

Internet-based remote health self-checker symptom data as an adjuvant to a national syndromic surveillance system

A. J. ELLIOT¹*, E. O. KARA¹, P. LOVERIDGE¹, Z. BAWA¹, R. A. MORBEY¹,
M. MOTH^{2,3}, S. LARGE³ AND G. E. SMITH¹

¹ Real-time Syndromic Surveillance Team, Public Health England, Birmingham, UK

² Digital Assessment Service, NHS Choices, Southampton, UK

³ Research and Clinical Audit, NHS Direct, Southampton, UK

Received 24 July 2014; Final revision 5 January 2015; Accepted 25 February 2015;
first published online 10 April 2015

SUMMARY

Syndromic surveillance is an innovative surveillance tool used to support national surveillance programmes. Recent advances in the use of internet-based health data have demonstrated the potential usefulness of these health data; however, there have been limited studies comparing these innovative health data to existing established syndromic surveillance systems. We conducted a retrospective observational study to assess the usefulness of a national internet-based ‘symptom checker’ service for use as a syndromic surveillance system. NHS Direct online data were extracted for 1 August 2012 to 1 July 2013; a time-series analysis on the symptom categories self-reported by online users was undertaken and compared to existing telehealth syndromic data. There were 3·37 million online users of the internet-based self-checker compared to 1·43 million callers to the telephone triage health service. There was a good correlation between the online and telephone triage data for a number of syndromic indicators including cold/flu, difficulty breathing and eye problems; however, online data appeared to provide additional early warning over telephone triage health data. This assessment has illustrated some potential benefit of using internet-based symptom-checker data and provides the basis for further investigating how these data can be incorporated into national syndromic surveillance programmes.

Key words: Internet; public health, syndromic surveillance, telehealth.

INTRODUCTION

Syndromic surveillance is the near real-time collation, interpretation and dissemination of data to underpin early identification of potential public health threats and their impact, enabling public health action [1]. Public Health England (PHE) is an Executive Agency of the Department of Health with a mission

to protect and improve the nation’s health and to address inequalities in health [2]. To support this role, a programme of syndromic surveillance is coordinated, capturing health data from a number of distinct healthcare settings including a telephone triage health advice line service, general practitioners (GPs) (in hours and out of hours) and emergency departments. Data are routinely analysed to monitor the emergence and spread of common infectious diseases in the community in near real-time and to support health protection activities during incidents with the potential to impact on public health [3–6].

* Author for correspondence: Dr A. J. Elliot, Real-time Syndromic Surveillance Team, Public Health England, Birmingham B3 2PW, UK.
(Email: alex.elliott@phe.gov.uk)

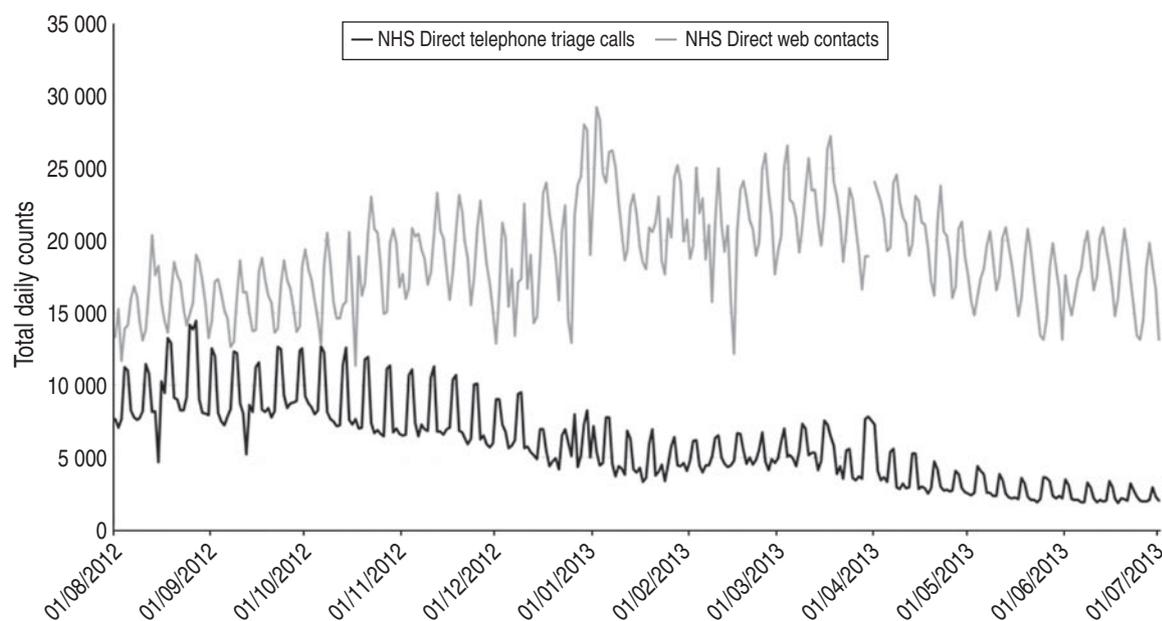


Fig. 1. Time-series analysis of total daily NHS Direct telephone triage calls and NHS Direct web contacts.

Symptom-based telephone triage calls to the NHS supports syndromic surveillance; the NHS Direct syndromic surveillance system has monitored the population of England and Wales from 2001 [3]; however, national NHS Direct services were replaced by a new telephone health service, NHS 111 during 2013/2014 [7]. NHS Direct used a series of clinical assessment algorithms to evaluate the symptoms of each patient, record the predominant presenting symptom or complaint and provide clinical advice to the patient regarding the need for further healthcare, including advice for self-care, referral to an emergency department, referral to urgent GP care, or referral to routine GP care. The NHS Direct service was accessible all day, every day and provided a reliable and continuous feed of data that were utilized by the PHE Real-time Syndromic Surveillance Team (ReSST) to form the basis of the NHS Direct syndromic surveillance system [3]. Following the completion of the phased national replacement of NHS Direct with NHS 111, ReSST has developed a new syndromic surveillance system using NHS 111 data [8].

In 2010, the NHS Direct digital services, including an online self-assessment service, were launched. The aim of the digital services was to enable choice for patients and to create the potential for greater cost-effectiveness by substantially increasing the number of people able to access NHS Direct at any one time [9]. The NHS Direct symptom checker was available to the patient in the form of a web platform providing

a number of different questions describing a range of different health concerns (e.g. problems with ear, nose and throat). A further series of more specific symptom assessments followed (e.g. symptoms of cold/flu) and a final suggestion of recommended further health advice was provided to the patient. The success of this service, and impact of increasing use of digital health resources on changes to the pattern of healthcare-seeking behaviour in the population, has implied that monitoring web-based health data will be an integral part of syndromic surveillance in years to come.

This paper reports the early findings of a collaborative effort to assess the usefulness of NHS Direct online service self-checker (NHS Direct web) data to augment existing syndromic surveillance systems. We present a preliminary analysis of this novel data source and assess its potential as an adjunct to the national PHE syndromic surveillance service.

METHODS

NHS Direct web and telephone triage call data

Daily NHS Direct web data were extracted from the NHS Digital Services database using an automated reporting routine. Fields within the dataset comprised basic demographics of the patient (web user) including age (in years) and sex, the symptoms reported by the patient, and the healthcare outcome (disposition) of the visit, i.e. the advice given to the patient with

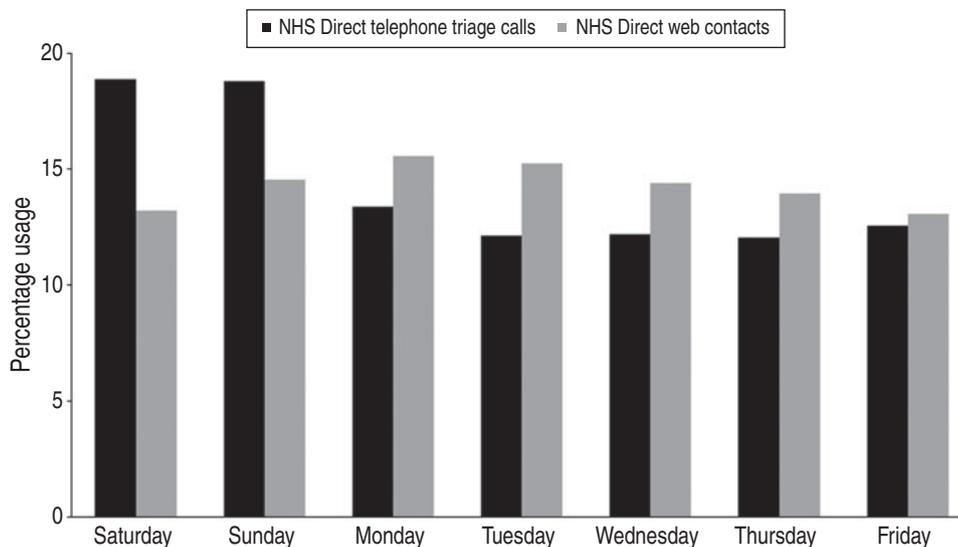


Fig. 2. Daily usage of NHS Direct telephone triage calls and NHS Direct web contacts. Data presented as the usage per day as a percentage of the total weekly load.

regard to further healthcare (no patient identifiable data were contained within the data extract). The symptom field was based upon the final endpoint of the patients' use of the symptom checker.

A comparable set of data was aggregated from the existing NHS Direct telehealth syndromic surveillance system [3].

Descriptive analysis

NHS Direct web data and NHS Direct telephone triage call data were compared for a period of 11 months, between 1 August 2012 and 1 July 2013. The underlying characteristics of the data were initially compared, including patient demographics and daily service usage. The age of the patient using the self-checker website was grouped into six age bands: <1, 1–4, 5–14, 15–44, 45–64, and ≥ 65 years. The disposition of each web visit, i.e. the health advice provided to the visitor based upon the health/symptom information provided, was also grouped to complement the existing phone call system. The disposition categories included: home care, GP 2 hours, GP 6 hours, GP >6 hours, emergency department, 999 (ambulance dispatch), and 'other' (e.g. emergency dental services, attend a local NHS walk-in centre).

Syndromic time series analysis

NHS Direct web symptom-checker protocols, where appropriate, were grouped into a common set of

syndromic indicators already used by the NHS Direct telehealth syndromic surveillance system, including those vital for tracking the emergence and spread of infectious diseases and for other incidents of public health importance. The syndromes chosen for comparison were: cold/flu, difficulty breathing, eye problems, diarrhoea and vomiting, and rash.

Statistical analysis

The daily percentage of self-checker indicator data, e.g. cold/flu were calculated using the daily total number of self-checker contacts as the denominator. The percentage time-series were checked for autocorrelation and appropriate high-order autoregressive models fitted to the data for each series. These models were used to remove autocorrelation from the data. After accounting for autocorrelation, the resulting data were compared using linear regression to test for correlations between the web and phone data across each syndrome; testing also for correlations when time lags were introduced. All analyses were undertaken using Stata v. 12 [10].

RESULTS

NHS Direct web data

A total of 3.37 million uses of the NHS direct web self-checker service were recorded over the 11-month period 1 August 2012 to 1 July 2013. The mean daily number of visits was 18 410, ranging from a

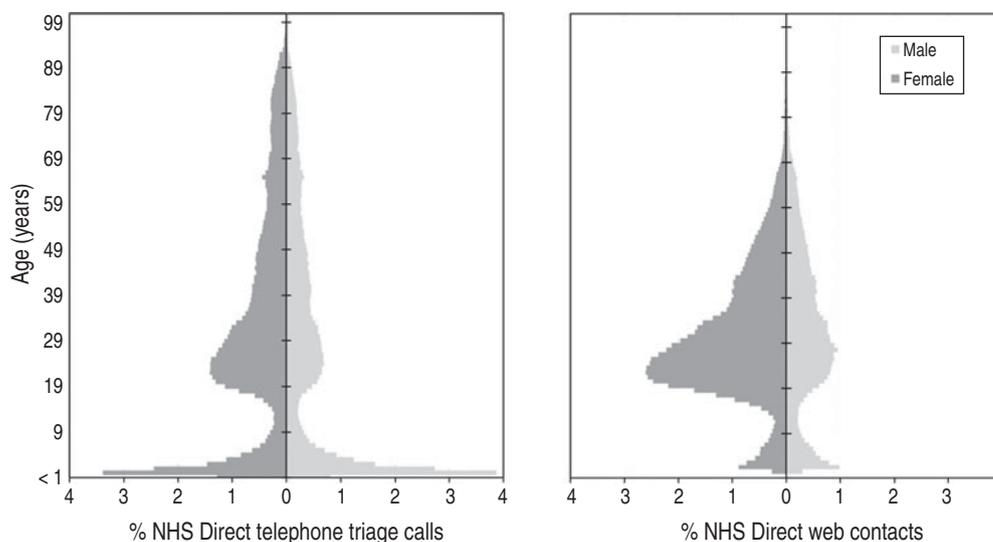


Fig. 3. Age and gender distribution of users of the NHS Direct telephone triage and NHS direct web data.

minimum of 11 358 to a maximum of 29 252; this was 136% higher than the telephone triage calls (1.43 million) where the mean daily number of calls was 7643 (Fig. 1). There was a general downward trend in the number of NHS Direct telephone calls recorded over this period corresponding to the gradual decommissioning of the NHS Direct service and switchover to NHS 111.

A larger percentage of symptom-checker users was observed at the beginning of the week; Monday with the highest percentage of hits [16.8%, 95% confidence interval (CI) 16.7–16.9] followed by Tuesday (16.1%, 95% CI 16.0–16.2), and Saturday with the lowest (8.45%, 95% CI 8.38–8.52). The NHS Direct telephone triage system experienced the highest level of usage at weekends (Fig. 2).

Data capture for the gender field from the web service users was excellent at 96%. Females were more likely to use the web service (ratio female/male = 1.9:1) mirroring similar statistics in the telephone triage service. Age of the user was captured in 99.1% of visits; the mean age for females was 30.3 years (95% CI 30.2–30.3, median 28) and for males 31.4 years (95% CI 31.3–31.5, median 30). After stratifying age into age groups, the most frequent age group utilizing the web services was the 15–44 years age group (72%) and lowest was those aged <1 year (Fig. 3). Comparing the NHS Direct web and telephone data, the web data were particularly underrepresentative in the very young and elderly age groups.

Health outcome (disposition) of the NHS Direct web and telephone triage call data were compared (Fig. 4). Overall, the outcomes of web hits and

telephone triage calls were comparable, with key dispositions, e.g. home care, GP, emergency department and '999' emergency calls at similar levels although there appeared to be a slight predominance of advice to consult a GP with the phone service compared to the web.

Times-series analysis of syndromic indicator data

NHS Direct web and telephone triage data were compared (Fig. 5). There were strong correlations found between the web data and the corresponding syndromes from the telephone triage data. There was a strong, but not quite significant correspondence (at 95% level) between the combined web diarrhoea and vomiting indicator ($P = 0.021$) and the separate diarrhoea ($P = 0.054$) and vomiting ($P = 0.071$) indicators from the telephone triage calls. For three indicators, cold/flu, rash and eye problems the strongest correlation occurred with a lag in the data, the web data prefiguring changes in the phone data. The cold/flu indicator showed a significant correspondence with a zero lag; however, the optimum fit occurred with an 8-day lag ($P = 0.001$); eye problems at a 2-day lag ($P = 0.007$); and rash at a 7-day lag ($P = 0.018$).

DISCUSSION

This preliminary analysis of data extracted from a national health website providing an online symptom-checker service for the population of England illustrates that the data collected through this online health service are comparable to existing telephone triage call

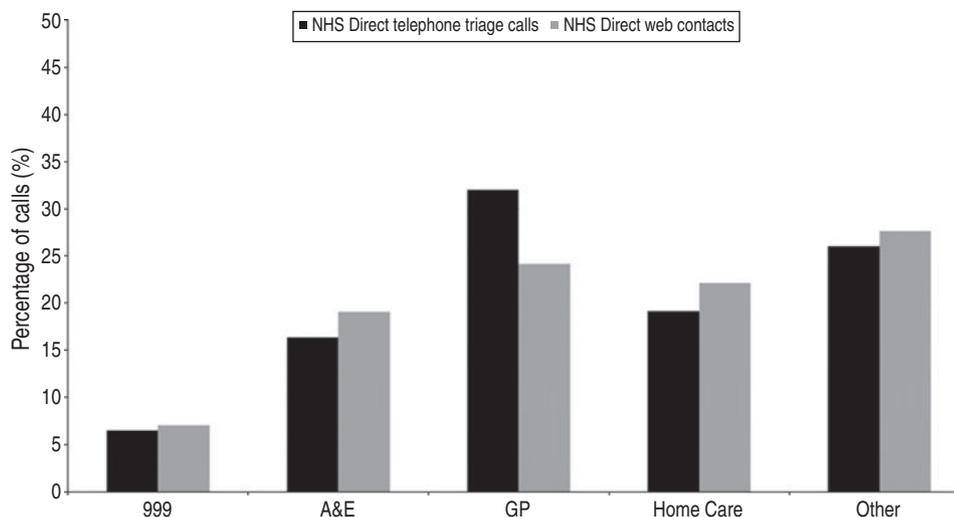


Fig. 4. The outcome (disposition) of NHS Direct telephone triage and web contact data.

data that have been successfully utilized for syndromic surveillance for over a decade [3]. The NHS Direct telehealth syndromic surveillance system traditionally provided early warning of increases in community-based influenza and norovirus activity [11, 12]. The findings from this work suggest that with respect to timeliness, the web data provide comparable, if not more timely early warning of an increase in syndromic signals.

These results are encouraging as they provide additional intelligence to support the national influenza surveillance programme in providing accurate information about the start of the influenza season or other outbreaks of infectious disease. However, although it would be unwise to generalize using data from a single winter, these findings do raise the possibility of NHS Direct web data providing an increased early warning signal for the start of the influenza season. Future work in this area could include the development of early warning thresholds to determine when influenza activity is increasing thereby providing better intelligence to public health authorities that measures should be put in place to prepare for the oncoming season.

A review of the UK public health lessons from the 2009 A(H1N1) influenza pandemic experience, focusing on evaluating the strengths and weaknesses of the data collected, recommended further mechanisms to elucidate the proportion of the population who are symptomatic, but do not consult a healthcare professional [13]. These data are important for improving the estimation of infection and transmission rates for use in real-time models to assist the public health

management and response to a pandemic. The NHS Direct web data collected through the symptom-checker website would appear to partially meet this recommendation, generating information on patients who do not consult a healthcare professional, but who self-diagnose using the specialist online service. In an increasingly digital age, the proportion of the population who are accessing digital health information and advice is increasing, and therefore the use of digital information in public health surveillance will become increasingly important.

Despite the apparent benefits of using online health data, there are several important limitations that have to be considered. The web data are biased: they are underrepresentative of certain populations, e.g. the elderly, who are less likely to have, or use, the internet for accessing health-related information [14]. These data are also underrepresentative of young infants, whose parents are more likely to request an immediate consultation with a healthcare professional rather than seek advice using online services. These two cohorts of the population are, however, more likely to be monitored by other surveillance systems, including GP and emergency department systems, both of which are routinely monitored by PHE [8]. We were also unable to undertake spacial analyses of the data as the information recorded by the web service included inconsistent use of location data including combinations of free text (e.g. city/town/district) and postcode (provided at varying levels) which proved unreliable for allocating any geography to the web visits.

A further limitation of these web data is the relative uncontrolled nature of data capture. Telephone calls

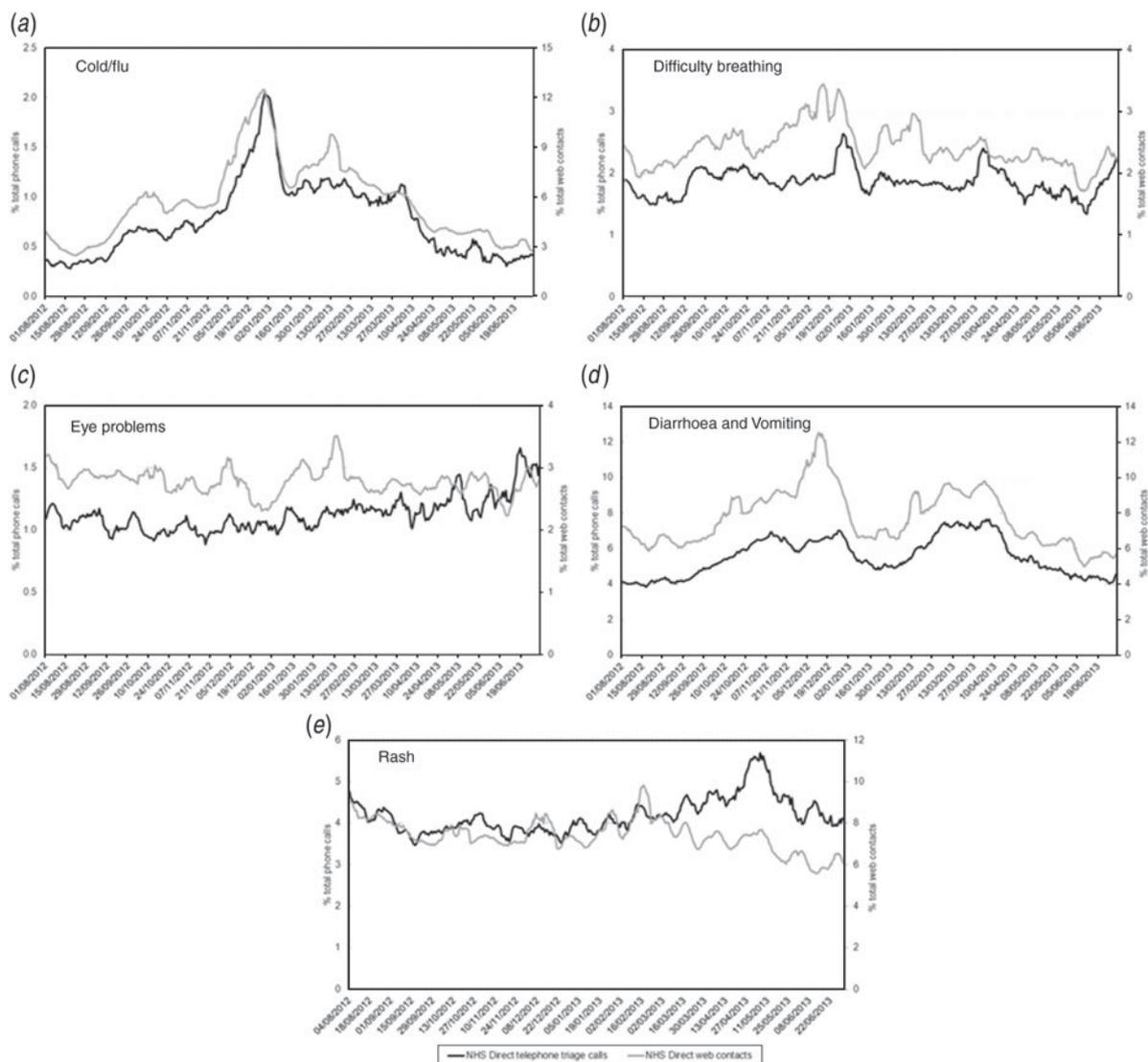


Fig. 5. Comparison of selected syndromic indicators from NHS Direct telephone triage call and web contact data: (a) cold/flu; (b) difficulty breathing; (c) eye problems; (d) diarrhoea and vomiting; (e) rash. Data presented as 7-day moving average of total percentage of calls/contacts.

managed through the NHS Direct telephone triage service (and now NHS 111) are directed by a call handler who utilizes clinical algorithms and leads the caller through the algorithm flow according to answers provided and the clinical judgement of the call handler. However, users of online health services are able to control their journey through the self-checker and able to navigate back through previous health questions to change answers to fulfil their requirements, although we assume that the proportion of users who do this is likely to be constant over time. There is also no confirmation that the users of the web services are symptomatic, with these data also potentially capturing asymptomatic patients who are seeking health information for themselves, or on behalf of others.

Over recent years there has been an increase in the utilization of internet-based health data for public health surveillance. Google and Yahoo internet searches have been used to demonstrate their potential usefulness for monitoring influenza trends; however, these initiatives are often focused on single disease groups rather than a range of public health syndromic indicators [15, 16]. The use of telephone triage and web-based healthcare syndromic surveillance is naturally limited to those countries that utilize such health systems: Sweden is one such example, where the Vårdguiden medical website (www.vardguiden.se) has been used for syndromic surveillance to respond to a number of different public health problems [17–19]. Other web-based surveys have utilized data

capture mechanisms whereby participants complete online diaries, recording the occurrence of illness and providing further information about each episode and their healthcare usage [20]. These systems are valuable in providing further information about the proportion of the population who do not seek medical consultation when ill. There are, however, limitations with these approaches including patient recall of symptoms during self-reporting of illness, high drop-out rates throughout the reporting period, problems calculating accurate denominators and bias in the age and geography of participants.

We are confident that augmenting our existing syndromic surveillance with the web-based symptom-checker data will complement our ability to monitor disease trends at the population level. We intend to develop and apply statistical tests to these national data to identify unusual peaks and trends to aid our ability to identify and respond to public health incidents. Following the transition of NHS services from NHS Direct we are continuing to work with the host organization of the new NHS online symptom checker to continue this work. A further aim of the work is to explore the usefulness of the 'search' facility on the NHS website, which although would not provide symptom-based information, would increase the range of data captured providing further information on the health-seeking behaviours of the population.

ACKNOWLEDGEMENTS

The authors acknowledge the contribution from the PHE Real-time Syndromic Surveillance Team and NHS Direct. We acknowledge technical support from Infermed.

This work was undertaken as part of the national surveillance function of Public Health England and received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

DECLARATION OF INTEREST

None.

REFERENCES

1. **Triple S Project.** Assessment of syndromic surveillance in Europe. *Lancet* 2012; **378**:1833–1834.
2. **Public Health England.** Public Health England (<https://www.gov.uk/government/organisations/public-health-england>). Accessed 10 October 2013.
3. **Baker M, et al.** Early warning and NHS Direct: a role in community surveillance? *Journal of Public Health Medicine* 2003; **25**: 362–368.
4. **Elliot AJ, et al.** Establishing an emergency department syndromic surveillance system to support the London 2012 Olympic and Paralympic Games. *Emergency Medicine Journal* 2012; **29**: 954–960.
5. **Harcourt SE, et al.** Developing a new syndromic surveillance system for the London 2012 Olympic and Paralympic Games. *Epidemiology and Infection* 2012; **140**: 2152–2156.
6. **Harcourt SE, et al.** Use of a large general practice syndromic surveillance system to monitor the progress of the influenza A(H1N1) pandemic 2009 in the UK. *Epidemiology and Infection* 2011; **140**: 100–105.
7. **NHS Choices.** NHS 111 service (<https://www.nhs.uk/111/>). Accessed 10 October 2013.
8. **Public Health England.** Real-time syndromic surveillance (<https://www.gov.uk/government/collections/syndromic-surveillance-systems-and-analyses>). Accessed 6 March 2015.
9. **NHS Direct.** Digital services (<http://webarchive.nationalarchives.gov.uk/20140220132333/http://www.nhsdirect.nhs.uk/en/Commissioners2/DigitalServices>). Accessed 13 June 2014.
10. **StataCorp.** Stata version 12. College Station, TX, USA: Stata Corporation, 2012.
11. **Cooper DL, et al.** Can syndromic thresholds provide early warning of national influenza outbreaks? *Journal of Public Health (Oxford)* 2009; **31**: 17–25.
12. **Loveridge P, et al.** Vomiting calls to NHS Direct provide an early warning of norovirus outbreaks in hospitals. *Journal of Hospital Infection* 2010; **74**: 385–393.
13. **Chief Medical Officer's Statistical Legacy Group.** Pandemic influenza preparedness programme: Statistical Legacy Group – a report for the Chief Medical Officer, 2010 (<https://www.gov.uk/government/publications/pandemic-influenza-preparedness-programme-statistical-legacy-group-a-report-for-the-chief-medical-officer>). Accessed 24 July 2014.
14. **Takahashi Y, et al.** Internet use for health-related information via personal computers and cell phones in Japan: a cross-sectional population-based survey. *Journal of Medical Internet Research* 2011; **13**: e110.
15. **Ginsberg J, et al.** Detecting influenza epidemics using search engine query data. *Nature* 2009; **457**: 1012–1014.
16. **Polgreen PM, et al.** Using internet searches for influenza surveillance. *Clinical Infectious Diseases* 2008; **47**: 1443–1448.
17. **Hulth A, et al.** Eye-opening approach to norovirus surveillance. *Emerging Infectious Diseases* 2010; **16**: 1319–1321.
18. **Hulth A, Rydevik G.** Web query-based surveillance in Sweden during the influenza A(H1N1)2009 pandemic, April 2009 to February 2010. *Eurosurveillance* 2011; **16**: pii=19856.
19. **Lindh J, et al.** Head lice surveillance on a deregulated OTC-sales market: a study using web query data. *PLoS ONE* 2012; **7**: e48666.
20. **Tilston NL, et al.** Internet-based surveillance of Influenza-like-illness in the UK during the 2009 H1N1 influenza pandemic. *BMC Public Health* 2010; **10**: 650.