



A Windows application for computing standardized mortality ratios and standardized incidence ratios in cohort studies based on calculation of exact person-years at risk

Karla Geiss*, Martin Meyer

Population Based Cancer Registry Bavaria, Registry Office, Erlangen, Germany

ARTICLE INFO

Article history:

Received 21 March 2012

Received in revised form

14 May 2013

Accepted 26 May 2013

Keywords:

Standardized mortality ratio

Standardized incidence ratio

Person-years

Software

ABSTRACT

Standardized mortality ratios and standardized incidence ratios are widely used in cohort studies to compare mortality or incidence in a study population to that in the general population on a age-time-specific basis, but their computation is not included in standard statistical software packages. Here we present a user-friendly Microsoft Windows program for computing standardized mortality ratios and standardized incidence ratios based on calculation of exact person-years at risk stratified by sex, age and calendar time. The program offers flexible import of different file formats for input data and easy handling of general population reference rate tables, such as mortality or incidence tables exported from cancer registry databases. The application of the program is illustrated with two examples using empirical data from the Bavarian Cancer Registry.

© 2013 Elsevier Ireland Ltd. All rights reserved.

1. Introduction

Rates and ratios are often used in demography and epidemiology to describe the occurrence of a health-related event. Standardization or adjustment of rates and ratios is employed to ensure comparability between heterogeneous populations with differences in their demographic structure regarding the main factors affecting health, most commonly sex, age, race, calendar period. The standardized mortality ratio (SMR) is a summary measure to assess the risk associated with a specific disease, particularly when the population under study is small or the event of interest is rare. SMRs are widely used in analyses of cohort studies, where the individuals of a defined study population are followed through time in order to

compare the number of occurred events to the number of events that would be expected based on sex-age-time-specific reference rates for the general population. In typical applications, SMRs are used to compare mortality from certain causes of death in a study population with that in the general population, for example in occupational cohort studies [1]. Similar names are adopted for morbidity, such as standardized incidence ratio (SIR), where the events of interest are new disease cases. The SIR is used, for example, to determine whether the occurrence of cancer is higher or lower than expected in a certain geographical area compared to a reference population. Another application to cancer data is the analysis of the risk of subsequent malignancies following a first cancer diagnosis.

The calculation of SMRs/SIRs in cohort studies is based on the stratification by age and calendar period of the

* Corresponding author at: Bevölkerungsbezogenes Krebsregister Bayern, Östliche Stadtmauerstr. 30, 91054 Erlangen, Germany. Tel.: +49 9131 8536063; fax: +49 9131 8536040.

E-mail address: karla.geiss@ekr.med.uni-erlangen.de (K. Geiss).
0169-2607/\$ – see front matter © 2013 Elsevier Ireland Ltd. All rights reserved.
<http://dx.doi.org/10.1016/j.cmpb.2013.05.028>

person-years of observation time for the study population. Although SMRs/SIRs are widely used, their computation is not readily available in standard statistical software packages. Publicly available software is scarce and most users have to develop their own macros or programs. Since the implementation is quite complex, the calculations are often simplified, for example approximate numbers of person-years are calculated or the calendar year is ignored and only one reference rate table for the general population is used for several calendar years.

Some software has been published, like SAS macros for computing exact person-years [2-4]. Using these macros requires not only the availability of the underlying software package but also sound knowledge about it since the input data and the algorithms have to be adjusted to fit each other. A tool for calculating SIRs for the analysis of subsequent cancers is provided by SEER*Stat [5]. However, SEER*Stat is specifically tailored to the analysis of cancer data of the SEER (Surveillance Epidemiology and End Results) Program. Analyzing own data with SEER*Stat requires the input data to be converted to a complex specific file format, which complicates the practical use of the software. PAMCOMP [6], a stand alone Microsoft Windows application for computing exact person-years and SMRs, can import different file formats. The calculation of person-years allows flexible stratification by sex and self-defined calendar periods and age groups. However, the rate file, which holds the mortality rates of the reference population for a specific cause of death, must have the same structure as the person-years table computed by the program for the study population, that is, the same number of age and calendar categories. In consequence, the user has to construct a suitable rate table for each calculation, which can be tedious.

This manuscript presents a user-friendly software application running under Microsoft Windows for computing SMRs/SIRs based on calculation of exact person-years at risk stratified by sex, age and calendar time. It can handle different file formats for input data for the study population and offers a convenient import of general population reference rate tables as, for example, reported from cancer registries.

2. Statistical methods

The standardized mortality ratio (SMR) was discussed by various authors [1,7,8]. The results apply to the standardized incidence ratio (SIR) as well. This section describes briefly the definition of the SMR, how it is calculated and how the confidence limits can be obtained.

The SMR is defined as the ratio of the observed number of deaths in the population being studied and the expected number of deaths in a comparable group of individuals from the general population, matched with respect to the main factors affecting mortality, commonly age, sex, and calendar period:

$$\text{SMR} = \frac{\text{observed number of deaths in study population}}{\text{expected number of deaths in study population}} \quad (1)$$

The expected number of deaths is obtained using general population mortality tables for the corresponding geographical region stratified by age, sex, calendar year and sometimes by race. In case of SIR, where the events of interest are new

disease cases, incidence rates for the general population are used for the calculation of the expected number of disease cases.

The SMR compares the observed number of deaths in the study population with an expected number of deaths obtained by applying the mortality rates for the general population to the demographic structure of the study population. Suppose that there are m strata defined by stratification variables like age group, sex and calendar year. The total number of observed deaths D in the study population is the sum of all observed deaths in the individual strata. Let n_i denote the person-years of observation time that the study subjects contribute to the i th stratum, and let λ_i^* represent the stratum-specific mortality rate for the general population. The total number of expected deaths E^* is calculated by applying these standard stratum-specific rates to the study population and the SMR can be written as

$$\text{SMR} = \frac{D}{E^*} = \frac{\sum_{i=1}^m d_i}{\sum_{i=1}^m n_i \lambda_i^*} \quad (2)$$

where d_i denotes the observed number of deaths in the i th stratum and $n_i \lambda_i^*$ is the expected number of deaths in the i th stratum.

Confidence intervals around the observed SMR give an indication of the precision of that estimate and are used to make statistical inferences about the SMR. The statistical power of a study depends on the size of the difference in mortality between study and background population, the incidence of the event of interest, the number of observed events and hence on the size of the study population and the observation time.

Confidence intervals provide a range of possible plausible values for the true SMR. A method for obtaining exact confidence limits was presented by Sahai and Khurshid [9]. In practice, Byar's approximation of the confidence limits is often used, being easy to calculate and sufficiently accurate [1,10]. For a $100(1 - \alpha)\%$ confidence interval, the approximate limits for the SMR according to Byar are

$$\text{SMR}_L = \frac{D}{E^*} \left(1 - \frac{1}{9D} - \frac{Z_{\alpha/2}}{3\sqrt{D}} \right)^3 \quad (3)$$

and

$$\text{SMR}_U = \frac{(D+1)}{E^*} \left(1 - \frac{1}{9(D+1)} + \frac{Z_{\alpha/2}}{3\sqrt{D+1}} \right)^3 \quad (4)$$

where $Z_{\alpha/2}$ denotes the $100(1 - \alpha/2)$ percentile of the standard normal distribution, that is, for $\alpha = 0.05$ it is $Z_{\alpha/2} = 1.96$.

The presented program provides confidence limits for the SMR according to Byar's approximation method.

Applying the formulas above to incidence data, that is, replacing the number of deaths with the number of new cases and the stratum-specific mortality rates with the stratum-specific incidence rates of the general population, the SIR and its confidence interval is obtained.

Download English Version:

<https://daneshyari.com/en/article/10344555>

Download Persian Version:

<https://daneshyari.com/article/10344555>

[Daneshyari.com](https://daneshyari.com)