Air Pollution and Hospital Admissions: Where are the Bodies?

Ross McKitrick
Financial Post March 2010

For many years we have heard that air pollution in Canada is responsible for thousands of annual deaths and hospitalizations. In 2004 Toronto Public Health claimed that 1700 premature deaths and 6000 hospitalizations occur each year in Toronto alone, due to air pollution. The Ontario Medical Association, Provincial and Federal Governments, lung associations and other groups regularly cite these kinds of figures in support of calls for new regulatory initiatives. These death and hospitalization rates are astonishing. It is like suffering a 9/11-style terrorist attack every 10 months.

But is it really true? The estimates are derived by taking correlations in the epidemiological literature between observed pollution levels and health indicators, like hospital admission rates, and then extrapolating across populations to estimate how many deaths and illness diagnoses can, in theory, be attributed to pollution. In other words, the numbers come from statistical models, not from direct observations. That means we need to pay close attention to how the statistical modeling is done.

Together with my coauthors Gary Koop of Strathclyde University and Lise Tole of the University of Edinburgh, I have just published a peer-reviewed study in the journal *Environmental Modelling and Software* that does just that. What we found gives us reason to believe that the kind of statistical modeling behind common claims about air pollution may need a careful second look.

There are hundreds of studies in the epidemiological literature that have reported correlations between air pollution and health measures. But there are some common weaknesses to this literature. First, the results are not consistent across studies. Some studies find particulate matter (PM) affects health, but not sulphur dioxide (SO_2) or carbon monoxide (CO). Others reported SO_2 ahs an effect, but not PM. Another reports CO has an effect but not ozone (O_3), while another finds O_3 matters in some cities but not others. One large US study found PM increased mortality risk a little bit across the US, except in 20 out of 88 cities in which it actually reduced mortality risk. These kinds of inconsistencies should not occur if the health effect is based on a real physiological response. This is a second puzzling aspect of the literature: despite decades of testing, clinical investigations have not found experimental support for the idea that current ambient air pollution levels cause lung disease or mortality.

A third weakness of the literature is that studies tend to use short panels of data from recent years (post-1990) when air pollution levels were rather low and steady. When constructing a statistical model of the effect of an explanatory variable (such as the effect of PM levels on hospital admissions) it is difficult to identify the correlation if the explanatory variable does not change much over the sample. Use of short panels with low variation can lead to unstable parameter estimates.

A fourth weakness of the literature is that few studies control for important factors like smoking, income levels and weather. Some recent studies have added in socioeconomic covariates, and after doing so the apparent effect of pollution vanished.

A fifth weakness has to do with the fact that a researcher needs to make a great many choices about how to approach the data. There are dozens of variables that could potentially be included in the statistical model. The number of possible combinations can potentially run into the billions. How should a researcher choose which model to use? In many cases the decision is arbitrary: just pick on model and report the results. But in some cases when people have gone back and tried different models on the same data, they get different results. So it is important to use a methodology that takes into account the uncertainty associated with the fact that the researcher needs to choose how to construct the model, and the data will tend to support some models over others.

Koop, Tole and I set out to re-examine the relationship between air pollution and health using a data set and methodology that would address these concerns. We were able to get a unique data set from Statistics Canada containing counts of hospital admissions for all lung-related ailments in 11 cities across the country from 1974 to 1994. This spans a time period when air pollution was initially higher than today, then steadily fell. If today's low air pollution sends thousands of people per year to the hospital in Toronto, then we should easily be able to identify the effect in previous years when pollution was double or triple its current concentrations. We then built a data set that included the urban air contaminants, average income, smoking rates, air temperature, wind speed and air pressure in each city over the same period. Finally, rather than picking one statistical model and relying on it, we used a technique called Bayesian Model Averaging that evaluates all possible model specifications, assesses the support each one gets in the data, and then constructs parameter and uncertainty estimates based on the whole distribution.

We found, not surprisingly, that smoking is bad for lung health. We found that regions with higher Gross Domestic Product (GDP) tend to have higher hospital admission rates, depending on the model specification, which may indicate that those regions have more hospital services. And we found evidence that hot days with high air pressure tend to produce more hospital admissions.

What we did not find was any evidence that increases in air pollution levels are associated with increased rates of hospital admissions. We looked at the data every which way imaginable. If we were to cherry pick, by looking only at a sub-sample of the time or by picking just the right form of the model, we could find evidence that CO or nitrogen dioxide (NO_2) have positive effects on lung disease, but those results do not get strong support in the data. The models that get consistent support either show no pollution effects or – paradoxically – negative effects. In other words, in some cases as air pollution rises, hospital admissions go down. As odd as that sounds, we are by no means the first to report negative coefficients in the literature. Nobody is trying to argue that air pollution is good for you: this is either just noise in the data, or it might be an effect from "averting" behaviour, where people who are susceptible to lung problems stay indoors on days with bad air quality.

Based on our analysis, we could estimate what the effect on hospital admissions would be if all the pollution currently observed in Toronto air were to disappear. Toronto Public Health claims about 6000 fewer hospitalizations would occur. But this claim gets no support in the data. We found that there would be no reduction in lung-related hospitalizations. If anything there might be somewhere between 20 and 200 more admissions, if we apply the statistical results in a mechanical fashion.

What can we make of all this? It would be a mistake to argue that, because our paper only weighs a few grams whereas the pile of papers on the other side weighs about seven kilograms, therefore 6,000 people get sick from air pollution in Toronto each year. If we are going to get into that sort of argument, I'd be happy to compare the weight of a printout of our data base against any of the ones used in the other stack. But hopefully the issue will be seen not in terms of tonnage, but in terms of quality and thoroughness of the analysis.

On that point, despite the large number of epidemiological studies on this topic in the past decade, very few have controlled for socioeconomic covariates (including smoking), fewer still have looked at long data panels back to the 1970s, and fewer still have dealt with model uncertainty. Those that have addressed one or more of these issues typically find the effect of air pollution shrinks or disappears outright. Thus our results are actually quite consistent with the relevant group of previous studies. The popular idea that current ambient air pollution has a powerful effect on lung health might look like it is based on a large empirical foundation, but on closer inspection the pile contains a lot of weak results.

So the bottom line is that, for the purpose of assessing the link between air pollution levels and hospital admissions, one needs to look closely at the kinds of studies being done, and how they did the statistical modeling. More studies need to be done using long time series that go back to the 1970s or earlier, more studies need to control for socioeconomic covariates, and more studies need to take account of model uncertainty. Based on evidence to date, as these things begin to happen, we should not be surprised if current estimates of the health effects of air pollution turn out to be in need of major revision.

Ross McKitrick is a Professor of Economics at the University of Guelph