Deriving relative risks from aggregate data. 2. An application to the relationship between unemployment and suicide

THOR NORSTRÖM

From the Swedish Institute for Social Research, University of Stockholm, S-10691 Stockholm, Sweden

SUMMARY   In Part 1 of this study, the mathematical relations between micro and macro measures of effect were derived. The formulas thus obtained can be used for cross level comparisons of findings concerning the effect of some specific risk factor on, eg, mortality. The approach is illustrated by means of an empirical example relating to the association between unemployment and suicide. This relationship is estimated on micro data as well as on aggregate time series data. The findings from the two levels are fairly consistent.

In Part 1 of this study, I pointed out the desirability of comparing findings from micro and macro studies on the relationship between mortality and risk factors. Such comparisons are hampered by the use of different effect measures, and it is thus necessary first to analyse how they relate to each other. The mathematical relations between micro measures (relative risk and attributable fraction) and the preferable macro measure (the unstandardised regression coefficient) were derived. In this paper, the relationship between unemployment and suicide is estimated on prospective micro data and on aggregate time series data. The micro and macro estimates are then compared by applying the formulas derived in Part 1. (The numbering of equations and formulas begins in Part 1).

Unemployment and suicide—previous studies

I have located three micro investigations of principally the same design as that employed below, ie, prospective studies. Two of these are based on census data from England and Wales, [1] [2] and the third on Danish census data.[3] All three studies report an elevated suicide risk among the unemployed. The same holds true for a large number of cross sectional micro studies as well[4] but, compared with the prospective studies, they are even more impaired by the problem of self selection. It is quite probable that personality traits such as mental illness are related to an excess suicide risk, as well as to an increased likelihood of becoming unemployed.

Based on Swedish panel data for the seventies, a study by Björklund[5] indeed lends support to the hypothesis that the unemployed are selective with regard to characteristics conducive to suicide. A correlation between symptoms of mental illness and unemployment was found in a cross sectional analysis of the data. However, this association vanished when the data were analysed longitudinally, thereby controlling for fixed effects (selection effects).

Significant relationships between unemployment and suicide have also been found in several time series analyses of aggregate data.[6–9] Such results, and in particular those reported by Brenner,[7] have evoked much controversy,[10] [11] for example with regard to the model and lag structure specification. Undoubtedly, much of the work carried out in this tradition is flawed by the application of inefficient estimation procedures, the failure to detrend the data, and somewhat dubious lag specifications.

The prospective study

The source of data for the individual level analysis is the Swedish registry on causes of death in 1961–70 linked to the census of 1960. The census comprised the entire population and contains information on employment status. The present analysis focuses on mortality from suicide (E950–959 in the 1965 International Classification of Diseases, 8th revision) among men 20–69 who were unemployed at the time of the census and had been so for a period exceeding 4 months (n = 4562). The absolute number of suicides

in this category was 25 during the follow up period 1961–65. The estimate of the age standardised relative risk, using the remaining male population aged 20–69 as reference category, is 3·04 (95% confidence interval 2·05 to 4·51). This estimate is very close to that reported by Iversen *et al.*[3] For the period 1963–68 (which is focused upon in a subsequent section) the associated attributable fraction is 0·028 (SE = 0·007). The period average of the fraction unemployed (males 20–69) is used as an estimate of θ in formula (2) (paper 1).

An apparent weakness of the data is that there is no information regarding unemployment during the follow up period. The reason for focusing upon the long term unemployed is that we can expect this category to have a substantially greater risk of experiencing unemployment after 1960, compared to other people. The price we have to pay, however, is that the selection problem is likely to be aggravated. Fox *et al*[12] have suggested the following method for addressing the selection problem in mortality studies of the present kind (adopted by Iversen *et al*[3]). The relative risk for two successive follow up periods, of say 5 years, are compared. The selection hypothesis would predict an increased risk in the first period which is nullified or much reduced in the second. If the elevated risk by and large persists in the second period, factors other than health selection should also be operating; selection is not expected to affect mortality after such a long period of time. The logic of this approach does not seem entirely convincing, mainly because the exposure, for example the excess risk of unemployment which is assumed subsequent to the time of measurement, should level off with the passage of time. In the present context there is yet another circumstance: suicidogenic characteristics (eg, depression) that operate as selection mechanisms are probably more enduring than health selection mechanisms in general.

However, it is still worth reporting the estimated relative risk for the follow up period 1966–70. Based on 23 cases, it is 2·41 (95% confidence interval 1·60–3·63). What bearing this has on the problem of selection is, however, an open question.

The aggregate time series analysis

In this section the relationship between unemployment and suicide is estimated on Swedish annual data for the period 1920–68. Previous analyses of these data suggest that, in addition to unemployment, per capita alcohol consumption (a proxy for the prevalence of alcohol abuse) is associated with suicide, while other plausible predictors, such as divorce and various indicators of economic growth are insignificant.[9]

Method

We apply the method for time series analysis that has been developed by Box and Jenkins.[13] A feature of this method which greatly reduces the risk of omitted variable bias is the removal of linear trends by means of differencing the data. The analysis is thereby performed on the yearly changes rather than on the raw data. The rationale of this procedure is that an omitted predictor is more likely to be correlated with an included predictor due to coinciding or diverging trends, than due to synchronisation in the yearly changes. Another feature of the Box-Jenkins method is that no assumption of independent error terms is necessary. The noise term (explanatory variables not included in the model) is allowed to have a temporal structure, which is modelled and estimated in terms of autoregressive or moving average parameters. The method lends itself to investigations of lag structures by inspection of the cross correlations. The cross correlation is approximately the product-moment correlation between two time series. It can be calculated for different lags of the input series (the independent variable), and is usually denoted CCF(k), where k indicates the lag. CCF(0) is thus the concomitant correlation between the two series; CCF(1) means that the input series is lagged one observation, which can be extended to higher lags. The presence of significant cross correlations at higher lags reveals the existence of a lag structure. However, before examining the cross correlations, the series need to be transformed by a filter which removes the structure (the autocorrelation) of the series (prewhitening), which otherwise would confound the coefficients.[13]

Results

Inspection of the cross correlations between the prewhitened unemployment and suicide series revealed a significant coefficient (0·27, SE = 0·14) at lag 0, while coefficients at higher lags were well below significance. Consequently, there is no indication of any lag structure, which may appear somewhat surprising. The evidence reported by Eyer[14] also suggest a quick response to job loss. A feasible interpretation here is that the unemployed who commit suicide are already at a high risk prior to losing their jobs. The unemployment experience then acts as a trigger. It is also reasonable to assume that indirect effects, resulting from, for example, fear of being laid off, are quite rapid.

We specify an additive model to simplify the ensuing exercises, although these are manageable with multiplicative models as well. There is reason to believe that the relationship between alcohol

consumption and suicide is convex,[15] while that between unemployment and suicide is more likely to be concave: the higher the unemployment rate, the more probable is it that "ordinary" people will experience unemployment, which should dampen the aggregate suicide response. On these premises, the following model was estimated (with alcohol consumption squared and unemployment logged to accommodate the above non-linearities):

$$\nabla S_t = b_1 \nabla A_t^2 + b_2 \nabla LnU_t + \nabla N_t \qquad (9)$$

where S is the age standardised suicide rate (males 20–69), A is alcohol consumption (litres per capita 15+), U is the unemployment rate, and N is the noise term. The operator ∇ denotes that the data are differenced (eg, $\nabla S_t = S_t - S_{t-1}$).

The estimates are displayed in the table. To check the stability of the estimated unemployment effect, the study period 1920–68 was also divided into two subperiods which are partly overlapping in order to secure a sufficient number of observations for estimation (1920–45, 1940–68). As can be seen, the unemployment parameter is significant in both of the subperiods, and approximately of the same magnitude. (For an explanation of the difference in the alcohol parameter between the two periods, see Norström.[15])

The χ^2 distributed diagnostic tests are satisfactory. None of the models have indications of autocorrelated residuals or significant cross correlations between residuals and input series.

Estimation of relative risk from the macro estimates

On the basis of the finding from the time series analysis, we will now estimate the relative risk of unemployment with respect to suicide. To this end, we require age specific unemployment data which are available from 1963 onwards. All calculations will thus be performed on the basis of figures for the period 1963–68. First the attributable fraction is calculated.

The period-average (1963–68) of the unemployment series which was used in the previous analysis is equal to 1·417%, and the average suicide rate, 39·77 per 100 000 (males 20–69). Using the unemployment parameter of 4·78 obtained from (9), and applying formula (6) we have:

$$AF = \frac{4 \cdot 78 * Ln1 \cdot 417}{39 \cdot 77} = 0 \cdot 042,$$

with an approximate standard error of 0·008 (based on $SE(\hat{b}_1)$). That is, if unemployment were removed, the suicide rate is expected to decrease by about 4%.

For the calculation of the relative risk (RR), we need an estimate of θ (the proportion unemployed among males 20–69, 1963–68). From external sources, we obtain $\hat{\theta}$ equal to 0·012. Using the above estimate of AF and applying formula (2), after rearranging we get:

$$RR = 1 + \frac{0 \cdot 042}{(1 - 0 \cdot 042)\ 0 \cdot 012} = 4 \cdot 65$$

Taking into account the standard errors of $\hat{\theta}$ and $\hat{b}_2$, the approximate 95% confidence interval is 3·16 to 6·18.

Comparison of the micro and macro estimates

Though the micro and macro estimates of the relative suicide risk associated with unemployment do not differ too much in magnitude, they differ with respect to potential sources of bias and interpretation. (For convenience, we term the micro estimates RR_{micro} and AF_{micro}, and the macro estimates RR_{macro} and AF_{macro}.) The most likely bias of RR_{micro} is selection effects, the magnitude of which are very difficult to assess—to some degree this bias should be counteracted by decreased exposure during the follow up period. The other measure, RR_{macro}, might suffer from omitted variable bias. The risk of this should be small in the present case, however. First, it is

Table *Estimates of time series model for suicide (equation 9). Estimated on differenced data for the periods 1920–68, 1920–45 and 1940–68. Q = test for residual autocorrelation; Sq = squared; Ln = Log (natural); $S_1 - S_2$ = test for cross correlations between residuals and prewhitened input (in parentheses)*

	1920–1968	1920–1945	1940–1968
Sq. alcohol consumption	0·44(0·08)‡	0·39(0·13)†	0·55(0·14)‡
Ln unemployment	4·78(0·98)‡	5·15(1·33)‡	4·42(2·01)*
Noise			
AR(1)	−0·51(0·15)	−0·51(0·21)	−0·32(0·21)
AR(2)	−0·49(0·15)	−0·42(0·21)	−0·28(0·21)
AR(3)	−0·26(0·15)		
Diagnostics			
Q	13·86, 21df, p>0·88	18·23, 22df, p>0·69	19·26, 22df, p>0·63
S_1 (alcohol consumption)	13·50, 23df, p>0·94	10·17, 23df, p>0·99	4·17, 23df, p>0·99
S_2 (unemployment)	21·01, 23df, p>0·58	16·18, 23df, p>0·85	10·56, 23df, p>0·98

* p<0·05; † p<0·01; ‡ p<0·001.

minimised by means of the differencing procedure. Secondly, and more important, it has been shown elsewhere,[16] that the included predictors (unemployment and alcohol consumption) seem to account for the systematic part of the temporal variation in suicide.

The interpretation of RR_{micro} is, of course, the excess suicide mortality of unemployed, ie, the direct effect of unemployment. In addition to the direct effect, RR_{macro} also captures indirect effects, eg, suicides triggered by the fear of losing one's job. Strictly, this renders the term "relative risk" a somewhat alien connotation, and it may be more appropriate to compare the other measures of impact, AF_{micro} and AF_{macro}. It may be recalled that AF_{micro} is equal to 0·028, and AF_{macro}, 0·042. The same difference in import applies here too: AF_{micro} includes direct effects only, and AF_{macro} comprises both direct and indirect effects. Taken at face value, these figures would indicate that about one third of the total impact of unemployment on suicide is indirect. However, we should not attach too much importance to the difference between the two estimates; it is not statistically significant (t = 1·32), and the direction of possible bias is not ascertained. A sounder statement would be that two quite different methods, each one with its own weaknesses, lead to very similar conclusions about the effect of unemployment on suicide.

Discussion

In Part 1 of this study, I derived the mathematical relations between the relative risk and the attributable fraction on the one hand, and the unstandardised regression coefficient on the other. By way of an empirical example, it was demonstrated in Part 2 how these relations can be utilised to broaden the empirical basis of the effect of a risk factor. It is somewhat surprising that these relations are not treated in epidemiological text books, or applied in empirical research. One reason might be that epidemiologists and sociologists are so preoccupied with their respective traditional methods. The only papers that present a somewhat related approach are those by Beral et al[17] and Morgenstern.[18] The micro-macro relation they rely upon is that between the relative risk on the one hand, and the parameters in an ecological regression model on the other: the intercept (a) and the unstandardised regression coefficient (b):

$$RR = 1 + \frac{b}{a} \qquad (10)$$

What makes this formula less feasible in most contexts is the use of the intercept (a). Since the estimate of the intercept depends on how many potential predictors

are included, the omission of *any* relevant predictor introduces bias in the estimate of RR. In contrast, the method I have outlined involves omitted variable bias only when a predictor is left out that in differenced form is correlated with both X and Y (in differenced forms). Further, ecological analysis carries a greater potential for bias in the estimation of variables than does the Box-Jenkins method for time series analysis.

Finally, we turn to some of the pros and cons of deriving risk measures from time series estimates, starting with the two main problems. First, time series estimates are often sensitive to misspecifications. Although the risk of omitted variable bias is minimised by the Box-Jenkins technique, it is not eliminated. In addition, there are other potential specification errors related, for example, to lag structure and functional form. Secondly, the requirement for long time series—at least 30 observations are desirable—makes the assumption of a time invariant association between X and Y vulnerable. The more physiological the character of the association, the more plausible this assumption should be. One way of checking the stability of the parameter estimate is to estimate the model for subperiods, although this increases the necessary length of the series.

Among the merits of the method, we may first note that it is inexpensive compared to observational studies. The fact that the risk measures derived from time series estimates are expected to be uncontaminated by selection effects is of course an attractive feature. Their capturing of indirect effects may in some situations be a disadvantage, but given the overall purpose of assessing the impact of a risk factor on a specific cause of death, this is a very essential merit of the approach. It would be most ideal, of course, if we could partition the total impact into direct and indirect effects. This implies a comparison between AF_{micro} and AF_{macro}; the direct effect would be indicated by the first measure, and the indirect effect by the difference between the two. It is quite likely, however, that most attempts along this line will fail to provide any trustworthy assessments. The typical situation is probably instanced by our attempt to separate the direct and indirect effects of unemployment on suicide, where the fairly large standard errors of AF_{micro} and AF_{macro} precluded any definite inferences. One way to get further is to perform replications on data for other countries, and to check whether there is any consistency in the proportion of direct and indirect effects.

What weight should be attached to an estimated RR_{macro} compared to an estimated RR_{micro}? It is often contended that a "large" relative risk based on observational data is likely to reflect a genuine excess risk.[19] Such a general rule can hardly be claimed for

risk estimates derived from aggregate time series data. The reliability of the findings are too dependent on how carefully the analysis is carried out, and how well the diagnostic checking is done. For instance, the failure to check for outliers in the bivariate plots between Y and the Xs may lead to the reporting of an association which is in fact artefactual. Taken separately, the RR_{micro} thus carries greater weight than does the RR_{macro}. It seems reasonable, though, that two micro estimates indicating an elevated risk should be less compelling than if a micro as well as a macro estimate did so. The rationale of this argument is that estimates from two different levels of aggregation are much less likely to be impaired by a common source of bias.

Another context in which the derived relative risk (or attributable fraction) can be employed is when assessing the plausibility of an estimated macro parameter. Consider, for example, the estimate obtained by Gove and Hughes[20] of the relationship between living alone and mortality from liver cirrhosis. Is the (unstandardised) regression coefficient—estimated from ecological data—of 0·223 plausible? If we calculate the attributable fraction by applying expression (6) to the figures given in the article, we get an AF equal to 0·77; that is, about 77% of the cirrhosis mortality would be attributable to living alone. The incredibility of this figure indicates that the regression coefficient is seriously biased. Gove and Hughes also estimate the effect of living alone on suicide. The AF implied by their estimate is 0·89, which barely adds any confidence to their analyses.

If researchers to a higher degree made such simple checks of their results instead of regarding a significant coefficient as a reward for hard work, we would have fewer obscure findings in our journals.

References

1 Moser KA, Jones DR, Fox AJ, Goldblatt PO. Unemployment and mortality: further evidence from the OPCS longitudinal study 1971–81. *Lancet* 1986; i: 365–7.

2 Moser KA, Goldblatt PO, Fox AJ, Jones DR. Unemployment and mortality: comparison of the 1971 and 1981 longitudinal study census samples. *Br Med J* 1987; **249:** 86–90.

3 Iversen L, Andersen O, Andersen PK, Christoffersen K, Keiding N. Unemployment and mortality in Denmark, 1970–80. *Br Med J* 1987; **295:** 879–84.

4 Platt S. Unemployment and suicidal behaviour: a review of the literature. *Soc Sci Med* 1984; **19:** 93–115.

5 Björklund A. Unemployment and mental health: some evidence from panel data. *J Hum Resour* 1985; **20:** 469–83.

6 Vigderhous G, Fishman G. The impact of unemployment and familial integration on changing suicide rates in the USA, 1920–1969. *Soc Psychiatry* 1978; **13:** 239–48.

7 Brenner MH. Mortality and the national economy: a review, and the experience of England and Wales, 1936–76. *Lancet* 1979; ii: 568–73.

8 Marshall JR, Hodge RW. Durkeim and Pierce on suicide and economic change. *Soc Sci Res* 1981; **10:** 101–14.

9 Norström T. A time series analysis of the Swedish suicide rate 1915–1968. *Sociological miscellany. Essays in honour of Gunnar Boalt.* Stockholm: Stockholm University, Dept of Sociology, 1988.

10 Kasl SV. Mortality and the business cycle: some questions about research strategies when utilizing macro-social and ecological data. *Am J Public Health* 1979; **69:** 784–8.

11 Gravelle HSE, Hutchinson G, Stern J. Mortality and unemployment: a critique of Brenner's time series analysis. *Lancet* 1981; ii: 675–9.

12 Fox AJ, Goldblatt PO, Jones DR. Social class differentials: artefact, selection or life circumstances? *J Epidemiol Community Health* 1985; **39:** 1–8.

13 Box GEP, Jenkins GM. *Time series analysis: forecasting and control.* London: Holden-Day Inc., 1976.

14 Eyer J. Does unemployment cause the death rate peak in each business cycle? A multifactor model of death rate change. *Int J Health Serv* 1977; **7:** 625–62.

15 Norström T. Alcohol and suicide in Scandinavia. *Br J Addict* 1988; **83:** 553–9.

16 Norström T. The ecology of suicide in Sweden. Stockholm: Swedish Institute for Social Research, 1988.

17 Beral V, Chilvers C, Fraser P. On the estimation of relative risk from vital statistical data. *J Epidemiol Community Health* 1979; **33:** 159–62.

18 Morgenstern H. Uses of ecological analysis in epidemiological research. *Am J Public Health* 1982; **72:** 1336–43.

19 Schlesselman JJ. Assessing effects of confounding variables. *Am J Epidemiol* 1978; **108:** 3–8.

20 Gove WR, Hughes M. Re-examining the ecological fallacy: a study in which aggregate data are critical in investigating the pathological effects of living alone. *Soc Forces* 1980; **58:** 1157–77.