

# A Method to Assess Seasonality of Urinary Tract Infections Based on Medication Sales and Google Trends

Louise Rossignol<sup>1,2,3,\*</sup>, Camille Pelat<sup>4,5</sup>, Bruno Lambert<sup>6</sup>, Antoine Flahault<sup>3,7</sup>, Emmanuel Chartier-Kastler<sup>8</sup>, Thomas Hanslik<sup>3,9</sup>

**1** Département de médecine générale, UPMC Univ Paris 06, Paris, France, **2** UMRS 707, UPMC Univ Paris 06, Paris, France, **3** U707, INSERM, Paris, France, **4** U738, INSERM, Paris, France, **5** UMRS 738, Université Paris Diderot, Paris, France, **6** IMS-Health, La Défense, France, **7** Descartes School of Medicine, Sorbonne Paris Cité, Paris, France, **8** Urologist hospital universitaire Pitié-Salpêtrière AP-HP, faculté de médecine Pierre et Marie Curie Paris VI, Paris, France, **9** Université Versailles-Saint-Quentin-en-Yvelines, Versailles, France

## Abstract

**Background:** Despite the fact that urinary tract infection (UTI) is a very frequent disease, little is known about its seasonality in the community.

**Methods and Findings:** To estimate seasonality of UTI using multiple time series constructed with available proxies of UTI. Eight time series based on two databases were used: sales of urinary antibacterial medications reported by a panel of pharmacy stores in France between 2000 and 2012, and search trends on the Google search engine for UTI-related terms between 2004 and 2012 in France, Germany, Italy, the USA, China, Australia and Brazil. Differences between summers and winters were statistically assessed with the Mann-Whitney test. We evaluated seasonality by applying the Harmonics Product Spectrum on Fast Fourier Transform. Seven time series out of eight displayed a significant increase in medication sales or web searches in the summer compared to the winter, ranging from 8% to 20%. The eight time series displayed a periodicity of one year. Annual increases were seen in the summer for UTI drug sales in France and Google searches in France, the USA, Germany, Italy, and China. Increases occurred in the austral summer for Google searches in Brazil and Australia.

**Conclusions:** An annual seasonality of UTIs was evidenced in seven different countries, with peaks during the summer.

**Citation:** Rossignol L, Pelat C, Lambert B, Flahault A, Chartier-Kastler E, et al. (2013) A Method to Assess Seasonality of Urinary Tract Infections Based on Medication Sales and Google Trends. PLoS ONE 8(10): e76020. doi:10.1371/journal.pone.0076020

**Editor:** Vishnu Chaturvedi, California Department of Public Health, United States of America

**Received:** June 18, 2013; **Accepted:** August 16, 2013; **Published:** October 25, 2013

**Copyright:** © 2013 Rossignol et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Funding:** The authors have no support or funding to report.

**Competing Interests:** E. Chartier-Kastler is a consultant for Zambon laboratory and B. Lambert is employed by a commercial company IMS-Health. This does not alter the authors' adherence to all the PLOS ONE policies on sharing data and materials.

\* E-mail: rossignol@u707.jussieu.fr

These authors contributed equally to this work.

## Introduction

Urinary tract infection (UTI) is considered to be the most common bacterial infection, with *Escherichia coli* (*E. coli*) involved in 70–80% of cases [1,2]. It affects mainly women. By the age of 24, one third of women have experienced at least one UTI [3].

Seasonal variation in the incidence of *E. coli* bloodstream infection has recently been reported in different countries [4]. UTI is the primary source of these infections [5]. Despite this fact, little is known about the seasonality of UTIs in the community. If seasonality was evidenced, it could help to implement infection control interventions and to orient research along pathophysiological pathways [6,7].

There is a lack of historical systems of UTI surveillance in the population. Indeed, to our knowledge, there is no available UTI incidence time series to test the seasonality of UTI in the community. Epidemiological studies based on microbiological databases are exposed to biases as many uncomplicated UTI are diagnosed and treated without microbiological confirmation. In such a context, non-clinical databases such as medication sales and

web searches can be used for studying seasonal trends. They have already proved to be useful proxies for other diseases [8,9,10,11,12].

Here, we analysed multiple time series of medication sales and Google search data 1) to test if a seasonality of UTI exists or not, and 2) if seasonality exists, to determine its magnitude.

## Materials and Methods

### Ethical considerations

This study was conducted by the Sentinel network. The Sentinel network is regulated by the mixed research unit UMR-S 707 of INSERM and University of Paris VI: Pierre and Marie Curie, in collaboration with the French Institute for Public Health Surveillance (Institut de veille sanitaire, InVS). It has obtained a research authorisation from the French independent administrative authority protecting privacy and personal data (CNIL), n°471 393. This study does not involve human participants and specific ethical approval is not required for this study.

**Table 1.** Seasonal variations in the eight time series between summer and winter.

	Mean in winter (sd)	Mean in summer (sd)	Proportion of increase	p value, Mann-Whitney's test
Fosfomycin and sulfamethizole sales in France <sup>#</sup> (2000–2012)	62.2 (10.9)	74.4 (13.0)	20%	<0.001
Google searches <sup>*</sup> (2004–2011)				
France	53.7 (13.4)	61.7 (14.1)	15%	<0.001
USA	77.0 (7.0)	84.6 (6.4)	10%	<0.001
Germany	50.3 (12.7)	54.2 (10.9)	8%	<0.001
Italy	59.5 (10.6)	70.7 (12.9)	19%	<0.001
China	54.2 (11.7)	62.7 (12.1)	16%	<0.001
Brazil	30.8 (10.7)	27.1 (5.6)	13%	0.003
Australia	55.5 (16.3)	53.2 (9.0)	4%	0.01

<sup>#</sup> expressed as number of boxes per week per 100,000 inhabitants.

<sup>\*</sup>Google searches containing the terms cystitis and/or urinary tract infection. Each is rescaled between 0 and 100 by Google Trends internal processes at download. sd = standard deviation.

doi:10.1371/journal.pone.0076020.t001

## Data

A database containing aggregated medication sales was provided by IMS-Health France, a private company collecting data from 59% of French community pharmacies [13]. Their medication sales data are provided for free, within an agreement between Inserm and IMS Health, and can only be used for scientific research. The provided data consisted of weekly aggregated sales of 593 therapeutic classes, following the Anatomical Classification of the European Pharmaceutical Marketing Research Association [14]. There are many therapeutic classes including antibiotics, many of them having several indications (i.e. fluoroquinolone can be used for UTIs or respiratory infections). However, therapeutic class G04A1 includes only antibiotics with no indication other than uncomplicated UTI in France [15]. The oral form of fosfomycin represented 97% of the sales in this class. This antibiotic is recommended by French guidelines as the first line treatment for uncomplicated UTIs in women [16]. The remaining 3% of the therapeutic class G04A1 was sulfamethizole (as single component substance), an antibiotic still prescribed only for patients with uncomplicated UTI [15,17]. Thus, therapeutic class G04A1 represented a specific indicator of UTI. The IMS-Health database covered the period between 4 September 2000 and 2 September 2012. Sales were measured in numbers of boxes, per week, per 100,000 inhabitants.

Temporal trends in web searches containing either *cystitis* or *urinary tract infection* (or both) were downloaded from Google Trends [18]. This is a public web facility of Google Inc., showing how often a particular search term is searched for, relative to the total number of searches, in various countries and languages. The queries from eight countries were studied: Germany, France, Italy, the USA, Australia, South Africa, China, and Brazil (countries were picked in both hemispheres, with at least one on each continent and, for Europe, one in the north, one in the middle and one in the south). Queries were translated into the native language of the country: *cystite* and *infection urinaire* for France, *cystitis* and *urinary tract infection* for the USA, Australia and South Africa, *blasenentzündung* and *harnwegsinfektion* for Germany, *cistite* and *infezioni delle vie urinarie* for Italy, *膀胱炎* and *尿路感染* for China, and *cistite* and *infecção urinária* for Brazil. Each point of the time series represents the search fraction of the requested query over all queries made the same week in the country. Google Trends does not directly report this fraction but instead rescales it between 0 and 100, 100 being the highest point of the time series. The

Google database covered the period between 11 January 2004 and 30 December 2012.

## Analysis

First, we calculated the mean and standard deviation (sd) of observations in the summer (i.e. week 25 to week 38, June to September) and the winter (i.e. week 51 to week 11, December to March), for each time series. Statistical differences were determined via the Mann-Whitney test. Then, seasonality was evaluated using the Harmonic Product Spectrum (HPS).

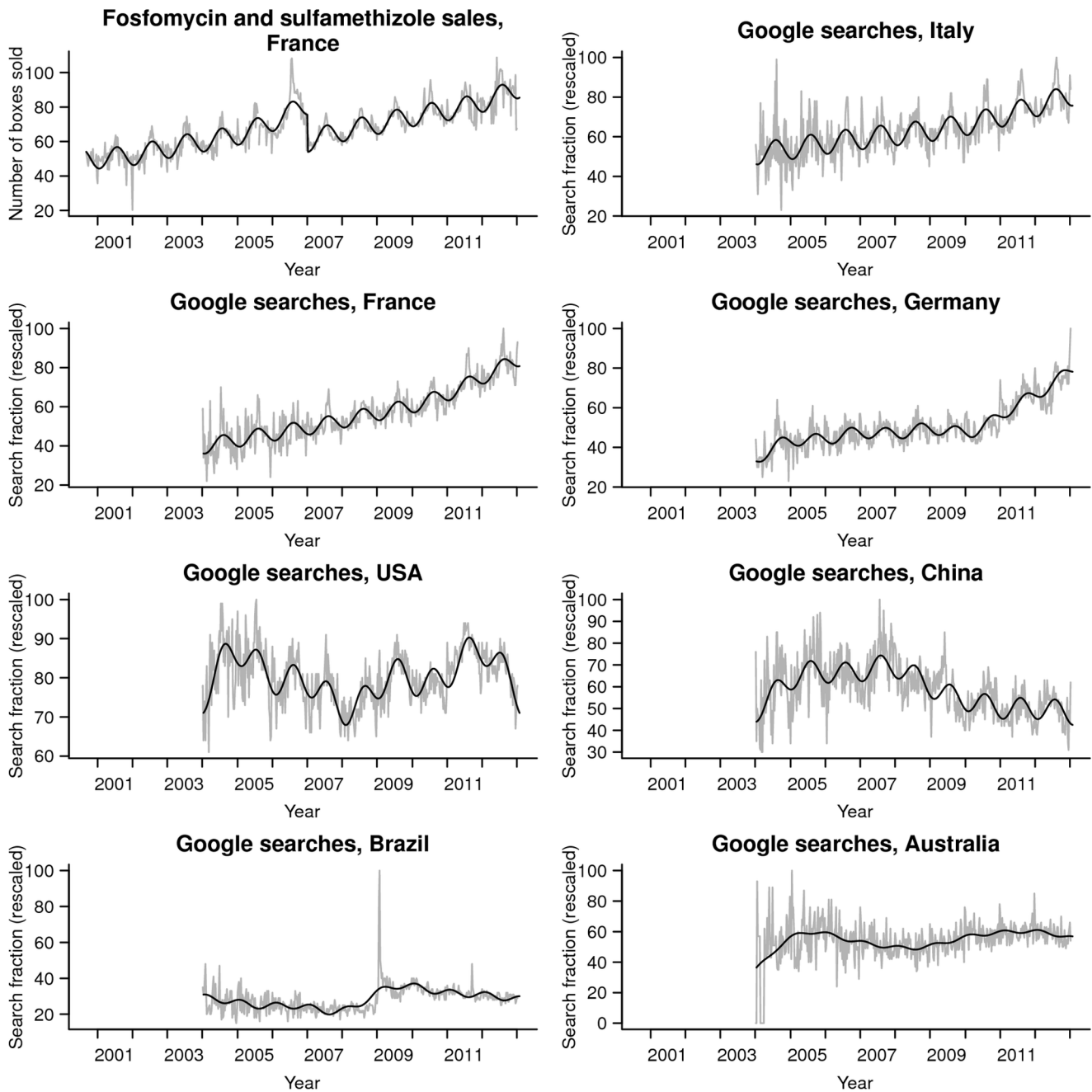
Most of the studied time series displayed some secular trends, often including a long-term tendency to increase. Thus, before looking for seasonal patterns, we estimated and removed these secular trends, using cubic smoothing splines [19]. The smoothing parameter was determined through generalised cross-validation, using the *mgcv* package of the R software package [20]. The frequency spectrum was obtained with Fast Fourier Transform (FFT) on medication sales data from 4 September 2000 to 2 September 2012, and on Google data from 11 January 2004 and 30 December 2012 [21]. We identified the fundamental frequency by applying the HPS method to the frequency spectrum of the detrended time series [22].

## Results

All the time series could be analysed with the exception of South Africa, where the volume of search data was not reported before 2011. We thus discarded this series.

Annual peaks were in the summer for UTI drug sales in France and Google searches in France, Germany, Italy, the USA, and China. They were in the winter for Google searches in Brazil and Australia, i.e. during the austral summer for this southern hemisphere country. The mean (sd) of drug sales in France during the summer was 74.4 (13.0) boxes per 100,000 inhabitants per week. It was 62.2 (10.9) boxes per 100,000 inhabitants per week during the winter, that is, a 20% increase ( $p < 0.001$ ). UTI-related searches on Google were more frequent in the summer compared with the winter, with increases ranging from 8% to 19% (Table 1).

The fundamental period found by HPS was one year for the eight series: 52.17 weeks for sales of the therapeutic class G04A1 in France, and 52.22 weeks for the Google search series in France, Italy, Germany, the USA, China, Brazil, and Australia. Figure 1



**Figure 1. Time series of antibiotic sales and Google search queries for urinary tract infections.** Using “fosfomycin and sulfamethazole” for antibiotic sales, 2000–2012, in France and “cystitis or urinary tract infection” for Google search queries, 2004–2012 in France, the USA, Brazil, Italy, Germany, China and Australia indicated in grey. Black lines show the model’s predictions. The horizontal axis represents the time of year, the vertical axis for antibiotic sales represents drug sales (in number of boxes per 100,000 inhabitants), and the vertical axis for Google search queries represents search fraction. Search fractions are rescaled between 0 and 100 by Google Trends internal processes at download.  
doi:10.1371/journal.pone.0076020.g001

plots the one-year pattern fitted to all time series: it clearly captures the major seasonal variations.

## Discussion

This study provides evidence of the seasonality of UTI, based on specific non-clinical indicators. Peaks were in the summer in France, Germany, Italy, the USA, and China, and during the austral summer in Brazil and Australia. The summer increase

ranged from 8% to 20%, compared to winter periods. The fundamental frequency analysis conducted here was not hypothesis-driven, yet consistently returned a 52-week fundamental period for all series.

The seasonality of UTIs was first suspected in Canada [23]. The authors conducted their study in a specific population, and only culture-positive cases of UTI were included. Also, the method used did permit an estimate of the magnitude of the seasonal variations. Time series analyses were later performed for bloodstream

infections in different countries [5,24,25,26]. All these studies suggested seasonality of *E. coli* bloodstream infections, with summer increases ranging from 12% to 35% compared to the winter or to the rest of the year. No explanation was provided to explain the summer increases. In some previous studies, it was shown that higher temperature was associated with increased frequency of *E. coli* bloodstream infections [5,24,26]. Other hypotheses than the effect of temperature have been proposed, such as reduced micturition due to relative dehydration or more frequent sexual activity due to behavioural changes in the summer [27]. Divergent results have been found, for community-acquired infections caused by extraintestinal pathogenic *E. coli*, including UTI in children, with peak periods during the winter months [28,29]. Such discrepancy could be explained by the fact that these studies did not include UTI only, but also any other kind of *E. coli* infections, or within specific populations (i.e. children). This could result in different pathophysiology and thus in different epidemiology.

During 2011, Google modified their geographical assignment, with retroactive application from 01/01/2011 [30]. Search trend data being given as fraction of the total search volume, this change in geographical assignment would have an impact only if the new defined areas were very discrepant from the old ones in terms of search patterns. This did not appear to be the case, as no irregularity like a sudden increase or drop in mean search fraction was noticed around the date of 01/01/2011 on the time series we downloaded. Also, during 2011 Google updated its categorization taxonomy [31]. However, this did not affect our data since we did not use categories when specifying our queries of interest, but used only keywords.

It is noteworthy here that consistent results were obtained from two different data sources: medication sales and Google search trends, and from seven different countries. When no historical systems of disease surveillance in the population are available, such proxies have already been validated for studying seasonal trends of other diseases, such as gastroenteritis, influenza, dengue, and depression [8,9,10,11].

This study has several limitations. First, we supposed that there was a stationary link between the incidences and the proxies we used. However, regarding the Google searches, any publicised event relating to the studied disease can prompt searches, as has been seen previously, for example, in the case influenza during the increase in avian flu cases in 2005 or during the 2009 pandemic [8,32,33]. In our study, the Google UTI searches in Brazil peaked in January 2009, in conjunction with the death of a Brazilian top model from a UTI, which was highly covered by the press [34].

## References

- Keating KN, Perfetto EM, Subedi P (2005) Economic burden of uncomplicated urinary tract infections: direct, indirect and intangible costs. *Expert Rev Pharmacoecon Outcomes Res* 5: 457–466.
- Stamm WE, Hooton TM (1993) Management of urinary tract infections in adults. *N Engl J Med* 329: 1328–1334.
- Foxman B, Barlow R, D'Arcy H, Gillespie B, Sobel JD (2000) Urinary tract infection: self-reported incidence and associated costs. *Ann Epidemiol* 10: 509–515.
- Richet H (2012) Seasonality in Gram-negative and healthcare-associated infections. *Clin Microbiol Infect* 18: 934–940.
- Al-Hasan MN, Lahr BD, Eckel-Passow JE, Baddour LM (2009) Seasonal variation in *Escherichia coli* bloodstream infection: a population-based study. *Clin Microbiol Infect* 15: 947–950.
- Fisman DN (2007) Seasonality of infectious diseases. *Annu Rev Public Health* 28: 127–143.
- Zeger SL, Irizarry R, Peng RD (2006) On time series analysis of public health and biomedical data. *Annu Rev Public Health* 27: 57–79.
- Pelat C, Turbelin C, Bar-Hen A, Flahault A, Valleron A (2009) More diseases tracked by using Google Trends. *Emerg Infect Dis* 15: 1327–1328.
- Pelat C, Boelle PY, Turbelin C, Lambert B, Valleron AJ (2010) A method for selecting and monitoring medication sales for surveillance of gastroenteritis. *Pharmacoepidemiol Drug Saf* 19: 1009–1018.
- Althouse BM, Ng YY, Cummings DA (2011) Prediction of dengue incidence using search query surveillance. *PLoS Negl Trop Dis* 5: e1258.
- Yang AC, Huang NE, Peng CK, Tsai SJ (2010) Do seasons have an influence on the incidence of depression? The use of an internet search engine query data as a proxy of human affect. *PLoS One* 5: e13728.
- Brownstein JS, Freifeld CC, Madoff LC (2009) Digital disease detection—harnessing the Web for public health surveillance. *N Engl J Med* 360: 2153–2155, 2157.
- Conseil de l'Ordre des pharmaciens (2008) Les Pharmaciens. Panorama au 1er janvier 2008.
- EphMRA (2012) Anatomical Classification. Available: <http://www.ephmra.org>. Accessed 2012 Aug 19.
- Agence Nationale de Sécurité du Médicament (2008) Résumé des caractéristiques du produit : Monuril Adulte. Available: <http://afssaps-prd.afssaps.fr/php/ecodex/frames.php?specid=67951706&typedoc=R&ref=R0152956.htm>. Accessed 2013 Jun 5.

The interpretation of Google searches must therefore always be cautious, and based on several years of observation to avoid such biases, as was done in the present study. Second, there may be a bias in the number of Google searches related, say, to the more frequent use of search engines in the summer. However, Google uses a method to rescale the fraction of searches. If people use search engines more in the summer, this is taken into account in their model. So, the time series reflected well the seasonal variation in specific queries. Moreover, we verified that the Google searches fit well with antibiotic sales in France, which provided reassurance on the pertinence of this proxy for studying seasonality (the lag was at two weeks, data not shown). Third, drug sales estimates were based on a therapeutic class that did not include all antibiotics given for UTI treatment. However, our aim was not to estimate UTI incidence, but only to test seasonality. Thus, completeness was not necessary here, as it was more important to have a specific indicator of UTI, as was the case with the therapeutic class used in this study. Fourth, we observed temporal trends in multiple UTI time series, which increased regularly between 2000 and 2012, either in drug sales or in Google searches. We have no explanations for these trends (the method we used does not permit us to conclude that they reflected an increase in the incidence of UTI). However, our analyses were adjusted based on these trends, which consequently did not alter seasonality estimation.

## Conclusions

To conclude, it was shown here that seasonality of UTI searches exists in various countries, resulting in increases in the summer. The magnitude of these increases has been estimated for the first time, to our knowledge. As UTI is a very frequent disease, any intervention that can help to decrease its incidence will have a useful impact. For example, a next step following this work would be to test control measures such as the reinforcement of UTI prevention measures during the summer.

## Acknowledgments

We thank IMS–Health France for providing their databases.

## Author Contributions

Conceived and designed the experiments: LR TH. Performed the experiments: BL CP LR. Analyzed the data: LR CP TH. Contributed reagents/materials/analysis tools: LR CP. Wrote the paper: LR CP TH AF BL ECK.

16. Agence française de sécurité sanitaire des produits de santé (2008) Recommandations de bonne pratique:diagnostic et antibiothérapie des infections urinaires bactériennes communautaires chez l'adulte. Available: [http://www.infectiologie.com/site/medias/\\_documents/consensus/afssaps-inf-urinaires-adulte-argumentaire.pdf](http://www.infectiologie.com/site/medias/_documents/consensus/afssaps-inf-urinaires-adulte-argumentaire.pdf). Accessed 2013 Jun 5.
17. Agence Nationale de Sécurité du Médicament (2011) Résumé des caractéristique du produit : Rufol. Available: <http://afssaps-prd.afssaps.fr/php/ecodex/frames.php?specid=63584814&typedoc=R&ref=R0190065.htm>. Accessed 2013 Jun 5.
18. Google (2012) Google insights for search. Available: <http://www.google.com/insights/search/#>. Accessed 2013 Jun 5.
19. Hastie T, Tibshirani R (1995) Generalized additive models for medical research. *Stat Methods Med Res* 4: 187–196.
20. Wood SN (2001) Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *Journal of the Royal Statistical Society (B)* 73: 3–36.
21. Singleton RC (1979) Mixed Radix Fast Fourier Transforms, in *Programs for Digital Signal Processing*; Committee IDSP, editor.
22. Schroeder MR (1968) Period histogram and product spectrum: new methods for fundamental-frequency measurement. *J Acoust Soc Am* 43: 829–834.
23. Anderson JE (1983) Seasonality of symptomatic bacterial urinary infections in women. *J Epidemiol Community Health* 37: 286–290.
24. Perencevich EN, McGregor JC, Shardell M, Furuno JP, Harris AD, et al. (2008) Summer peaks in the incidences of gram-negative bacterial infection among hospitalized patients. *Infect Control Hosp Epidemiol* 29: 1124–1131.
25. Chazan B, Colodner R, Edelstein H, Raz R (2011) Seasonal variation in *Escherichia coli* bloodstream infections in northern Israel. *Clin Microbiol Infect* 17: 851–854.
26. Eber MR, Shardell M, Schweizer ML, Laxminarayan R, Perencevich EN (2011) Seasonal and temperature-associated increases in gram-negative bacterial bloodstream infections among hospitalized patients. *PLoS One* 6: e25298.
27. Freeman JT, Anderson DJ, Sexton DJ (2009) Seasonal peaks in *Escherichia coli* infections: possible explanations and implications. *Clin Microbiol Infect* 15: 951–953.
28. George DB, Manges AR (2010) A systematic review of outbreak and non-outbreak studies of extraintestinal pathogenic *Escherichia coli* causing community-acquired infections. *Epidemiol Infect* 138: 1679–1690.
29. Stansfeld JM (1966) Clinical observations relating to incidence and aetiology of urinary-tract infections in children. *Br Med J* 1: 631–635.
30. Google (2011) Note: An improvement to our geographical assignment was applied retroactively from 1/1/2011. Available: <https://support.google.com/trends/answer/1383240?hl=en>. Accessed 2013 Jul 31.
31. Google (2012) Update to Vertical Taxonomy. Available: <https://support.google.com/trends/answer/2370165?hl=en>. Accessed 2013 Jul 31.
32. Valdivia A, Lopez-Alcalde J, Vicente M, Pichiule M, Ruiz M, et al. (2010) Monitoring influenza activity in Europe with Google Flu Trends: comparison with the findings of sentinel physician networks - results for 2009–10. *Euro Surveill* 15.
33. Valdivia A, Monge-Corella S (2010) Diseases tracked by using Google trends, Spain. *Emerg Infect Dis* 16: 168.
34. FoxNews.com (2009) Model dies after losing hands, feet to urinary tract infection. Available: <http://www.foxnews.com/story/0,2933,482548,00.html>. Accessed 2013 Jun 5.