Common errors in disease mapping

Ricardo Ocaña-Riola

Escuela Andaluza de Salud Pública, Granada, Spain

Abstract. Many morbid-mortality atlases and small-area studies have been carried out over the last decade. However, the methods used to draw up such research, the interpretation of results and the conclusions published are often inaccurate. Often, the proliferation of this practice has led to inefficient decision-making, implementation of inappropriate health policies and negative impact on the advancement of scientific knowledge. This paper reviews the most frequent errors in the design, analysis and interpretation of small-area epidemiological studies and proposes a diagnostic evaluation test that should enable the scientific quality of published papers to be ascertained. Nine common mistakes in disease mapping methods are discussed. From this framework, and following the theory of diagnostic evaluation, a standardised test to evaluate the scientific quality of a small-area epidemiology study has been developed. *Optimal* quality is achieved with the maximum score (16 points), *average* with a score between 8 and 15 points, and *low* with a score of 7 or below. A systematic evaluation of scientific papers, together with an enhanced quality in future research, will contribute towards increased efficacy in epidemiological surveillance and in health planning based on the spatio-temporal analysis of ecological information.

Keywords: small-area, ecological design, disease mapping, incidence, mortality, morbidity, geographical information systems, Bayesian model.

Introduction

Statistics plays a crucial role in research, planning and decision-making in the healthcare sector (Nicholls, 1999; Zeger et al., 2004). Progress in technologies and continued research in computational statistics has enabled us to implement sophisticated mathematical models using software that can be readily utilized by non-specialists (Buchan, 2000). Such accessibility has undoubtedly made a major contribution towards the dissemination and transfer of mathematical know-how to other disciplines and, in particular, towards practical applications within health sciences research.

As a result, over the last decade medical journals have published a host of papers that used novel statistical methods. However, the application of statistical techniques and the interpretation of results are not always appropriate. Recent studies have shown that a number of articles published in high impact-factor journals contain errors in data analysis or interpretation of results, with the ensuing repercussions on the validity and efficiency of the research conducted (Armitage et al., 2002; Basskin, 2003; Garcìa and Alcaraz, 2004). Inappropriate use of statistical methods poses a serious problem affecting both the quality of medical publications and the advancement of scientific knowledge (Wang and Zhang, 1998).

Journal editors are not unaware of this issue. Some editors even acknowledge that many studies published today have serious methodological flaws that lead to unfounded conclusions (Smith, 2001; Editorial, 2004). Small area epidemiological studies are not immune to this problem. The study of the geographical distribution of diseases or *disease*

Corresponding author:
Ricardo Ocaña-Riola
Escuela Andaluza de Salud Pública
Campus Universitario de Cartuja
Cuesta del Observatorio, 4
Apdo. de Correos 2070, 18080 Granada, Spain
Tel. +34 958 027400; Fax +34 958 027503
E-mail: ricardo.ocana.easp@juntadeandalucia.es

mapping is a major area of research today. Advances in information technology, the availability of powerful geographical information systems (GIS) and the implementation of complex mathematical models in specialised software have all led to a range of morbid-mortality atlases and small-area studies over the last decade (Gundersen, 2000). However, the methods used to draw up such research, the interpretation of results and the conclusions published are not always accurate.

The scant critical discussion on this kind of study has also meant that new research hypotheses, health policy implementations and the risks conveyed to the population are based on spurious conclusions arising from ecological fallacies and other methodological issues (Monmonier, 1996).

Often, the proliferation of this practice has led to inefficient decision-making, alarm and scientific frustration as well as political, economic and social deterioration that evidently all have a negative impact on the advancement of scientific knowledge and the implementation of appropriate health policies (Smith and Neutra, 1993).

This paper reviews the most frequent errors in the design, analysis and interpretation of small-area epidemiological studies and proposes a diagnostic evaluation test that should enable the scientific quality of published papers to be ascertained. A systematic evaluation of such research, together with an enhanced quality in future publications, will contribute towards increased efficacy in epidemiological surveillance, and in health planning based on the spatio-temporal analysis of the available information.

Misuses of adjusted rates

Many researchers believe that a single summary measure is easier to interpret than several specific rates. As a result, most small-area epidemiological studies use age- and sex-adjusted rates. The intended aim is to compare morbidity and mortality in geographic areas with a distinct population pyramid. However, for this comparison to be meaning-ful, the specific rates for each area need to be proportional to the specific rates for any other area, with the same proportionality factor between them (Fleiss et al., 2003). This indispensable condition has been sorely overlooked and very few studies bear this in mind, giving rise to inaccurate results and conclusions.

Table 1 shows a situation where it is inappropriate to compare two areas using the age-adjusted rate (scenario I) as well as another situation where this is appropriate (scenario II). For this purpose, the adjusted rate is defined as:

$$r_m = \frac{\sum_{i=1}^{I} P_i \, r_{im}}{\sum_{i=1}^{I} P_i} = \sum_{i=1}^{I} \omega_i \, r_{im}$$

where r_{im} is the specific mortality rate for age group i in the geographic area m, while P_i is the standard or reference population for the same group.

In fact, the adjusted rate is merely a weighted mean of its specific rates, where the weighting for each age group, ω_i, is equal to the proportion of people from that group who make up the reference population.

In the first scenario, the raw mortality rate is 0.028 and 0.015 in areas 1 and 2, respectively, which means an excess mortality of 87% in the first area *versus* the second. However, when the age-adjusted rates are calculated, the result is the reverse, which suggests now that area 1 has a 30% lower mortality rate than area 2. Supposedly, the effect of age has been eliminated by using adjusted rates. This would suggest that the difference in mortality between areas is due to other causes.

Most researchers would propose that global health plans should be arranged in area 2, regardless of the age group, arguing that their age-adjusted mortality rate is higher than in area 1. However, the real situation is rather different.

If we study specific rates, mortality in area 2 is higher than in area 1 in the age group 14 years and below, while the situation is reversed for the 65 years and older group. In this case, health plan-

Table 1. Age-adjusted rates and rates ratio in two scenarios.

Scenario I Age group	Standard population Population (P_i)	Weight (ω_i)	Area 1 Cases	Population	Specific rate (r_{i1})	Area 2 Cases	Population	Specific rate (r_{i2})	Comparison area 1/area 2 rates ratio
≤14	2,000	0.250	7	1,000	0.007	67	1,000	0.067	0.10
15-44	3,000	0.375	45	5,000	0.009	90	10,000	0.009	1.00
45-64	2,000	0.250	161	7,000	0.023	92	4,000	0.023	1.00
≥65	1,000	0.125	268	4,000	0.067	14	2,000	0.007	9.57
Total	8,000	1.000	481	17,000		263	17,000		
Raw rate					0.028			0.015	1.87
Age-adjusted rate					0.019			0.027	0.70
Scenario II Age group	Standard population Population (P_i)	Weight (ω_i)	Area 1 Cases	Population	Specific rate (r_{i1})	Area 2 Cases	Population	Specific rate (r_{i2})	Comparison area 1/area 2 rates ratio
≤14	2,000	0.250	7	1,000	0.007	14	1,000	0.014	0.50
15-44	3,000	0.375	45	5,000	0.009	180	10,000	0.018	0.50
45-64	2,000	0.250	161	7,000	0.023	184	4,000	0.046	0.50
≥65	1,000	0.125	268	4,000	0.067	268	2,000	0.134	0.50
Total	8,000	1.000	481	17,000		646	17,000		
Raw rate					0.028			0.038	0.74
Age-adjusted rate					0.019			0.038	0.50

ning should address both areas with specific strategies for each, given that the differences in mortality rates occur in different age groups. As the specific rates ratio is not constant in all groups, using the weighted mean of these rates to compare areas does not reflect the real status of mortality.

In the second scenario, the specific rates ratio between areas is kept constant, showing that mortality in area 1 is 50% lower than in area 2 in all age groups. When the adjusted rate is calculated, the outcome is the same, stating that area 1 has half the mortality rate of that in area 2, regardless of age group. In this case, it would be appropriate to use the adjusted rate to compare mortality in both populations.

When specific mortality rates are proportional between areas, then $r_{i1} = k r_{i2}$, where k is a constant factor of proportionality for all age groups. In this case, the adjusted rates ratio (RR), given by

$$RR = \frac{\sum_{i=1}^{I} \omega_i \, r_{i1}}{\sum_{i=1}^{I} \omega_i \, r_{i1}} = k$$

will be a sound indicator of comparative mortality rates between the two areas.

To compare the adjusted rate for all areas involved in the ecological study, this condition of proportionality must be checked for each area. It is generally difficult to ensure compliance with this optimal situation. Consequently, the geographic distribution of the specific rate for each stratum of population would need to be studied separately (Woolsey, 1959; Elveback, 1966). So, the specific rate should be the optimal measure of frequency in small-area epidemiological studies (Pickle and White, 1995; Choi et al., 1999; Julios et al., 2001).

In most studies, it would be meaningless to rank geographic areas according to the value for the adjusted rate. If the specific rates are not proportional, then the areas with a higher adjusted rate will not necessarily have to exhibit greater morbid-mortality for all population groups. The same applies when analysing temporal trends.

Unfortunately, the indiscriminate use of rate adjustment in scientific publications is often leading to spurious interpretations and conclusions that, in turn, give rise to inappropriate health policies and decision-making.

Misuse of standardised mortality ratio and standardised incidence ratio

The afore mentioned rates adjustment is known as the direct method. The specific morbidity and mortality rate for each population group must be known. However, sometimes only the total number of deaths is available, but without a group-wise distribution. As a result, the indirect method of rates adjustment is generally used in this case (Fleiss et al., 2003). This method is a particular case of the above, but has become especially widespread and distorted in small-area studies that it requires special discussion.

The standardised mortality ratio (SMR) is the quotient between the number of cases seen in a given geographic area, o_m, and the number of cases that are expected in an outside area if the population pyramid is the same as in the area studied, e_m. This would be expressed as:

$$SMR_m = \frac{o_m}{e_m} = \frac{o_m}{\sum_{i=1}^{I} p_{im} \dfrac{O_i}{P_i}}$$

where p_{im} is the population in stratum i in geographic area m, and P_i and O_i are the population and number of cases seen in the outside area for the same population stratum, respectively. The study area is usually called the *exposed population* and the outside area the *non-exposed population*. There will be an excess mortality, not due to a different population pyramid, if the number of cases observed in the exposed area is higher than the number of expected cases for the non-exposed area, with the same age structure in both areas after standardisation. The same method is applied in the case of the standardised incidence ratio (SIR), so the following comments apply to both SMR and SIR.

SMR can also be expressed as follows:

$$SRM_m = \frac{\dfrac{\sum_{i=1}^{I} p_{im} \dfrac{o_{im}}{p_{im}}}{\sum_{i=1}^{I} p_{im}}}{\dfrac{\sum_{i=1}^{I} p_{im} \dfrac{O_i}{P_i}}{\sum_{i=1}^{I} p_{im}}} = \frac{\sum_{i=1}^{I} \omega_{im}\, r_{im}}{\sum_{i=1}^{I} \omega_{im}\, R_i}$$

where r_{im} and R_i are the specific rates for the exposed and non-exposed areas, respectively, and ω_{im} is the proportion of people from stratum i who make up area m.

Seen in this light, SMR is simply a RR adjusted by the direct method, where the reference population is the area m or *exposed population*. This feature is pivotal to highlight two key issues that are often overlooked in most published studies:

(i) SMR is a quotient of adjusted rates. As such, it is only meaningful to calculate this ratio when the specific rates for the exposed and non-exposed areas are proportional (Woolsey, 1959; Elveback, 1966). This condition is particularly difficult to check when using the indirect method, since the specific rates for the exposed area are not used. Because of this, the study of the geographic distribution of mortality by SMR may be attenuating important issues for each population stratum and showing biased results (Breslow and Clayton, 1983; Armstrong, 1995).

(ii) In the calculation of SMR, the standard or reference population is the population in the area m or the *exposed population*. The weightings used to find the weighted mean for specific rates is derived from them, such that the reference population is never the outside area or *non-exposed population* as many researcher believe. SMR for different geographic areas always have different reference populations and cannot, therefore, be compared. Consequently, it is not correct to state that those geographic areas with a high SMR have greater mortality than areas with low SMRs. If they are not comparable, then their values cannot be ranked as a synonym of the frequency of mortality in these areas (Armstrong, 1995).

The misuse of SMRs in small-area studies also affects temporal-related analysis of mortality and ecological correlations. Most epidemiological studies model the logarithm of the mean number of cases seen according to the logarithm of the number of expected cases, which acts as an *offset,* plus a linear combination of explanatory variables, time

sometimes being one of these (Lawson, 2006). The parameters of the model are estimated using frequentist or Bayesian methods, such that the exponential of the linear combination of explanatory variables is the risk or estimated SMR (Lawson et al., 2003). However, given that the order of these values is meaningless from the epidemiological standpoint, it would be inappropriate to speak of a percentage increase or decrease in SMR between two time units or two values for an explanatory variable, as being equivalent to an increase or decrease in mortality (Breslow et al., 1983; Greenland and O'Rourke, 2008).

The indirect method of standardisation is well known. Its limitations and conditions for use were set at the beginning of the 20th century (Wolfenden, 1923; Yule, 1934), the precursors of modern-day epidemiology already alerted against its indiscriminate application and, over the years, its use has been widely questioned (Breslow et al., 1983; Rothman et al., 2008). In spite of this, however, SMR and SIR are still incomprehensibly being used to study the spatio-temporal distribution of mortality and incidence in small areas, and are welcomed by scientific journals with widespread international coverage.

Misuse of total population of an area as the denominator for mortality rates from a specific cause

The denominator of any raw or specific rate must be the number of person-years who run the risk of disease, in the case of incidence rates, or risk of death for mortality rates (Greenland and Rothman, 2008).

When calculating the incidence rate for a disease that may affect the population as a whole, the denominator coincides with the number of inhabitants in the geographic area. The same occurs when dealing with the mortality rate from all causes.

However, when the mortality rate for a particular cause is being examined, the denominator is not the whole population in the given geographic area. In this case, the number of people at risk only includes those living inhabitants who presented the disease during the study period, i.e., the prevalent cases (Greenland and Rothman, 2008).

Most published studies use the mortality rate for the whole population in an area as the denominator instead of the number of prevalent cases. Thus, the mortality rate for a specific cause is generally expressed as:

$$r_m = \frac{o_m}{p_m}$$

where o_m and p_m are the number of deaths from the disease under study and the population in the area m, respectively. However, the real mortality rate, calculated with prevalent cases, would be

$$r'_m = \frac{o_m}{p'_m}$$

where p'_m is the number of people with that disease.

It is easy to check that both rates verify the relationship $r_m = r'_m \, \omega'_m$, where

$$\omega'_m = \frac{p'_m}{p_m}$$

is the prevalence of disease in the area, with a value ranging between 0 and 1. Several major questions may be deduced from this ratio:

a) using the whole population as the denominator leads to an underestimation of the real mortality rate. This underestimation will be greater, the lower the disease prevalence is for that disease in the geographic area;

b) it would be right to calculate mortality rates with the total population as the denominator only when all the inhabitants of an area are prevalent cases, in other words, $p'_m = p_m$. As a result, it would be appropriate to use the total population to study all-cause mortality, but not mortality from specific causes, particularly for low-prevalence diseases.

This widespread error affects both the value for the rate as well as the comparison of mortality between two areas. To check this effect, we define

$$RR = \frac{r_1}{r_2}$$

as the rates ratio using the total population as the denominator and

$$RR' = \frac{r'_1}{r'_2}$$

as the real rates ratio, i.e. the ratio using the number of prevalent cases as the denominator in areas 1 and 2, respectively. Both mortality ratios verify the relationship

$$RR = RR' \frac{\omega'_1}{\omega'_2}$$

where ω'_1 and ω'_2 are the respective prevalence for disease in the two areas. From this, the following results may be deduced:

(i)   the rates ratio calculated using the total population as the denominator gives rise to skewed results regarding the real rates ratio. The bias factor is equal to the quotient between prevalence for disease in both areas;

(ii) the rates ratio calculated using the total population as the denominator will be correct only if the areas being compared have the same disease prevalence;

(iii) two areas with the same real mortality rate, i.e. $RR' = 1$, may reveal a false difference in mortality if the rates are calculated using the total population as the denominator. This would occur if both areas simply have different disease prevalence; in this case, the calculated rates ratio (i.e. RR), will be equal to the quotient of both prevalence.

Most studies published use the whole population to calculate the mortality rate from specific causes, and then use this to calculate the mortality ratio between areas. The differences found in mortality may be due to using the inappropriate denominator.

Many authors justify using the whole population as the denominator by claiming it is impossible to ascertain or estimate the prevalent cases for the disease under study. In this case, perhaps it is preferable not to examine the geographic distribution of mortality from specific causes, given that the comparison between areas with different disease prevalence will give results that may suggest unfounded geographic inequities in mortality.

Misuse of aggregate data in a long time period

Most small-area studies describe the geographic distribution of morbid-mortality using a single time period that groups all cases occurring over a period of years. In some studies, this period may span over 20 years, offering a static view of morbid-mortality that assumes an unchangeable geographic distribution. However, health outcomes are dynamic, and there has been a fast pace of change in most countries over the past few decades (Frenk, 1997; EPHA, 2007).

Recent methodological studies have shown that using long time periods may give rise to a bias in estimates for morbid-mortality rates and relative risks, such that the excess seen in some areas may simply be a reflection of a past situation that still appears due to the aggregation of information (Ocaña-Riola, 2007).

Any small-area study using long-term data or different cuts in time has one spatial component and one temporal component. Just as there is a geographic correlation between areas, there may also be a temporal correlation between time units that should not be overlooked (Ocaña-Riola, 2007).

Given that the morbid-mortality rate changes over time, it is not a sound practice to assign the morbid-mortality rate for the whole period to an area, unless it has been proven that the rate is constant throughout the time units making up that period.

Many researchers believe that grouping several years of information will confer greater stability on the rates, given the higher numerators and denominators. Far from producing an enhanced design, this practice could give rise to biased results. As will be seen below, the small number of cases does not hinder drawing valid conclusions in small-area studies.

Confusion between sample size and values for an ecological variable

In a small-area study, the unit for analysis is the geographic area. The sample size, or more aptly, the number of study units is made up of all the geographic areas under research.

Each of those units for analysis has certain characteristics that distinguish it from others and these can be measured with demographic, economic, social, health or environmental indicators. These indicators are simple random variables with a probability distribution or density function that enable us to study variability between geographic areas (Ebdon, 1985).

When the spatio-temporal distribution of morbid-mortality is being studied, the variable of interest is the number of cases observed in an area. These are usually low values in small-area studies, especially if the information is studied by age groups and gender separately. For this reason, some authors have claimed that specific rates may be less accurate and reliable for detailed comparisons (Fleiss et al., 2003). A poor interpretation of this issue has led many researchers to believe erroneously that the results of the analysis are not consistent due to the *small sample* size. However, these claims need to be qualified if we are to avoid major methodological errors.

In an ecological study, the number of cases is simply the value for one of the random variables in the study that may be large or small, in line with the nature of the variable itself. In fact, it is precisely the small number of cases seen in each area that makes it possible to model this indicator through Poisson's distribution (Feller, 1957). This is one of the most widely used distributions and forms the basis for any disease mapping model (Lawson, 2006). The number of cases observed in all geographic areas could be 0, which is one of the most frequent values seen in Poisson distributions considering rare diseases. This does not invalidate the study. It merely reflects the homogeneity of geographic areas regarding this particular variable.

Let's suppose that in a geographic area with a population of 10,000, there is one death in a given age or gender group. If the reference rate for the same population stratum is 1 x 10,000 inhabitants, then the geographic area will not show an excess in mortality. However, if by chance there were one more death in the area, the specific rate would be 2 x 10,000 and the excess in mortality *vis à vis* the ref-

erence rate would be 100%. Some authors define this situation as unstable, inaccurate or rather unreliable rates and again allude to small sample size as the cause of this effect.

What is actually happening is that, sometimes, the variance in the number of cases has a structured spatial heterogeneity and another unstructured heterogeneity which gives rise to the extra-Poisson variability or overdispersion (Lawson, 2006). In such cases, the estimation of rates requires more sophisticated smoothing statistical techniques, many of which are included in the so-called generalised linear mixed models (Breslow and Clayton, 1993; Fernández and Green, 2002). The Besag, York and Mollié model is currently one of the most widely used for this purpose, enabling a smoothing of the rates and the study of geographic distribution of morbid-mortality in small areas (Besag et al., 1991).

The confusion over the concepts of small sample size and the value of a random variable has led many researchers to link together several geographic areas or to aggregate data from several years, believing that the increase in what they consider to be *sample size* will achieve more reliable results. However, sometimes this practice is just the start of a chain of methodological errors that will give rise to a result that is the exact opposite of the desired outcome.

Excessive confidence in information sources

Over the past few years, authors have raised the alert over errors and biases that are arising in information registries (Polissar, 1980; Boyle, 2004). Some studies also point to the difficulty in obtaining data that is totally error-free and to the need to improve both population data sources as well as health registries (Librero et al., 1993; Gómez et al., 1994; Arts et al., 2002).

Studies conducted in Spain show that between 17% and 84% of deaths assigned to certain municipalities refer to people who were not registered as residents in these municipalities (Librero and Benavides, 1995). Further studies in the USA revealed that 24% of deaths studied stated an incorrect residence code

on the death certificate (Williams et al., 1990).

Errors of this nature have not only been found in mortality studies. Research conducted on hospital admissions, cancer incidence and other health have revealed some kind of error in information recording. For instance, 24% of hospital admissions at a hospital in Almeria (southern Spain) were patients who were living in the city but were not registered residents, leading to an overestimation of hospital admissions (Salas et al., 2003). Also, some studies warn of major differences in calculating cancer incidence depending on the source of information used (Roos, 1993; Phipps et al., 2005).

Simulations have recently been run to understand how migratory flows that are not included in official population registers can affect morbid-mortality and relative risk rates for geographic areas (Ocaña-Riola et al., 2009). Results showed that there was an underestimation of morbid-mortality rates and relative risk over 8% and 11%, respectively, in areas with a high degree of unregistered emigration. Moreover, areas with large numbers of unregistered immigrants showed an overestimation of morbid-mortality rates and relative risk above 19% and 15%, respectively (Ocaña-Riola et al., 2009).

These findings show that viewing areas with high morbid-mortality rates or excess risk may actually only be a consequence of accumulated errors in the information source.

Today there is growing interest in health inequities using small area epidemiology designs. Most of these publications still posit hypotheses related to social, economic or environmental issues in an attempt to account for differences seen between geographic areas. However, very few studies assess the errors or biases in information sources linked to the geographic mobility of the population that may well be causing these differences.

Overlooking numerical legend in interpreting choropleth maps

Many authors have given consideration to the best way of setting the cut-off points for a morbid-

mortality rate, the optimum number of intervals and the most appropriate range of colours for graphic representation in a choropleth map (Slocum et al., 2009). The powerful visual impact of these graphs makes the researcher focus on the darker colours that usually represent the highest values for the indicator (MacEachren, 1995). Immediately, these areas on the map are named as high-risk areas in scientific publications and then the media convey the social alarm to the population. However, very few studies discuss the issue of whether the highest values for morbid-mortality are very different in epidemiological terms to the lowest values (Pickle and Herrmann, 1999; Monmonier, 2005).

As for any other random variable, the morbid-mortality rate always shows variability between geographic regions that can be measured using dispersion measures. Of these, the interquartile range (i.e. difference between the third and the first quartile) is the soundest. A small value will indicate that the rate is relatively homogeneous in all geographic areas and, as a result, any division in coloured intervals will merely be a classification offering scant epidemiological value (Ebdon, 1985).

Figure 1 shows the smoothed mortality rate for males in municipalities throughout Andalusia (southern Spain) for ages ranging between 15 and 44 years (Ocaña-Riola et al., 2008; Ocaña-Riola and Mayoral-Cortés 2010). The rate is divided into quartiles, using the darkest colour for values in the last quartile. On the graph, municipalities are seen to group together in the eastern area with higher mortality rates. However, the interquartile range of 3.18 deaths per 10,000 inhabitants suggests a certain degree of homogeneity for these values throughout the region. The differences in mortality between areas are scarcely relevant from the epidemiology standpoint, particularly if we bear in mind possible unregistered migratory flows, their repercussion on rate calculations and the small, though not impossible, variation in outcome depending on the smoothed method chosen (Best et al., 2005; Ocaña-Riola et al., 2009).
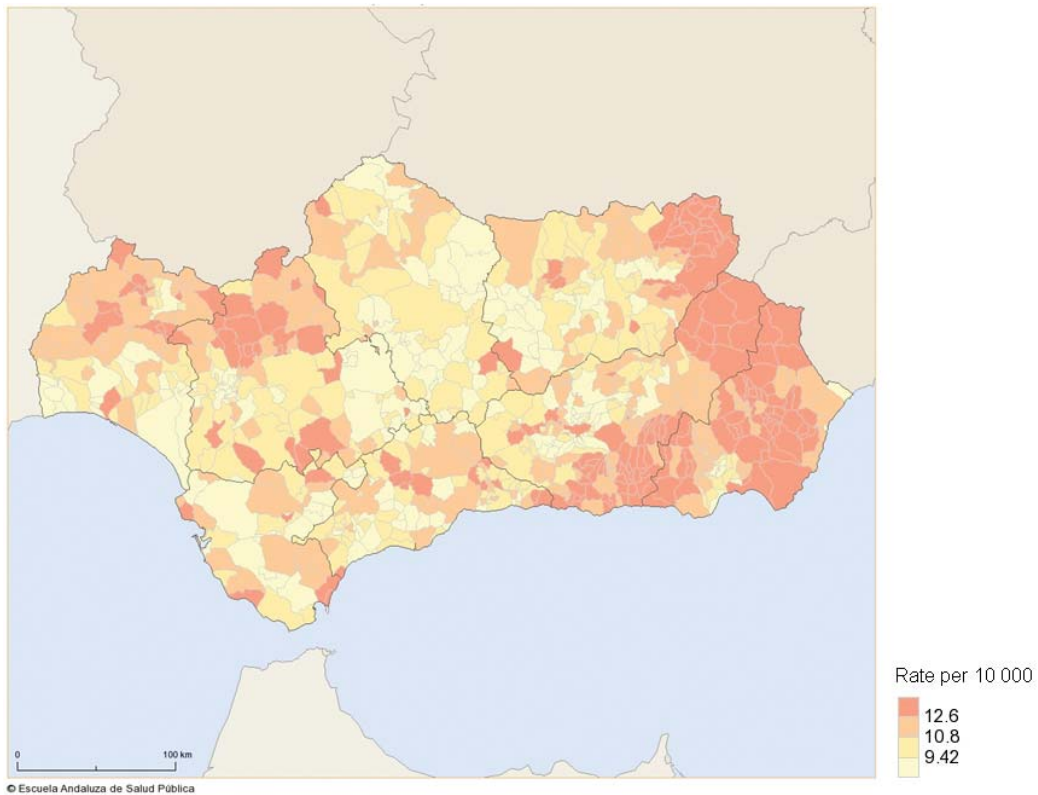
As a result then, the range of colours is not

Fig. 1. Geographical distribution of smoothed male mortality rate in Andalusia (southern Spain) in 2006 in the 15-44 year age group. Source: Interactive Mortality Atlas for Andalusia (AIMA).

enough to interpret the geographic distribution of a rate or a RR correctly. We must also bear in mind the interquartile range for the indicator or any other dispersion measure. If not, chloropleth maps will generate misleading information and unwarranted alarm (Monmonier, 1996). Unfortunately, most published articles tend to overlook this essential issue when analysing the geographic variability of morbid-mortality.

Clinging to the reference population as the model

In general terms, there will always be variability between the morbid-mortality rates for different geographic areas. Since these are random variables, the laws of probability that govern their distributions or density functions will render it highly unlikely that all areas will present the same value.

So, any health strategy intended to match the rates for all the areas in a region or country will inevitably be destined to failure. The least that could be achieved would be to homogenise these values around a pre-set reference, such that the variability will be small enough to assume that the whole region has rates that are comparable epidemiologically speaking.

However, there are two problems that have not yet been solved for this purpose, namely, to ascertain the acceptable reference value for an incidence or mortality rate, and to decide the tolerated range of values surrounding that reference.

As there is no epidemiological criterion, most researchers resort to statistical reasoning. As a result, the reference morbid-mortality rate is usually internal, taking the sum of cases and the population in all geographic areas in the study, or external, tak-

ing the morbid-mortality rate for the country. Both methods have advantages and disadvantages and are used indistinctly (Greenland, 1987). The ratio for specific rates in the area and the reference population, together with certain confidence or probability intervals, is usually the most widely used criterion to decide whether there really is statistically significant excess morbid-mortality (Lawson, 2006; Rothman et al., 2008).

Despite its popularity, this criterion is arbitrary, since a given area may show significant excess risk or not depending on both the measure of magnitude used "RR or rate difference" (Moser et al., 2007) and the reference population selected (Morin and Tremblay, 1981; Wilczyńska et al., 1990; Sorlie et al., 1999). Moreover, as any other geographic area, the reference region is subject to unregistered migratory flows in official population figures, such that the incidence or mortality rate will be skewed according to such population movements (Ocaña-Riola et al., 2009). Consequently, keeping the same morbid-mortality rate, a given geographic area may reveal a RR higher than for preceding years simply because the estimation of the reference rate for that year is lower, and not because the morbid-mortality has actually risen in that area.

For the above reasons, the most appropriate way of ascertaining the behaviour of morbid-mortality in a given area is to compare it with itself over time, keeping track of the trend in specific rates. As a complementary measure, a reference interval must also be set to mark the acceptable values for morbid-mortality from the epidemiological, political and social standpoints for each country, in line with the level of technological and healthcare development in each. The definition of this interval should also form part of national health policies for the country's government, who would alter this as certain strategic goals are met. In this way, a decreasing specific morbid-mortality rate, set within the pre-established interval for values, will indicate the appropriate evolution of the geographic area.

Epidemiological studies published to date have not innovated in this field. Most replicate previous stud-

ies with different reference populations. As a result, the scientific advancement concerning the excess morbid-mortality is very limited and, sometimes, of little epidemiological interest (Rothman, 1990).

Incorrect interpretation of ecological correlations

Publications on small-area epidemiology studies often show a map representing the morbid-mortality rates and, next to it, another figure describing the geographic distribution of air pollution, per capita income, unemployment or any other ecological indicator. If the areas with the highest morbid-mortality rates coincide with the areas with greatest pollution or least socio-economic development, the authors tend to conclude that no causal relationship can be established between both indicators so as to avoid falling into the well-known ecological fallacy. However, they do state that the geographic correlation between the most disadvantaged areas and those with the worst health status suggest a research hypothesis that should be addressed in future work with designs devised on an individual basis.

The ecological fallacy entails assuming that the ecological correlation seen between geographic areas also exists on an individual scale (Morgenstern, 2008). Despite acknowledging the existence of the fallacy, most authors mention that ecological correlation is a clue that leads to the discovery of the same correlation on an individual scale. However, this statement fall into the ecological fallacy mentioned earlier. This convoluted expression is a serious error in interpretation that easily slips through the net of referees and editors alike.

Figure 2a shows a perfect positive correlation between unemployment and mortality percentages in three geographic areas. The areas with highest unemployment also show the greatest mortality. However, Table 2 shows that excess mortality is due to a higher number of deaths among the working population. In all three areas, the percentage of deaths is lesser among the unemployed. The risk of death, measured with the odds ratio is 62% lower

amongst the unemployed than for workers in area 1, 89% lower in area 2 and 98% lower in area 3. In this case, the result on an individual scale is completely the opposite of what it may appear to be on an ecological scale.

Contrary to the preceding descriptions, Figure 2b shows a negative correlation between unemployment and mortality. The areas with highest unemployment show lower mortality, a result which would possibly surprise most researchers. However, Table 2 shows that in the three areas, the percentage of deaths is lower in the group in employment. In this case, the risk of death among the unemployed is double that for those working, in all three areas. Again, the result on an ecological scale cannot be extrapolated to the individual scale.

As we are dealing with aggregate data, the level of exposure for individuals to any risk factor is unknown. As a result, any hypothesis suggesting that there is a relationship between the excess mortality seen in certain areas and social inequities, use of health services, environmental exposures or occupational hazards for individuals are falling into the ecological fallacy (Wakefield and Shaddick, 2006).

Apart from this frequent error in interpretation, some studies use scatter plots where SMR or SIR act as a dependent variable. However, this is not an appropriate procedure, since the order of the values arranged along the vertical axis is not equivalent to an order in the morbid-mortality for the areas, unless the specific rates are proportional to the population rates used for the comparison (Breslow et
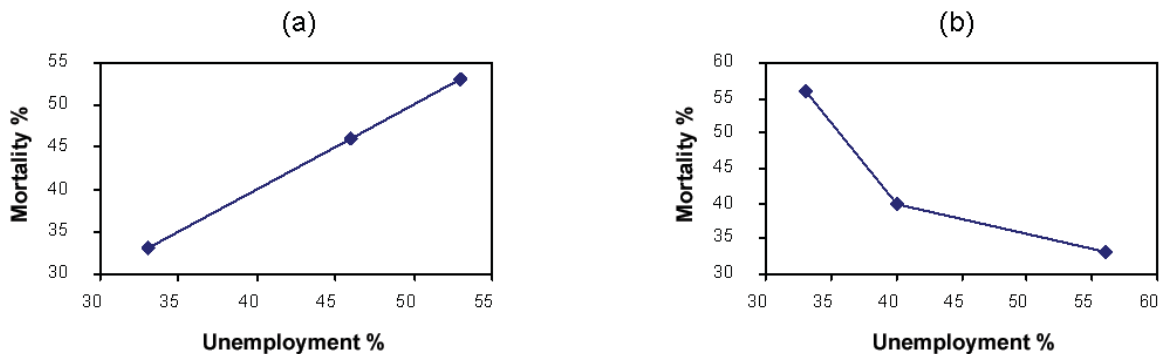

Fig. 2. Geographical correlation between unemployment and mortality from all causes percentages.

Table 2. Ecological correlation between unemployment and mortality from all causes in working age people.

	Positive ecological correlation – Figure 2a								
	Area 1			Area 2			Area 3		
	Deaths	Alive	Total	Deaths	Alive	Total	Deaths	Alive	Total
Unemployed	100 (20%)	400	500 (33%)	175 (20%)	700	875 (46%)	130 (20%)	530	660 (53%)
Employed	400 (40%)	600	1,000	700 (69%)	308	1,008	530 (91%)	50	580
Total	500 (33%)	1,000	1,500	875 (46%)	1,008	1,883	660 (53%)	580	1,240
	Negative ecological correlation – Figure 2b								
	Area 1			Area 2			Area 3		
	Deaths	Alive	Total	Deaths	Alive	Total	Deaths	Alive	Total
Unemployed	200 (67%)	100	300 (33%)	200 (50%)	200	400 (40%)	200 (40%)	300	500 (56%)
Employed	300 (50%)	300	600	200 (33%)	400	600	100 (25%)	300	400
Total	500 (56%)	400	900	400 (40%)	600	1,000	300 (35%)	600	900

al., 1983). Graphic representation and modelling of specific rates would be a more suitable approach, although this would not avoid ecological fallacy.

An excellent paper by Morgenstern (2008) makes interesting comments on geographic correlation studies and biases appearing in estimates of effects. In spite of the warnings that several authors have issued over the years, many researchers still propose conjectures based on geographic correlation studies that can cause biased conclusions.

Conclusion

The design and analysis of small-area epidemiological studies should include an appropriate selection of the territorial unit for analysis, a systematic validation of the sources of information, a painstaking analysis of specific rates and careful interpretation of the results. If we are to reach this goal, as in other areas of scientific inquiry, then a suitable level of training is required (Altman et al., 2002).

The quality of geographic studies which form the basis for devising health policies will play a vital role towards meeting these strategic goals. For this purpose, there must be a procedure that enables us to ascertain what contribution published studies can make to decision-making. In this context, the theory of diagnostic evaluation provides a suitable framework in which to develop standardised tests that would enable us to ascertain the level of excellence for each of these research studies (Wyatt-Smith and Cumming, 2009).

Following this method, Table 3 shows 12 questions to evaluate the scientific quality of a small-area epidemiology study. Each question may have more than one answer that the evaluator should mark on the second column. The third column shows the maximum score for each question (Hogan, 2005).

The overall score for the study will be derived from the sum of all the scores noted in the second column. *optimal* quality is achieved with the maximum score (16 points), *average* with a score between 8 and 15 points, and *low* with a score of 7 or under. Any study not achieving an optimal quality should be treated with caution, both in terms of the dissemination of results as well as when devising health plans.

The systematic application of this or other evaluation tools will contribute towards an enhanced quality of theoretical and applied research, and, consequently, to the advancement of scientific understanding in this field.

Table 3. Diagnostic evaluation test for small-area epidemiological studies.

Question	Score	Maximum
1. How is the geographic area under consideration defined in the study? (mark only one option)		1
a) It is the basic territorial unit that is to be described or where interventions are to be implemented	1	
b) It is a grouping of basic territorial units, with justification that the value of the rate is the same for all units	1	
c) It is a grouping of basic territorial units, but there has been no check on whether the value for the rate is the same for all units	0	
2. How is the time period under consideration defined? (mark only one option)		1
a) It is the smallest unit of time with available information	1	
b) Various time units have been grouped, with justification that the geographic distribution of the inidcator is the same for all units	1	
c) Various time units have been grouped, but the invariability of the values for the indicator throughout the time units has not been checked	0	
3. What actions have been taken to ensure the quality of the information sources? (mark one or more options)		3
a) None	0	
b) Check on whether the residence address for each case agrees with census records	1	
c) Check that the underlying disease or basic cause of death assigned to each case is correct	1	
d) Check there are no unregistered migrations in the official population data or correct for this effect in each area	1	

(continued)

Question	Score	Maximum
4. Is the raw rate analysed? (mark only one option)		1
a) No	1	
b) Yes, with justification there is no confounding factors for age or gender groups	1	
c) Yes, but there is no check on whether there is confusion for age or gender groups	0	
5. Is the adjusted rate analysed by the direct method? (mark only one option)		1
a) No	1	
b) Yes, with prior justification that the specific rate for each area is proportional to that in the remaining areas	1	
c) Yes, but there is no check on the proportionality between specific rates in all areas	0	
6. Is the Standardised Mortality Rate (SMR) or the Standardised Incidence Rate (SIR) considered? (mark only one option)		1
a) No	1	
b) Yes, with prior justification that the specific rate for each area is proportional to that of the outside population	1	
c) Yes, but no check on the proportionality between specific rates for each area and the outside population	0	
7. Is each specific rate for each population stratum analysed independently? (mark only one option)		1
a) No	0	
b) Yes	1	
8. Is the mortality rate analysed by causes? (mark only one option)		1
a) No	1	
b) Yes, using the prevalent population as the denominator	1	
c) Yes, using the whole population in the area as the denominator and with justification that all areas share the same disease prevalence	1	
d) Yes, using the whole population in the area as the denominator, but there is no check on whether all areas share the same prevalence	0	
9. Is there a criterion to define the excess or deficit of morbid-mortality in each area? (mark only one option)		1
a) No	0	
b) Yes, checking the rate for the area against the rate for the country or region the area belongs to	0	
c) Yes, comparing the rate for the area with a range of acceptable values set previously by experts	1	
10. Is the correlation between the morbid-mortality rate and other ecological indicators considered? (mark only one option)		1
a) No	1	
b) Yes, using the specific rate (or the adjusted rate, checking for the specific rates proportionality) as the dependent variable	1	
c) Yes, using SMR, SIR or a direct-method adjusted rate (without checking for the specific rates proportionality) as the dependent variable	0	
11. Are chloropeth maps interpreted? (mark one or more options)		2
a) No	0	
b) Yes, using the colour range from the legend	1	
c) Yes, using the interquartile range or another dispersion measure for the indicator	1	
12. What is mentioned in the discussion and conclusions sections of the study? (mark one or more options)		2
a) The limitations of the small-area design, including the ecological fallacy	1	
b) Caution on the interpretation of results	1	
c) An individually based research hypothesis that may account for differences between areas	0	
	Total (*)	16

(*) Sum of all scores marked in the columns. Quality of the study: low (0-7 points), average (8-15 points), high (16 points).

References

Altman DG, Goodman SN, Schroter S, 2002. How statistical expertise is used in medical research. JAMA 287, 2817-2820.

Armitage P, Berry G, Matthews JNS, 2002. Statistical methods in medical research (4th edition). Blackwell Scientific Oxford, UK.

Armstrong BG, 1995. Comparing standardized mortality ratios. Ann Epidemiol 5, 60-64.

Arts DGT, De Keizer NF, Scheffer GJ, 2002. Defining and improving data quality in medical registries: a literature review, case study, and generic framework. J Am Med Inform Assoc 9, 600-611.

Basskin L, 2003. Statistical interpretation can also bias research evidence. BMJ 327, 752.

Besag J, York J, Mollié A, 1991. Bayesian image restoration with applications in spatial statistics (with discussion). Ann I Stat Math 43, 1-59.

Best N, Richardson S, Thomson A, 2005. A comparison of Bayesian spatial models for disease mapping. Stat Methods Med Res 14, 35-59.

Boyle P, 2004. Population geography: migration and inequalities in mortality and morbidity. Prog Hum Geogr 28, 767-776.

Breslow NE, Clayton DG, 1993. Approximate inference in generalized linear mixed models. J Am Stat Assoc 88, 9-25.

Breslow NE, Lubin JH, Marek P, Langholz B, 1983. Multiplicative models and cohort analysis. J Am Stat Assoc 78, 1-12.

Buchan IE, 2000. StatsDirect-statistical software for medical research in the 21st century. BMJ 321, 7275.

Choi BCK, de Gguia NA, Walsh P, 1999. Look before you leap: stratify before you standardize. Am J Epidemiol 149, 1087-1096.

Ebdon D, 1985. Statistics in Geography (2nd Edition). Blackwell Scientific, Oxford, UK.

Editorial, 2004. Published research contains high level of statistical errors. Health Med Week 21, 887.

Elveback LR, 1966. Discussion of Indices of mortality and tests of their statistical significance. Hum Biol 38, 322-324.

EPHA, 2007. Diez afirmaciones sobre el futuro de la salud pública en Europa. Gac Sanit 21, 349-354.

Feller W, 1957. An Introduction to Probability Theory and its Applications, vol. I. Wiley, New York, USA.

Fernández C, Green PJ, 2002. Modelling spatially correlated data via mixtures: a Bayesian approach. J Roy Stat Soc B 64, 805-826.

Fleiss JL, Levin B, Paik MC, 2003. The standardization of rates. In: Fleiss JL, Levin B, Paik MC (eds). Statistical methods for rates and proportions (3rd Edition). John Wiley & Sons, New Jersey, USA, pp. 627-647.

Frenk J, 1997. La salud de la población: hacia una nueva salud pública. Fondo de Cultura Económica, México.

García E, Alcaraz C, 2004. Incongruence between test statistics and P values in medical papers. BMC Med Res Methodol 4, 13.

Gómez JA, Mateos A, Reolid M, Almar E, 1994. Padrón municipal: ¿Es útil como base poblacional de un estudio de prevalencia? Revista ROL de Enfermería 190, 85-88.

Greenland S, 1987. Bias in indirectly adjusted comparisons due to taking the total study population as the reference group. Stat Med 6, 193-195.

Greenland S, O'Rourke K, 2008. Meta-analysis. In: Rothman KJ, Greenland S, Lash TL (eds). Modern Epidemiology (3rd Edition). Lippincott Williams & Wilkins, Philadelphia, USA, pp. 665.

Greenland S, Rothman KJ, 2008. Measures of occurrence. In: Rothman KJ, Greenland S, Lash TL (eds). Modern Epidemiology (3rd Edition). Lippincott Williams & Wilkins, Philadelphia, USA, 32-50.

Gundersen L, 2000. Mapping it out: using atlases to detect patterns in health care, disease and mortality. Ann Intern Med 133, 161-164.

Hogan TP, 2005. Educational Assessment: a practical introduction. John Wiley & Sons, New York, USA.

Julious SA, Nicholl J, George S, 2001. Why do we continue to use standardized mortality ratios for small area comparisons? J Public Health Med 23, 40-46.

Lawson AB, 2006. Statistical methods in spatial epidemiology (2nd Edition). John Wiley & Sons, Chichester, USA.

Lawson AB, Browne WJ, Vidal Rodeiro CL, 2003. Disease mapping with WinBUGS and MlwiN. John Wiley & Sons, Chichester, USA.

Librero J, Benavides FG, Godoy C, 1993. Análisis de la mortalidad en áreas pequeñas: el problema de la residencia. Gac Sanit 7, 169-175.

Librero J, Benavides FG, 1995. La validez del municipio de

residencia en las estadísticas de mortalidad: hallazgos a partir de la actualización de los padrones municipales en dos municipios de la comunidad valenciana. Gac Sanit 9, 232-236.

MacEachren AM, 1995. How Maps Work: Representation, Visualization, and Design. The Guilford Press, New York, USA.

Monmonier M, 1996. How to lie with maps (2nd Edition). University of Chicago Press, London, UK.

Monmonier M, 2005. Lying with maps. Stat Science 20, 215-222.

Morgenstern H, 2008. Ecologic studies. In: Rothman KJ, Greenland S, Lash TL (eds). Modern Epidemiology (3rd Edition). Lippincott Williams & Wilkins, Philadelphia, USA.

Morin JJ, Tremblay D, 1981. The choice of health priorities: the use of standardized mortality rates. Cah Que Demogr 10, 467-472.

Moser K, Frost C, Leon DA, 2007. Comparing health inequalities across time and place-rate ratios and rate differences lead to different conclusions: analysis of cross-sectional data from 22 countries 1991-2001. Int J Epidemiol 36, 1285-1291.

Nicholls D, 1999. Statistics into the 21st century (with discussion). Aust & NZ J Stat 41, 127.

Ocaña-Riola R, 2007. The misuse of count data aggregated over time for disease mapping. Stat Med 26, 4489-4504.

Ocaña-Riola R, Fernández A, Mayoral JM, Toro S, Sánchez-Cantalejo C, 2009. Uncontrolled migrations as a cause of inequality in health and mortality in small area studies. Epidemiology 20, 411-418.

Ocaña-Riola R, Mayoral-Cortés JM, 2010. Spatio-temporal trends of mortality in small-areas of Southern Spain. BMC Public Health 10, 26.

Ocaña-Riola R, Mayoral JM, Sánchez-Cantalejo C, Toro S, Fernández A, Méndez C, 2008. Interactive mortality atlas in Andalusia, Spain (AIMA). Rev Esp Salud Pública 82, 379-394.

Phipps AI, Clarke CA, Ereman RR, 2005. Impact of intercensal population projection and error of closure on breast cancer surveillance: examples from 10 California counties. Breast Cancer Res 7, 655-660.

Pickle LW, Herrmann DJ, 1999. Cognitive research for the design of statistical rate maps. Proceedings of the Survey Research Methods Section (The American Statistical Association 1, 186-191.

Pickle LW, White AA, 1995. Effects of the choice of age-adjustment method on maps of death rates. Stat Med 14, 615-627.

Polissar L, 1980. The effect of migration on comparison of disease rates in geographic studies in the United States. Am J Epidemiol 111, 175-182.

Roos LL, 1993. Registries and administrative data: organization and accuracy. Med Care 31, 201-212.

Rothman KJ, 1990. A sobering start for the cluster busters' conference. Am J Epidemiol 132, S6-S13.

Rothman KJ, Greenland S, Lash TL, 2008. Modern Epidemiology (3rd Edition). Lippincott Williams & Wilkins, Philadelphia, USA.

Saavedra JM, Bello LM, Núñez D, Ortega P, Medrano MJ, 2001. Mortalidad por enfermedad isquémica del corazón en las Islas Canarias: Errores en la certificación de la variable residencia de los fallecidos en zonas turísticas. Boletín Epidemiológico 9, 161-168.

Salas J, Díez F, Puerta J, Zambrana JL, Delgado M, Rivera F, Gallego F, 2003. Tasa de empadronamiento real de los pacientes hospitalizados en el Hospital de Poniente almeriense. Gac Sanit 17, 312-315.

Slocum TA, McMaster RB, Kessler FC, Howard HH, 2009. Thematic Cartography and Geovisualization (3rd Edition). Pearson Prentice Hall, New Jersey, USA.

Smith D, Neutra RR, 1993. Approaches to disease cluster investigations in a state health department. Stat Med 12, 1757-1762.

Smith R, 2001. Medical editor lambasts journals and editors. BMJ 323, 651.

Sorlie PD, Thom TJ, Manolio T, Rosenberg HM, Anderson RN, Burke GL, 1999. Age-adjusted death rates: consequences of the year 2000 standard. Ann Epidemiol 9, 93-100.

Wang Q, Zhang B, 1998. Research design and statistical methods in Chinese medical journals. JAMA 280, 283-285.

Wakefield J, Shaddick G, 2006. Health-exposure modelling and the ecological fallacy. Biostatistics 7, 438-455.

Wilczyńska U, Szeszenia-Dabrowska N, Szymczak W, 1990. Effect of the selection of reference population on the evaluation of the degree of the risk of death among industrial cohorts. Med Pr 41, 175-180.

Williams AN, Johnson RA, Bender AP, 1990. Use of coded

mortality data to assess area cancer rates: impact of residence reporting and coding errors. Am J Epidemiol 132, 178-182.

Wyatt-Smith C, Cumming J, 2009. Educational assessment in the 21st Century: connecting theory and practice. Springer.

Wolfenden HH, 1923. On the methods of comparing the mortalities of two or more communities, and the standardization of death rates. J R Stat Soc 86, 399-411.

Woolsey TD, 1959. Adjusted death rates and other indices of mortality. In: Linder FE, Grove RD (eds). Vital statistics rates in the United Status, 1900-1940. Washington DC, US Government Printing Office.

Yule GU, 1934. On some points relating to vital statistics, more specially statistics of occupational mortality. J R Stat Soc 97, 1-84.

Zeger SL, Diggle PJ, Liang KY, 2004. A Cox model for biostatistics of the future. Department of Biostatistics working paper. Johns Hopkins University, Berkeley, USA.