Journal of
Epidemiology &
Community Health

Google search patterns monitoring the daily health impact of heatwaves in England: how do the findings compare to established syndromic surveillance systems from 2013 to 2017?

Journal:	*Journal of Epidemiology & Community Health*
Manuscript ID	Draft
Article Type:	Short Report
Date Submitted by the Author:	n/a

Complete List of Authors:	Green, Helen; Public Health England, Real-time Syndromic Surveillance Edeghere, Obaghe; Public Health England, Real Time Syndromic Surveillance, National Infection Service; Public Health England, Field Epidemiology Service, National Infection Service; Kings College London / Public Health England, NIHR Health Protection Research Unit in Emergency Preparedness and Response Elliot, Alex; Public Health England, Real-time Syndromic Surveillance Team Cox, Ingemar; University College London Department of Computer Science; Kobenhavns Universitet Datalogisk Institut Morbey, Roger; Public Health England, Real-time Syndromic Surveillance Team Pebody, R. G.; Publ Hlth England Bone, Angie; Public Health England, McKendry, Rachel; London Centre for Nanotechnology and Division of Medicine, University College London Smith, Gillian; Public Health England, Real-time Syndromic Surveillance Team

Keywords:	SURVEILLANCE, TIME-SERIES, TEMPERATURE, EPIDEMIOLOGY

SCHOLARONE™
Manuscripts

Google search patterns monitoring the daily health impact of heatwaves in England: how do the findings compare to established syndromic surveillance systems from 2013 to 2017?

Authorship
Helen K Green[1], Obaghe Edeghere[1], Alex J Elliot[1], Ingemar J. Cox[2,3], Roger Morbey[1], Richard Pebody[4], Angie Bone[5], Rachel A McKendry[6], Gillian E Smith[1*]

[1]Real-time Syndromic Surveillance Team, Public Health England, Birmingham, UK
[2]Department of Computer Science, UCL, London, UK
[3]Department of Computer Science, University of Copenhagen, Copenhagen, Denmark
[4]Respiratory Diseases Department, Public Health England, London, UK
[5]Extreme Events, Public Health England, London, UK
[6]London Centre for Nanotechnology and Division of Medicine, UCL, London, UK

*Corresponding author
Professor Gillian Smith, Consultant Epidemiologist, Real-time Syndromic Surveillance Team, National Infection Service, Public Health England
5 St Philip's Place, Birmingham, B3 2PW
Tel: 0344 225 3560 [option 4 → option 2]
Email: gillian.smith@phe.gov.uk

Word count (excluding title page, abstract, references, figures and tables): 1,547

Abstract

Background

The impact of heatwaves on the health of the population is captured through real-time syndromic healthcare surveillance systems monitored daily in England during the summer months. Internet search data could potentially provide improved timeliness and help to assess the wider population health impact of heat by capturing a population sub-group who are symptomatic but do not seek healthcare.

Methods

A retrospective observational study was carried out from June 2013 to September 2017 in England to compare trends in validated syndromic surveillance heat-related morbidity indicators against symptom-based heatwave-related Google search terms. Spearman correlation coefficients were calculated and lag assessment was carried out to determine timeliness.

Results

Daily increases in frequency in Google search terms during heatwave events correlated well with validated syndromic indicators. Correlation coefficients between search term frequency and syndromic indicators from 2013-17 were highest with the telehealth service NHS 111 (range of 0.684 to 0.900 by search term). Lag analysis revealed a similar timeliness between the data sources, suggesting Google data did not provide a delayed or earlier signal.

Conclusion

This preliminary work highlights the potential benefits for countries which lack established public health surveillance systems and the use of internet search data to assess the wider population health impact of exposure to heat.

What is already known?

Real-time syndromic surveillance systems based on healthcare utilisation data are in operation in a number of countries and have been validated for monitoring the health impact of heatwaves.

The utility of online user-generated content such as internet search queries in surveillance of infectious and non-infectious diseases has been explored but there has been limited investigation into the use for monitoring the health impact of heatwaves.

What does this paper add to the literature?

In England, Google search term frequency for heat-related morbidity demonstrated very good correlation with comparable validated indicators across a suite of established daily syndromic surveillance systems when assessment was limited to the heatwave period.

These findings have important implications for countries where there are no such healthcare-based syndromic surveillance systems in place as monitoring search term frequency may provide a low-cost option for monitoring heat-related morbidity, and provides a potential opportunity to determine the health impact on a wider population.

BACKGROUND

Heatwaves can significantly impact on the health of the population. The highest vulnerability is typically seen in the elderly, children, and individuals with pre-existing medical conditions[1,2], although all age groups may be affected, particularly in hot countries where people are working outdoors without adequate protection[3] or cannot seek shelter (e.g. homeless[4]).

The Heatwave Plan for England was launched in 2004[5] in response to the 2003 European heatwave which resulted in an estimated 70,000 deaths across Europe[6]. As part of this Plan, from 1st June to 15th September each year, Public Health England (PHE) routinely monitors heat-related morbidity across a suite of syndromic surveillance systems daily[7,8]. Indicators monitored within these systems in England have been demonstrated to provide a sensitive measurement of the health impact of heat[9,10] and information can facilitate optimal public health response during the heatwave through informing appropriate messaging[11,12].

The utility of novel data sources in health surveillance such as social media and internet search queries is gaining increasing interest[13]. The potential advantages of online user-generated content include the provision of more timely information and identification of illness in symptomatic individuals, including those who do not seek healthcare. For data sources such as Google Trends[14], this comes at a minimal cost and may therefore be useful in countries with sufficient internet coverage without an established public health surveillance infrastructure. However, there are notable limitations with these data sources, including uncertainty of the underlying reason for an individual searching and of their characteristics.

The utility of internet search data to monitor heat-related morbidity was demonstrated in Shanghai[15], with strong correlation between *heat stroke* internet searches and heatstroke deaths and hospitalised cases. However, availability of health outcome data was not as timely as in England and it is not known how internet search data for heat-related illness compares to outputs from real-time syndromic surveillance systems. During 2017 in England, there was a Level 2 heatwave alert issued (defined as a risk of reaching high temperatures) from 15 - 21st June, with areas reaching high temperatures (a Level 3 heatwave alert) across most of the country from 17-20th June[16]. This short report outlines the findings from a retrospective observational study comparing trends in heat-related syndromic surveillance indicators with heatwave-related internet search terms during 2017 and for previous heatwaves.

METHODS

Data

Daily data for heat impact indicators[9] was accessed from the four national PHE syndromic surveillance systems; NHS 111 (a telehealth service), an in-hours general practitioner (GP) system (GPIH), an out of hours GP system (GPOOH), and a sentinel emergency department (ED) surveillance system[8].

Google search volume data was available from the Google Health Trends Application Programming Interface (API), a tool to explore Google search data (like the publicly available Google Trends website). Data is taken from a daily updated uniformly distributed random sample of 10-15% of Google web searches. Daily probabilities are defined as the probability of searching for a specified term on a given day in England, multiplied by ten million to be human readable. Relevant search terms (Appendix 1) were identified through the following process:

1. Descriptions of clinical codes underlying the heat syndromic indicators (e.g. *heatstroke*) were collated and synthesised into a list of code terms.
2. Each code term was entered into Google Trends and related internet search terms retained if also searched for at least half as often[14] to identify further terms used (e.g. individuals searching for *heatstroke* also searching for *sunstroke*) and associated terms (e.g. *symptoms*).
3. Where variations of the search terms were used (e.g. *heatstroke*, *heat stroke*), there was generally one that had a larger search probability and this was retained in the final list of terms.

Daily data for the final Google search terms and heat-related syndromic indicators were extracted from 01/06/2013 – 15/09/2017.

Analysis

Time series plots of heat-related syndromic indicators and Google search data were produced to visually compare patterns from 2013 to 2017 using daily data and seven day moving averages to minimise the impact of day of the week effects[17]. Correlation was quantified through Spearman correlation coefficients as the data was not normally distributed. This comparison was also separately done for 2017 from 08/06/2017 – 27/06/2017 to focus on the known heatwave event, encompassing one week either side of the official heatwave alerts. Lagged time series plots were produced for the 2017 heatwave to assess when the two datasets best correlated. The Google time series was shifted forwards or backwards in time up to ten days to determine if search activity was leading or lagging syndromic data respectively. In line with Google data availability, analysis was restricted to national level.

RESULTS

The key Google search terms identified were *heat exhaustion*, *prickly heat*, *heat rash*, *heat stroke*, *heatwave*, *sunstroke* and *sunburn*. Each of the search terms showed an increase in frequency during periods of hot weather in each summer, with the largest increase seen in 2013 and 2017, corresponding with longer heatwave periods (Figure 1, Supplementary Figure 1). In 2017, the most frequently searched term out of this list was *sunburn* (peak probability 18/06/2017), followed by *heatwave* (18/06/2017), *heat stroke* (19/06/2017), *heat rash* (20/06/2017), *sunstroke* (18/06/2017), *prickly heat* (19/06/2017) and *heat exhaustion* (19/06/2017). These increases coincided with heatwave alerts (Supplementary Table 1).

There was a positive correlation observed between daily Google and syndromic data (Figure 1). Correlation coefficients between search term frequency and syndromic indicators from 2013-17 for the different systems demonstrated the highest correlation with NHS 111 (range of 0.684 to 0.900 by search term) and the lowest with GPIH (0.410 to 0.526, Table 1). When focusing on the June 2017 heatwave event, correlation coefficients increased and were all statistically significant except for GPIH daily data which was only significantly correlated with *heatwave*.

Lag analysis demonstrated maximal correlation between Google and syndromic data within two days of each other across search terms (Table 1). By surveillance system, for most search terms GPIH was preceded by Google data by a day or two. There was no consistent pattern by search term.

DISCUSSION

Rapidly detecting the health impact of heatwaves is of increasing importance with the implications of climate change and predicted increases in temperature likely leading to more frequent severe heatwaves[18]. To the best of our knowledge, this is the first study to look at daily Google heat-related search term frequency and compare these to established syndromic surveillance indicators. The increases in frequency across search terms in England during heatwave periods from 2013-2017 correlated well with NHS 111's heat-related morbidity indicator and correlations improved for other syndromic surveillance systems when focusing on the 2017 heatwave event.

Limitations of Google search data include the representativeness of internet users (e.g. lower use in the elderly), the inability to look at subnational data to correspond to regional heatwave alerts and uncertainty around who is searching and the intention behind the search, with the possibility of media-driven searches. It may not therefore be as indicative of poor health as existing syndromic data sources which have subnational data and information on the person characteristics (e.g. age). However, by selecting symptom-based terms, a consistent increase seen across search terms and correlation with syndromic indicators suggests people were searching for information because of a health complaint for themselves or someone else. Furthermore, associated search terms identified in Google Trends were *signs*, *symptoms* and *treatment* for all queries except for *heatwave* where they were *UK* and *weather*, suggesting the term was used to search for information about the event. Exploratory qualitative research with a representative sample of Google users will help to determine the reasoning behind searching for such terms and validate if because of a health complaint.

Lag analysis revealed a similar timeliness by day between Google and syndromic datasets during the June 2017 heatwave, suggesting Google data did not provide a delayed or earlier signal relative to syndromic data. The exception to this was GPIH where Google data provided a slightly earlier signal of one or two days. This may be explained by the unavailability of routine GP services during weekends leading to a delay in patients consulting, with the peak of the June 2017 heatwave occurring at a weekend.

These findings have highlighted two key potential benefits. Firstly, Google-based syndromic surveillance will be of interest to countries lacking established public health surveillance systems able to monitor the near real-time health impact of heat but could access freely available Google data. Secondly, this could contribute to assessing the wider population health impact of exposure to heat. Indicators monitored through syndromic surveillance systems, in conjunction with excess mortality surveillance[19], are focussed on outcomes severe enough to require healthcare to rapidly determine if the impact of a heatwave is more severe than expected. Monitoring internet search query frequency could complement existing surveillance by supporting estimation of the proportion of the population whose health may be affected but may not present to healthcare. Further work is required to fully develop a more specific validated search term list using natural language processing methods[20] and to use these outputs to estimate the proportion.

This forms part of a wider programme of work being carried out by PHE to assess the utility of Google search data for the surveillance of a range of health outcomes in the context of existing syndromic surveillance. It demonstrates the potential utility in England to contribute to existing heat health surveillance systems, with further work needed to validate the findings.

ACKNOWLEDGMENTS

We acknowledge support from: NHS 111 and NHS Digital for their assistance and support in the NHS 111 system; OOH providers submitting data to the GPOOHSS and Advanced Heath & Care; ED clinicians and NHS Trust staff supporting EDSSS, the Royal College of Emergency Medicine and technical support provided by EMIS Health and L2S2 Ltd; TPP and participating SystmOne practices and University of Nottingham, ClinRisk, EMIS and EMIS practices submitting data to the QSurveillance database and Google for access to their Google Health Trends API. The authors would like to thank Mr Paul Loveridge of the PHE Real-time Syndromic Surveillance Team (ReSST) and Dr Bill Lampos of the Department of Computer Science, UCL for their technical expertise. IJC and RAM are supported by the i-sense EPSRC IRC in Early Warning Sensing Systems for Infectious Diseases (EP/K031953/1) and RAM is also supported by the Royal Society Wolfson Merit Award. AJE and GES are supported by the National Institute for Health Protection Research Health Protection Unit (NIHR HPRU) in Emergency Preparedness and Response at King's College London in partnership with PHE. The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR, the Department of Health or Public Health England.

ETHICAL APPROVAL

PHE has approval under Section 251 of the NHS Act 2006 to process confidential patient information for public health purposes; separate ethical approval was not required for this study.

Figure 1. Daily (pale blue) and seven day moving average (dark blue) probabilities[1] searching for a heat-related term[2] in Google and syndromic surveillance heat-related health indicators, England, 2013-2017[3]

[1] *Daily probability is defined as searching for a specified term on a given day in England, multiplied by ten million*
[2] *Heat exhaustion* or *heat rash* or *heat stroke* or *heatwave* or *prickly heat* or *sunburn* or *sunstroke*
[3] *NHS 111 data was only available from September 2013 onwards. Periods with heatwave alerts greater than level 1 are indicated with vertical grey bars.*

Table 1. Correlation coefficients between Google search term probability[4] and syndromic surveillance system heat-related indicator, England, 2013-2017[5]

Syndromic Surveillance system	Google Search Term	Spearman correlation coefficient			Lag analysis during 2017 heatwave										
		2013 to 2017	June 2017 Heatwave		Google dataset lags (days behind)[6]						Google dataset leads (days ahead)[7]				
			Daily	7dma	-5	-4	-3	-2	-1	0	1	2	3	4	5
111	Heat exhaustion	**0.754**	**0.721**	**0.918**	0.097	0.331	**0.571**	**0.680**	**0.736**	**0.721**	**0.725**	**0.567**	0.365	0.128	-0.162
GPIH	Heat exhaustion	**0.526**	0.280	**0.488**	-0.078	0.060	0.372	**0.392**	0.286	0.280	0.322	0.285	0.271	0.274	0.151
GPOOH	Heat exhaustion	**0.628**	**0.673**	**0.855**	0.126	0.341	0.326	**0.560**	**0.599**	**0.673**	**0.704**	**0.538**	**0.432**	0.215	0.021
ED	Heat exhaustion	**0.548**	**0.531**	**0.863**	-0.002	0.187	0.247	**0.503**	**0.593**	**0.531**	**0.514**	**0.566**	**0.484**	0.263	-0.056
111	Heat rash	**0.856**	**0.828**	**0.928**	0.132	0.358	**0.570**	**0.708**	**0.800**	**0.828**	**0.770**	**0.573**	0.293	-0.010	-0.243
GPIH	Heat rash	**0.480**	0.218	**0.480**	0.158	0.232	0.230	0.336	0.317	0.218	0.345	**0.393**	0.214	0.110	0.106
GPOOH	Heat rash	**0.643**	**0.758**	**0.874**	0.130	0.280	**0.472**	**0.656**	**0.718**	**0.758**	**0.617**	**0.467**	0.348	0.105	-0.145
ED	Heat rash	**0.533**	**0.594**	**0.833**	0.149	0.293	0.330	**0.464**	**0.572**	**0.594**	**0.523**	**0.412**	0.327	0.144	-0.061
111	Heat stroke	**0.684**	**0.916**	**0.958**	-0.082	0.122	**0.413**	**0.616**	**0.779**	**0.916**	**0.873**	**0.757**	**0.535**	0.285	0.008
GPIH	Heat stroke	**0.410**	0.336	**0.482**	0.001	0.021	0.207	0.283	0.240	0.336	**0.405**	0.328	0.228	0.279	0.156
GPOOH	Heat stroke	**0.574**	**0.680**	**0.923**	0.037	0.205	**0.394**	**0.530**	**0.709**	**0.680**	**0.684**	**0.607**	**0.552**	0.266	0.113
ED	Heat stroke	**0.476**	**0.537**	**0.819**	-0.030	0.184	0.221	0.335	**0.515**	**0.537**	**0.527**	**0.498**	**0.497**	0.335	0.214
111	Heatwave	**0.774**	**0.450**	**0.667**	0.043	0.266	0.387	**0.461**	**0.515**	**0.450**	**0.480**	**0.394**	0.289	0.218	0.169
GPIH	Heatwave	**0.458**	**0.396**	**0.427**	-0.100	0.033	0.112	0.161	0.311	**0.396**	**0.409**	0.351	0.310	0.172	0.075
GPOOH	Heatwave	**0.514**	**0.522**	**0.704**	0.133	0.324	0.293	**0.454**	**0.514**	**0.522**	**0.534**	**0.609**	**0.423**	**0.450**	0.228
ED	Heatwave	**0.497**	**0.575**	**0.625**	-0.101	0.062	0.215	0.267	**0.452**	**0.575**	**0.518**	**0.527**	0.339	0.286	0.051
111	Prickly heat	**0.900**	**0.812**	**0.929**	0.168	0.358	**0.643**	**0.749**	**0.778**	**0.812**	**0.729**	**0.562**	0.276	0.048	-0.182
GPIH	Prickly heat	**0.504**	0.344	**0.482**	0.167	0.014	0.294	**0.404**	0.244	0.344	**0.414**	0.301	0.137	0.192	0.118
GPOOH	Prickly heat	**0.638**	**0.642**	**0.825**	0.200	0.304	**0.509**	**0.534**	**0.730**	**0.642**	**0.595**	**0.534**	0.355	0.179	-0.146
ED	Prickly heat	**0.531**	**0.652**	**0.869**	0.092	0.219	0.359	0.347	**0.587**	**0.652**	**0.565**	**0.450**	0.301	0.157	-0.008
111	Sunburn	**0.898**	**0.844**	**0.956**	0.095	0.201	**0.421**	**0.560**	**0.722**	**0.844**	**0.844**	**0.696**	**0.402**	0.108	-0.162
GPIH	Sunburn	**0.506**	0.137	**0.493**	0.232	0.126	0.253	0.313	0.182	0.137	**0.374**	**0.444**	0.215	0.248	0.155
GPOOH	Sunburn	**0.629**	**0.693**	**0.908**	0.097	0.276	0.421	**0.492**	**0.691**	**0.693**	**0.618**	**0.489**	**0.402**	0.071	-0.129
ED	Sunburn	**0.528**	**0.563**	**0.835**	0.135	0.283	0.297	0.321	**0.501**	**0.563**	**0.493**	**0.375**	0.344	0.189	0.015
111	Sunstroke	**0.884**	**0.811**	**0.943**	-0.143	-0.012	0.264	**0.452**	**0.653**	**0.811**	**0.869**	**0.776**	**0.556**	0.325	0.069
GPIH	Sunstroke	**0.476**	0.109	**0.464**	0.156	-0.063	0.079	0.219	0.118	0.109	**0.448**	**0.480**	0.161	0.200	0.236
GPOOH	Sunstroke	**0.634**	**0.655**	**0.902**	-0.105	0.073	0.243	**0.394**	**0.645**	**0.655**	**0.555**	**0.524**	**0.575**	0.322	0.049
ED	Sunstroke	**0.547**	**0.555**	**0.825**	-0.084	0.123	0.170	0.212	**0.445**	**0.555**	**0.458**	**0.395**	**0.467**	0.345	0.161

[4]Daily probability is defined as searching for a specified term on a given day in England, multiplied by ten million. [5]Correlation coefficients are emboldened where significant and shaded pale yellow (p<0.05). Cells shaded yellow contain the highest correlation coefficient for each row in the lag analysis. 7dma = seven day moving average. [6]The Google dataset was lagged backwards in time. E.g. The highest correlation coefficient with a lag of -1 implies Google data needed to be shifted back a day to best correlate with syndromic data and it had a delay of one day relative to syndromic data. [7]The Google dataset was shifted forwards in time. E.g. The highest correlation coefficient with a lag of +1 implies Google data needed to be shifted forwards a day to best correlate with syndromic data and it was one day ahead of syndromic data.

APPENDIX ONE. SELECTION OF GOOGLE SEARCH TERMS

a) Terms identified from syndromic surveillance (additional terms and comments from Google Trends)

- Heat exhaustion (Heat stroke)
- Heat rash (Prickly heat)
- Heatstroke (Heat stroke, Sunstroke)
- Heat stroke
- Heatwave
- Prickly heat (Heat rash)
- Sunburn (No additional terms identified)
- Sunstroke
- Sun stroke (heat stroke, sunstroke)
- Heat syncope (Not enough searches carried out in Google to provide data)
- Heat fatigue (Not enough searches carried out in Google to provide data)
- Heat oedema (Not enough searches carried out in Google to provide data)
- Heat prostration (Not enough searches carried out in Google to provide data)

b) Final list

- Heat exhaustion
- Heat rash
- Heat stroke
- Heatwave
- Prickly heat
- Sunburn
- Sunstroke

Where multiple variations of the search terms were available (e.g. *heatstroke*, *heat stroke*), there was generally one that had the largest search probability and this was retained in the final list of terms.

REFERENCES

1. Benmarhnia T, Deguen S, Kaufman JS, et al. Review Article: Vulnerability to Heat-related Mortality: A Systematic Review, Meta-analysis, and Meta-regression Analysis. *Epidemiology* 2015;26(6):781-93.

2. Bouchama A, Dehbi M, Mohamed G, et al. Prognostic factors in heat wave related deaths: a meta-analysis. *Arch Intern Med*. 2007;167(20):2170-6.

3. Nelson NG, Collins CL, Comstock RD, et al. Exertional heat-related injuries treated in emergency departments in the U.S. 1997–2006. *Am J Prev Med* 2011;40(1):54–60.

4. Ramin B, Tomislav S. Health of the Homeless and Climate Change. *J Urban Health* 2009;86(4):654–664.

5. Public Health England. 2015. Heatwave plan for England. Protecting health and reducing harm from severe heat and heatwaves. Available online: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/429384/Heatwave_Main_Plan_2015.pdf

6. Robine JM, Cheung SL, Le Roy S, et al. Death toll exceeded 70,000 in Europe during the summer of 2003. *C R Biol*. 2008;331(2):171-8.

7. Triple S Project. Assessment of syndromic surveillance in Europe. *Lancet*. 2011;378(9806):1833-4.

8. Public Health England. 2017. Syndromic surveillance: systems and analyses. Available online: https://www.gov.uk/government/collections/syndromic-surveillance-systems-and-analyses

9. Smith S, Elliot AJ, Hajat S, et al. Estimating the burden of heat illness in England during the 2013 summer heatwave using syndromic surveillance. *J Epidemiol Community Health*. 2016;70(5):459-65.

10. Elliot AJ, Bone A, Morbey R, et al. Using real-time syndromic surveillance to assess the health impact of the 2013 heatwave in England. *Environ Res*. 2014;135:31-6.

11. Josseran L, Fouillet A, Caillère N, et al. Assessment of a syndromic surveillance system based on morbidity data: results from the Oscour network during a heat wave. *PLoS One*. 2010;5(8):e11984.

12. Pascal M, Laaidi K, Wagner V, et al. How to use near real-time health indicators to support decision-making during a heat wave: the example of the French heat wave warning system. *PLoS Curr*. 2012;4:e4f83ebf72317d.

13. Nuti SV, Wayda B, Ranasinghe I, et al. The use of google trends in health care research: a systematic review. *PLoS One*. 2014;9(10):e109583.

14. Google. 2017. Google Trends. Available online: https://trends.google.com/trends/

15. Li T, Ding F, Sun Q, et al. Heat stroke internet searches can be a new heatwave health warning surveillance indicator. *Sci Rep*. 2016;6:37294.

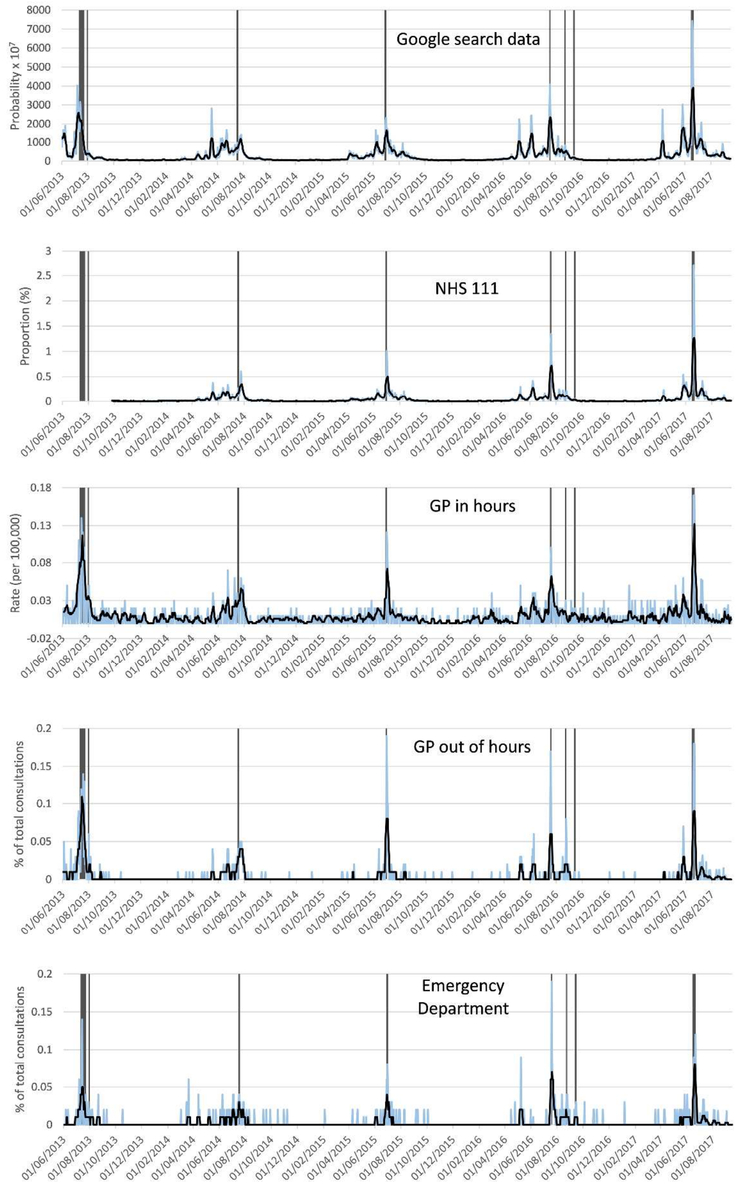
16. Met Office. 2017. Heat-health watch. Available online: http://www.metoffice.gov.uk/public/weather/heat-health/#?tab=heatHealth

17. Buckingham-Jeffery E, Morbey R, House T, et al. Correcting for day of the week and public holiday effects: improving a national daily syndromic surveillance service for detecting public health threats. *BMC Public Health*. 2017;17:477.

18. Hajat S, Vardoulakis S, Heaviside C, et al. Climate change effects on human health: projections of temperature-related mortality for the UK during the 2020s, 2050s and 2080s. *J Epidemiol Community Health*. 2014;68(7):641-8.

19. Green HK, Andrews N, Armstrong B, et al. Mortality during the 2013 heatwave in England--How did it compare to previous heatwaves? A retrospective observational study. *Environ Res*. 2016;147:343-9.

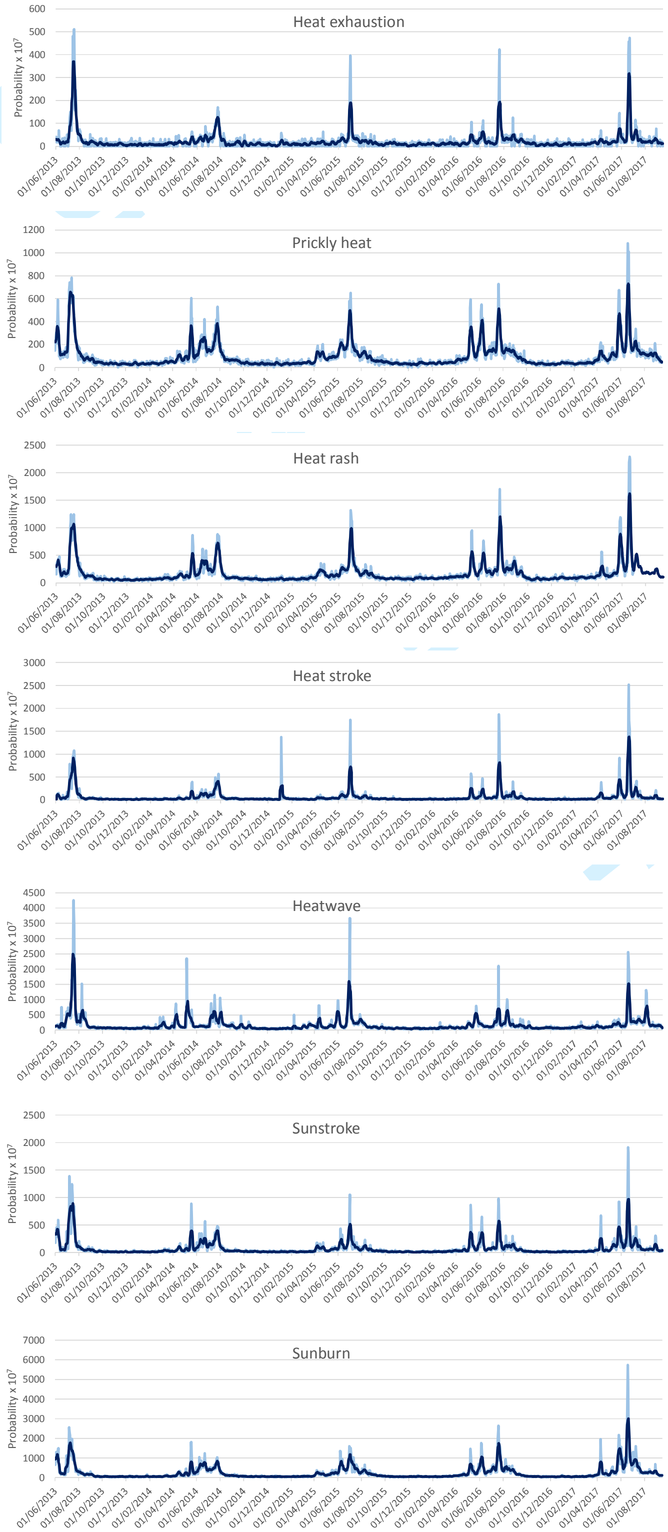
20. Lampos V, Zou B, Cox IJ. Enhancing Feature Selection Using Word Embeddings: The Case of Flu Surveillance. Proceedings of the 26th International Conference on World Wide Web. In: International World Wide Web Conferences Steering Committee; 2017. p. 695–704.



[1]Daily probability is defined as searching for a specified term on a given day in England, multiplied by ten million.┬ [2]Heat exhaustion or heat rash or heat stroke or heatwave or prickly heat or sunburn or sunstroke [3]NHS 111 data was only available from September 2013 onwards. Periods with heatwave alerts greater than level 1 are indicated with vertical grey bars.

1257x1997mm (96 x 96 DPI)

Supplementary Figure 1. Daily (pale blue) and seven day moving average (dark blue) probabilities (x 10 million)[A] searching for a heat-related term in Google, England, 2013-2017



[A]Daily probability is defined as searching for a specified term on a given day in England, multiplied by ten million

Supplementary Table 1. Daily probabilities (x 10 million) searching for a heat-related term in Google, England, June 2017[B]

Date	Google search term(s)						
	Heat exhaustion	Prickly heat	Heat rash	Heat stroke	Heatwave	Sunstroke	Sunburn
01/06/2017	31.73	369.45	555.45	216.81	95.31	306.48	1560.47
02/06/2017	31.80	285.98	440.51	194.11	95.59	238.58	1232.58
03/06/2017	46.08	215.10	461.69	156.12	135.67	233.78	1098.85
04/06/2017	53.55	154.48	411.37	105.76	77.66	130.91	748.45
05/06/2017	19.45	174.57	218.75	52.58	170.62	92.31	406.82
06/06/2017	14.54	125.62	198.37	32.33	213.25	43.56	279.95
07/06/2017	15.07	204.60	204.86	26.51	165.31	67.26	328.84
08/06/2017	29.20	116.61	116.91	25.55	88.04	29.24	305.84
09/06/2017	14.66	92.72	186.08	13.03	102.88	53.77	263.39
10/06/2017	18.63	130.38	236.45	89.13	74.86	173.91	743.91
11/06/2017	18.58	129.02	332.68	101.65	92.32	227.94	1001.07
12/06/2017	14.74	142.42	191.83	113.21	152.80	113.08	735.37
13/06/2017	22.59	149.95	270.53	94.00	270.93	125.16	594.28
14/06/2017	20.40	167.99	398.09	194.21	158.61	244.64	854.51
15/06/2017	45.93	244.46	372.86	185.93	322.19	239.76	804.56
16/06/2017	56.16	187.38	364.85	178.09	273.96	117.91	727.57
17/06/2017	152.14	427.60	864.14	780.71	504.76	911.76	3004.34
18/06/2017	407.57	840.06	1623.46	2013.51	**2555.64**	**1913.08**	**5741.17**
19/06/2017	**456.62**	**1084.70**	2189.72	**2518.22**	2428.68	1808.59	4835.66
20/06/2017	404.78	908.24	**2290.17**	1732.61	2120.14	1066.07	2993.85
21/06/2017	473.38	1009.79	2242.39	1548.77	1944.02	650.30	2179.58
22/06/2017	239.08	557.94	1387.89	793.30	778.45	325.24	1271.89
23/06/2017	89.15	287.75	760.47	249.70	336.05	120.53	1024.24
24/06/2017	47.34	352.49	618.47	145.45	252.54	157.52	1069.23
25/06/2017	49.65	297.27	595.90	115.36	279.82	136.59	1051.89
26/06/2017	30.76	219.99	445.79	127.96	200.11	158.73	863.18
27/06/2017	19.82	128.62	307.34	53.68	243.10	94.00	681.94
28/06/2017	14.91	148.49	149.02	40.40	268.45	54.66	396.12
29/06/2017	20.66	154.43	175.47	60.59	274.21	66.99	385.84
30/06/2017	40.29	139.77	258.73	33.60	356.42	37.73	365.47

[B]Daily probability is defined as searching for a specified term on a given day in England, multiplied by ten million allowing for results to be directly compared over time. Cells shaded yellow correspond to a Level 2 heatwave alert (at risk of reaching high temperatures) and cells shaded orange correspond to a Level 3 heatwave alert (have reached high temperatures, Met Office 2017). Emboldened cells correspond to the peak probability.