How to calculate incidence rates from proportionate data

Olaf Chresten Jensen1, 3, 4, Agnes Flores2, Despena Andrioti Bygvraa1, Fereshteh Baygi1, George Charalambous3

1Centre of Maritime Health and Society, Institute of Public Health, University of Southern Denmark, Esbjerg, Denmark
2Caja Seguro Social, Rep. of Panamá, Vacamonte, Panamá
3Graduate School, Frederick University “Nicosia”, Cyprus
4University of Panamá, Faculty of Health, Panama

ABSTRACT
This paper describes the methodological aspects of calculation of incidence rates from incomplete data in occupational epidemiology. Proportionate measures in epidemiological studies are useful e.g. to describe the proportion of slips, trips and falls compared to other types of injury mechanisms within single age-strata. However, a comparison of proportions of slips, trips and falls among the different age-strata gives no meaning and can hamper the conclusions. Examples of a constructed example and some selected studies show how estimates of incidence rates can be calculated from the proportionate data by applying estimates of denominators available from other information. The calculated examples show how the risks based on the incidence rates in some cases differ from the risks based on the proportionate rates with the consequence of hampering the conclusions and the recommendations for prevention. In some cases the proportionate rates give good estimates of the incidence rates, but in other studies this might cause errors. It is recommended that estimates of the incidence rates should be used, where this is possible, by estimation of the size of the population. The paper is intended to be useful for students and teachers in epidemiology by using the attached Excel training file.

Key words: epidemiology, injury, occupational, proportionate, maritime

DEFINITIONS
The basic measurements in epidemiology are incidence and prevalence rates. The unique feature of the incidence and prevalence rates is that they express a risk that is needed to know for prioritisation of the prevention. Further, the risk expressed as the incidence and prevalence is comparable between different populations and between segments of the population, e.g. between the age groups [1]. To obtain the incidence and prevalence rates we need two types of information, one is the denominator, the number of persons or better the sum of person-time at risk and the number of incidents, injuries or diseases, the numerator. In most studies we have access to both types of data but sometimes we have no access to the denominator, the number of persons (time) at risk and then we can only calculate the proportionate rate of e.g. per cent of slips trips and falls. The proportionate measures are useful as the first step of the analysis of the risk but they are not comparable directly within different age-groups, which is the problem we address here.

Proportionate measurements or, with other words, the percentages in epidemiology are very useful to study the overall distribution of the variables. However the use of only the percentages is of limited use for identification of the relative risks needed for the prioritisation of the prevention. As the percentages do not present the relative risks, they cannot be used for comparison of the risks in different strata of the population and for comparison with other studies. Still, the proportionate rates are used and give meaning, e.g., for comparison of the percentages of slips, trips, and falls within the single-study strata, such as the age-groups. Also the proportional mortality ratio, which is the percentage...
The proportion of a specific type of disease or injury y in stratum x
IR = Incidence rate per 10^7 man-days

(P_{Injury type y in stratum x}) \times IR_{(All injuries in stratum x)} = IR_{(Injury type y in stratum x)}

P = The proportion of a specific type of disease or injury y in stratum x
IR = Incidence rate per 10^7 man-days

Figure 1. Mathematical relationship between the proportions and the incidence rates

Figure 2. Constructed example of slips, trips and falls’ proportions and incidence rates

CONSTRUCTED EXAMPLE

The constructed example gives a clear illustration of how the incidence rates can be calculated from proportionate data by applying estimates of the size of the populations at risk (Fig. 2). The denominators, e.g., national population data, are applied (in this case the number of man-days at
risk). In this constructed example the sizes of the popula-
tions at risk in four departments are selected to illustrate
the problem most clearly and the proportions of slips, trips,
and falls (STF) are chosen to be equal 70% also for the best
illustration.

The incidence risks of STF-related injuries based on
man-days across the departments are significantly differ-
ent from the proportionate data; percentages are almost
equal in the four groups, while the incidence rates vary
significantly. The number of cases and the proportions
of the specific sub-groups of cases need to be available
together with estimates of the population data. The risks in
the strata are calculated by multiplying the incidence rate
for all types of injuries in the specific strata (e.g., all injuries
on Deck = 20/1000 man-days) by the proportion of STF.
In this example the proportion of STF-related injuries was
chosen to be 70%, and the incidence rate of STF injuries
in this strata = 0.7 × 20 = 14/1000 man-days. Concerning
the ethical issues, no personal information is included, so
approval from the Ethics Committee or written informed
consent was not necessary.

**STUDY EXAMPLES**

The first example is an analysis of the proportional
STF-related injuries in a sample of data for 582 cases of
injured commercial fishermen from an emergency depart-
ment. Information about the number of all injuries and the
number of STF-related injuries was available, but there
was no information about the population at risk [3]. The
proportions of injuries from STF by age was U-shaped and
constituted around 40% for men under 20 and over 50, and
around 20% for those between these ages. These estimates
do not reflect the relative risk for STF-related injuries in the
age groups. However, incidence rates for comparison can
be calculated by using estimates of the person-time in the
age groups as shown in Figure 3.

This type of methodological problem was seen in
a study about repatriations of seafarers from the sea.
The numbers of repatriations are taken from Figure 3
in the article [4]. There is no information about the age
structure among the Filipino seafarers, only the number of
seafarers and the number of repatriations. In an attempt
to calculate the best estimates of the incidence rates,
an estimate of the population of seafarers at risk was
calculated by multiplying the percentages for each age
group. The Filipino Population Pyramid 2017 for men was
applied to calculate the approximate incidence rates [5].
The pattern of the numbers is clearly different from the
incidence rates of repatriations with the impact on the
results and conclusions (Fig. 4). In the same study, the
standardised mortality ratios. The expected number of fatalities for each age-group was summed up seafaring and fishing [6, 7]. In these studies the observed methods that were used e.g. in the studies of fatal injury in fishing vessels, the information on the denominators in the age groups was not available (Fig. 3). The methodological weakness of the study using only proportionate measures of STF in the age groups was not discussed either [3]. However, an estimation of the distribution of the number of days at sea in the age groups among the fishermen could have been applied based on an earlier epidemiological study on injuries among fishermen [8]. The numbers in Figure 6 are adjusted to show that the risk differences of STF-related injuries in the age groups may be even greater than the differences seen in the proportionate measures.

In a Mexican study the overall incidence rates of fatal industrial injuries declined from 1980 to 1995. However, the risks of fatal injuries in the industrial branches (15% in construction, 14% in oil and gas production, and 11% in farming) were based on the proportionate measures and then less useful for the prevention [9].

In a study of repatriations of seafarers, those from India were more frequently repatriated than the Philippines based on proportionate rates [10]. But if information of the populations at risk were applied to calculate estimates of the incidence rates, then the Philippines might have less repatriation incidence rates than the Indians.

In a recent study from Latin America on snakebites, the data were collected from hospital emergency room records. The authors argue that farmers are most affected [11]. But as the occupations of the patients were not registered in the hospital no relative risks for agricultures or other occupations were calculated. Further, when proportionate measures for each age group and no incidence rates are given, this reduces the usefulness of these data for the prevention.

Use of only proportionate rates for age groups was lately also seen in a study among Finnish fishermen [12]. The lack of the age distribution of the population is men-
Olaf Chresten Jensen et al., How to calculate incidence rates from proportionate data

The incidence rates in Figure 3 are recalculated by using an estimation of the distribution of the number of days at sea in the age groups based on an previous study [8]. The risk differences of slips, trips, and falls (STF)-related injury expressed as the incidence rates in the age groups may be even greater than the differences seen in the proportionate measures.

Figure 5. Numbers and incidence rates of repatriations by average stay at sea [4]

<table>
<thead>
<tr>
<th>Average stay at sea</th>
<th>Numbers of seafarers repatriated</th>
<th>Numbers of seafarers</th>
<th>Days at risk</th>
<th>Risk/100000 days</th>
</tr>
</thead>
<tbody>
<tr>
<td>150</td>
<td>6612</td>
<td>388963</td>
<td>58344450</td>
<td>11</td>
</tr>
<tr>
<td>250</td>
<td>840</td>
<td>51840</td>
<td>12960000</td>
<td>6</td>
</tr>
</tbody>
</table>

The numbers of repatriations depending on average stay on board and the persons at risk per 100,000 days on board are estimated based on the study data. The contrasts of the numbers of repatriated seafarers in the two groups are clearly different from the incidence rates.

Figure 6. Proportions and incidence rates of slips, trips, and falls-related injuries by age groups [3]

<table>
<thead>
<tr>
<th>Age-groups</th>
<th>-19</th>
<th>20–29</th>
<th>30–39</th>
<th>50–59</th>
</tr>
</thead>
<tbody>
<tr>
<td>Man-days at risk</td>
<td>9000</td>
<td>30000</td>
<td>30000</td>
<td>6000</td>
</tr>
<tr>
<td>Total number of injuries</td>
<td>1000</td>
<td>800</td>
<td>700</td>
<td>700</td>
</tr>
<tr>
<td>Number of STF-related injuries</td>
<td>400</td>
<td>240</td>
<td>210</td>
<td>280</td>
</tr>
<tr>
<td>Proportion of STF-related injuries</td>
<td>40</td>
<td>30</td>
<td>30</td>
<td>40</td>
</tr>
<tr>
<td>Incidence of STF/1000 man-days</td>
<td>44</td>
<td>8</td>
<td>7</td>
<td>47</td>
</tr>
<tr>
<td>Incidence rates of all injuries/1000 man-days</td>
<td>111</td>
<td>27</td>
<td>23</td>
<td>117</td>
</tr>
</tbody>
</table>

The incidence rates in Figure 3 are recalculated by using an estimation of the distribution of the number of days at sea in the age groups based on an previous study [8]. The risk differences of slips, trips, and falls (STF)-related injury expressed as the incidence rates in the age groups may be even greater than the differences seen in the proportionate measures.
tioned but the impact of the lack of incidence rates was not discussed.

In a review of studies of fishery health research in Latin America, only studies on occupational diseases and no studies of occupational injuries in fishing were found [13]. The lack of studies on occupational injuries in the Latin American fishery with 2–3 million, mainly small-scale fishermen is probably due the absence of register data regarding occupational diseases and injuries in the countries [14]. This situation will be improved when the infrastructure and data systems are modernised and the needed data sources to produce incidence rate studies are available.

**RECOMMENDATIONS**

The proportionate rates can only give some unprecise estimates of the risks. Supervisors, teachers, reviewers and editors should advice and help the students to get estimates of the populations at risk for calculation of the incidence rates where this is possible. Further, when the training in epidemiology is globally widespread and the availability of data for registers of the workforce is improved, the use of only injury proportionate rates will be reduced and incidence rates be more frequently used. Finally, the public health researchers from the universities can help and inspire governments, industries, Maritime Authorities, Insurances and Unions to compile valid and complete data for registers of the workforce as well as effective reporting systems of diseases and injuries to calculate comparable incidence rates and trends over time.

Excel file for training is available from this Dropbox: https://www.dropbox.com/s/53txv6jjn3p786i/Excel_Table%201-3%20for%20training.xlsx?dl=0.

**REFERENCES**