THEORY AND METHODS

A simple approximate mathematical model to predict the number of severe acute respiratory syndrome cases and deaths

B C K Choi, A W P Pak

Background: Severe acute respiratory syndrome (SARS) is currently spreading in many countries. This paper proposes a simple approximate mathematical model for public health practitioners to predict the number of SARS cases and deaths.

Methods: The model is based on four parameters: R_o (basic reproductive number), F (case-fatality rate), i (incubation period), and d (duration of disease). The calculations can be done by hand or by using a computer spreadsheet.

Results: The best parameters to fit Canadian data as of 6 April 2003 (before infection controls took effect) are $R_o = 1.5$, F = 30%, i = 5 days, d = 14 days. On 6 April (day 40) there were 74 cases and 7 deaths. If this trend continues, SARS numbers in Canada are predicted to be as follows: 387 cases and 34 deaths by 26 April (day 60), 4432 cases and 394 deaths by 26 May (day 90), and 50 500 cases and 4489 deaths by 25 June (day 120). By comparison, the best parameters to fit Hong Kong data as of 10 April 2003 are $R_o = 2.0$, F = 20%, i = 5 days, d = 14 days.

Conclusions: Using the proposed mathematical model, it was estimated that about 1.5 to 2 new infectious cases were produced per infectious case every five days. Also, about 20% to 30% of the cases die within 14 days. The case-fatality may therefore be considerably higher than initially thought. The model indicates that SARS can spread very fast when there are no interventions.

See end of article for authors' affiliations
....

Correspondence to:
Dr B C K Choi, 432
Pleasant Park Road,
Ottawa, Ontario, Canada
K1H 5N1;
Bernard.Choi@utoronto.ca

Accepted for publication
4 July 2003
....

Severe acute respiratory syndrome (SARS), a contagious and rapidly progressive infectious disease, is only months old but is already spreading in many countries.[1] The epidemic of SARS has caused a lot of concern in the media and the general public.[2–4] The World Health Organisation issued a global health alert on 12 March 2003[1 5] and set up a daily registry of reported cases on 17 March.[6] Within a matter of weeks since the first cases were reported, research results have started appearing in scientific journals on the epidemic,[1 5] epidemiological and clinical features of patient clusters in Hong Kong[7–9] and in Canada,[10] and a coronavirus as possible cause.[1 2 9–12] Scientists and health practitioners are acting fast, but still may not be fast enough.[1] Currently, public health practitioners lack a mathematical tool to assist them in predicting the number of cases and deaths in the short-term, in order to plan resources and to evaluate the effectiveness of intervention strategies.

This paper proposes a simple approximate mathematical model for public health practitioners to predict the number of SARS cases and deaths arising in the first months of an epidemic under the assumption that no intervention takes place to cut its spread. It is intended to be a user friendly epidemiological paper dealing with a highly time sensitive problem. The proposed epidemiological framework for predicting SARS morbidity and mortality is meant to be clear, simple, and applicable even in developing countries with limited resources. The calculations can be done by hand or by using a computer spreadsheet. The model is illustrated by publicly available data on the World Health Organisation web site.[6]

THE MODEL

The model, based on the standard SIR (susceptible-infected-removed) epidemic model,[13 14] assumes that transmission of SARS is contagious from person to person[1 10 11] and not point source.[15] It is further assumed that, at an initial stage of the SARS epidemic, the proportion of the population with immunity to SARS is negligible.[14]

At the initial stage of a contagious epidemic, a small number of infected people start passing the disease to a large population. Individuals can go through three states. They start out susceptible (S), then become infected and infectious (I), after which they are removed (R), either by recovery with immunity or by death. Schematically,

$$S \rightarrow I \rightarrow R.$$

The proposed model is based on four parameters: R_o (basic reproductive number—that is, the expected number of new infectious cases per infectious case), F (case fatality rate—that is, the proportion of cases who die within the symptomatic period), i (incubation period—that is, the time from infection to symptom), and d (duration of disease, or symptomatic period—that is, the time from symptom to recovery or death).[14 15]

Assumptions made in formulating the model are:

- The population at risk is large enough and time period of concern is short enough that over the time period of interest, very close to 100% of the population is susceptible.

- The epidemic is at an early stage and has not reached the point where the susceptible population decreases so much due to death or post-infection immunity that the average number of secondary cases falls.

- Unprotected contact results in infection.

- The epidemic in the population of interest begins with a single host. (Note that the equations and Excel formulas

used in computing cases and deaths are easily modified if this is not the case.)

- There is no intervention to prevent disease from spreading.
- There is homogenous mixing among the infectives and susceptibles, such that every infected person will pass the disease to exactly R_o susceptible individuals simultaneously within an incubation period of i days.
- Infectivity occurs during the incubation period only.
- The models are deterministic—that is, the four parameters take on constant values.

Predicting the number of SARS cases

This model requires only the following input parameters: R_o (basic reproductive number) and i (incubation period).

After one incubation period (i), one infectious case produces R_o new infectious cases. The cumulative total number of cases at this time is $1+R_o$. After two incubation periods (2i), there are R_o^2 cases produced by the previous R_o cases. The total number of cases is $1+R_o+R_o^2$.

Mathematically, the predicted number of incident cases on day t·i, that is, $C_{t·i}$, where t is time expressed in the number of incubation periods, is

$$C_{t·i}$$

The predicted total number of cases (C) is

$$C = \Sigma\ C_{t·i}$$

Table 1 illustrates the application of the model to calculate the predicted number of SARS cases, using $R_o = 3$ and i = 5 days. It is predicted that on day 15 there are 27 new cases and that the total number of cases by day 15 is 40. The value for i can be determined from epidemiological studies of patients. The optimal values for R_o in a particular situation, for example, a country or a local area, can be determined by trying out several values to see which combination of R_o and i produces the predicted total number of SARS cases that most closely matches the observed total number of SARS cases.

Predicting the number of SARS deaths

This model requires the following input parameters: R_o (basic reproductive number), F (case fatality rate), i (incubation period) and, d (duration of disease).

After one disease duration (d), the cases are removed (by death or recovery). F percentage of them die, while $1-F$ percentage recover.

Mathematically, the predicted number of deaths on day t·i+d—that is, $(D_{t·i+d})$, where t is time expressed in the number of incubation periods, is

$$D_{t·i·d}\ = C_{t·i}\ \times F.$$

Table 1 Predicted number of SARS cases using $R_o = 3$ and i = 5 days

Number of incubation period (t)	Day (t·i)	Predicted incident cases ($C_{t·i}$)	Predicted total cases (C)
0	0	1*	1
1	5	3 (= R_o)*	4 (= 1+3)
2	10	9 (= R_o^2)	13 (= 1+3+9)
3	15	27 (= R_o^3)	40 (= 1+3+9+27)

*Mathematically, $1 = R_o^0$, $3 = R_o^1$.

Table 2 Predicted number of SARS deaths using $R_o = 3$, i = 5 days, F = 10% and d = 14 days

Number of incubation period (t)	Day	Predicted incident cases ($C_{t·i}$)	Predicted new deaths ($D_{t·i+d}$)	Predicted total deaths (D)
0	0 (= t×i)	1	0	0.0
1	5	3	0	0.0
2	10	9	0	0.0
	14 (= t×i+d)	0	0.1 (= 1×F)*	0.1
3	15	27	0	0.1
	19	0	0.3 (= 3×F)*	0.4

*Cases from day 0 are removed on day 14, after one disease duration (14 days). The one case from day 0 is expected to produce 0.1 death (1×10%) and 0.9 recovered people. Similarly, the three cases from day 5 are expected to produce on day 19 (after 14 days) 0.3 death and 2.7 recovered people.

The predicted total number of deaths (D) is

$$D = \Sigma\ D_{t·i+d}.$$

Table 2 illustrates the application of the model to calculate the predicted number of SARS deaths, using $R_o = 3$, i = 5 days, F = 10%, and d = 14 days. It is predicted that on day 19 there are 0.3 new deaths and that the total number of deaths by day 19 is 0.4. The value for d can be determined from epidemiological studies of patients. The optimal value for F in a particular situation, given optimal values for R_o, i and d, can be determined by trying out several values to see which combination of F, R_o, i and d produces the predicted total number of SARS deaths that most closely matches the observed total number of SARS deaths.

METHODS

The mathematical model was used to fit the observed numbers of cases and deaths posted on the World Health Organisation web site[6] available from 17 March 2003 onwards. For Canada, additional data on observed numbers of cases and deaths from 25 February (first Canadian case) to 16 March were available from Poutanen et al.[10]

Early epidemiological studies suggest that the incubation period ranges from 1 to 11 days, with a median of about 5 days[5][7]; therefore i = 5 days was used in the model. From the two deaths in Hong Kong reported in Tsang et al[7] and the three deaths in Canada reported in Poutanen et al,[10] the time from symptoms to death ranges from 8 days to 23 days, with a median of 14 days; therefore d = 14 days was used in the model.

The fitting of Canadian data was based on reported number of probable cases and deaths attributable to SARS from 25 February up to 6 April 2003 only. On 26 March 2003 Canada (Ontario) declared a public health emergency and implemented infection controls such as: strict rules on masks and protective clothing in hospitals, quarantine of suspected cases at home for 10 days, restriction of visitor access to hospitals, closing down of admissions, emergency, and non-urgent services, screening of travellers at airports, and closing of schools with suspected cases.[2][3] If these infection controls were effective, according to our model, effects on R might be seen after i+d days—that is, by 14 April (after 5+14 days), or at the earliest, by 4 April (after 1+8 days). Inspection of the observed data in Canada suggests that infection controls might have taken effect around 6 April 2003.

RESULTS

The best parameters to fit Canadian SARS data as of 6 April 2003 are $R_o = 1.5$, F = 30%, i = 5 days, d = 14 days. If this

Table 3 Predicted number of probable cases and deaths attributable to SARS in Canada since 25 February 2003 based on the observed trend from 25 February to 6 April 2003, and observed numbers from 25 February to 26 May

Date	Day	Predicted incident cases	Predicted total cases	Predicted new deaths	Predicted total deaths	Observed total cases	Observed total deaths
Feb 25*	0	1	1	0	0.00	1	0
Mar 2	5	1.5	2.5	0	0.00	3	0
Mar 7	10	2.25	4.75	0	0.00	5	1
Mar 11	14	0	4.75	0.30	0.30	8	1
Mar 12	15	3.38	8.13	0	0.30	8	1
Mar 16	19	0	8.13	0.45	0.75	8	2
Mar 17	20	5.06	13.19	0	0.75	8	2
Mar 21	24	0	13.19	0.68	1.43	9	2
Mar 22	25	7.59	20.78	0	1.43	9	2
Mar 26	29	0	20.78	1.01	2.44	19	3
Mar 27	30	11.39	32.17	0	2.44	28	3
Mar 31	34	0	32.17	1.52	3.96	44	4
Apr 1	35	17.09	49.26	0	3.96	53	4
Apr 5	39	0	49.26	2.28	6.23	74	7
Apr 6	40	25.63	74.89	0	6.23	74	7
Apr 10	44	0	74.89	3.42	9.65	97	10
Apr 11	45	38.44	113.33	0	9.65	98	10
Apr 15	49	0	113.33	5.13	14.78	100	13
Apr 16	50	57.67	171.00	0	14.78	103	13
Apr 20	54	0	171.00	7.69	22.47	132	12
Apr 21	55	86.50	257.49	0	22.47	132	12
Apr 25	59	0	257.49	11.53	34.00	140	15
Apr 26	60	129.75	387.24	0	34.00	142	18
Apr 30	64	0	387.24	17.30	51.30	148	20
May 1	65	194.62	581.86	0	51.30	147	20
May 5	69	0	581.86	25.95	77.25	148	22
May 6	70	291.93	873.79	0	77.25	148	22
May 10	74	0	873.79	38.92	116.17	145	22
May 11	75	437.89	1311.68	0	116.17	145	22
May 15	79	0	1311.68	58.39	174.56	142	23
May 16	80	656.84	1968.52	0	174.56	140	23
May 20	84	0	1968.52	87.58	262.14	140	23
May 21	85	985.26	2953.78	0	262.14	140	23
May 25	89	0	2953.78	131.37	393.50	140	23
May 26	90	1477.89	4431.68	0	393.50	148	26
May 30	94	0	4431.68	197.05	590.56		
May 31	95	2216.84	6648.51	0	590.56		
Jun 4	99	0	6648.51	295.58	886.14		
Jun 5	100	3325.26	9973.77	0	886.14		
Jun 9	104	0	9973.77	443.37	1329.50		
Jun 10	105	4987.89	14961.66	0	1329.50		
Jun 14	109	0	14961.66	665.05	1994.55		
Jun 15	110	7481.83	22443.48	0	1994.55		
Jun 19	114	0	22443.48	997.58	2992.13		
Jun 20	115	11222.74	33666.22	0	2992.13		
Jun 24	119	0	33666.22	1496.37	4488.50		
Jun 25	120	16834.11	50500.34	0	4488.50		

*The first SARS case in Canada, a 78 year old woman who had returned to Toronto from Hong Kong, first noted symptoms on 25 February 2003.[10] Paremeters: R_o (basic reproductive number, the number of new infectious cases per infectious case) = 1.5, F (case fatality rate) = 30%, i (incubation period, time from infection to symptom) = 5 days, and d (duration of SARS, from symptom to recovery or death) = 14 days. Predicted numbers are based on the mathematical model. Observed numbers from 25 February to 16 March are given by Poutanen et al[10]; observed numbers since 17 March are from the WHO web site (http://www.who.int/csr/sarscountry/en/).

early trend continues, SARS numbers in Canada are predicted to be as follows: 387 cases and 34 deaths by 26 April (day 60), 4432 cases and 394 deaths by 26 May (day 90), and 50 500 cases and 4489 deaths by June 25 (day 120) (table 3).

Plots of the predicted and observed numbers of SARS cases and deaths on linear and log scales are shown in figures 1–4.

The proposed model is easy to use. Table 3 and figures 1–4 were generated, using Excel (Microsoft Corporation), in less than an hour. The model can also be done by hand, with a calculator and graph paper; it took about four hours to generate the table and the four figures.

DISCUSSION

The method described in this paper is easy to understand and to use. It can be useful in at least two situations. Firstly, a public health officer can estimate the size of an outbreak before control measures are effectively put into practice.

Secondly, it may be useful for public education to illustrate what could happen in a population where no action is taken to stop the epidemic.

Under the contagious epidemic assumption, SARS is passed from person to person and the initial rise in the number of cases and deaths is slow. However, the model indicates a possible devastating effect in just a few months' time, if proper measures to control the epidemic are not available or enforced to reduce the number of new infectious cases per infectious case (R_o) to below 1.0. Exponential growth of the SARS epidemic in the absence of interventions has also been suggested by other authors.[16–18]

The model can be used to evaluate the success of interventions by monitoring the reduction of R_o (success in controlling the spread) and the reduction of F (effectiveness of the treatment) needed to produce the best fitting model for new observed data. Furthermore, if the actual numbers of
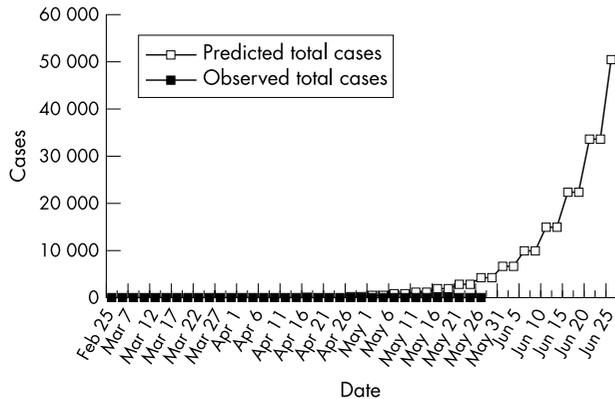
Figure 1 Predicted and observed probable cases of SARS in Canada, February to June 2003.
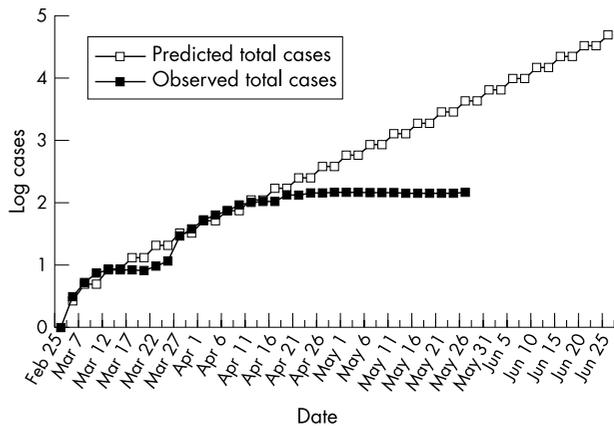
Figure 2 Predicted and observed probable cases (log scale) of SARS in Canada, February to June 2003.

cases and deaths are increasingly lower than the predicted numbers, as current Canadian data appear to show, health practitioners may be reassured that the interventions are doing well. The gradual fall off of the observed number of cases and deaths after 6 April 2003 from the expected curves based on the initial trend indicates that control measures in Canada implemented since 26 March 2003 probably have taken effect. In evaluating success, however, confounding by natural intervention must also be considered, for example, it is possible that R_o may change and diminish regardless of control measures, simply as the weather becomes warmer.[18]

The model can also be used to provide a more accurate estimate for the case fatality rate than the traditional method. Under the traditional method, which is based on a cross sectional approach, case fatality is simply the number of deaths divided by the number of cases in a specified time period. For example, the traditional method estimates that, as of 3 April 2003, the case fatality is 10% (6 of 62) for Canada and 2% (17 of 734) for Hong Kong.[19] However, at the initial stage of an epidemic, there is an accelerating increase of daily new cases. These new cases are not likely to die within the same day. Their inclusion results in an under-estimation of case fatality. Using our model, which is based on a cohort approach, it is estimated that the case fatality is 30% in Canada and 20% in Hong Kong, assuming a disease duration of 14 days. The case fatality estimate varies depending on the disease duration, for example, it becomes 20% for Canada and 5% for Hong Kong if a disease duration of seven days is used. Our predictions are confirmed by a

recent report based on 1425 cases in Hong Kong that SARS death rate is higher than WHO estimate; and that about 20% of the SARS cases in Hong Kong are dying (13% for cases younger than 60 years; 43% for cases 60 years and older).[18 20] After the study, the WHO has revised its SARS death rate from 6% to 15%.[21]

An interesting observation from using the model is that, for fitting the spread of SARS in Hong Kong as of 10 April 2003, the best parameters are $R_o = 2.0$, F = 20%, i = 5 days, d = 14 days (data not shown). It is understandable that, because of the higher population density, R_o in Hong Kong is higher than in Canada. The lower case fatality in Hong Kong (20%), as compared with that in Canada (30%), may be attributable to a difference in attribution of deaths to SARS, as different rules may govern the coding of a death when a SARS patient dies of a non-SARS related cause.

Our model has been developed for the rapid calculation of predicted cases and deaths for the short term at this initial stage of the SARS epidemic. It is intended to help front line public health practitioners in their planning. The model is not intended to replace more sophisticated mathematical methods at a later stage when more data on the epidemic pattern of SARS are available. To keep the model simple and user friendly for the average public health officer, a number of assumptions are made that may reduce the validity, compared with more sophisticated models. As the model is simple to use, it may result in situations where assumptions are not fulfilled. As an illustration, according to current knowledge, mortality risk seems to be strongly dependent on age[18 21] but this dependency is not taken account into the
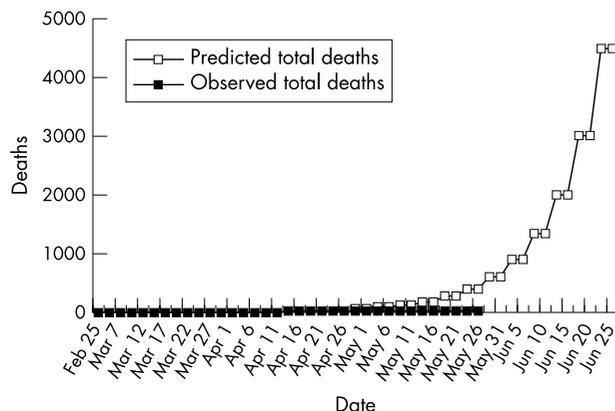
Figure 3 Predicted and observed SARS deaths in Canada, February to June 2003.
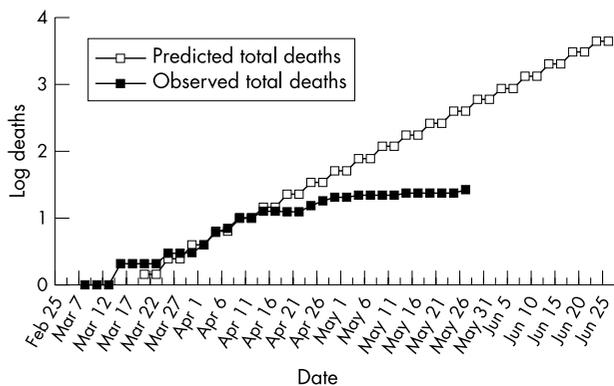
Figure 4 Predicted and observed SARS deaths (log scale) in Canada, February to June 2003.

model. Also, the basic reproductive number may not be homogenous, for example, it may be higher in certain subgroups (in particular hospitals) and lower in others.[17]

For simplicity, the model fitting is carried out by "trying out several values" (eyeballing) instead of some formal fitting algorithm that then requires extensive computer programming. The models are deterministic to avoid the complex stochastic models where, for example, i and d take values from a specific distribution with mean and variance. Also, 95% confidence intervals for the curves or sensitivity analyses are not suggested. The model does not address issues such as evaluating the efficacy of interventions by shortening the period between onset of symptoms and hospital admission.

When this paper was first developed based on available data up to mid-April 2003, it was our hope that our estimates for Canada will be off in a month or two. This will then show that the recent infection control measures have been effective. The predicted numbers based on the early trend only illustrate the hypothetical situation if the early trend continues, for example, when control measures are either not available or not enforced. By late May, when this paper was finalised, the estimates are already off. The departure of observed numbers from the predicted can be considered a measure of the success of infection controls, in terms of the number of potential cases and deaths prevented.

. . . .

Authors' affiliations

B C K Choi, Department of Epidemiology and Community Medicine, Faculty of Medicine, University of Ottawa, Ottawa, Ontario, Canada and Department of Public Health Sciences, Faculty of Medicine, University of Toronto, Toronto, Ontario, Canada

A W P Pak, Institutional Research and Planning, University of Ottawa, Canada

The views expressed in this paper are solely those of the authors, and do not necessarily represent those of the University of Ottawa, University of Toronto, or any agencies or organisations.

REFERENCES

1 **Gerberding JL**. Faster…but fast enough? Responding to the epidemic of severe acute respiratory syndrome. *N Engl J Med* 2 Apr 2003;[Published on line ahead of print] (2 pages) http://www.nejm.org/(accessed 9 Apr 2003).
2 **Hawaleshka D**, MacQueen K, Macklem K, *et al.* Is this your best defence? As Canadians anguish over how to protect themselves from SARS, health authorities take note of what went wrong. *Macleans Magazine* 14 Apr 2003;[Published on line ahead of print] http://www.macleans.ca/xta-asp/storyview.asp?viewtype=browse&tpl=browse_frame&vpath=/2003/04/14/Cover/58105.shtml(accessed 9 Apr 2003).
3 **CBC News**. Health matters: SARS special. Aired on 13 April 2003, 2:00-3:00 pm Eastern Time on CBC Newsworld (Cable 26). Canadian Broadcasting Corporation, Toronto, Ontario, Canada. (email: healthmatters@cbc.ca.).
4 **Kickbusch I**. A wake-up call for global health. International Herald Tribune–The IHT online. 29 Apr 2003; http://www.iht.com/articles/94675.htm(accessed 3 May 2003).
5 **Drazen JM**. Case clusters of the severe acute respiratory syndrome. *N Engl J Med* 31 Mar 2003;[Published online ahead of print] (2 pages) http://www.nejm.org/(accessed 9 Apr 2003).
6 **World Health Organisation**. Cumulative number of reported cases (SARS) since March 17, 2003.http://www.who.int/csr/sarscountry/en/(accessed 17 Apr 2003).
7 **Tsang KW**, Ho PL, Ooi GC, *et al.* A cluster of cases of severe acute respiratory syndrome in Hong Kong. *N Engl J Med* 31 Mar 2003;[Published online ahead of print] (9 pages) http://www.nejm.org/(accessed 9 Apr 2003).
8 **Lee N**, Hui D, Wu A, *et al.* A major outbreak of severe acute respiratory syndrome in Hong Kong. *N Engl J Med* 7 Apr 2003;[Published online ahead of print] (9 pages) http://www.nejm.org/(accessed 9 Apr 2003).
9 **Peiris JSM**, Lai ST, Poon LLM, *et al.* Coronavirus as a possible cause of severe acute respiratory syndrome. *Lancet* 8 Apr 2003;[Published online ahead of print] (7 pages) http://image.thelancet.com/extras/03art3477web.pdf(accessed 9 Apr 2003).
10 **Poutanen SM**, Low DE, Henry B, *et al.* Identification of severe acute respiratory syndrome in Canada. *N Engl J Med* 4 Apr 2003;[Published online ahead of print] (11 pages) http://www.nejm.org/(Accessed 9 Apr 2003).
11 **Kiazek TG**, Erdman D, Goldsmith C, *et al.* A novel coronavirus associated with severe acute respiratory syndrome. *N Engl J Med* 10 Apr 2003;[Published online ahead of print] (12 pages) http://www.nejm.org/ (accessed 11 Apr 2003).
12 **Drosten C**, Günther S, Preiser W, *et al.* Identification of a novel coronavirus in patients with severe acute respiratory syndrome. *N Engl J Med* 10 Apr 2003;[Published online ahead of print] (10 pages) http://www.nejm.org/ (accessed 11 Apr 2003).
13 **Anderson RM**, May RM. *Infectious diseases of humans: dynamics and control.* Oxford: Oxford University Press, 1991.
14 **Halloran ME**. Concepts of infectious disease epidemiology. In: Rothman KJ, Greenland S, eds. *Modern epidemiology*, 2nd edn. Philadelphia, PA: Lippincott-Raven, 1998:519–54.
15 **Beaglehole R**, Bonita R, Kjellström T. *Basic epidemiology. Communicable disease epidemiology.* Geneva: WHO, 1993:97–105.
16 **Falsey AR**, Walsh EE. Novel coronavirus and severe acute respiratory syndrome. *Lancet* 8 Apr 2003;[Published online ahead of print] (2 pages) http://image.thelancet.com/extras/03cmt87web.pdf(accessed 9 Apr 2003).
17 **Razum O**, Becher H, Kapaun A, *et al.* SARS, lay epidemiology, and fear. *Lancet* 2 May 2003;[Published online ahead of print] (2 pages) http://image.thelancet.com/extras/03cor4133web.pdf(accessed 11 May 2003).
18 **Donnelly CA**, Ghani AC, Leung GM, *et al.* Epidemiological determinants of spread of causal agent of severe acute respiratory syndrome in Hong Kong. *Lancet* 7 May 2003;[Published online ahead of print] (6 pages) http://image.thelancet.com/extras/03art4153web.pdf(accessed 11 May 2003).
19 **CDC**. SARS web site. Atlanta, GA, USA, US Centers for Disease Control and Prevention. http://www.cdc.gov/ncidod/sars/ppt/globalsars.ppt(accessed 15 Apr 2003).
20 **CBC News**. SARS death rate higher than WHO estimate: study. 7 May 2003; http://www.cbc.ca/stories/2003/05/07/sars_study030507(accessed 10 May 2003).
21 **CBC News**. SARS more lethal for older people. 8 May 2003; http://www.cbc.ca/stories/2003/05/08/sars_who030508(accessed 10 May 2003).

PostScript

Siberian live birth sex ratios and the SPrOO hypothesis

Melnikov and Grech found a highly significant seasonal pattern of the sex ratio (SR) at birth in western Siberia, namely, a peak in the second and a trough in the fourth quarter of the year.[1] This peak and trough are in line with the seasonal preovulatory overripeness ovopathy (SPrOO) hypothesis, which states (1) an approach to gender equity at the peak of the "ovulatory" seasons, (2) preferential fertilisation of non-optimally matured oocytes by Y-bearing sperm during the transitional stages between them, and (3) SR reversal because of excess of male biased fetal loss during the "anovulatory" seasons.[2] The mentioned peak would correspond, just like in non-human mammals with the breakdown of the "ovulatory" season in spring; the trough with SR reversal during the "anovulatory" season in winter.

The authors wonder about the annual secular trend in a fall and then rise of the SR with the turning point in early 1980s, being different from the continuous decline in industrial countries over the past half century. We have argued that the rise in SRs before the first world war in Finland[3] and in many other countries concurred with continuous improvement in living conditions, education, and reproductive hygiene, and, thus, a decrease of conceptopathology rate, and in turn, increase of male surviving fetuses.[4] The fall of the SRs after the turning point around the second world war was interpreted as consequence of further amelioration of the ovulatory and conception pattern, reflected by the concurrent decrease in pregnancy wastage. The same reasoning accounts for the initial very low and then increasing and again decreasing SRs in developmental countries that are in demographic transition going hand in hand with amelioration of socioeconomic conditions. This may be compared with the rise and fall in socially upward family conditions.[5] The odds for delivering a male child increases (while pregnancy loss diminishes) when the socioeconomic level increases from low to moderate up to a plateau and then decreases (despite continuation of decreasing pregnancy wastage) when this level increases further from moderate to higher.

The relatively low SRs in western Siberia, when compared with other countries in west Europe, may also be related to higher rates of conceptopathology because of the extreme climatic variations and inherent stronger seasonal variation in reproduction further away from the equator. This suggests higher rates of male biased pregnancy loss as the underlying mechanism. The SR increase in the early 1980s would mean that this rate is diminishing and that further progress in socioeconomic conditions will result in still higher SRs and ultimately in a decrease, in analogy with those in west European countries.

P H Jongbloet
Department of Epidemiology and Biostatistics, University Medical Centre Nijmegen, PO Box 9101, 6500 HB Nijmegen, Netherlands; p.jongbloet@epib.umcn.nl

References

1 **Melnikov VN**, Grech V. Seasonality of live birth sex ratio in south western Siberia, Russia, 1959–2001. *J Epidemiol Community Health* 2003;**57**:471–2.
2 **Jongbloet PH**. Overripeness ovopathy, Sex ratio increase and sex ratio reversal—a challenging hypothesis for sex ratio modulation. *Hum Reprod* 2004;**19**:769–74.
3 **Vartiainen T**, Kartovaara L, Tuomisto J. Environmental chemicals and changes in sex ratio: analysis over 250 years in Finland. *Environ Health Perspect* 1999;**107**:813–15.
4 **Jongbloet PH**, Zielhuis GA, Groenewoud HMM, et al. The secular trends in male:female ratio at birth in postwar industrialized countries. *Environ Health Perspect* 2001;**109**:749–52.
5 **Teitelbaum MS**, Mantel N. Socio-economic factors and the sex ratio at birth. *J Biocsoc Sci* 1971;**3**:23–41.

Siberian pros and cons of the SPrOO hypothesis

I thank Dr Jongbloet for thoughtful response to our article. According to the SPrOO hypothesis, human females have a fundamental seasonal variation in ovulatory pattern, one of several factors explaining differences in sex ratio (SR). The hypothesis presumes inherent ovulatory and anovulatory seasons and suggests that secondary SR varies over the year from the femininity, or an approach to gender equality, coinciding with the zenith of birth frequency, to the excess of male births concentrating at the beginning and the end of this optimum. The question is what are these seasons in Siberia?

An analysis of our data by month yielded the seasonal pattern for SR variation (p<0.1, Edwards' test). According to the hypothesis, the birth optimum should occur in February–March when the small trough in SR is observed. This period of gender equality corresponds to the "ovulatory" season in May–June. This trough in turn was preceded by a sharp peak in SR occurring in January and followed by the major peak in April–June. These two peaks might seem to reflect "transitional stages".

What are the arguments against the hypothesis?

Huntington's 1938 report of the seasonal variation in both the SR and the total number of births in seven countries, based on analysis of about 52×10^6 births, showed an *inverse* relation between the SR and the number of births.[1] Just the contrary is apparent in Siberia. These two curves agree closely at least during the first half of the year. In other words, the total number of births and the number of male births vary correspondingly. This direct relation between SR and birth number seems to be a characteristic feature of the Siberian population.

An analysis of 1989 Novosibirsk census data shows that January born males and females comprise 20.8% and 24.2%, respectively, of men and women aged more than 80 years, whereas the expected proportions according to the uniform distribution would be $100/12 = 8.3\%$. This means that January as a month of birth and April as a month of conception are strong predictors of longevity in Siberia. However, in accordance with the hypothesis, April ought to be a month of conceptopathology, associated with "preferential fertilisation of non-optimally matured oocytes by Y-bearing sperm" and would consequently not seem to be associated with the surprisingly long span of life seen in this cohort.

In summary, while I agree that some aspects of the Siberian data are "in line with" the SPrOO hypothesis, I do not find this concept to be the most satisfying explanation of our findings. While the hypothesis has been supported by animal studies, its reliance on mechanisms such as ovopathy and differential pregnancy loss makes it difficult to establish (or refute) on the basis of studies such as our own. We remain persuaded that A Lerchl's hypothesis of the temperature dependence of SR better explains our peak of total and male births in January.[2]

V N Melnikov
Siberian Independent Institute, P O Box 175, Novosibirsk, 630060, Russia; mevlanic@nsk.ru

References

1 **Huntington E**. Sex, season and climate. In: *Season of birth: its relation to human abilities*. New York: Wiley, 1938:192–214.
2 **Lerchl A**. Sex ratios at birth and environmental temperatures. *Naturwissenschaften* 1999;**86**:340–2.

Ethnicity and epidemiological research: not so black and white

The analysis by Ahern *et al* of risk factors for preterm birth among African American and white women in San Francisco[1] concluded that pregnant African American smokers are more prone to preterm delivery than white pregnant smokers. This conclusion is misleading. Firstly, the evidence of interaction between smoking and ethnicity was unconvincing—the difference in the odds ratios

(ORs) was modest, and confidence intervals (CI) overlapped considerably (African American women: OR 1.77, 95% CI 1.12 to 2.79; white women: OR 1.25, 95% CI 1.01 to 1.55). Secondly, the authors did not consider residual confounding by factors such as maternal infection and previous preterm birth, which differ by ethnic group.[2 3] Their assumption that the smoking-preterm birth association is linear seems biologically unlikely and problematic, as African American women in their study population smoked more than white women.

Such analyses raise a more fundamental issue: the limitations of using ethnicity/race in epidemiological studies of causality. There is no consensus as to what "ethnicity" means. To some, "ethnicity" describes cultural differences between populations definable by phenotype, while "race" signifies differences strictly under genetic influence. This separation is imperfect—"ethnic groups" differ in genetic mix, culture, and socioeconomic situation. The "catch all" quality of ethnicity makes its use in epidemiological research attractive, but is an obstacle to causal inference. Ethnic groups have qualitative and contextual differences that do not make them directly comparable. Ethnic differences in biological effects will be difficult to uncover, as cultural, social, and economic factors that act as causal intermediates are too numerous and divergent to be adequately controlled.

Finally, ethnicity is not a singular exposure that can be turned on and off. The observation of higher preterm birth rates among black than among white women smokers is not simply the answer to the question "what would rates of preterm birth among white women who smoke be, if they were black?" Social inequity and deprivation, however, are amenable to change in a way that ethnicity is not. In the paper by Ahern et al, African American women delivering preterm clearly were more often single, working class, receiving public insurance, and had lower level of education. Yet their analysis stratified by ethnic group does little to explain these great disparities between African American and white women, concentrating instead on variations within ethnic groups. An analysis of their data comparing preterm birth between social strata rather than ethnicity would shed some light on the role of ethnicity in research: useful tool or epidemiological distraction.

C C Tam, S J Lee, L C Rodrigues
Department of Infectious and Tropical Diseases, London School of Hygiene and Tropical Medicine, London, UK

C C Tam
Gastrointestinal Diseases Department, Health Protection Agency Communicable Disease Surveillance Centre, London, UK

Correspondence to: Mr C C Tam, Infectious Disease Epidemiology Unit, Department of Infectious and Tropical Diseases, London School of Hygiene and Tropical Medicine, Keppel Street, London WC1E 7HT, UK; clarence.tam@lshtm.ac.uk

References

1 **Ahern J**, Pickett KE, Selvin S, et al. Preterm birth among African American and white women: a multilevel analysis of socioeconomic characteristics and cigarette smoking. *J Epidemiol Community Health* 2003;**57**:606–11.
2 **Fiscella K**. Racial disparities in preterm births. The role of urogenital infections. *Public Health Rep* 1996;**111**:104–13.
3 **Adams MM**, Elam-Evans LD, Wilson HG, et al. Rates of and factors associated with recurrence of preterm delivery. *JAMA* 2000;**283**:1591–6.

Authors' reply

In their letter, Tam et al present several critiques of our analysis of smoking and preterm delivery in African American and white women and then broadly criticise the use of race and ethnicity as factors in epidemiological research. We address each issue in turn.

Firstly, Tam et al object to the fact that the odds ratios (ORs) for the effects of smoking on preterm delivery for African American women and white women in our analysis were only modestly different and that their confidence intervals overlap. However, even a small difference in risk can translate to a very large difference in the number of people affected on the population level. In addition, the ORs presented in our paper were for 10 cigarettes per day. The ORs for a pack of 20 cigarettes per day were 3.13 for African American women and 1.55 for white women, a more substantial difference. More importantly, the key difference worth examining is that of the parameter estimates on the log odds scale. The change in the log odds of preterm delivery for every cigarette smoked per day was 0.057 for African American women, while it is only 0.022 among white women—a difference of more than twofold. Ultimately, while there was some overlap in the confidence intervals presented, this observation is not equivalent to a statistical test of the difference between the values. Such a test was not used because of the stratified nature of our sample; we chose a stratified sample for methodological reasons that we address here. We believe the interpretation of point estimates is still appropriate in our study.

Next, the authors suggest our effect estimates may be subject to residual confounding. This clearly may be the case with any observational study. However, in this particular instance, the factors that Tam et al discuss as omissions on our part are potentially on the causal pathway between our exposures and outcomes of interest, and thus, should not be adjusted for in our models. If prior cigarette smoking contributed to an earlier preterm delivery, adjusting for prior preterm delivery might have washed out the very effect we were interested in studying. The issue of maternal infections was explicitly discussed in our paper as a factor that may interact with smoking in its relation to preterm delivery. While we were unable to examine this in our analysis, as we lacked data on maternal infections, we hope that this work and our discussion will encourage others to look at this question.

Tam et al also question the linearity of the relation between smoking and preterm delivery. In our analysis, we found no evidence of non-linearity in the relation between smoking and preterm delivery. The fact that African American women smoked more on average does not affect the validity of the risk estimates, as cigarette smoking was modelled as cigarettes per day and the overall range of values was similar for African American and white women.

Finally, Tam et al criticise our decision to examine effects of neighbourhood and individual risk factors for preterm delivery within, rather than across, racial groups. Our decision to do so was inspired by the lack of such prior research and the fact that an expert conference on the problem of racial disparities in preterm delivery had called for just such an analysis.[1]

We agree with Tam et al that race encompasses social, economic, and cultural factors. Of course this makes the study of race complex. However, complexity has never before been a reason to abandon the study of an issue. In the context of the United States, the history of racial oppression and strife makes this a crucial issue to study in the hope of understanding its effects on people's lives. The differences in the rates of preterm delivery between African American and white women are independent of socioeconomic status[2 3] making racial differences even more important to understand. The complexity of the question, far from scaring us away, should draw us in. While race is not amenable to change, there may be underlying reasons for a racial disparity in health that may in fact be amenable to change. A differential effect of smoking on preterm delivery by race, suggested in our analysis, may be one such underlying factor contributing to racial disparities. We hope our analysis will lead to studies that help us intervene on one aspect of this important health problem and cause of racial disparities in health.

J Ahern, K E Pickett
New York Academy of Medicine, 1216 5th Avenue, Room 553, New York 10029, USA; jahern@nyam.org

References

1 **Rowley DL**, Hogue CJR, Blackmore CA, et al. Preterm delivery among African-American women: a research strategy. *Am J Prev Med* 1993;**9**:1–6.
2 **Berkowitz GS**, Papiernik E. Epidemiology of preterm birth. *Epidemiol Rev* 1993;**15**:414–43.
3 **Foster HW**, Wu L, Bracken MB, et al. Intergenerational effects of high socioeconomic status on low birthweight and preterm birth in African Americans. *J Natl Med Assoc* 2000;**92**:213–21.

Spending more, feeling worse: medical care expenditures and self rated health

International comparisons show that the United States spends more on health care than other industrialised nations.[1] According to 2000 data, the United States led the way in per capita healthcare spending at $4631, more than double the Organization for Economic Cooperation and Development (OECD) median of $1983 (in purchasing power parities based on the US dollar).[2]

Despite massive medical care expenditures, the US lags behind its industrialised counterparts in major indicators of population health. For example, researchers have reported that the US has lower life expectancy at birth and higher maternal and infant mortality.[3 4] Equally important, cross national studies consistently show that US citizens are less satisfied with their healthcare system than Canadians or Europeans.[2]

Another key indicator of population health, not previously used in cross national comparisons, is self rated health (SRH). As a broad measure of health status, SRH focuses on a person's subjective perceptions of their
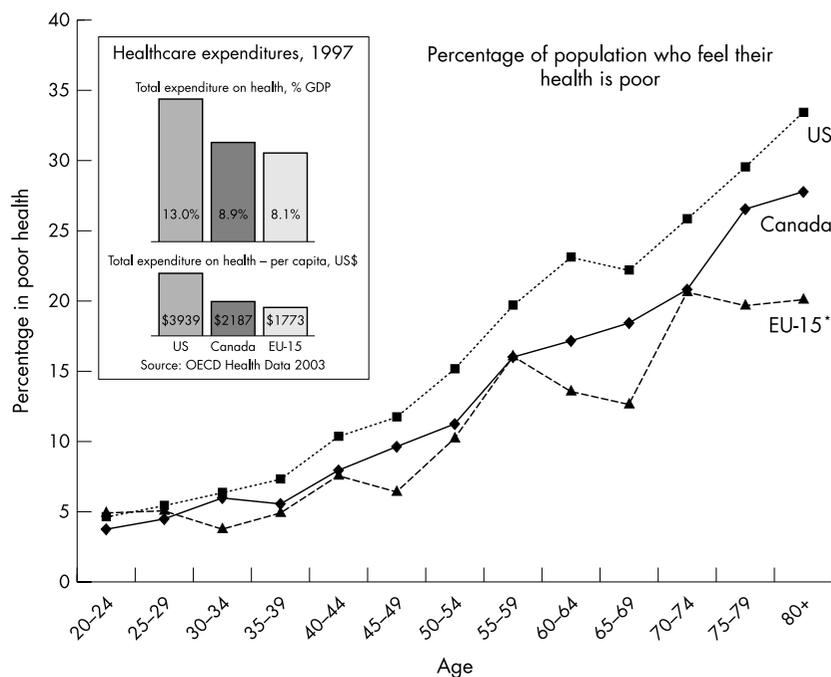
Figure 1 National healthcare expenditures and self rated health. Data sources: National Population Health Survey 1996–97 (Canada), National Health Interview Survey 1996 (US), and European Commission Eurobarometer 1996 (EU-15). *Austria, Belgium, Denmark, Finland, France, Germany, Greece, Ireland, Italy, Luxembourg, Netherlands, Portugal, Spain, Sweden, and United Kingdom.

health. This popular single item measure on both sides of the Atlantic is related to mortality and to rates of recovery from episodes of illness across the life span.[5] Drawing on comparable datasets, we asked whether the prevalence of poor self rated health is higher in the US than in Canada or the European Union.

Data (weighted) were obtained from the 1996/97 Canadian National Population Health Survey—NPHS (n = 66 435), the 1996 US National Health Interview Survey—NHIS (n = 63 402), and the 1996 European Commission Eurobarometer (n = 16 235). Age specific self rated health was assessed using the NPHS and NHIS questions "In general would you say your health is excellent, very good, good, fair, or poor?" In the Eurobarometer, self rated health was assessed with "Over the last 12 months, would you say your health has on the whole been very good, good, fair, bad or very bad?" In our analysis, "poor health" refers to responses of "fair" or "poor" in the NPHS and NHIS and to "bad" or "very bad" in the Eurobarometer. Data for national healthcare expenditures were obtained from the OECD (http://www.oecd.org/health/healthdata).

Figure 1 shows that the US, Canada, and the European Union (EU-15) differ in terms of the prevalence of poor SRH. While outspending Canada on a per capita basis by 1.8 to 1 and the EU-15 by 2.2 to 1,[6] the US had higher prevalences of poor self rated health across the age span.

Overall, Canadians and Europeans spend about half of what Americans spend on health care yet feel better in nearly all age categories. As an approximate significance test for differences in average SRH, we computed *t* values pooling the standard errors across studies. The results indicated there were no significant differences between samples at the youngest age groups, but for middle age and older groups, beginning with age 30, Americans had significantly ($p < 0.05$) higher prevalence of poor SRH. Interestingly, disparities in SRH seem to increase with advancing age, underscoring the need for better care and more complete coverage across the entire life span.

Our study confirms and complements earlier research indicating that the expensive and technologically advanced healthcare system in the US does not yield population health outcomes (including SRH) comparable to those in countries with much lower spending whether measured per capita or as a percentage of gross domestic product (GDP).[2 3 7] Unequal and uncoordinated provision of care along with other inefficiencies in the US health system[7] may explain why Americans spend more but feel worse.

Acknowledgements

M S Kaplan
School of Community Health, Portland State University, USA

B H McFarland
Department of Psychiatry, Oregon Health and Science University, USA

J T Newsom
Institute on Aging, Portland State University

N Huguet
School of Community Health, Portland State University

Correspondence to: Dr M S Kaplan, School of Community Health, Portland State University, PO Box 751, Portland, Oregon 97207, USA; kaplanm@pdx.edu

References

1 **Huber M**. Health expenditure trends in OECD countries, 1970–1997. *Health Care Financ Rev* 1999;**21**:99–117.
2 **Anderson GF**, Reinhardt UE, Hussey PS, *et al*. It's the prices, stupid: why the United States is so different from other countries. *Health Aff* 2003;**22**:89–105.
3 **Miringoff M**, Miringoff M-L. *The social health of the nation.* New York: Oxford University Press, 1999.
4 **United Nations**. *Demographic yearbook 2000.* New York: Department of Economic and Social Affairs, Statistical Office, United Nations, 2002.
5 **Bowling A**. *Measuring health: a review of quality of life measurement scales.* 2nd edn. Buckingham: Open University Press, 1999.
6 **Organisation for Economic Cooperation and Development**. *OECD health data 2003.* Paris: OECD, 2003.
7 **Blendon RJ**, Schoen C, DesRoches C, *et al*. Common concerns amid diverse systems: health care experiences in five countries. *Health Aff* 2003;**22**:106–21.

BOOK REVIEWS

Reducing the burden of headache

Although not specifically stated this book represents the proceedings of the 11th international headache research seminar held in Copenhagen in March 2002. This 11th volume in the series *Frontiers in Headache Research* aims to bring the reader up to date on studies of the burden of headache. All the books in the series follow a similar structure: "Each topic is introduced by one or more overview chapters by leading experts. Thereafter, data from new studies are presented in short articles. Finally, each topic closes with a summary of discussion by the chairpersons."

This volume of 68 chapters is divided into six sections. In the preface the editors state that migraine, according to WHO, is number 20 of all diseases with regard to years lived with disability. They also state that migraine has more impact on quality of life than many other diseases generally considered to be more serious. In section I and II, headache related disability and quality of life are discussed in more detail. Section III deals with family burden, comorbidities, and healthcare utilisation for headache. Section IV discusses economic aspects of headache, whereas the two last sections describe and analyse guidelines and intervention aimed at improving health care for headache.

The studies presented at the 2002 Copenhagen meeting, and subsequently published in this volume, are described comparatively briefly and with variable quality. The most interested readers will probably need to search for more detailed information from

those studies also published elsewhere. However, an interesting summary discussion highlights key points at the end of each section. From my point of view, this book should be on the reading list of everyone involved in headache research, and may also be of interest to epidemiologists as well as primary healthcare workers.

K Hagen

The making of a global controversy

Edited by M W Bauer and G Gaskell. Cambridge University Press, Cambridge, 2002. (Pp 411; $75.00). ISBN 0-52177-317-2

Experts in different social and scientific fields have been coordinated by editors Martin W Bauer and George Gaskell with the aim to analyse, from different angles, one of the more important and dynamic socioeconomic phenomenon of the recent years: the biotechnology. As the editors expose in the volume presentation: *"This book takes up themes explored at a conference at the Science Museum London, in 1993, which was convened to explore the structures and functions of resistance in the development of new technologies"*. The volume holds the main thesis of the conference, which could be resumed as follow: *"resistance is not a problem of the public opinion rather it is a signal that acts as a catalyst for organisational and institutional learning"*.

The book comprises a total of 13 chapters gathered in five parts, which include the results of a four year international research project conducted between 1996 and 1999 and called "Biotechnology and the European public". In part I, named "The framing of a new technology: 1973–1996", the authors go over bioethics debate and biotechnology regulation during the past 25 years in Europe, throughout five contributions. Part II, deals with "Public representation in 1996"; it includes four chapters about traditional blue and green resistance, the structure of public perceptions, and the image of the genetic engineering such as a way of nature manipulation. Part III highlights "The watershed years 1996–97" describing two emblematic examples: the modified soya and the cloned Dolly. In part IV, and under the attractive title of "The transatlantic puzzle", the authors analyse how genetically modified seed, food, and pharmaceuticals became a reality, in which great companies like Mosanto emerged from USA, although when extended to Europe, they found a controversial socio-political atmosphere. Finally, part V includes a unique chapter dealing with the biotechnology movement that integrates all the book contents under the highline "Towards a social theory of new technology".

In short, the book attempts to highlight a social representation about biotechnology and public opinion in occidental countries during the past decade. Appropriately, its contents could be resumed as *"the making of a global controversy"*.

M A Peinado

CORRECTION

Two typographical errors in the equations occurred in this paper by Drs Choi and Pak (2003;**57**:831–5).

Wrong equation	Correct equation
$C_{t \cdot i}$	$C_{t \cdot i} = R_0{}^t.$
$D_{t \cdot i \cdot d} = C_{t \cdot i} \; X \; F.$	$D_{t \cdot i+d} = C_{t \cdot i} \; X \; F.$

The complete set of correct equations is therefore:

PREDICTING THE NUMBER OF SARS CASES

Mathematically, the predicted number of incident cases on day $t \cdot i$, that is, $C_{t \cdot i}$, where t is time expressed in the number of incubation periods, is

$$C_{t \cdot i} = R_0{}^t.$$

The predicted total number of cases (C) is

$$C = \Sigma C_{t \cdot i}.$$

PREDICTING THE NUMBER OF SARS DEATHS

Mathematically, the predicted number of deaths on day $t \cdot i+d$, that is, $D_{t \cdot i+d}$, where t is time expressed in the number of incubation periods, is

$$D_{t \cdot i+d} = C_{t \cdot i} \; X \; F.$$

The predicted total number of deaths (D) is

$$D = \Sigma D_{t \cdot i+d}$$