Reviewer's report

Title: Internet-based surveillance of Influenza-like-illness in the UK during the 2009 H1N1 influenza pandemic

Version: 1 Date: 31 August 2010

Reviewer: Cecile Viboud

Reviewer's report:

In this paper, Tilston and colleagues present the results of an internet-based influenza surveillance system established during the 2009 pandemic in the UK. While similar systems have been established in other European countries to monitor seasonal influenza activity, this is the first instance of such a system in the UK. I very much enjoyed reading this paper and have a couple of comments:

Major compulsory revisions

1) The system was initiated during the 2009 pandemic and children (and seniors) were underrepresented. Incidence rates during the pandemic were dramatically skewed towards children. Is this a problem and is it possible that the system would do better in a typical seasonal epidemic where the age distribution of cases is perhaps less skewed? Also, recruitment of participants may have been relatively easy in 2009 given the media attention and public interest generated by the pandemic, but it may be different in the case of seasonal influenza epidemic. It would be useful to discuss how this system could be expanded to monitor seasonal influenza.

2) The authors enrolled 5,000 participants to participate in this survey. What was this number based on? Is there anything like a sample size calculation for internet-based surveillance systems?

3) In the censored internet sample (the most reliable), 2,368 participants completed 17,532 questionnaires, resulting in about 7.3 reports per participant over a 24-week study period. This suggests that participants do not report symptoms on a weekly basis as instructed, and presumably they are more inclined to return their questionnaire when they are sick (or one of their contacts is). It would be interesting to plot the weekly number of questionnaires returned, overlaid with the weekly number of ILI reports (unweighted), to gauge whether reporting is strongly linked with disease. Also, a histogram displaying the distribution of number of reports per participant would be nice.

4) Can the authors provide an estimate of reinfection rates in their system (=2 or more reports of ILI between July and December 2009)?

Minor essential revisions

1) Abstract/results: would be nice to include a quantitative measure supporting the agreement between the internet system and the HPA estimates (eg, the correlation measures)
2) Last sentence of intro: RCGP not defined
3) Rapid onset of fever (top of p.7): how was this defined?
4) The internet surveillance system was initiated in July 2009 and caught only half of the pandemic summer wave in the UK. Is this taken into account in the estimates of attack rates presented in Fig 6?
5) Add P-values to correlation coefficients presented in results

Discretionary revisions

1) It would be interesting to compare the attack rate estimates generated by the internet system with those of serosurveys conducted in the UK after the first wave (Miller et al, Lancet 2010). It would be also interesting to compare the merits of internet-based and telephone-based disease surveillance during the 2009 pandemic (eg, telephone surveys in NYC during the spring wave).
2) I understand where the authors are going with the SIS and SIR concepts p.11; however, traditional influenza GP surveillance systems consider that influenza is a SIS infection (ie, such systems use a denominator based on the number of GPs and their catchment population, and tally the number of ILI cases for a given period of interest)... In the interest of comparing with GP-based or HPA estimates, isn’t it best to use the SIS definition?

Level of interest: An article of importance in its field

Quality of written English: Acceptable

Statistical review: Yes, and I have assessed the statistics in my report.

Declaration of competing interests:

'I declare that I have no competing interests