Commentary: Mortality from environmental factors, but which ones?

WR Keatinge

The paper by Gouveia et al.\(^1\) reaches two main conclusions, that in this sub-tropical city the relationship of mortality to temperature was U-shaped, and that it did not differ significantly between high and low socioeconomic groups. The first conclusion is in line with many reports from other, usually colder, countries. The second conclusion, that the relationship did not differ between socioeconomic classes, is at first sight surprising, but it too is in line with work in other countries, at least in respect to cold-related mortality.\(^2,3\) Regional differences in cold-related mortality depend on the protective measures taken locally against cold stress, and those in turn depend mainly on the different perceived needs for protection in different climates.\(^4,5\) The main questions relate to methodology. This paper has been written with attention to some possible pitfalls, and to methodology used in previous work. Its conclusions are likely to be broadly correct, but since the questions apply to a great many papers in this area, they are worth close inspection.

Multiple regressions to assess the effect of environmental factors on mortality contain collinear variables that readily attribute effects of one variable incorrectly to another. If measurement error differs between two closely collinear explanatory variables, the multiple regression tends to attribute mortality change to the one that has least measurement error, rather than to one which is causing the mortality change. Two-stage regressions tend to attribute effects to variables in the first stage, and suppress effects of those in the second stage. Misspecification of the true relationship tends to transfer the apparent effect from the misspecified variable. For example, if a log-linear relationship is assumed when the real relationship is linear or exponential, the effect tends to be transferred to another variable whose crude relationship to mortality happens to be closer to log-linear. Generalized additive Poisson regression models (GAM) modelling, used in the present paper, reduces that problem by seeking the best-fit relationship for each explanatory variable, but it requires subjective judgement of the minimum degree of smoothing needed to produce a convincing relationship for each explanatory variable. This can place strain on the judgement of a researcher who finds that a particular degree of smoothing will generate the result he is seeking, while a neighbouring degree of smoothing will not.

Such problems are well documented\(^6–8\) but are frequently ignored. The existence of numerous prior papers that used similar methods is of little help in establishing validity unless those papers include a rigorous search for possible errors in their approach. Current targeting of large-scale funding on chosen research topics tends to generate numerous papers over a short period, all using a similar approach and generating the same systematic error. Substantial short-term mortality has, for example, frequently been attributed in the past to pollution by sulphur dioxide (SO\(_2\)), but SO\(_2\) is absorbed in the upper airways and does not reach the lungs in measurable amounts.\(^9\) The apparent mortality ascribed to it seems to have been due to collinear factors, such as unusual patterns of cold weather associated with air pollution.\(^10\)

There is no absolute solution to these problems. This makes it important for authors, editors, and referees routinely to make a systematic search for any factors that could have spuriously generated the relationship described, and to assess frankly any such possibilities of error that they cannot eliminate. Application of a few simple principles can greatly facilitate this.

The reader must be able to see exactly what was done. This paper states that it used standard time series methods that have been developed for air pollution studies. Since the precise approach used in such papers has varied enormously, it is hard to see exactly what was done in this case. The paper states, for example, that daily values of a wide variety of pollutants were obtained, but it does not say whether all were included in all mortality analyses, whether any or all of pollutants appeared to have significant effects, and whether any that did not were dropped from the analysis.

It can also be extremely helpful to show simple relationships between key variables as well as residual relationships after processing. The more complex the processing, the more numerous the distortions it can produce and the more difficult they are to exclude. This paper does give ‘unadjusted’ mortality/temperature plots for pooled population mortality, as well as processed ones. Assuming that the former were not subjected to any form of processing, these plots greatly strengthen confidence in the first conclusion, that there was a U-shaped relationship between these variables. No such mortality/temperature plots are presented for the different socioeconomic groups. These would be extremely helpful in assessing the second conclusion in the paper. They might confirm the conclusion of the complex analyses that the relationship was similar in all the groups. Without them, the possibility remains that there were real differences in the mortality/temperature relationship between different socioeconomic groups, and that the subsequent analysis falsely attributed these to some other variable, such as pollution. That could happen if people in the low socioeconomic groups lived in areas of relatively high pollution, which is likely, and if such pollution increased in cold weather, as it usually does elsewhere.
References