Modélisation en temps réel des épidémies de grippe : une analyse de régression linéaire

J.D. Mooney[1], E. Holmes[2], P. Christie[1]

[1] Scottish Centre for Infection & Environmental Health, Glasgow, Royaume-Uni
[2] University of Strathclyde, Glasgow, Royaume-Uni

Les épidémies saisonnières de grippe mobilisent d'énormes ressources des services de santé. Or elles sont connues pour leur variabilité d'un hiver à l'autre. En utilisant les données historiques du système écossais de surveillance sentinelle de la grippe depuis 1972, un modèle potentiel de prédiction des épidémies de grippe a été établi en utilisant une analyse de régression linéaire simple. Il a été appliqué avec un certain succès au cours de la saison hivernale 1999–2000.

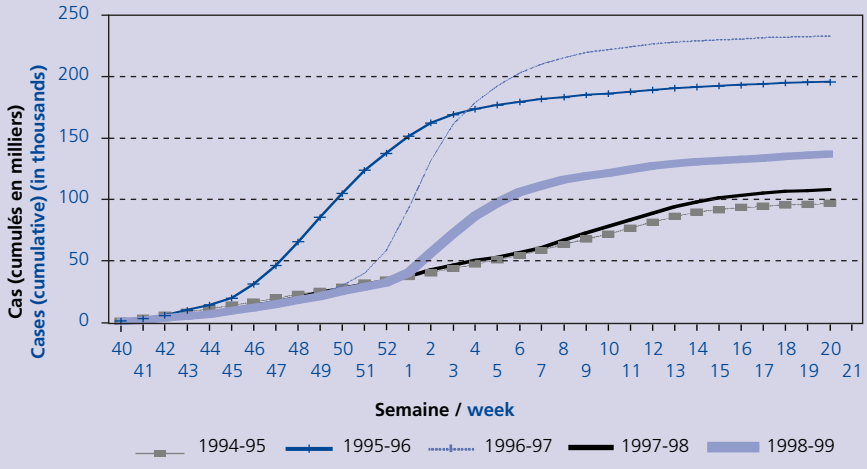
Introduction

Il est difficile de prédire l'ampleur et la sévérité des épidémies de grippe au niveau individuel ou communautaire, même lorsqu'une flambée saisonnière est en cours (1). Deux épidémies de grippe associées à une même souche virale peuvent en effet évoluer de façon très différentes dans leur survenue et dans leur étendue.

En Ecosse, l'indicateur principal de suivi en temps réel du niveau de l'activité grippale est fourni par le réseau sentinelle de médecins généralistes volontaires (2). Ce système de surveillance compte actuellement 90 cabinets de douze régions sanitaires couvrant au total 10% de la population écossaise. Les cabinets participants transmettent chaque semaine le nombre total des consultations pour un syndrome grippal, à partir desquels sont extrapolés des taux de consultations pour 100 000 habitants, selon les projections des déclarations de l'échantillon à la population. Bien que ce réseau fonctionne sur la base du volontariat et qu'il ne couvre pas tous les services sanitaires, ce système de surveillance a prouvé, depuis sa création en 1972, qu'il donnait un indicateur précoce fiable d'un début d'épidémie de grippe saisonnière.

Figure 1

Estimations des consultations cumulées pour symptômes grippaux en Ecosse, 1994–99
Estimated cumulative consultations for flu-like illness in Scotland, 1994–99



Pour les données, voir l'annexe 1 / For data table see Appendix 1

L'analyse des données cumulées durant les saisons antérieures produit des schémas classiques de courbes sigmoïdes, qui permettent d'illustrer la variabilité des épidémies saisonnières de grippe, dans leur survenue et leur amplitude (figure 1). On considère que le taux d'augmentation en milieu d'épidémie, quand l'augmentation des déclarations de cas de syndrome grippal est la plus forte, permet de prédire le total probable de cas pour cette saison, selon les estimations des totaux cumulés de grippe. A partir des données cumulées, la meilleure estimation qui peut être effectuée en milieu d'épidémie saisonnière correspond au taux maximun d'augmentation entre deux semaines consécutives.

Real-time Modelling of Influenza Outbreaks - A Linear Regression Analysis

J.D. Mooney[1], E. Holmes[2], P. Christie[1]

[1] Scottish Centre for Infection & Environmental Health, Glasgow, United Kingdom
[2] University of Strathclyde, Glasgow, United Kingdom

Seasonal outbreaks of influenza exert a considerable burden on health services, and are notorious for their variability from year to year. Making use of historical data from the Scottish sentinelle surveillance since 1972, a potential candidate model has been derived based on simple linear regression. It was applied with a measure of success in the 1999–2000 winter season.

Introduction

Influenza outbreaks are notoriously difficult to predict, even when a seasonal outbreak is underway, both in their likely time course and severity at an individual and a population level (1). Even two subsequent annual outbreaks caused by an identical strain of the virus can have very different impacts on both the timing and the levels of resulting illness in the population.

The major real time indicator of influenza activity in Scotland comes from the sentinel network of volunteer general practices (2). This spotter scheme currently involves 90 practices in 12 health board areas covering a total of 10% of the Scottish population. Participating practices submit weekly totals for the approximate number of consultations for 'flu-like illness' from which can be derived a consultation rate per 100,000 based on population projections from the sample reporting. Although essentially a voluntary set-up, in which not all health boards are represented, the flu-spotter network has proven to be a consistently reliable early indicator of the onset of seasonal influenza illness, since the scheme's inception in 1972.

As well as serving to illustrate the wide between season variability of influenza outbreaks in both timing and magnitude, an examination of cumulative plots for spotter data from past seasons reveals classic sigmoid curves (see Figure 1). It was postulated that the rate of increase at the midpoint of the outbreak, where the rise in reporting for flu like illness is greatest, may be used to predict the likely total number of cases for that season as estimated by the cumulative flu spotter totals. From the cumulative plot, the best approximation that can be measured for the midpoint of a seasonal outbreak would be the maximum rate of increase between any two consecutive weeks.

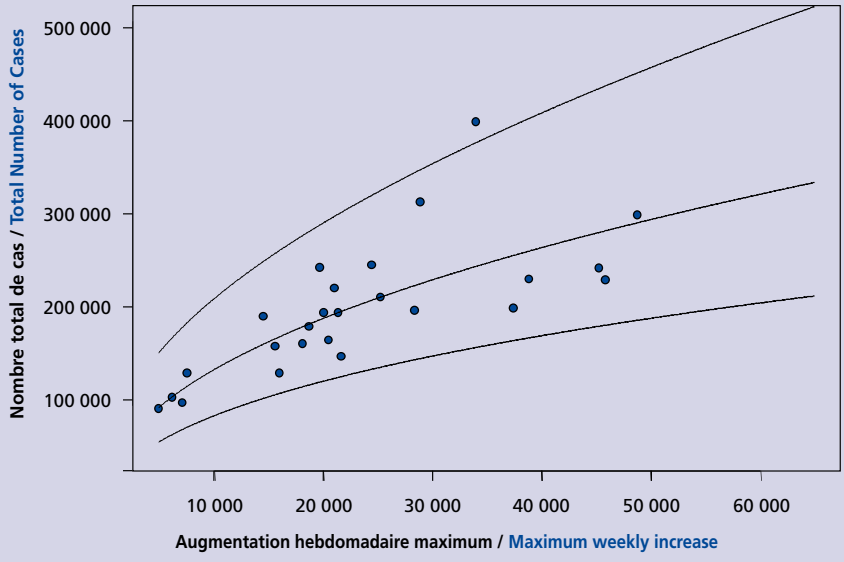
Méthodes

Les données sur les consultations pour un syndrome grippal provenaient du système de surveillance écossais basé sur les déclarations de médecins généralistes de 1972 à 1999. Le nombre total de cas reçus a été estimé chaque semaine pendant la saison de surveillance (des semaines 40 à 20 de l'année suivante) en multipliant le taux global pour 100 000 par 51,2 en Ecosse (population de 5,12 millions d'habitants).

Les différences entre les nombres de cas d'une semaine à l'autre ont été calculées pour chaque semaine, et l'augmentation maximale a été notée pour chaque année. Les logarithmes des séries de données ont été considérés, et un modèle de régression linéaire a été ajusté au nombre total de cas versus l'augmentation maximale observée pour chaque saison entre deux semaines consécutives. Un intervalle de prédiction à 95 % a été calculé pour les nombres totaux de cas attendus en fonction de l'augmentation maximale. Le modèle a ensuite été appliqué pour obtenir des estimations hebdomadaires des nombres totaux de cas attendus à partir de la semaine 20 des saisons grippales de 1999-2000 et 2000-01.

Figure 2

Diagramme de dispersion des totaux cumulés de cas de grippe et augmentation saisonnière maximum de 1972 à 1998 / Scatter diagram of estimated total cumulative flu cases and maximum seasonal increase for years 1972 to 1998



Pour les données, voir l'annexe 2 / For data table see Appendix 2

Résultats

Appliquer une régression linéaire simple avec le total cumulé estimé des cas cumulés pour chaque saison versus l'augmentation maximale (d) (qui correspond à l'augmentation la plus forte du taux) (en prenant les logarithmes de ces deux valeurs) aboutit à une corrélation positive significativement ($p<0.005$, $R^2 = 72$ %), qui peut être décrite ainsi (avec un intervalle de prédiction de 95 %*) (figure 2) :

log (total attendu) = 7,5134 + 0,4693 x log (augmentation maximale)

ce qui donne :

total attendu = exp (7,5134) x augmentation maximale ^{0,4693}

Bornes supérieure/inférieure de prédiction =
exp (7,1534 +/- 1,96 x 0,1998) + augmentation maximale $^{0.4693}$

[*IP 95 % basé la déviation standard résiduelle de la ligne ajustée].

Application du modèle

L'utilité du modèle a ensuite été évaluée pour la saison grippale 1999–2000. La hausse la plus forte des taux de consultations des médecins généralistes est survenue entre les semaines 52 et 53, soit un total attendu de 169 057 consultations avec un intervalle de prédiction de 95% ([114 277– 250 096]) pour toute la saison. A la fin de la saison, l'estimation réelle basée sur les chiffres cumulés des semaines 40 à 20 était de 175 787, soit une différence de moins de 5% par rapport ➤

Methods

Data on consultations for influenza like illness was available from Scottish GP spotter practices for the years 1972 to 1999. Estimates for the total numbers of cases seen in each week were derived during the flu spotter season (weeks 40 to 20 of the following year) by multiplying the overall Scottish rate per 100 000 by 51.2 (population 5.12 million).

The differences between the numbers of cases one week and the next week were calculated for each week of the season and the maximum increase for each year was noted.

The dataset was log-transformed and a linear regression model was fitted to the total number of cases vs. the maximum increase seen for each season between any two consecutive weeks. A 95% prediction interval was calculated for the expected total numbers of cases dependent on the maximum increase. The resulting model was then used to provide weekly estimates of the total numbers of expected cases by week 20 during the ongoing flu seasons of 1999–2000 and 2000–2001.

Results

Performing a simple linear regression with the total estimated cumulative cases for each season versus the maximum increase (d) (corresponding to the sharpest rise in the rate) (both log transformed) gives rise to a significant positive correlation ($p < 0.005$, R2 = 72%), which can be described as follows (with 95% prediction interval*) (Figure 2):

log (expected total) = 7.5134 + 0.4693 x log (max. increase)

Giving

Expected total = exp(7.5134) x max increase $^{0.4693}$

Upper / lower PI = exp (7.1534+/-1.96*0.1998) + max increase $^{0.4693}$

[*95% PI based on the residual standard deviation about the fitted line].

Application of the model

The utility of the model was then investigated for the winter flu season of 1999/2000. The sharpest increase in the GP spotter rates occurred between week 52 and 53 and gave rise to an expected total of 169 057 consultations with 95% prediction interval ([114 277–250 096]) for the whole season. At the end of the season, the actual estimate based on cumulative figures from week 40 to week ➤

➤ à la prédiction. Puisqu'il n'y a aucun moyen de connaître à l'avance le changement maximum, la prédiction du total des consultations probables a été revue chaque semaine au cours de la saison, en fonction de l'amplitude des variations par rapport à la semaine antérieure (annexe 3). Les prédictions révisées en continu ont permis de prévoir à partir de la semaine 53 que la grippe de la saison 1999–2000 serait probablement plus sévère que celle de 1998–99, avec une probabilité de 84% (selon la déviation standard de l'intervalle de prédiction) avec un nombre de cas total estimé à la fin de la saison à 137 336.

Le modèle a été revu en incluant les résultats de la saison 1999–2000 :

log (total attendu) = 7,526 + 0,468 x log (augmentation maximum)

ce qui donne :

total attendu = attendu (7,526) x augmentation maximum ^{0,468}

Il a ensuite été appliqué pendant la saison 2000–01, présentant l'activité grippale la plus faible depuis 1972, avec des taux de consultations qui ont rarement dépassé les niveaux de base de 50 consultations pour 100 000 habitants (5). Même avec ce taux très faible d'activité, le total final cumulé des consultations (54 033) restait dans l'intervalle de prédiction (prédiction d'un total de 46 556 consultations ; 95 % PI = [7089–305 775]).

Discussion

Il est difficile de faire des estimations sur les épidémies saisonnières de grippe pour plusieurs raisons : variations antigéniques d'une saison à l'autre, introduction de souches virales nouvelles, proportions élevées des infections asymptomatiques, et controverse permanente sur les facteurs de transmission. Tous ces paramètres empêchent la modélisation ou la définition d'une épidémie de grippe « standard ». Cependant, puisque même une petite épidémie de grippe peut lourdement grever les services de santé, il serait utile de disposer d'un modèle facile à utiliser pour prédire le cours d'une épidémie.

Les principaux inconvénients de ce modèle, s'il devait être utilisé comme outil de prédiction, résident tout d'abord dans les intervalles de prédiction de l'amplitude de l'épidémie, qui sont très larges. Deuxièmement, comme tous les modèles de régression linéaire, il devient moins fiable aux valeurs extrêmes de la série des données sur lesquels il est basé (3). Puisque dans les intervalles de prédiction, la dispersion des données autour de la tendance ajustée est moindre, les intervalles sont toujours plus larges que l'intervalle de confiance équivalent pour les valeurs ajustées (6). En théorie, il devrait également être possible d'affiner le modèle après chaque saison, bien que la nature des intervalles de prédiction implique que leur réduction probable soit faible. La disponibilité croissante de tests virologiques rapides permet d'identifier rapidement les types de virus en circulation qui contribuent à l'augmentation des consultations (par exemple, infection par le virus grippal de type A seul, par le virus de type B seul ou les deux A+B). Etant donné que les différences entre les grippes de type A et B en terme de gravité de la maladie et d'impact sur la santé de la population ont été bien établies (7), l'introduction de termes d'interaction à la régression linéaire selon le type épidémique pourrait améliorer la capacité de prédiction du modèle (8). Un modèle qui tiendrait compte du type de virus impliqué pourrait également mieux prédire l'évolution la plus probable d'une épidémie en cours. Ces considérations comptent souvent dans la planification des services de santé autant que le taux d'attaque dans la population générale.

Bien que les limites du modèle entravent son adoption comme outil de prédiction définitif, son utilité tient plus de sa capacité à fournir une évaluation hebdomadaire dynamique et révisable de la gravité probable d'une épidémie de grippe en cours. Bien que le modèle actuel ne per-

➤ 20 was 175 787, less than 5% difference from the predicted total. Since there is no way of knowing in advance what the maximum change will be, the estimate of total likely consultations was revised weekly throughout the season, based on the extent of change over the previous week (see appendix 3). The continuously revised estimate made it possible to expect by week 53 that 1999–2000 was likely to be more severe than the flu season of 1998–99 with a probability of 84% (based on the standard deviation of the prediction interval), where the total estimated cases at the end of the season was 137 336.

A revised model (incorporating the results of the 1999–2000 season – revised expression:

Log (expected total) = 7.526 + 0.468 x log (max. increase),

giving:

Expected total = exp(7.526) x max increase $^{0.468}$

was then applied during the 2000/01 season, a winter that saw the lowest flu activity since 1972, and spotter rates that rarely exceeded the baseline threshold level of 50 consultations per 100000 population (5). Even at this very low level of activity, the final cumulative total for consultations (54 033) was still within the predicted range (Predicted total = 46 556; 95% PI = 7089,305775).

Discussion

Seasonal outbreaks of influenza are difficult to predict for a number of reasons. The continual antigenic changes between seasons, the introduction of new viral strains, the high proportions of subclinical infections and continuing controversy over factors which affect transmission all combine to frustrate attempts to model or define a 'typical' influenza outbreak. Since even modest influenza outbreaks can exert additional pressures on health services however, the benefits for planning and healthcare purposes of a model that is simple to apply and has some capacity to predict the course of an ongoing outbreak are self-evident.

The main drawbacks to the above model as a predictive tool are firstly the very wide prediction intervals which accompany the estimated eventual size of the outbreak and secondly, like all linear regression models, it becomes less reliable at the extreme ends of the range of the source data on which it is based (3). Since in prediction intervals, the scatter of the individual data about the fitted line becomes more directly relevant, they are invariably much wider than the equivalent confidence interval for the fitted values (6). In theory it should be possible also to refine the model with each additional season, although the nature of prediction intervals means again that their likely reduction will be small. The increasing availability of rapid virological testing also makes it possible to identify quickly the underlying virus types that are contributing to an increase in illness presentation (eg: A alone, B alone or A + B). The well established differences in severity and population health impact between A and B strains (7) may mean that introducing interaction terms to the regression, according to the epidemic type as suggested by Dab et al, could improve the predictive capability of the model (8). A model which took account of virus type may also be able to begin to address the likely time course of an ongoing outbreak, often as important a consideration with regard to health service planning as overall population attack rate.

Although the limitations of the model prevent its adoption as a definitive predictive tool, its usefulness relates more to the capacity to provide a dynamic weekly revisable estimate of the likely severity of an ongoing flu outbreak. While the current model does not specifically address the timing of any peak, large increases in

mette pas de prévoir spécifiquement quand aura lieu le pic d'activité grippale, il n'en reste pas moins que les fortes augmentations des taux de consultations seront vraisemblablement suivies d'une charge de travail plus lourde pour les services de santé. De plus, bien que les taux des consultations ne constituent en aucun cas les seuls indicateurs de l'activité grippale, ils sont certainement les plus rapides. Les réseaux de surveillance comme celui de l'Ecosse sont largement utilisés dans toute l'Europe (9). Des variantes du modèle présenté peuvent donc intéresser d'autres pays ayant des séries de données historiques importantes.

Conclusion

Parmi d'autres indicateurs, Tillet et Spencer ont déjà souligné le pouvoir potentiel de prédiction des totaux cumulés des consultations en médecine générale, pour décrire l'étendue des épidémies de grippe en Angleterre et au Pays de Galles (4). Le modèle présenté dans cet article montre qu'il est possible de décrire le lien entre les nombres totaux cumulés des consultations et l'augmentation hebdomadaire maximale des épidémies saisonnières de grippe, en utilisant une régression linéaire simple qui permet de faire des prédictions sur l'ampleur éventuelle d'une épidémie, avec des corrections au cours de la saison. Le grand intervalle de prédiction observé pendant la saison grippale inhabituellement douce de 2000–01, bien que lié à la difficulté d'appliquer les modèles de régression aux valeurs périphériques, ne constitue probablement pas une limitation sérieuse en pratique, car ce modèle servirait à détecter les épidémies potentiellement très importantes le plus tôt possible. La disponibilité croissante de tests virologiques rapides pourrait aider à affiner des modèles tels que celui présenté ici, en intégrant les données sur les virus en circulation durant une saison donnée. ■

consulting rates are likely to be followed with higher workloads in secondary health services. Additionally, although consulting patterns are not by any means the only indicator of influenza activity, they are certainly the timeliest and sentinel practice networks like that in Scotland are used widely throughout Europe 9. Variations of the presented model may also therefore be of interest to other countries that have a significant historical dataset.

Conclusion

Tillet and Spencer have previously highlighted the potential of cumulative totals of GP consultations, among other indicators, for describing the extent of influenza outbreaks in England and Wales (4). The model presented here demonstrates that it is possible to describe the relationship between cumulative total numbers of consultations and the maximum weekly increase for seasonal outbreaks of influenza using simple linear regression, allowing predictions for the eventual size of an outbreak to be revised as the winter season progresses. The wide ranging prediction interval seen during the exceptionally mild influenza season of 2000–01, although in keeping with the diminishing applicability of regression models at the extremes of their range, is probably not a serious practical limitation in that the main use of the model would be to flag up potentially large epidemics as early as possible. The increased availability of rapid virological testing may make it possible to further refine models such as that presented here, on the basis of the type(s) of influenza in circulation in any one season. ■

References

1. Cliff AD. Statistical modelling of measles and influenza outbreaks. *Statistical Methods in Medical Research* 1993; **(2)**: 43-73.
2. Christie P, Mooney J. Surveillance Report on Flu Spotters data 1999-2000. *SCIEH Weekly Report* 2000;**34**(36):218-219
3. Kirkwood B.R. Correlation and Linear Regression. Chapter 9 in *Essentials of Medical Statistics*. Blackwell Science Ltd. Oxford 1998; p57-64.
4. Tillett HE, Spencer IL. Influenza surveillance in England and Wales using routine statistics. Development of 'cusum' graphs to compare 12 previous winters and to monitor the 1980/81 winter. *J Hygiene* 1982 Feb;**88**(1):83-94.
5. Christie P, Mooney J, Smith A. Surveillance Report on Flu Spotters data and SERVIS scheme 2000-2001. *SCIEH Weekly Report* 2001; **35**(24): 154.
6. Altman D. Relationship between two continuous variables. Chapter 11 in *Practical Statistics for Medical research*. Chapman & Hall. London 1995; p277-234.
7. Monto AS. Individual and community impact of influenza. *Pharmacoeconomics* 1999;**16** Suppl 1:1-6
8. Dab W, Quenel P, Cohen, JM, Hannon C. A new influenza surveillance system in France: the Ile-de-France "GROG".2. Validity of indicators (1984-89). *Eur J Epidemiol* 1991; **7**(6):579-87.
9. Zambon M. Sentinel surveillance of influenza in Europe, 1997/1998. *Eurosurveillance* 1998; **3**: 29-31.