CM Wong 黃淅明
TQ Thach 石國順
S Ma

A pilot study to elucidate and validate public health–related environmental effect estimates from statistical modelling of daily health outcome counts

Key Messages

1. Results from statistical modelling of short-term health effects of air pollution can be used to derive the number of health outcomes avoidable for a given reduction in air pollutant concentrations. Effect estimates obtained from daily time-series studies can be used to produce cost curves for reduction in pollutants and benefit in terms of the avoidance of bad health outcomes. This forms the basis for setting air quality guidelines and assessment of air quality interventions.
2. In studies involving time-series data and complex statistical modelling, simulation can be used to obtain data, which mimic the complex structure of the data and the power functions for various effect sizes. This supports the estimation of sample sizes required for the studies, which is often neglected at present.

Hong Kong Med J 2008;14(Suppl 3):S9-12

Department of Community Medicine, 5/F, William MW Mong Block, Faculty of Medicine Building, The University of Hong Kong, 21 Sassoon Road, Hong Kong SAR, China
CM Wong, TQ Thach, S Ma

HCPF project number: 212920

Principal applicant and corresponding author:
Dr CM Wong
Department of Community Medicine,
5/F, William MW Mong Block, Faculty of
Medicine Building, The University of Hong
Kong, 21 Sassoon Road, Hong Kong SAR,
China
Tel: (852) 2819 9109
Fax: (852) 2855 9528
E-mail: hrmrwcm@hkucc.hku.hk

Introduction

Air pollution is a major public health hazard in nearly all cities today. Health effect estimates of air pollutants provide information for setting air quality guidelines and for assessing the cost-benefit of air pollution abatement measures. While longitudinal and intervention studies are the best designs in terms of control for confounding factors, they are time consuming and expensive. Routine daily data on hospital admissions, mortality, air pollution, and meteorological conditions have been used: (1) to assess the health effects of air pollution[1] and the effectiveness of government interventions in improving environmental and public health,[2] and (2) to support the prediction of air pollution–associated health service utilisation over time.[3]

Ecological time-series studies using Poisson regression modelling on daily mortality and hospital admissions are increasingly used for health effect estimates. This method estimates the association between variations in daily health outcomes with daily air pollutant concentrations and covariates.

The statistical methods involve modelling of aggregated daily counts of hospital admissions and deaths with observable daily variables, such as daily meteorological measures and identifiable covariates.

Aims and objectives

In this study we proposed to develop a statistical and computational procedure to simulate daily environment and health data sets to assess the following research questions:

(1) Are statistical modelling methods for daily counts of health outcomes (eg hospital admissions and deaths due to cardiovascular and respiratory diseases) valid in making adjustment for the various sources of unobservable and observable factors?
(2) How valid are simulation methods for elucidating variation in daily counts of health outcomes and facilitating better understanding by public health physicians and policy makers?
(3) Can simulation data be used to assess the power of the statistical procedure for a given sample size, in terms of the number of years for which the data was required?

Methods

This study was conducted from July 2001 to June 2002.

Data

Over 3 years (1995-97), the following data were retrieved and aggregated into daily counts or means: (a) hospital admission data from the Hospital Authority and mortality data from the Census and Statistics Department, for cardiovascular disease (International Classification of Diseases version 9 [ICD9] codes: 350-459), respiratory disease (ICD9: 460-519) and chronic obstructive pulmonary

Table 1. Daily variables used in air pollution time-series study

Variable	Description	Symbol
Outcome	Hospital admissions/mortality	Y
Observed confounders	Temperature	Z
	Relative humidity	
	Day of the week (dummy variables for Monday, Tuesday, Wednesday, Thursday, Friday, and Saturday)	
	Holiday	
	Day after holiday	
	Influenza epidemics (dummy variables for weeks in the 4th quartile of weekly influenza admissions each year)	
Unobserved confounders	Time trend (*t*)	Z
	Seasonal variations ($\sin 2\pi kt/365$, $\cos 2\pi kt/365$, where *k* are numbers of cycles)	
Pollutants	Concentration of nitrogen dioxide, sulphur dioxide, respirable suspended particulate matter with aerodynamic diameter <10 μm, and ozone	X

Table 2. Parameter estimates for cardiovascular admissions

Estimate	Parameter estimate	Standard error	t value
Intercept	4.54	4.62×10^{-2}	98.27
t (day number)	3.69×10^{-4}	4.09×10^{-5}	9.03
t^2	-2.53×10^{-7}	3.61×10^{-8}	-6.99
$\cos(2\pi t/365)$	6.16×10^{-2}	1.05×10^{-2}	5.85
$\cos(4\pi t/365)$	2.68×10^{-3}	4.64×10^{-3}	0.58
$\cos(6\pi t/365)$	-1.04×10^{-2}	4.63×10^{-3}	-2.26
$\cos(8\pi t/365)$	-4.11×10^{-3}	4.65×10^{-3}	-0.89
$\sin(2\pi t/365)$	6.39×10^{-2}	6.90×10^{-3}	9.27
$\sin(4\pi t/365)$	2.07×10^{-2}	4.82×10^{-3}	4.29
$\sin(6\pi t/365)$	-1.31×10^{-2}	4.61×10^{-3}	-2.85
$\sin(8\pi t/365)$	-1.14×10^{-2}	4.46×10^{-3}	-2.56
Monday	4.49×10^{-1}	1.24×10^{-2}	36.29
Tuesday	3.35×10^{-1}	1.27×10^{-2}	26.26
Wednesday	3.90×10^{-1}	1.25×10^{-2}	31.12
Thursday	3.38×10^{-1}	1.27×10^{-2}	26.63
Friday	2.99×10^{-1}	1.28×10^{-2}	23.34
Saturday	1.79×10^{-2}	1.36×10^{-2}	1.31
Holiday	-3.33×10^{-1}	1.54×10^{-2}	-21.62
Day after holiday	7.34×10^{-2}	1.89×10^{-2}	3.89
Temperature	5.35×10^{-3}	1.56×10^{-3}	3.44
Humidity	-3.54×10^{-4}	3.72×10^{-4}	-0.95
Influenza epidemic	3.52×10^{-2}	8.99×10^{-3}	3.91
Nitrogen dioxide	9.61×10^{-4}	2.01×10^{-4}	4.78

disease (ICD9: 490-496); (b) pollutant concentrations for four iconic pollutants: nitrogen dioxide, sulphur dioxide, respirable suspended particulate matter with aerodynamic diameter <10 μm, and ozone from the Environmental Protection Department; (c) meteorological data including temperature and relative humidity from the Hong Kong Observatory.

Statistical modelling

Poisson regression models using variables as shown in Table 1 were fitted to obtain a core model (ie a model without a pollutant variable). Pollutant concentrations were then entered into the core model to obtain the effect estimate (Table 2).

Simulation models

We simulated daily time series of a pollutant concentration variable (X) and a set of variables (Z) mimicking all the observable and unobservable confounding variables mentioned in Table 1. We then derived the corresponding expected daily health outcome, E(Y), according to X (parameter β) and Z (parameter γ) and using equation[1] below.

$$\text{Log E(Y)} = \alpha + \beta X + \gamma Z \ldots [1]$$

Finally we simulated Y according to a Poisson distribution with mean E(Y).

Validation

We estimated parameter β in the simulated data using the APHEA[1] approach and assessed the coverage probability. This is a probability that the 95% confidence interval (95% CI) contains the true β. We repeated each trial with 1000 sets of simulated data.

The full model achieved coverage probability close to 95%, bias within 2%, and standard errors from 12 to 55%. The validity was weaker for estimates obtained from reduced models (ie models using fewer variables than those developed according to the APHEA approach).

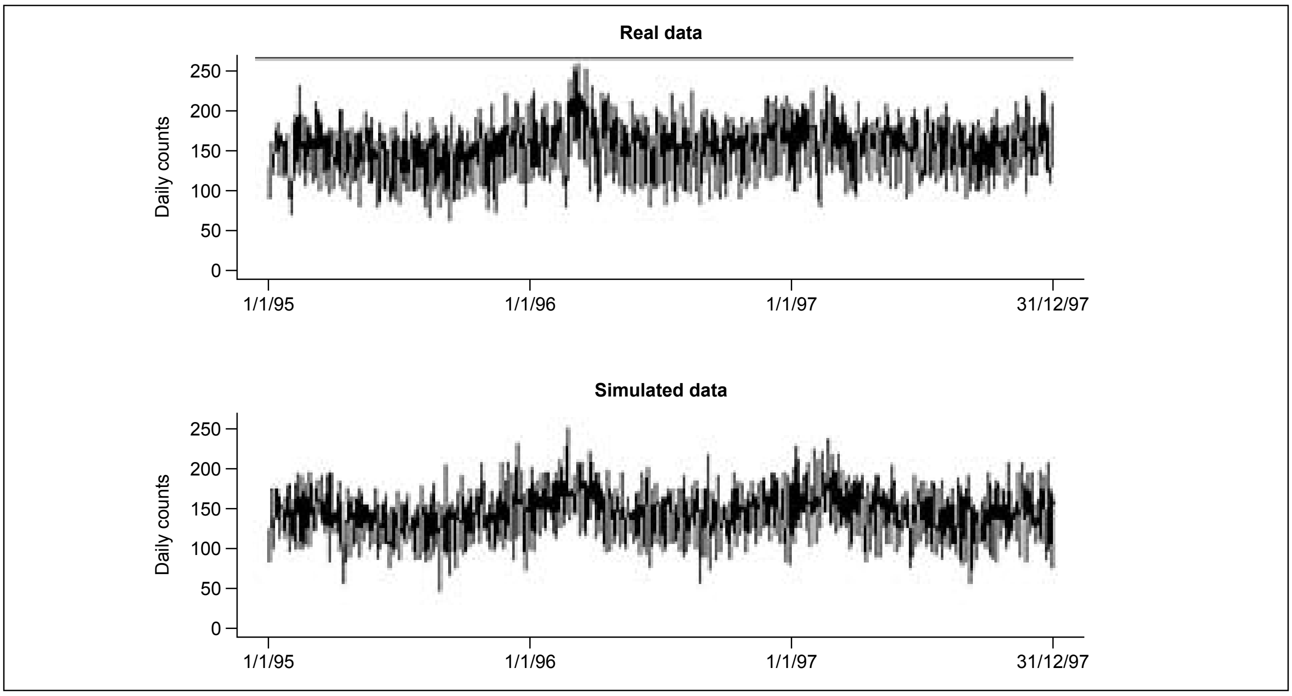
Elucidation of statistical methods

We adopted a set of simulated data mimicking the real data, applied the APHEA approach to obtain the core model, derived the estimates, and presented the results to assess whether simulation can be used to elucidate variations in daily health outcomes.

The real and simulated data are shown in Figure 1. The parameters estimated from the real data are shown in Table 2.

Study power

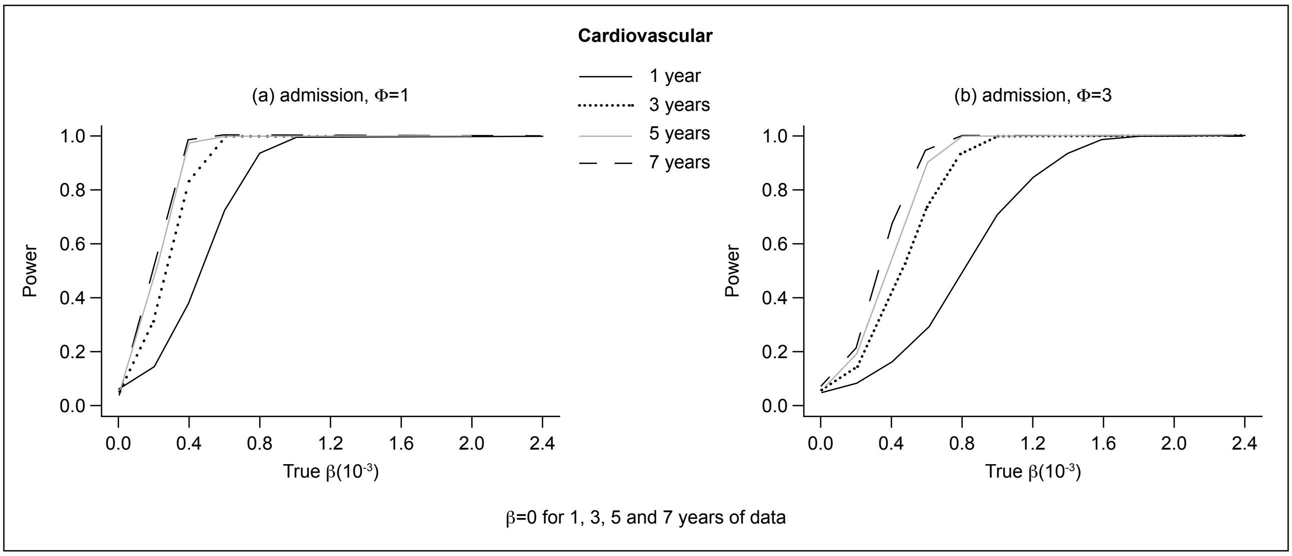
We simulated 1000 replicates, each with a β parameter ranging from 0 to 2.4×10^{-3}, of a series of 365n daily data, where n=1, 3, 5 and 7 are numbers of years in the simulation. Under the null hypothesis that β=0 we obtained the power function which is the proportion for which $|\hat{\beta}/\text{se}(\hat{\beta})| > 1.96$ where $\text{se}(\hat{\beta})$ is the standard error for the estimate $\hat{\beta}$ of β. As expected the bigger the true β and the greater the number of years, the higher the statistical power. In addition, the larger the dispersion parameter, the smaller the power of the test (Fig 2).



Real data

Simulated data

Fig 1. Daily counts of hospital admissions due to cardiovascular disease



Cardiovascular

(a) admission, Φ=1

(b) admission, Φ=3

β=0 for 1, 3, 5 and 7 years of data

Fig 2. Power function in testing the effect of nitrogen dioxide under the null hypothesis

Discussion

In assessing short-term health effects of air pollutants from daily data, the other determinants of daily health outcomes together with seasonality are regarded as confounders in the observable patterns of daily health outcomes. The major task in simulating the impact of air pollution on health is to develop the core model (with variables and terms not including the air pollutant concentration variable), which can explain all discernible patterns in the daily health outcomes. We simulated the data with parameters known-effect daily health outcomes and demonstrated that the data were consistent with the effects of air pollution. We also demonstrated the usefulness of simulation data to assess the effects of air pollution on health. Simulation strategies using hypothetical parameters for a number of years can be produced and the power for a given number of years can be determined.

Conclusions

Simulation can be applied to elucidate and validate complex statistical procedures and be utilised to design the sample size for studies, which require statistical modelling.

Acknowledgements

This study was supported by the Health Care and Promotion Fund (#212920). We thank the Hospital Authority, the Environmental Protection Department and the Census and Statistics Department for provision of data; Mr KP Chan for data analysis; and Ms Marie Chi for secretarial support.

References

1. Katsouyanni K, Schwartz J, Spix C, et al. Short term effects of air pollution on health: a European approach using epidemiologic time series data: the APHEA protocol. J Epidemiol Community Health 1996;50(Suppl 1):12S-18S.
2. Hedley AJ, Wong CM, Thach TQ, Ma S, Lam TH, Anderson HR. Cardiorespiratory and all-cause mortality after restrictions on sulphur content of fuel in Hong Kong: an intervention study. Lancet 2002;360:1646-52.
3. Wong C, Ma S, Lam T, Hedley A. Coronary artery disease varies seasonably in subtropics. BMJ 1999;319:1004.