Chapter 7: Age-standardisation and denominators
M. Boniol and M. Heanue

Objective

The key determinant of cancer incidence is age, as the risk of cancer increases exponentially with increasing age. The crude rate of incidence is therefore influenced by the population structure, and cannot be used alone to evaluate whether the burden of cancer differs between populations.

To compare the incidence of cancer, the summary rates should be independent of age. A common way to take into account the age structure of a population is to standardise incidence rates for age using an external (standard) population.

Crude incidence rate

The crude incidence is *the rate at which new cases occur in a population during a specific period*. This rate is classically expressed as the average number of cases occurring per 100 000 persons each year or 100 000 person-years. In *Cancer Incidence in Five Continents (CI5), Volume IX*, numbers of cancer cases are reported for the period 1998–2002. Hence the crude incidence rate is computed with the following formula:

$$\text{Crude incidence rate} = \frac{\text{Number of new cancer cases observed in the period 1998-2002}}{\text{Total population in the period 1998-2002}}$$

The total population in the period 1998–2002 corresponds to the sum of population size for each year between 1998 and 2002. This rate is called crude because it relates to each population as a whole and is influenced by the age structure of each population. It cannot be used for comparison purposes.

Age-standardised incidence rate

The age-standardised rate is a summary of the individual age-specific rates using an external population called a standard population. This is the incidence that would be observed if the population had the age structure of the standard population, and corresponds to the crude incidence rate in the standard population. The age-standardised incidence rate is expressed, as is the crude incidence rate, as the number of new cases per 100 000 person-years.

The calculation is a weighted average of age-specific rates.

$$\text{Age-standardised rate} = \sum_i \frac{d_i w_i}{y_i}$$

Such that i represents each age group,
d_i the number of cases in the ith age group,
y_i the population size in the ith age group,
and w_i the weight applied for the ith age group, with
d_i/y_i being the age-specific rates for each ith category
and the sum of w_i being equal to 100 000 to express the age-standardised rate per 100 000 person-years.

It should be stressed that the objective of age standardisation is essentially to establish rates for comparison purposes. Any population standard would enable comparison, even considering equal weights for each age group.

The most commonly used standard worldwide is the Segi standard population (Segi, 1960). Modified by Doll et al. (1966), it was based on the age structure of the pooling of populations

---

Table 7.1 Computation of age-standardised incidence rates (Melanoma cancer, Denmark, males, 1998–2002)

Age group	No. of cases	Person-years at risk	Age-specific incidence (per 10^5 years)	Standard world population	Expected cases in standard population
i	d_i	y_i	$10^5(d_i/y_i)$	w_i	$d_i w_i/y_i$
0–4	0	869 668	0.00	12 000	0.00
5–9	0	878 535	0.00	10 000	0.00
10–14	0	779 915	0.00	9 000	0.00
15–19	7	723 313	0.97	9 000	0.09
20–24	23	840 592	2.74	8 000	0.22
25–29	59	961 907	6.13	8 000	0.49
30–34	83	1 042 749	7.96	6 000	0.48
35–39	117	1 041 800	11.23	6 000	0.67
40–44	153	953 731	16.04	6 000	0.96
45–49	188	930 210	20.21	6 000	1.21
50–54	262	990 187	26.46	5 000	1.32
55–59	303	871 969	34.75	4 000	1.39
60–64	272	649 423	41.88	4 000	1.68
65–69	223	523 486	42.60	3 000	1.28
70–74	217	439 676	49.35	2 000	0.99
75–79	162	342 435	47.31	1 000	0.47
80–84	119	210 084	56.64	500	0.28
85+	115	140 073	82.10	500	0.41
Total	2 303	13 189 753	17.46	100 000	11.94

from 46 countries. The use of this standard population allows international comparison and evaluation of changes in incidence by comparing them to rates published in previous volumes of CI5. An example of this age standardisation appears in Table 7.1.

In this example, the crude rate is 2303 / 13 189 753 = 17.46 cases per 100 000 person-years, and the age-standardised rate is 11.94 per 100 000 person-years.

The age-standardised rate is lower than the crude rate because the standard population is on average younger than the Danish population. However, the rate of 11.94 can be compared with other rates standardised on the world population.

Reference population

A new standard population was proposed by WHO in 2000 (Ahmad *et al.*, 2000). This new standard allows the estimation of incidence for the theoretical world population. This new population is, however, not of great utility for cancer incidence standardisation for two main reasons:

This new standard population is supposed to better represent the theoretical world population compared to Segi's population, but still does so insufficiently, as population data is lacking for many regions in the world. Secondly, there is no additional statistical property from this standard to the Segi population for comparison purposes (Bray, 2002). For data in the present volume, for all cancer incidences excluding non-melanoma skin cancer, we estimated the Spearman correlation (r) between the rates using Segi's standard compared to the new standard. The percentage of variation not explained by the new standard compared to the old standard $(1-r^2)$ was 0.14% for men and 0.11% for women. When looking at ranking of registries by level of incidence, the new standard produces an average change of 2.9 rank levels for men and 1.9 for women in data including all registries; this corresponds to an average of 0.7% change for men and 0.5% change for women. The lack of clear difference between incidence rates computed with Segi's population compared to the new world standard does not justify the use of the new population.

For these reasons, and to allow international comparisons with continuity from the previous volumes, we use here only Segi's world standard.

Cumulative rate

The cumulative rate is an approximation of the probability to develop a cancer during a certain period—for example, a lifetime. We present the cumulative rates for the life spans of 0–64 and 0–74 years. This has the advantage of summarising the age-specific rates independently of the age structure of the population, and gives the probability of an individual developing a cancer. This calculation is theoretical and assumes that no death occurs in the period, and that the age-specific incidence rates will be stable for an individual.

$$\text{Cumulative rate between 0 - 74 years old} = 5 \times \sum_{i=1}^{15} \frac{d_i}{y_i}$$

Such that i = 5-year age groups,
d_i=number of cases in the ith age group, and
y_i=number of person-years in the ith age group.

The cumulative incidence rate ranges between ages 0 and 75, and by its construction overestimates the probability of developing a cancer during life when incidence rates are high. However when incidence rates are sufficiently small, apart from all cancer sites combined, the cumulative rate is not significantly affected by this bias.

In the previous example for melanoma in men in Denmark (Table 7.1), the cumulative rate can be estimated from the age-specific rates:

The sum of the age specific rates between ages 0–74 is: (0.97 +2.74+6.13+7.96+11.23+16.04+20.21+26.46+34.75+41.88+42.60+49.35)/100 000 = 260.33/100 000. Hence the cumulative rate is (260.33/100 000) × 5, or 1.3%.

The age-specific incidence rates calculated in a registry are derived from different birth cohorts; i.e. populations born at different periods who experienced different levels of exposure to carcinogens. The cohort effect is an important determinant of the observed incidence for most cancer sites. For example, mesothelioma incidence is driven by asbestos exposure that took place in the past (until the 1990s in most developed countries), and the cumulative rates calculated in CI5 Volume IX would not reflect the lifetime risk of mesothelioma of a population born in 2000.

Nevertheless, the cumulative rate reflects the burden of cancer in a place, and has the added advantage over the age-standardised rate of not imposing an external population structure.

Statistical tests

Statistical tests, the results of which are not published here, were used to flag certain registries as 'unusual' or possibly inconsistent with previous published data.

Calculation of variance of ASRs

Variances of age-standardised rates (ASR) are used to assess the statistical significance of the difference between rates in Volume VIII and Volume IX and to compute confidence intervals of rates in Volume IX. We based the estimation of variance of ASR on Breslow and Day's method modified to use a binomial assumption (Keyfitz, 1966) for the variance of the crude age-specific rates.

$$Var(ASR) = \frac{\sum_{i=1}^{18} \left(\frac{d_i * (y_i - d_i) * w_i^2}{y_i^3} \right)}{\left(\sum_{i=1}^{18} w_i \right)^2}$$

We estimated 95% confidence intervals (CI) of rates based on an assumption of the normal distribution of the rates.

$$CI = \pm 1.96 * \sqrt{Var(ASR)}$$

Comparison of ASRs from Volume IX with the values from Volume VIII

To identify significant changes in incidence between Volume VIII and Volume IX we estimated the Comparative Incidence Figure (CIF), which corresponds to the ratio of the age standardised incidence rate in Volume IX divided by the age-standardised rate in Volume VIII.

$$\text{Comparative Incidence Figure} = \frac{ASR_{\text{from Volume IX}}}{ASR_{\text{from Volume VIII}}}$$

The upper (CI_u) and lower (CI_l) 95% confidence interval limits of CIF were computed using the following formula:

$$CI_{u/l} = e^{Ln(CIF) \pm 1.96 * \sqrt{\frac{Var(ASR_{\text{from Volume IX}})}{ASR_{\text{from Volume IX}}^2} + \frac{Var(ASR_{\text{from Volume VIII}})}{ASR_{\text{from Volume VIII}}^2}}}$$

We considered the difference between rates in Volume XI compared to rates in Volume VIII to be significant if the value 1.0 was not included within the 95% confidence interval of the CIF.

Comparison of percentage of microscopically verified cases (MV%) from volume IX with the values from volume VIII
We estimated the Z statistic as:

$$Z = \frac{\left| Ln\left(\frac{V_{VIII}}{nV_{VIII}}\right) - Ln\left(\frac{V_{IX}}{nV_{IX}}\right) \right|}{\frac{1}{V_{VIII}} + \frac{1}{nV_{VIII}} + \frac{1}{V_{IX}} + \frac{1}{nV_{IX}}}$$

With V_{VIII} and V_{IX} being the number of cases microscopically verified in Volumes VIII and IX, respectively, and nV_{VIII} and nV_{IX} being the number of cases not microscopically verified in Volumes VIII and IX, respectively.

We considered the difference in the proportion of microscopically verified cases to be significant if Z was greater than 1.96.

Population at risk
Population at risk figures are used as denominators in the formulas for the calculation of incidence rates, crude or adjusted. These denominators are often disregarded, and readers tend to only focus on the final incidence rates and discuss the variability in incidence data gathering (completeness, classification). This notion is driven by the age standardisation process, which tends to eliminate the effect of population age structure and gives the reassuring but false impression that population structure is no longer influencing the incidence rates. It must be stressed that denominators are an essential part of the calculation of incidence rates, and have their own biases and limitations that must be kept in mind. The age standardisation process is not intended to solve imprecise population size estimation.

Numbers of population at risk are routinely collected by registries from official statistical offices for example and registries provided these figures to Cancer Incidence in Five Continents along with their incidence figures.

Comparability
Population size estimates are frequently calculated by the official statistical office using the results from a census and a projection for the years following the most recent census. The closer the year a population estimate is to a census date, the more accurate the estimate will be. Therefore when countries do not have the same census year, there will be differences in the precision of their population estimate. Moreover, population estimates are often revised once a new census is available but not all cancer registries collect these updated figures. Hence the drift and even a significant change in incidence rate between the year of the census and the preceding year could occur simply because of the imprecision of the population estimate. This is especially true when population sizes are small and susceptible to more variability, for example, when there is migration within a country.

Despite these considerations, for most countries contributing to CI5 a census occurred during the period 1998–2002. Hence, we can expect that the biases described earlier could be largely alleviated. However, in some registries the population figures were estimated from a census performed more than 10 years before the period used in *Cancer Incidence in Five Continents, Volume IX*.

Another limitation in the population figure is the choice of the date in the year for population estimate. The best estimate of the population at risk for a specific year in which cancer incidence was observed is considered to be at mid-year, and some registries provided yearly population estimates for the 1st of July. However, this is not a systematic practice and many registries provided populations statistics estimated at the beginning of the year, using the 1st of January. In a population that increases, the use of 1st of January for population size could induce a systematic overestimation of incidence rates. It is expected to have only a minor influence in *Cancer Incidence in Five Continents*, which uses a period of 5 years combined.

Migration
Population size could also be overestimated or underestimated if migration flux patterns are not considered. Some registries collected population data corrected for immigration and emigration. Between-country migration is often better handled than within country migration, which is of major concern to regional registries tasked with maintaining optimal coverage of their population. Indeed, populations of a highly nomadic nature make the production of inter censal population estimates a real challenge. Hence incidence figures for these populations must be evaluated with care, and their comparability is to be questioned.

A few words of caution
Cancer is a disease that mainly occurs at older ages, and standard populations such as Segi have the tendency to give more weight to incidence in young age groups and hide differences observed in older populations. Therefore it cannot be stressed enough that neither the age-standardised rate nor the cumulative rate are alternatives to the age-specific incidence rates, which should always be the starting point and foundation of any thorough analysis of the incidence data.

References

Ahmad O., Boschi-Pinto C., Lopez A., Murray C., Lozano R., Inoue M. (2000). Age standardization of rates: a new WHO standard. In *GPE Discussion Paper Series: no31*. Geneva:World Health Organization.

Bray F., Guilloux A., Sankila R., Parkin D.M. (2002). Practical implications of imposing a new world standard population. *Cancer Causes Control* 13:175-182.

Breslow N.E., Day N.E. (1987). *Statistical Methods in Cancer Research, Volume II: The Design and Analysis of Cohort Studies*. Lyon: IARCPress:59.

Doll R., Payne P., Waterhouse J. (1966). *Cancer Incidence in Five Continents: A Technical Report*. Berlin: Springer-Verlag (for UICC).

Keyfitz N. (1966). Sampling variance of age standardised mortality rates. *Human Biology* 38; 309-317.

Segi M. (1960). *Cancer Mortality for Selected Sites in 24 Countries (1950-57)*. Sendai, Tohoku University School of Public Health.