

# Head Lice Surveillance on a Deregulated OTC-Sales Market: A Study Using Web Query Data

Johan Lindh<sup>1,2\*</sup>, Måns Magnusson<sup>1</sup>, Maria Grünewald<sup>1</sup>, Anette Hulth<sup>1</sup>

<sup>1</sup> Swedish Institute for Communicable Disease Control, Solna, Sweden, <sup>2</sup> Department of Microbiology, Tumor and Cell Biology, Karolinska Institutet, Stockholm, Sweden

## Abstract

The head louse, *Pediculus humanus capitis*, is an obligate ectoparasite that causes infestations of humans. Studies have demonstrated a correlation between sales figures for over-the-counter (OTC) treatment products and the number of humans with head lice. The deregulation of the Swedish pharmacy market on July 1, 2009, decreased the possibility to obtain complete sale figures and thereby the possibility to obtain yearly trends of head lice infestations. In the presented study we wanted to investigate whether web queries on head lice can be used as substitute for OTC sales figures. Via Google Insights for Search and Vårdguiden medical web site, the number of queries on “huvudlöss” (head lice) and “hårlöss” (lice in hair) were obtained. The analysis showed that both the Vårdguiden series and the Google series were statistically significant ( $p < 0.001$ ) when added separately, but if the Google series were already included in the model, the Vårdguiden series were not statistically significant ( $p = 0.5689$ ). In conclusion, web queries can detect if there is an increase or decrease of head lice infested humans in Sweden over a period of years, and be as reliable a proxy as the OTC-sales figures.

**Citation:** Lindh J, Magnusson M, Grünewald M, Hulth A (2012) Head Lice Surveillance on a Deregulated OTC-Sales Market: A Study Using Web Query Data. PLoS ONE 7(11): e48666. doi:10.1371/journal.pone.0048666

**Editor:** D. William Cameron, University of Ottawa, Canada

**Received:** May 25, 2012; **Accepted:** September 28, 2012; **Published:** November 7, 2012

**Copyright:** © 2012 Lindh 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:** These authors have no support or funding to report.

**Competing Interests:** The authors have declared that no competing interests exist.

\* E-mail: johan.lindh@smi.se

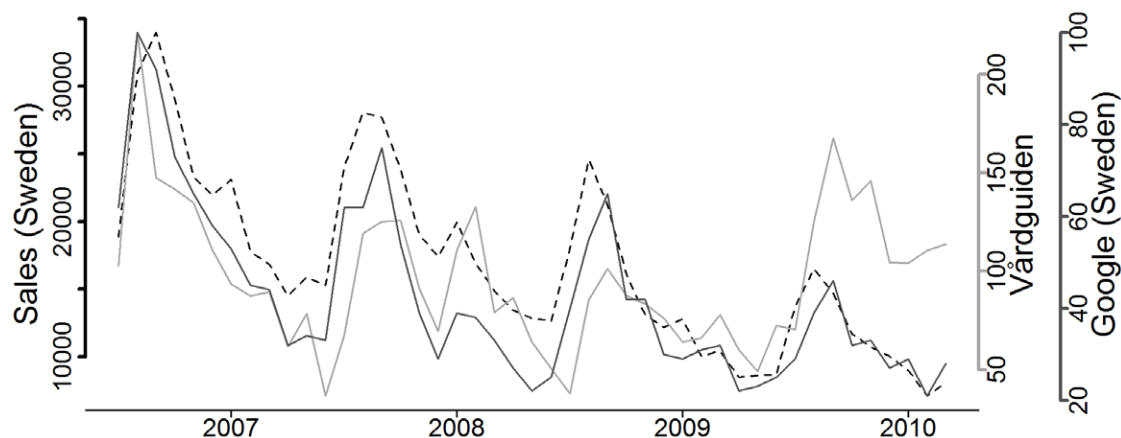
## Introduction

The head louse, *Pediculus humanus capitis*, is an obligate ectoparasite that causes infestations of humans. The parasite has been with us since before the modern man, *Homo sapiens*, diverged, and it has been proposed that both *Homo erectus* and *Homo neanderthalis* were infested with lice [1]. Human head lice live only on the human head and quickly die if they are removed [2]. Head lice infestations are mostly asymptomatic, although skin irritation and occasional secondary infection from scratching may occur [3]. Recently it has also been demonstrated that head lice *in vivo* can contain the bacteria, *Bartonella quintana* [4]. Even if no epidemiological data can be found, it is not debatable to say that head lice are the most common parasite in Europe and that they mainly infest children [5,6,7]. The parasites are spread between hosts via direct contact and are more common among girls [7]. In a study from Germany a clear seasonal variation in sale reports of number of pediculocides was seen during August/September and January/February [8]. This can be explained by the increases of contacts during school start, mainly between children [9]. The following decrease in sales can be explained by increased awareness, rapid diagnosis and treatment with pediculocides [8,10]. The number of pediculocides sold should therefore follow the demand and reflect the actual occurrence of head lice in the society over time [8]. The Swedish Institute for Communicable Disease Control (SMI) follows the prevalence of head lice in society. This surveillance aids in tracking the level of treatment resistance, recommending adequate preventive measures and evaluating the effect of these recommendations.

Previously, due to cost, Over-the-Counter (OTC) sales figures for treatment products were the only practically feasible source of information for head lice prevalence in Sweden. However, with the deregulation of the Swedish pharmacy market on July 1st 2009, the number of different outlets went from one to several. Therefore, the possibility to obtain complete sales figures for the whole country disappeared.

At SMI, a surveillance system based on queries submitted to a Swedish medical web site ([www.vardguiden.se](http://www.vardguiden.se)) has previously been developed [11]. The Vårdguiden website is owned by the Stockholm County Council and provides various kinds of medical information in Swedish for the county's citizens. SMI has been given access to the query logs for the Vårdguiden website. The web query-based system is used at SMI as a complement to the regular surveillance of influenza [12,13] and of norovirus [14]. Another part of the web query-based system allows for time series to be generated for any query terms. The system contains data from June 2005.

In the presented study we wanted to investigate whether web queries on head lice can be used as a substitute for OTC sales figures, as these figures are no longer comprehensive. Furthermore, we investigated whether there is a correlation between usages of different search engines with terms on head lice, and sales of pediculocides in Sweden and whether there is any difference between the Swedish web search engine “Vårdguiden” and Google web searches with respect to this query.



**Figure 1. Web queries and sales (January 2006–June 2010).** Dotted line – Sales (Sweden), light grey line – Vårdguiden, black line – Google (Sweden).

doi:10.1371/journal.pone.0048666.g001

## Materials and Methods

### Data Sources

In the presented study, we used query data from two different kinds of websites; the medical website Vårdguiden and the general purpose search engine (Google). As the Vårdguiden logs do not contain any geographical information, the data are aggregated on a national level. Google Insights for Search is a service provided by Google where a user can obtain the relative search load on queries submitted to the Google search engine [15].

We extracted the number of queries on “huvudlöss” (head lice) and “hårlöss” (lice in hair) submitted to the Vårdguiden medical web site for the period July 2006 until March 2010 from the query logs. Through Google Insights for Search, we obtained the relative search load on Google for Sweden on the same two queries for the same time period. The time series were based on data aggregated by month. OTC sales figures per month for pediculicides for lice treatment for the same period for Sweden were obtained from Apotekens Service AB.

### Statistical Analysis

The relationship between the sales and the web search series was analyzed with linear regression. Since the data analyzed are a time series, one of the main problems is auto correlated errors. The residuals were studied, and autocorrelation of lag one could be identified in all models. To account for this generalized least squared regression (GLS) with an AR (1)-model were used to account for auto correlated error in the statistical testing [16]. Seasonal components were used with a period of one year. Amplitude and phase were estimated from the data.

## Results and Discussion

Both the Google time series and the Vårdguiden time series were studied with regard to how well they could identify the trend and the seasonal pattern of the sales. This was done by comparing Akaike information criteria (AIC) between non nested models of sales with Google and Vårdguiden time series as covariates (Table 1). The trend was defined as a linear trend and the seasonal factor was defined as a sine/cosine function. The correlation between Google and sales was much larger than the correlation

**Table 1.** AIC and parameter estimates of time regression models (AR(1)).

Model	Time series comp.	AIC	(Intercept)	Trend (t)	Seasonal comp. (cos(t))	Seasonal comp. (sin(t))	Web queries	Phi			
No web queries	None	852	15988	***	–	–	–	0.89			
	Trend	838	24530	***	–348	**	–	0.72			
	Trend and season	793	24076	***	–325	***	–797	4631	***	0.50	
Google	None	799	5167	**	–	–	263	***	0.76		
	Trend	787	8812	***	–130	*	–	–	253	***	0.61
	Trend and season	756	10552	***	–142	*	–452	1783	*	218	***
Vårdguiden	None	833	9851	*	–	–	59	***	0.91		
	Trend	819	18368	***	–337	***	–	–	61	***	0.70
	Trend and season	779	19284	***	–308	***	–114	3736	***	45	**

Significance codes: \*\*\*\*:  $p < 0.001$ , \*\*\*:  $p < 0.01$ , \*\*:  $p < 0.05$ .

doi:10.1371/journal.pone.0048666.t001

**Table 2.** Correlation matrix, web queries and sales in Sweden, for lags 0, 1 and 2.

Correlation	Vårdguiden	Google	Sales	Vårdguiden, lag 1	Google, lag 1	Sales, lag 1	Vårdguiden, lag 2	Google, lag 2	Sales, lag 2
<b>Vårdguiden</b>	1								
<b>Google</b>	0.629	1							
<b>Sales</b>	0.474	0.920	1						
<b>Vårdguiden, lag 1</b>	0.600	0.367	0.286	1					
<b>Google, lag 1</b>	0.529	0.799	0.796	0.629	1				
<b>Sales, lag 1</b>	0.426	0.769	0.846	0.474	0.920	1			
<b>Vårdguiden, lag 2</b>	0.395	0.114	0.090	0.600	0.367	0.286	1		
<b>Google, lag 2</b>	0.365	0.584	0.561	0.529	0.799	0.796	0.629	1	
<b>Sales, lag 2</b>	0.313	0.576	0.617	0.426	0.769	0.846	0.474	0.920	1

doi:10.1371/journal.pone.0048666.t002

between Vårdguiden and sales (Table 2, Figure 1). The differences between the two series were tested statistically by including both series in a GLS model. The analysis showed that both Vårdguiden series and the Google series were statistically significant ( $p < 0.001$ ) when added separately, but if the Google series were already included in the model the Vårdguiden series was not statistically significant ( $p = 0.5689$ ). The two series were studied with regarding to how well they could explain the trend and seasonal variation in the sales data. When including trend and seasonal variables in each model the model using Vårdguiden data reduced the AIC much more than the Google data series. The Vårdguiden AIC was reduced with 19 when trend was included and an additional 32 in AIC when seasonal variables were included, while the model with Google as a covariate reduced the AIC with 14 (trend) and additional 23 (seasonal variables).

In this study we investigated how well web queries can be used to follow the spread of lice in the population, by using sales as a proxy for spread of lice. Our main finding was that Google web searches are a better covariate to explain both the trend and the seasonal patterns of the sales, but that the Vårdguiden data can be used as well. A possible explanation of the outcome can be that the sales are decreasing for other reasons than a decrease in lice. Google queries and Vårdguiden queries can have different effect on the sales. For example, if one searches for lice at Google Sweden, different products are found while searching for lice on Vårdguiden rather gives the guidance of using, the relatively cheap, louse comb. This is a natural result of the two business models for the two search engines; Google is selling advertisements, while Vårdguiden has the purpose of informing the Swedish citizens. In conclusion, the queries can detect if there is an increase or decrease of head lice infested humans in Sweden over a period of years, but not the number of head lice infested people

(Figure 1). The queries can also be as reliable a proxy as the OTC-sales figures for pediculicides previously were.

Google search query data have previously been shown to correlate with epidemiological data for communicable diseases, for example listeriosis [17], salmonella [18], West Nile virus [19], MRSA [20], and influenza [21,22]. The Vårdguiden data, which originate from a medical web site, have been shown to be a valuable complement to the surveillance of communicable diseases [11,12,13,14]. Web queries have also been shown to predict sales data time series for, for example, video games and cinema [23]. In this study we explored whether web queries can replace OTC sales data on head lice treatment as a proxy for head lice prevalence in society. It is our belief that there are a large number of other diseases for which web queries can complement other surveillance methods for communicable diseases. It would also be interesting to further explore the differences between query data from a general purpose search engine and a medical web site.

## Conclusions

Web queries on head lice can detect an increase or decrease of head lice infested humans in Sweden over a period of years, and be as reliable a proxy as the OTC-sales figures.

## Acknowledgments

Thanks to Vårdguiden for granting access to the query logs and to Euroling for implementing parts of the automatic transfer. Thanks also to the anonymous reviewers for valuable comments.

## Author Contributions

Conceived and designed the experiments: JL AH. Performed the experiments: AH MM. Analyzed the data: JL MM MG AH. Wrote the paper: JL MM MG AH. Quality assured the article: JL MG AH.

## References

1. Leo NP, Barker SC (2005) Unravelling the evolution of the head lice and body lice of humans. *Parasitol Res* 98: 44–47.
2. Lebwohl M, Clark L, Levitt J (2007) Therapy for head lice based on life cycle, resistance, and safety considerations. *Pediatrics* 119: 965–974.
3. Bonilla DL, Kabeya H, Henn J, Kramer VL, Kosoy MY (2009) *Bartonella quintana* in body lice and head lice from homeless persons, San Francisco, California, USA. *Emerg Infect Dis* 15: 912–915.
4. Angelakis E, Rolain JM, Raoult D, Brouqui P (2011) *Bartonella quintana* in head louse nits. *FEMS Immunol Med Microbiol* 62: 244–246.
5. Buczek A, Markowska-Gosik D, Widomska D, Kawa IM (2004) *Pediculosis capitis* among schoolchildren in urban and rural areas of eastern Poland. *Eur J Epidemiol* 19: 491–495.
6. Clore ER, Longyear LA (1990) Comprehensive pediculosis screening programs for elementary schools. *J Sch Health* 60: 212–214.
7. Speare R, Buettner PG (1999) Head lice in pupils of a primary school in Australia and implications for control. *Int J Dermatol* 38: 285–290.
8. Bauer E, Jahnke C, Feldmeier H (2009) Seasonal fluctuations of head lice infestation in Germany. *Parasitol Res* 104: 677–681.
9. Mossong J, Hens N, Jit M, Beutels P, Auranen K, et al. (2008) Social contacts and mixing patterns relevant to the spread of infectious diseases. *PLoS Med* 5: e74.
10. Rukke BA, Birkemoe T, Soleng A, Lindstedt HH, Ottesen P (2012) Head lice in Norwegian households: actions taken, costs and knowledge. *PLoS One* 7: e32686.

11. Hulth A, Rydevik G (2011) GET WELL: an automated surveillance system for gaining new epidemiological knowledge. *BMC Public Health* 11: 252.
12. Hulth A, Rydevik G (2011) Web query-based surveillance in Sweden during the influenza A(H1N1)2009 pandemic, April 2009 to February 2010. *Euro Surveill* 16(18). Available: <http://www.eurosurveillance.org/images/dynamic/EE/V16N18/art19856.pdf>. Accessed 2012 Oct 15.
13. Hulth A, Rydevik G, Linde A (2009) Web queries as a source for syndromic surveillance. *PLoS One* 4: e4378.
14. Hulth A, Andersson Y, Hedlund KO, Andersson M (2010) Eye-opening approach to norovirus surveillance. *Emerg Infect Dis* 16: 1319–1321.
15. Google (2012) Google Insights for Search.
16. Fox J, Weisberg S (2010) Time-Series Regression and Generalized Least Squares in R, An appendix to An R companion to applied regression, Second edition. Available: <http://socserv.mcmaster.ca/jfox/Books/Companion/appendix/Appendix-Timeseries-Regression.pdf>. Accessed 2012 Oct 15.
17. Wilson K, Brownstein JS (2009) Early detection of disease outbreaks using the Internet. *CMAJ* 180: 829–831.
18. 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.
19. Carneiro HA, Mylonakis E (2009) Google trends: a web-based tool for real-time surveillance of disease outbreaks. *Clin Infect Dis* 49: 1557–1564.
20. Dukic VM, David MZ, Lauderdale DS (2011) Internet queries and methicillin-resistant *Staphylococcus aureus* surveillance. *Emerg Infect Dis* 17: 1068–1070.
21. Ginsberg J, Mohebbi MH, Patel RS, Brammer L, Smolinski MS, et al. (2009) Detecting influenza epidemics using search engine query data. *Nature* 457: 1012–1014.
22. Eysenbach G (2006) Infodemiology: tracking flu-related searches on the web for syndromic surveillance. *AMIA Annu Symp Proc*: 244–248.
23. Goel S, Hofman JM, Lahaie S, Pennock DM, Watts DJ (2010) Predicting consumer behavior with Web search. *Proc Natl Acad Sci U S A* 107: 17486–17490.