

# Measures of health inequalities: part 2

Enrique Regidor

*J Epidemiol Community Health* 2004;**58**:900–903. doi: 10.1136/jech.2004.023036

This is the second part of a two part glossary on measures of health inequalities.

## MEASURES OF ASSOCIATION

The first part of this overview showed measures based on frequency ratios. This second part presents three other measures of socioeconomic inequality based on associations.

### Absolute difference in frequencies

This measure can be calculated as the difference between the observed frequency of the health event in each category of the socioeconomic variable and the reference category, based on data in simple contingency tables, or by using the estimated frequency based on log-linear regression models. As is the case with frequency ratios, when we obtain a single estimate it can be called a summary measure of socioeconomic inequality in health. It has the same advantages and limitations noted for the frequency ratios in comparing and monitoring socioeconomic inequality in health.

An additional consideration is that the size of the absolute difference may vary even when the relative difference remains constant. Or the absolute difference may decrease, as frequently occurs when the frequency of the health event decreases and, at the same time, the relative difference may increase (table 1). There is some evidence showing that ranking countries based on the size of the socioeconomic inequality in health may vary depending on whether absolute or relative differences are used.<sup>1 2</sup> However, there is no clear criterion as to whether a relative or absolute difference in frequencies is more appropriate to reflect the size of socioeconomic inequality in health. A frequency ratio of 1.5 between two categories of a socioeconomic variable may be very important if the frequency of the health problem is 20% and 10%, but much less so if the frequency is 2% and 1%, respectively. Some authors support the use of the absolute difference to evaluate the effect of public policies on health inequality as their objective is to reduce the number of cases of a health problem. However, relative differences are also appropriate to evaluate the strength of the relation between an intervention and the reduced frequency of the health problem. It may be that the measures of association used to make comparisons of socioeconomic inequality in health should include both estimates: the frequency ratio and the difference in frequencies.

### Regression coefficient

The regression coefficient represents the increase (or decrease) in the absolute magnitude of the dependent variable for each unit of increase in the socioeconomic variable. For example, in the equation  $Y = 39 - 2.1X$ , where  $Y$  represents the body mass index (BMI) in  $\text{kg/m}^2$ , and  $X$  represents monthly income in thousands of euros, the regression coefficient shows that BMI decreases by 2.1  $\text{kg/m}^2$  for each additional thousand euros of income. In calculating this statistic, both the dependent variable representing the health event and the independent variable representing the socioeconomic characteristic should be measured on an interval scale. Sometimes, however, it has been used with social class defined on an ordinal scale.<sup>3–6</sup> The regression coefficient can be transformed into a measure of relative difference. It is necessary to previously perform a log transformation of the dependent variable. In this case, the exponent of the regression coefficient minus 1 represents the proportion (or percentage if it is multiplied by 100) of increase in the dependent variable for each unit of increase in the socioeconomic variable.<sup>6</sup> The regression coefficient is a summary measure of health inequality as it is a single estimate and, therefore, can easily be used to compare various populations. Its disadvantage is that it is not appropriate when the adjustment of the regression function shows deviations from linearity.

### Pearson's correlation coefficient

This coefficient measures the degree of linear relation between a socioeconomic characteristic and the health event when the two variables are measured on an interval scale. It ranges in magnitude from  $-1$  to  $+1$ . The closer the data are to a straight line, the larger the absolute value of the correlation coefficient. This coefficient is very sensitive to the variation of each variable, therefore it is not an appropriate measure to estimate the relation between two variables when the number of observations is large, as it may be small in size even when the regression coefficient is important. This occurs most frequently with individual observations, therefore it is more appropriately used for group observations where the sample size is usually smaller.

## MEASURE OF POTENTIAL IMPACT

### Population attributable proportion

This measure represents the proportional decrease that would occur in the frequency of

**Abbreviations:** BMI, body mass index; RII, relative index of inequality; SII, slope index of inequality

Correspondence to:  
Dr E Regidor, Department  
of Preventive Medicine and  
Public Health, Faculty of  
Medicine, Universidad  
Complutense de Madrid,  
Ciudad Universitaria s/n,  
28040 Madrid, Spain;  
enriqueregidor@hotmail.  
com

Accepted for publication  
25 March 2004

the health problem in a population in the hypothetical case that the frequency of the health problem in all individuals was the same as for individuals in the highest socioeconomic category. The frequency of the health problem in individuals in the highest socioeconomic category is assumed to be lower than in the rest of the subjects in the population. It is calculated as the difference between the frequency of the health problem in the population and the frequency of the health problem in individuals in the highest socioeconomic category, expressed as a proportion or percentage of the frequency of the health problem in the population (table 1). There is another way to estimate this measure, which involves the frequency ratio of each socioeconomic category and the percentage of the total population represented by each category.<sup>7, 8</sup> The result obtained, in any case, is the same.

The socioeconomic variable can be dichotomic, polytomic, or measured on an interval scale. In the third case it is necessary to establish the value of the socioeconomic variable that will serve as the reference category and to use regression models to estimate the frequency of the health problem in subjects with this value. Although this measure has traditionally been used with the dependent variable measured as a binary variable, it could equally well be used with the dependent variable measured on an interval scale. For example, we could determine the proportional decrease in mean BMI in the population, in the hypothetical situation that all individuals had the same BMI as those in the highest socioeconomic category.

The population attributable proportion is a function of two types of information: (a) the association between the socioeconomic variable and the frequency of the health problem, and (b) the distribution of subjects across each category of the socioeconomic variable. The larger the association between the socioeconomic variable and the health problem and/or the larger the variation in the distribution of the socioeconomic variable, the larger the magnitude of the population attributable proportion. It is the measure of choice when the objective is to reduce the impact of socioeconomic circumstances on the burden of the health problem in the population. Thus, given that we often do not know the mechanism mediating the relation between socioeconomic circumstances and health, modifying the distribution of the population in the different socioeconomic

categories could become the objective of the policy intervention. For example, if the frequency ratio is 2, the population attributable proportion is 0.47 if the individuals in the socioeconomic category with the lowest frequency of the health problem represent 10% of the population, but drops to 0.39 if these individuals represent 35% of the population.

This is a summary measure of health inequality because it gives a single estimate. It has the advantage that its calculation implicitly takes into account the whole range of values of the socioeconomic variable and the population distribution across the different socioeconomic categories. The main disadvantage in comparative studies is that it requires that the reference category be similar in all the populations being compared and that its size should represent the same percentage of individuals. This is not always easy to achieve because the definition of the variable can vary across populations and, even when comparing the same population over time, the percentage of the population represented by the reference socioeconomic category usually increases.

## INDICES BASED ON RANKING OF THE SOCIOECONOMIC VARIABLE

### Concentration index

This measure of socioeconomic inequality was proposed by Wagstaff *et al.*<sup>9</sup> The value of the health variable assigned to each individual is a function of the socioeconomic category to which the individual belongs. This index is based on what these authors call the "concentration curve," where the x axis represents the cumulative proportion of individuals by socioeconomic level, beginning with those who have the lowest socioeconomic level and ending with those whose level is highest, while the y axis represents the cumulative total proportion of health in these individuals. The value of the concentration index ranges from -1 to +1. Although the concentration curve sounds like the Lorenz curve, statistically speaking, this index is not a measure of inequality in the strict sense, because individuals are ranked, not by the magnitude of the health variable, but by socioeconomic level.

If the concentration curve coincides with the diagonal, all individuals have the same level of health. If the curve is under the diagonal, this means that health is concentrated in persons of higher socioeconomic level, and if the curve is

**Table 1** Fictitious example of trend in mortality inequalities by educational level between 1990 and 2000

(1) Basic data								
Educational level	Number of people		Population share		Hierarchical rank		Mortality rate (per 100000 people)	
	1990	2000	1990	2000	1990	2000	1990	2000
Third level	5000	7500	0.10	0.15	0.05	0.08	80	40
Secondary, upper level	10000	12500	0.20	0.25	0.20	0.28	110	58
Secondary, lower level	12500	12500	0.25	0.25	0.43	0.53	128	80
First level	22500	17500	0.45	0.35	0.78	0.83	160	91
Total	50000	50000					134	72
(2) Summary measures of mortality inequalities by educational level								
Rate ratio*							1.48	1.73
Difference ratio (per 100000 people)*							48.6	36.7
Population attributable proportion (%)†							40.3	44.4
Concentration index							-0.10	-0.14
Slope index of inequality (per 100000 people)‡							104.4	70.4
Relative index of inequality§							2.40	3.03
Relative index of inequality (%)¶							140	203

\*First level and secondary, lower level compared with secondary, upper level, and third level. †Third level as reference category. ‡Predicted rate difference between those at the bottom and the top of the educational hierarchy. §Predicted rate ratio for bottom compared with top of educational hierarchy. ¶Percentage of predicted rate for bottom with relation to predicted rate for top of educational hierarchy.

above the diagonal, it means that health is concentrated in those with a lower socioeconomic level. The farther the curve is from the diagonal, the greater the degree of health inequality: the first case is known as health inequality in favour of individuals with higher socioeconomic level and the measure has a positive value, while the second case is known as health inequality in favour of individuals of lower socioeconomic level and the value of the measure is negative.<sup>10</sup> If all health is concentrated in the person with the highest socioeconomic level, the index will have a value of +1, and if all health is concentrated in the individual with the lowest socioeconomic level, the index will have a value of -1.

This index incorporates the socioeconomic dimension in the estimation of health inequality. All individuals in the population are included in its calculation, and it is sensitive to changes in the distribution of the population across the different socioeconomic categories. Because individuals are ordered by socioeconomic level, it does not have the disadvantage of the Gini index, as the size and sign of the concentration index depend on the gradient observed between socioeconomic level and health. This makes it possible to compare socioeconomic inequality in health over time and among different places. If the observations are ordered in the same way whether they are ranked by the magnitude of the health variable or by socioeconomic level, the concentration index and the Gini index will have the same value. Its disadvantage is that it can be applied only in those cases in which the socioeconomic categories can be ordered in accordance with a strict hierarchical ranking.

#### The slope and relative indices of inequality

The slope index of inequality (SII) represents the linear regression coefficient that shows the relation between the level of health or the frequency of a health problem in each socioeconomic category and the hierarchical ranking of each socioeconomic category on the social scale.<sup>11</sup> For this purpose, a variable is created from a series of values assigned to the different socioeconomic categories with reference to a range. For example, if the socioeconomic variable is educational level, and the category with the highest educational level includes 10% of the population, the range of the individuals in this category would be from 0 to 0.10, giving a mean of 0.05, which would be the value assigned to this category; if the next highest educational level category includes 20% of the population, its range is from 10% to 30%, thus it would be assigned a value of 0.20, and so on.

With this index, the hierarchical ranking in any population studied will have the same amplitude: the highest level has a value of 0, and the lowest level has a value of 1. The SII can be interpreted as the absolute change in health level or in the frequency of a health problem when one goes from the highest level in the social hierarchy (range = 0) to the lowest level (range = 1). The SII reflects the experience of all individuals in the population and is sensitive to changes in the distribution of the population among the different socioeconomic categories. Its disadvantage is that it can only be applied to socioeconomic variables which can be ordered hierarchically. In addition, the regression estimate has not to show significant deviations from linearity; otherwise, the magnitude of the index would be biased.

Because this is an absolute measure, it is sensitive to changes in the mean level of population health or changes in the frequency of the health problem being studied. If the mean level of health increases in the same proportion in all the socioeconomic categories, the SII will increase, whereas the relative differences remain constant. This limits, for example, comparisons of trends in socioeconomic inequality in a health problem across different populations if the

frequency of the problem has been reduced more in some populations than in others. One alternative that has been proposed is the relative index of inequality (RII), which can be estimated in two ways: one way is to divide the SII by the mean level of population health or by the frequency of the health problem in the population<sup>11</sup>; the other way is to divide the predicted value of the regression at the highest point (range = 1) by the predicted value of the regression at the lowest point (range = 0).<sup>12 13</sup>

The second method noted for the RII is quite frequently calculated by log-linear—or logistic—regression after the logarithmic—or logit—transformation of the dependent variable.<sup>14 15</sup> In this case the exponent of the regression coefficient represents the RII, which is merely the frequency—or the odds—predicted at the lowest point of the social hierarchy divided by the frequency—or the odds—predicted at the highest point of the social hierarchy. This is the most frequent way of presenting the RII. However, this can raise difficulties for persons not accustomed to using this index, because it can be interpreted as a measure of association, either a frequency ratio or an odds ratio. One way to facilitate the interpretation of this measure may be to express the RII as a percentage by subtracting 1 from it and multiplying the result by 100 (table 1).

The same as occurs with measures of impact, a larger RII in one population than in another may be attributable to a larger socioeconomic gradient in health and/or to a larger inequality in the way individuals are distributed across the different socioeconomic categories.

## CONCLUSIONS

### Individuals or areas

The choice of the unit of observation and/or the definition of the socioeconomic variable depends on the objective to be attained. For example, the evaluation of interventions aimed at individuals to reduce or socioeconomic inequalities in health requires individual observations, whereas interventions focusing on whole areas—from neighbourhoods to regions or provinces—require group observations or individual observations with group variables.

### Health inequality or socioeconomic inequality in health

When the objective is to measure health inequality, it is necessary to use univariate measures of inequality in the distribution of health: Gini index or index of dissimilarity. But if the objective is to estimate socioeconomic inequality in health, there are two options. The first is to incorporate the socioeconomic dimension in the previously mentioned measures. The problem with these measures is that they may give similar results even when the relation between health and socioeconomic status is different. The second option is to use the other three types of measures mentioned: association, potential impact, or based on the ranking of the socioeconomic variable. In this case, there is no unanimously accepted criterion about which measure is the most appropriate. A limitation of most of these measures is that they can only be used to reflect socioeconomic inequalities in health when the socioeconomic variable is ranked hierarchically.

### Relative and absolute differences

From the point of view of monitoring health inequalities and evaluating policy interventions, it is very important to estimate both relative and absolute differences, as relative differences may increase while absolute differences decrease if the frequency of the health problem declines. What is not appropriate is to use absolute differences for some health events and relative differences for others, as has been done in some studies,<sup>3</sup> because the results obtained cannot be

compared. Any of the summary measures shown in table 1 is appropriate, including the range, if there is a linear gradient. If not, the estimates of the differences in each category of the socioeconomic variable should be shown.

### Possibility of modifying the distribution of the population across socioeconomic categories through public policies

The measures of impact and the indices based on ranking of the socioeconomic variables are the most suitable for evaluating those policies. The latter measures require a linear gradient in the estimated association. The most commonly used one is the RII in the form of a ratio, although it may be more appropriate to express it as a percentage to avoid its possible confusion with a frequency ratio. This index has sometimes been used incorrectly to compare the strength of the association between two socioeconomic variables and different health measures.<sup>16, 17</sup> In this case, the interpretation of the findings could be biased, as the association could be similar for the two socioeconomic variables, but the estimated RII could be different if the population is distributed differently across the various socioeconomic categories.

To sum up, although in the final analysis ethical and political considerations will determine the importance of the health inequalities measured at any given moment, the challenge is to provide the fewest possible number of estimates, which will nevertheless permit a complete and accurate interpretation of the data.

### REFERENCES

- 1 **Vagero D**, Erikson R. Socioeconomic inequalities in morbidity and mortality in western Europe. *Lancet* 1997;**350**:516.
- 2 **Mackenbach JP**, Cavelaars AEJM, Kunst AE, *et al*. Socioeconomic inequalities in morbidity and mortality in western Europe. *Lancet* 1997;**350**:517.
- 3 **Ferrie JE**, Shipley MJ, Davey Smith G, *et al*. Change in health inequalities among British civil servants: the Whitehall study. *J Epidemiol Community Health* 2002;**56**:922–6.
- 4 **Blane D**, Hart CL, Davey Smith G, *et al*. Hawthorne VM. Association of cardiovascular disease risk factors with socioeconomic position during childhood and during adulthood. *BMJ* 1996;**313**:1434–8.
- 5 **Lawlor DA**, Ebrahim S, Davey Smith G. Socioeconomic position in childhood and adulthood and insulin resistance: cross sectional survey using data from British women's heart and health study. *BMJ* 2002;**325**:805–9.
- 6 **Brunner E**, Davey Smith G, *et al*. Childhood social circumstances and psychological and behavioural factors as determinants of plasma fibrinogen. *Lancet* 1996;**347**:1008–13.
- 7 **Hahn RA**, Eaker E, Barker D, *et al*. Poverty and death in the United States—1973 and 1991. *Epidemiology* 1995;**6**:490–7.
- 8 **Gutiérrez-Fisac JL**, Regidor E, Banegas JR, *et al*. The size of obesity differences associated with educational level in Spain. *J Epidemiol Community Health* 2002;**56**:457–60.
- 9 **Wagstaff A**, Paci P, van Doorslaer E. On the measurement of inequalities in health. *Soc Sci Med* 1991;**33**:545–57.
- 10 **Wagstaff A**. Socioeconomic inequalities in child mortality: comparisons across nine developing countries. *Bull World Health Organ* 2000;**78**:19–29.
- 11 **Pamuk ER**. Social class inequality in mortality from 1921 to 1972 in England and Wales. *Population Studies* 1985;**39**:17–31.
- 12 **Mackenbach JP**, Kunst AE. Measuring the magnitude of socio-economic inequalities in health: an overview of available measures illustrated with two examples from Europe. *Soc Sci Med* 1997;**44**:757–71.
- 13 **Mille Schalick L**, Hadden WC, Pamuk E, *et al*. The widening gap in death rates among income groups in the United States from 1967 to 1986. *Int J Health Serv* 2000;**30**:13–26.
- 14 **Kunst AE**, Groenhouf F, Andersen A, *et al*. Occupational class and ischemic heart disease mortality in the United States and 11 European countries. *Am J Public Health* 1999;**89**:47–53.
- 15 **Dalstra JAA**, Kunst AE, Geurts JJM, *et al*. Trends in socioeconomic health inequalities in the Netherlands, 1981–1999. *J Epidemiol Community Health* 2002;**56**:927–34.
- 16 **Davey Smith G**, Hart C, Hole D, *et al*. Education and occupational social class: which is the more important indicator of mortality risk? *J Epidemiol Community Health* 1998;**52**:153–60.
- 17 **Brunner E**, Shipley MJ, Blane D, *et al*. When does cardiovascular risk start? Past and present socioeconomic circumstances and risk factors in adulthood. *J Epidemiol Community Health* 1999;**53**:757–64.