Use of computerised general practice data for population surveillance: comparative study of influenza data

Neil Johnson, David Mant, Lesley Jones, Tony Randall

Abstract

Objective—**To assess the potential for using routine computerised general practice data for surveillance of illness.**

Design—**Comparison of the incidence of influenza during the 1989 epidemic derived from a computerised database with that derived from the Royal College of General Practitioners's weekly returns service—a well established predominantly manual surveillance system.**

Setting—**433 general practices throughout the United Kingdom that used a commercial computer system linked to a central databank.**

Main outcome measure—**Incidence of influenza.**

Results—**The slope of the influenza epidemic curve was essentially the same whether derived from the routine computerised data or royal college's weekly returns service data, and the computerised data were geographically consistent. Throughout the study period, however, the computer derived incidence was between one third and one quarter of that derived from the royal college's system (which is served by practitioners trained in surveillance methods). The peak weekly rates were 164 cases per 100 000 for the computerised system and 583 cases per 100 000 for the royal college's surveillance system.**

Conclusions—**The apparent underreporting in the routine computerised data probably reflects lack of motivation and experience in disease surveillance and haphazard computer entry (particularly of consultations that took place outside of the surgery and consultations that did not result in a prescription), along with overestimation of the population under surveillance. Nevertheless, routine computerised surveillance allows rapid data collection from a large number of practices over a wide geographical area and would greatly augment existing methods.**

Introduction

Surveillance of illness through general practice provides rapid information on changes in the incidence of disease and is well established in the United Kingdom.[1 2] It is achieved by regular intermittent recording of morbidity data at consultations with general practitioners in so called "spotter" practices. The largest surveillance scheme is the weekly returns service coordinated by the Royal College of General Practitioners research unit in Birmingham,[3] but there are similar groups in Wales and Scotland (coordinated by the Public Health Laboratory Service Communicable Disease Surveillance Centre (Welsh Unit) and the Communicable Disease (Scotland) Unit) and some local groups such as the Oxford sentinel practices.[1] In most spotter practices the data are collated and transferred by hand, which inevitably leads to delay in the availability of the information. This delay may be important in the identification of the early phases of an epidemic,[4] and McCormick has suggested that such delay is avoidable by the use of computer recording and transfer of electronic data.[5]

About half the general practices in the United Kingdom are now using computers[6] and many are entering morbidity data into their computers. These data are still incomplete but may already be of sufficient quality to be useful for public health surveillance. Although data from the average practice are unlikely to have yet achieved the precision achieved by the current sentinel practices, they may be adequate to indicate a change in incidence. There are also many more "average" than sentinel practices, and those using commercial systems have the potential to transfer information to a central database with minimum delay. AAH Meditel, the company that supplied the data analysed in this paper, was collecting data electronically each night from almost 550 practices scattered across the United Kingdom at the time of the study. It therefore seemed possible that the data routinely collected by Meditel and other commercial companies could provide warning of changes in the incidence of disease earlier than the sentinel system because of more rapid data transfer, less random error due to small numbers, and better geographical coverage. To assess this possibility reports of influenza on the AAH Meditel database during the last 10 weeks of 1989 were compared with data collected by trained sentinel practices as part of the Royal College of General Practitioners' weekly returns service.

Methods

AAH Meditel computers are currently installed in about 850 practices throughout England, Scotland, and Wales. All general practitioners providing data have received free computers under the no cost option and have a monitor and keyboard on their desk. Although they are required to use the computer only for prescribing and to enter the medical diagnosis for which the prescription was issued, many general practitioners collect morbidity data on all consultations, and some enter home visits. Practice computers are linked by telephone line to a mainframe computer based at AAH Meditel. Each night (except once a week when maintenance takes place) the practice computer downloads to a network from which the mainframe computer collects the data added since the last contact.

We considered the period from 18 October to 26 December 1989. During that period 548 practices had full telephone links with the network and thereby with the mainframe computer and were thus eligible for inclusion in the study. It was clear, however, that not all of these practices were providing full data. To ensure maximum accuracy of the data the study was divided into four week periods. The data from a practice was included only if during the four weeks two

Department of Public Health and Primary Care, University of Oxford, Radcliffe Infirmary, Oxford OX2 6HE
Neil Johnson, MRCGP, *research training fellow*
David Mant, MRCGP, *lecturer in general practice*
Lesley Jones, BA, *research assistant*
Tony Randall, MRCGP, *research fellow*

Correspondence to: Dr Johnson.

BMJ 1991;302:763-5

criteria were met. Firstly, to ensure accuracy of the denominator only practices whose list size was available for that period were included. Thus the incidence of influenza calculated for any one date was based on the denominator for the four week period in which that date fell. This made some allowance for the changes in individual practice list sizes that occur with time and for the inclusion (or exclusion) of the few practices that began (or stopped) providing morbidity data during the study period. Secondly, to maximise the accuracy of the numerator only practices that had recorded at least one episode of any respiratory tract disorder during that period were included. We thought that practices that did not record any respiratory illness during a four week period were unlikely to be recording useful morbidity data.

In all, 115 practices failed to meet one or other of these criteria for the whole of the study period and could not be included at any stage of the study (26 (23%) did not meet the list size criterion and 89 (77%) did not meet the recording criterion). This left 433 practices available for inclusion in the study. The mean denominator (mean total population under surveillance) during the study was 2 613 998. The table gives the variability about this mean during the study period. Additionally, the practices in England were separated into north (of a line drawn between Chester and Lincoln), south (of a line between Bristol and Colchester), and midlands (between these two lines).

Mean population denominators and variability about mean values during study, for data recorded on AAH Meditel practice computers

	Mean	Variability above and below mean (%)*
England (north)	754 213	+4, −4
England (midlands)	839 110	+2, −3
England (south)	891 852	+2, −3
Scotland	43 584	+3, −5
Wales	85 239	+11, −17
Total	2 613 998	+2, −2

*The difference between the mean and the maximum denominator (positive value) or the minimum denominator (negative value) expressed as a percentage of the mean denominator.

The table also gives the mean population figures and the variability for each region. The population base for the Royal College of General Practitioners's data was about 450 000.[7]

For this study the diagnostic categories included from the royal college's weekly returns service were epidemic influenza (International Classification of Diseases code 487) and influenza-like illness (ICD code 460(rdr)). The Read code[8] used to identify influenza on the Meditel database was .H35 (influenza and its complications), which included the comparable Read code for influenza-like illness (.H35z). The diagnoses were based on clinical grounds alone. Only first consultations were included. Thus if a patient had several consultations because of influenza data were included only once.

The data are presented at half weekly intervals. Each point on the graphs represents the rate of reports of first consultations for influenza per 100 000 population for the seven days up to and including the date indicated on the graph. Thus a weekly rate is plotted for every half week. The royal college's data for the same time periods were abstracted[7] and were available only on a weekly basis.

Results

Figure 1 gives the weekly influenza rate for the computer derived data and the royal college's weekly returns service data. Week 42 began on 18 October and week 51 ended on 26 December. Figures are plotted weekly for the royal college's data and every half week for the Meditel data. Figure 1 (top) shows the incidence (rate per 100 000) on a linear vertical axis plotted against time. The peak rate was 3·54 times greater for the royal college's data (583·4) than for the computer generated data (164·5). The curves begin to rise at the same time (week 45), although the peak for the royal college's data occurs earlier—during week 49 rather than week 50. Figure 1 (bottom) shows the same information on a logarithmic vertical axis to compare the rate of change in incidence recorded by both systems. The incidence of influenza according to the royal college's data remained consistently three to four times higher both before and during the epidemic, but the slopes of the curves are essentially the same.

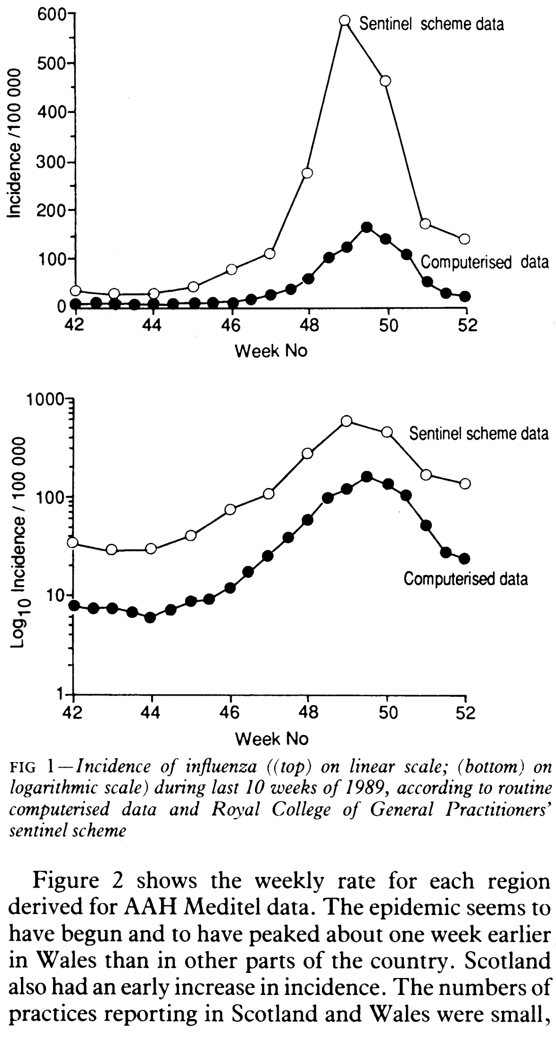

FIG 1—*Incidence of influenza ((top) on linear scale; (bottom) on logarithmic scale) during last 10 weeks of 1989, according to routine computerised data and Royal College of General Practitioners' sentinel scheme*

Figure 2 shows the weekly rate for each region derived for AAH Meditel data. The epidemic seems to have begun and to have peaked about one week earlier in Wales than in other parts of the country. Scotland also had an early increase in incidence. The numbers of practices reporting in Scotland and Wales were small,

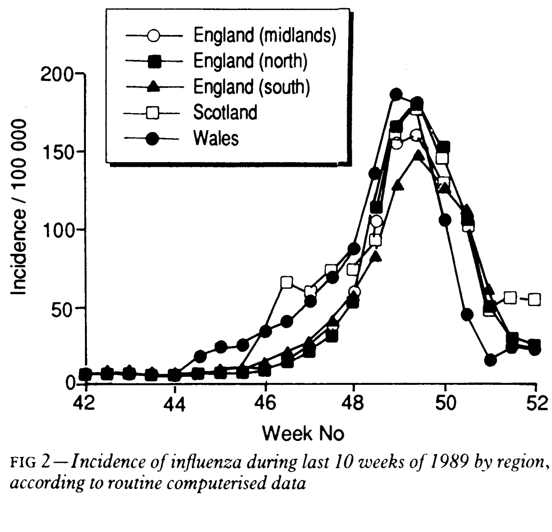

FIG 2—*Incidence of influenza during last 10 weeks of 1989 by region, according to routine computerised data*

however, and consequently the epidemic curves are less stable than for other regions. The most striking feature is that the peak incidence of influenza in each region occurred within one week. This not only shows the homogeneous nature of the epidemic in the United Kingdom but also the geographical consistency of the data.

The rates for individual practices varied widely. In week 44 (which had the lowest mean rate) the range of rates for individual practices was 0-347 per 100 000. The 10th centile and median rates were 0, the 90th centile rate was 32. The rate of 347 per 100 000 represented 12 cases in a practice with a list size of 3460. In week 49 (with the highest mean rate) the range of rates for individual practices was 0-1897 per 100 000. The 10th centile rate was 0, the median 101, and the 90th centile 410. The rate of 1897 per 100 000 represented 38 cases in a practice with a list size of 2010.

Discussion

It is clearly possible to monitor changes in the incidence of influenza by using routinely available general practice data from a commercial database. The shapes of the epidemic curves produced by the computer derived data and by the Royal College of General Practitioners's weekly returns service data were essentially the same. The rates from the computer derived data were considerably lower, however, than those from the royal college. There are several likely reasons for this.

The most important reason is the differing motivation for general practitioners in each of the two systems. The data for the royal college's service are provided by general practitioners who volunteer to complete manual records for all consultations. They are likely to be highly motivated to maintain accurate surveillance data and are therefore likely to include most, if not all, events. The computer derived data presented here were recorded by general practitioners who were likely to have installed computers for organisational reasons; epidemiological reasons were not likely to be foremost. They may, therefore, be less likely to be motivated to include all events. The second possible reason is that general practitioners experienced in providing morbidity data are likely to be experienced in deciding the disease category into which a non-specific presentation of illness fits. Those without this experience may be reluctant to put down a specific illness category, and information may be lost in this way. Continued use of the system should result in more confidence and less loss of data. Thirdly, because the main commercial drive is the provision of prescribing data, patients who present with illness but do not receive a prescription may not get included. This may be the case for many people with influenza. Finally, consultations taking place outside the surgery are likely not to be recorded on the computer (which is based at the practice premises).

All of the above factors reduce the recorded incidence by reducing the numerator. The use of computerised data may also lead to an overestimation of the denominator. This is particularly likely with the simple method used in this study of including a practice in the denominator if it reported a minimum of one respiratory illness and had a known list size during a four week period, as some doctors in a practice may be recording information whereas others are not. Thus to improve the accuracy of the denominator alternative methods may be necessary for recording the list size on which the morbidity data of a practice are based. Alternatively, it may be possible to develop a more sophisticated model which looks for variation in the levels of reporting by practices (beyond that expected from a Poisson distribution) and to use such a model to explore why a practice falls beyond the expected distribution.

Two other areas of the study also cause concern. Firstly, it was necessary to exclude 21% of eligible practices because they failed to meet one or both of the two simple criteria for inclusion. If routine computerised data collected by commercial companies is to be useful for illness surveillance the companies must encourage participating practices to include both accurate list sizes and morbidity data that are as complete as possible. Secondly, the enormous variation in the incidence of influenza between individual practices (0-1897 per 100 000) exemplifies the problems of using small numbers of events occurring in small list sizes. If any useful information about the geographical variation of communicable disease is to be gained from computer derived data practices will need to be grouped into suitable units with larger denominators. One such model (using an algorithm for grouping geographically contiguous zones)[9] is under consideration.[10]

Nevertheless, despite the routine computer based system giving an imprecise estimate of incidence, the slopes of the influenza epidemic curves were remarkably similar to those obtained by the weekly returns service, suggesting that the same process was being measured by both systems. In addition, the computer system does have some advantages over the manual system: it requires no specific training of doctors in surveillance methods; it allows immediate data collection and rapid data transfer so that there is virtually no time delay; and it allows rapid processing (and potentially rapid interpretation) of data from large numbers of practices over large geographical areas at low cost in terms of staffing.

In conclusion, a surveillance system using routinely collected data from 10-20% of general practices in the United Kingdom seems to be feasible and may offer some advantages over the systems in use at present.

We thank Mr Vaughan Reed of AAH Meditel for his help and expertise and all the general practitioners who entered information into their computers. Neil Johnson is funded by the Royal College of General Practitioners; David Mant, Lesley Jones, and Tony Randall are funded by the Imperial Cancer Research Fund.

1 Fleming DM, Crombie DL, Mayon-White RT, Fowler GH. Comparison between the weekly returns service and the Oxford regional sentinel practice scheme for monitoring communicable diseases. *J R Coll Gen Pract* 1988;**38**:461-4.
2 Campbell DM, Paixao MT, Reid D. Influenza and the 'spotter' general practitioner. *J R Coll Gen Pract* 1988;**38**:418-21.
3 Fleming DM, Crombie DL. The incidence of common infectious diseases: the weekly returns service of the Royal College of General Practitioners. *Health Trends* 1985;**17**:13-6.
4 Birmingham Research Unit of the Royal College of General Practitioners. Influenza. *J R Coll Gen Pract* 1977;**27**:544-51.
5 McCormick A. French lessons on surveillance of communicable diseases. *BMJ* 1987;**294**:74-5.
6 Statistics and Management Information Division, Department of Health. *Computing survey 1990.* London: Department of Health, 1990.
7 Birmingham Research Unit of the Royal College of General Practitioners. Newly diagnosed episodes of communicable and respiratory diseases, weekly numbers and rates per 100 000 at risk. *OPCS Monitor* 1989-90; **89**:43-90. (Report No 1.)
8 Read JD, Benson TJR. Comprehensive coding. *British Journal of Health Care Computing* 1986;**3**:22-5.
9 Huel G, Petiot J-F, Lazar P. Algorithm for the grouping of contiguous geographical zones. *Stat Med* 1986;**5**:171-81.
10 Allen ME. Influenza surveillance [dissertation]. Reading: University of Reading, 1990.

(*Accepted 6 February 1991*)