokers, non-smokers) and age groups (35–39, 40–44, 45–49, 50–54, 55–59, 60–64, 65–69, 70–74, 75–79, 80–84, 85–) is known from the National Health Interview Surveys (NHIS 2000, 2003) and from the European Health Interview Surveys (EHIS 2009, 2014). For intermediate years, a statistical interpolation approximation procedure was used. The percentage of non-smokers in the total population is p_0. Furthermore, as it was a survey-type data collection, the 95% confidence interval was always indicated.

Knowing the relative risks and the prevalence of smoking, the rate of smoking-attributable excess mortality, i.e. the so-called smoking-attributable fraction (*SAF*) can be defined within a given disease group.

$$SAF = \frac{(p_0 + p_1(RR1) + p_2(RR2)) - 1}{p_0 + p_1(RR1) + p_2(RR2)}$$

If this rate is multiplied by the number of deaths due to a certain cause of death (*D*), we will get the absolute number of smoking-attributable mortality (*SAM*) of those who died in the given cause of death.

$$SAM = SAF * D$$

Potential years of life lost (*PYLL*) attributable to smoking

The indicator sums up the (not lived) lifetimes of the deceased people in a certain population group until a fixed age. The selection of the age limit is optional, the present study analyses two age limits: the 70 years of age applied in international practice and the life expectancy at the age of death. If someone dies, for example, at the age of x, the number of years of life lost is $70 - x$ in case $x < 70$, while if he/she dies at the age of 70 or over, the indicator is 0 (*Wéber–Faragó* [2014] p. 22.).

The total number of years of life lost from the potential 70 years is:

$$PYLL_{70} = \sum (70 - x_i) * D_i ,$$

where

x_i – the average age of deceased people in the various age groups,
D_i – the number of deceased people by age group.

The number of years of life lost compared to the life expectancies by age group is:

$$PYLL_{Life\,\mathrm{exp}} = \sum (Life_{\mathrm{exp}_i} - x_i) * D_i ,$$

where

$Life_{\mathrm{exp}i}$ – life expectancy of the given age group.

The age-specific rate of years of life lost compared to life expectancies by age group is:

$$\sigma_i = \frac{\sum (Life_{\mathrm{exp}_i} - x_i) * D_i}{P_i} ,$$

where

P_i – mid-year population.