Statistics Netherlands

Discussion Paper

Complementing official health statistics with internet search indices

2015 | 08

Carin Reep
Bart Buelens

Official statistics and internet search behaviour are compared for themes related to health and lifestyle. While search indices – the frequency with which internet search queries are used – provide a different quantitative measure than survey based statistics on the same theme, they are found to be related and to complement each other. The present study uses internet search data from Google Trends, results from the Health Survey conducted by Statistics Netherlands, and data collected by the National Institute for Public Health and the Environment. Reported incidences of measles during the 2013-2014 measles epidemic in the Netherlands correlate well with the rapidly available data of regional internet search behaviour for measles – a finding closely related to the well-known example 'Google Flu'. Search indices exhibit seasonal patterns, for example on lifestyle topics including smoking cessation, dieting and physical exercise. Such temporal details are often not possible to achieve with existing surveys due to the relatively small sample sizes. Similarly, small sample sizes prevent an accurate estimation of uncommon phenomena such as the use of various specific types of treatments that are generally classed as 'alternative medicine'. The large amount of data underpinning the internet search queries, however, does provide quantitative insight into these matters. Since search queries are anonymous, search indices may provide a better picture of sensitive topics than surveys in which questionnaires are administered by an interviewer. An example of sexually transmitted diseases is given. Combining internet search behaviour with official statistics published by National Statistical Institutes provides valuable additional information on themes important for policy making and monitoring. Internet search indices have the potential to become more closely integrated into the regular survey process, for example by assisting the development of appropriate survey based measurement instruments, or by serving as covariate information in models to quantify discontinuities due to survey redesigns or to improve the precision of regional estimates.

Keywords
Health trends, seasonal patterns, medical consumption, spatial distribution, Google Trends

Contents

1. Introduction

The potential association between internet search behaviour and real world phenomena has received considerable attention following the 2009 Nature article reporting a strong correlation between the incidence of flu and Google search intensity for flu related terms (Ginsberg et al, 2009; Google.org Flu Trends, 2013). In regions where the flu incidence was higher, more online searches were performed for information about flu. This is one of the most famous examples that show that the Internet search behaviour patterns associate with real phenomena in society.

Google offers a service called Google Trends, making it easy to analyse search queries. This study investigates if and how internet search indices can contribute to a quantitative picture of health (related) topics in the Netherlands. Official health statistics are limited and mainly cover the most important and well known diseases. Internet search trends can possibly supplement the existing figures. Besides, a very positive characteristic of the search trends is that they are faster and more frequently available than the existing data. Currentness, speed, and volume are typical characteristics of 'Big data' sources, which usually register events, in this case searches (Daas and Van der Loo, 2013).

This report is structured as follows. Section 2 outlines the background of this research. In paragraph 3, the operation and capabilities of Google Trends are described. Section 4 briefly discusses existing health statistics from Statistics Netherlands and other agencies. Section 5 provides a description of some sample analyses with Google Trends on health in general, incidence of certain diseases, lifestyle-related topics, and care. Section 6 ends this report with conclusions and recommendations.

2. Background

2.1 Big data

This research on Google Trends can be situated within the context of big data, in which non-traditional, large datasets are used for making relevant quantitative output. Big data files are large, contain up-to-date information, and usually contain 'events'. It is often not possible to link these events to persons: this information is often not available or will not be used or disclosed. Big data is often too selective with respect to for official statistics relevant populations, for example all Dutch (see section 3.4).

Examples where big data is examined for possible use at Statistics Netherlands are data over traffic routing, social media messages, and mobile telephone data (Daas et al, 2013). Each involves large numbers of events, which explore a possible connection with real social or economic phenomena. In most cases it is a challenge to store and process these data, before analysis can be performed. The advantage of Google Trends is that the storage of and access to the files of events (here queries) are controlled by Google. The user of Google Trends can therefore deal with these big data in a simple way and instantly perform analysis. A disadvantage is that Google has complete control over the design of this system, the underlying methods, and the interface (see section 3.3) and that ambiguities are not easily detected and sometimes cannot be detected at all.

2.2 Literature

Particularly in response to the article in Nature about the flu trends (Ginsberg et al, 2009), a number of studies have been undertaken that seek to establish a relation with Google on search behaviour and actual social phenomena. It is not the intention here to give an exhaustive list, but some examples are interesting and relevant, especially regarding the health theme. Apart from health, seeking behaviour is already extensively analysed in other domains, for example in market research on the effects of advertising, interest in products, and purchasing behaviour.

Carr and Dunsiger (2012) examined user search for terms related to lifestyle, such as smoking, obesity, dieting and exercise. Many preventable deaths are caused by unhealthy lifestyles. On the basis of search terms they studied the interest in these issues in the (US) population, with the aim to measure the effect of government campaigns.

Ayers et al. (2011) compared search behaviour for stopping smoking between US states where an increase in tax on tobacco was implemented and US states where the tax increase was not implemented. They found that the introduction of the tax led to 12% more searches for quitting smoking, and a 28% increase in searching for illegal providers of tobacco products.

Seasonal trends in internet search queries have been investigated in the area of mental health (Ayers et al., 2013).

A link between search activity on Google, and results of a health survey is discussed in Murphy et al. (2011). They compared the use of the drug salvia divinorum as reported in a survey with search behaviour related to this drug on Google, and popularity on Twitter. The authors observed a correlation, but also expressed reservations . They concluded that that Google and Twitter analyses can provide complementary information but cannot (yet) replace sampling studies . They use the term infoveillance to indicate that it is surveillance in which online information is used. This term is first used by Eysenbach (2009) in the early phases of the research on flu trends.

The UK Office for National Statistics (ONS) has already reported on research into the use of Google Trends as a data source (Ayoubkhani, 2012). The theme in this case is not related to health; they study the Retail Sales Index, an index of retail turnover. The intention was to

determine if trends in search behaviour can be used to monitor the plausibility and quality of the official figures. For certain goods this indeed was the case, especially where the turnover is quite volatile, such as furniture and lighting, and audio and video material.
The literature quoted here concerns exploratory research, and includes hardly any references to previous research. This shows that the use of Google Trends is being studied in recent years, but there still is no consensus on how the results can be used. The present report contributes to this investigation, and contains mostly exploratory analyses from Google Trends that is where possible compared to results from regular sources.

3. Google Trends

3.1 Trends in search queries

Google Trends is a service, offered by Google, that provides information about trends in search queries registered in to their search engine. Google Trends is available on-line via a web browser (Google Inc., 2014). This tool can be used at no cost. Google gives data from January 2004 onwards and has offered the tool since 2006. In 2008 Google produced a more advanced version of the former 'Google Trends' which was named 'Insights for Search'. Since 2012 both variants are combined into 'Google Trends' (Wikipedia, 2013). Google Trends can be used for personal, educative and research goals. When used for other goals permission is needed beforehand (Google, 2012).

The most important functionality of Google Trends is presenting the development in time of the relative number of search queries of specified words or phrases. It is also possible to compare the relative trends between different search queries, or compare the relative popularity between different levels of regions. Countries can be compared with each other (which is only relevant when the same language is used) and regions or cities within the same country can be compared. The location information that Google uses can be based on a variety of technologies including IP address and GPS of the device (computer or mobile device) used for the search. For every search query Google Trends also gives the terms that are most similar and their popularity. Google Trends also integrates relevant news items in some peaks of the trends when these news items can account for the raised popularity shown. In this way one might see a connection between media-attention and search volume.

3.2 How to use Google Trends

For every query in Google Trends one can choose from a number of options:

- The search word(s).
    - When a line has to represent the searches that contain one of a series of words, the words have to be separated by a plus sign (+). If searches should not contain a specific word, that word has to be preceded by a minus sign (-).For example, if a word is often misspelled and the researcher wants to include more than one spelling variant for 'Mexican' they have to enter **Mexican + Meksican + Mexikan.** If one is interested in a number of words that all could represent the topic of interest, for example 'flu', one has to enter **sneezing+cold+blocked nose+ fever+ flu.** The maximum number of words per trend is 30.
    - If a line has to represent the search trend of an exact combination of words in the same order, like 'Mexican flu', the words have to be enclosed in double quotes **"Mexican flu".** Without the double quotes all search queries that comprises **Mexican** as well as **flu** are given in whatever order they were entered. Single quotes, apostrophes and brackets are neglected by Google Trends, writing in capital or small letters give the same results and the order of the words doesn't matter (if not enclosed in double quotes). However, diacritics have to be entered. So **'Mexican flu'** and **Mexican flu** and **(Mexican flu)** and **flu Mexican** and **mexican flu** all give the same results. But **Méxican flu** is different.

- The search domain. Default is ‘Web Search’, but trends can also be asked in the domains ‘Image Search’, ‘News Search’, ‘Google Shopping’ or ‘YouTube Search’.
- Location. Default is ‘Worldwide’, but almost all countries can be chosen. Most of them also in sub regions; in the Netherlands provinces can be chosen.
- Period. Default is ‘2004-present’. The most recent period that can be chosen is ‘Past 7 days’. The most recent number is that of the day before yesterday or yesterday. When information is wanted about the period before the past 90 days, only monthly periods can be chosen.
- Category. Default is ‘All categories’, so trends are given over the whole search engine. The user can choose from 25 main categories and within those also subcategories are chosable. If no search term is entered and but only a category, Google gives the trend of that category compared to the whole search engine.

Google Trends presents its results directly in graphs and regional maps. The trends are given in line graphs. Per task or ‘query’ different trends can be compared to one another. This may be different (combinations) or search terms, or different periods, or different regions. There is a maximum of 5 different trends that can be compared per query.
The underlying data can be downloaded as a CSV file. In case a period of 3 months or less is selected Google Trends provides graphs and data on a daily basis. When selecting a period between 3 months and two years graphs and data are provided on a weekly level. For longer periods the line graph is given on a monthly level, but the downloaded data are only given on a weekly level. For search terms with low popularity also the downloaded data are given by month instead of by week. Regional maps and its downloaded data are always given for the whole period that is entered.
Google Trends shows only integers in both the graphs and maps as in the downloads. The numbers in the graphs and maps from Google Trends can be viewed by navigating with the mouse.
Google Trends provides no absolute search volume numbers, but normalized and scaled figures. For this purpose the number of searches for the search term is divided by the total number of searches in the chosen region and category, in a given month (or week or day). The resulting numbers are scaled in proportion to the period with the highest relative search volume, which gets the value 100. The consequence is that the numbers will change if a different period is chosen in which the peak time falls in a different month (or week day).
If an entire class is viewed Google Trends provides a different type of graph (and other figures). This chart illustrates the changes in the course of time as a percentage of growth in relation to the first date in the graph (or the first date for which data are available). The y-axis of the category comparison chart then ranges from -100% to +100%, and the graph starts at 0.
Regional differences are shown through "heat maps", in this case maps in which the differences between the regions in the selected period are displayed through colour strength. When interpreting heat maps over several periods, some caution is needed. The variation in colour intensity shows the regional differences in search behaviour over the relevant period. But the colour strength cannot be compared over different periods or different heat maps. So if a province has changed colour between successive periods it does not necessarily mean that the relative seeking behaviour in that province has changed. It just means that the search behaviour compared to the other provinces has changed.
Google Trends distinguishes different (sub)categories that can be selected. The search terms are classified by Google, presumably by means of a machine-learning method where broader search patterns are examined. The trend for the specified search term is displayed for the selected category. In choosing a category Google Trends recommends the three most relevant

categories, based on the categories in which the term is most common. If no category is selected, the trend is calculated on the total. Google.nl trends distinguishes 25 categories, namely:

Arts & Entertainment
Autos & Vehicles
Beauty & Fitness
Books & Literature
Business & Industrial
Computers & Electronics
Finance
Food & Drink
Games
Health
Hobbies & Leisure
Home & Garden
Internet & Telecom
Jobs & Education
Law & Government
News
Online Communities
People & Society
Pets & Animals
Property
Reference
Science
Shopping
Sports
Travel

Those categories are divided in subcategories and those are subdivided again in other categories. The category 'Health' is divided into 21 sub-categories. Those are:

Ageing & Geriatrics
Alternative & Natural Medicine
Health Conditions
Health Education & Medical Training
Health Foundations & Medical Research
Health News
Medical Devices & Equipment
Medical Facilities & Services
Medical Literature & Resources
Men's Health
Mental Health
Nursing
Nutrition
Oral & Dental Care
Pediatrics
Pharmacy
Public Health
Reproductive Health
Substance Abuse
Vision Care
Women's Health

3.3 Google Trends methodology

Although the interface of Google Trends is relatively simple and the results are clearly presented, the methodology on which this tool is based is not so simple and neither is it fully documented. This section lists a number of points that emerged during our analysis.

Interpretation

The mechanism used by Google Trends would appear to be straight forward. However, interpretation of the results has to be done with caution. When a downward trend is observed for a given search term, this does not necessarily mean that the absolute number of searches for that term has decreased since the results are relative. A declining trend could theoretically be caused by an increase in the total search volume caused by other terms, while the search volume for the specific term remained constant. Usually this is not a main concern as there are rarely terms that have such a massive influence on the total search volume. When the studied term has only one predominant meaning it is advisable not to limit it to a specific category when analysing its trend in order to minimise the influence of other search terms. Relative figures offer advantages over absolute numbers, because this corrects for the increase or decrease in the use of Google in general.

Database searches

Results from Google Trends are based on a sample of all searches conducted through Google (Choi and Varian, 2011). The total number of searches since 2004 will probably reach billions. Therefore it is not feasible for every single analyses in Google Trends to search through all these web searches and combine these to the aggregated output. It is understandable that Google Trends is based on a subset of all searches but the exact operationalization of this is not well

documented. For instance, the sample size is not clear; is it in total, or by region and period? It is not clear if there are multiple samples, if the methods have changed over time or whether and how often the sample is changed.
Relative values instead of absolute numbers are presented in Google Trends. Relative values are as a rule very useful but this often makes comparison of different Google Trends analyses difficult or impossible. During research for this report some anomalies were observed which seem to be inconsistencies. As example, an analysis of a search term over the period of the year 2012 gives a peak (100) in the month of March. A similar analysis for the same search term over the period of 2011 to 2013 also shows the peak (100) in March 2012 but the values of the other months in 2012 differ slightly. This phenomenon could be caused by an underlying sampling mechanism. A consequence is that research results are not exactly reproducible. So can an analysis that is exactly the same give slight different results when it is repeated at a later time- in some other underlying sample.

Month, week or day

Behind the scenes, days are aggregated into weeks and months. In Google Trends' definition of day, it is not clear whether and how it takes account of time zones. For longer periods of time Google Trends shows the graph in months. The downloaded file however nearly always gives data on a weekly level. It is therefore impossible to exactly calculate the graphically displayed monthly numbers. This is for two reasons. Firstly weeks nearly always straddle two months and the weekly rate cannot be divided between those two months without figures given on a daily basis- figures that Google Trends does not provide. Secondly, this is because the weekly figures themselves are also scaled to the range 0-100. If the popularity of the peak month is mainly due to one particular week, the whole trend decreases in level compared with the figure of Google Trends. In the Google figure the level is determined by the peak month, in which an extreme week has less impact. With weekly numbers in which an extreme week occurs, the numbers of many other weeks will be rounded to 0. This hardly occurs in the Google-month figures. So it can happen that the downloaded file for all four weeks of a particular month gives a 0, while the value of the month in the graph is still higher than 0.

3.4 Representativeness

The extent to which search trends say something about the overall Dutch population depends on several factors: the proportion of the population that has access to the internet, the proportion of it that uses the Google search engine, and the proportion of it that uses the internet to find information about health.
Use of the internet by the Dutch population has been investigated by Statistics Netherlands. According to our agencie's research into ICT in 2012, 93% of the population aged 12 years or over used the internet in the three months prior to the survey (Statistics Netherlands, 2014a). Most people use a search engine (91%). Unsurprisingly this figure was less in preceding years, an effect especially pronounced among the elderly. In 2005 only 30% of those over 65 used a search engine on the internet; in 2012 this had grown to 67% and for young people (12-25 years) the percentage since 2005 is above 95%. This creates a selectivity in Google Trends, in the sense that older people in the earlier years are underrepresented. This would distort in particular the trends in search behaviour for age-related diseases. In addition, less educated people search less over the internet than the higher educated which could also bias the results. According to the National search engine Monitor, Google is the most used search engine in the Netherlands (iProspect, 2014). It has had a market share of around 93 percent for many years, from 2006 until at least the time of writing this report. In 2004 it was about 70%. So most searches conducted by the Dutch are included in the Google Trends analyses. Google has a

monopoly in the Netherlands that is currently not threatened because there do not seem to be any serious parties interested in investing in a Dutch language search engine; there is a preference of first making the English language version in order.

In addition, there is no reason to expect that people would select a different search engine to search for health related topics. Finally, there could well be some selectivity in the types of people (age, level of education, etc.) who look for information on the internet specifically related to health although we have no possibility of analysing this. Assuming such selectivity remains constant through time, trend analysis using Google Trends remains relevant.

4. Health statistics

In the next section the analysed and presented results based on Google Trends are, where possible and relevant, linked with existing figures from official sources. (Official) figures on health are published by both our own Statistical Agency and other (non-)government agencies.

Statistics Netherlands collects information for its health statistics from almost 50 different data sources (Statistics Netherlands, 2013a). Many of them are registrations, but a major source is the Health Survey. Topics on which Statistics Netherlands reports include use of health care services, care, mortality and causes of death, height and weight, smoking and drinking behaviour, sick leave, working conditions, and so on. A complete list is available on the website (Statistics Netherlands, 2013b, in Dutch, the English version is not complete), and the figures are available through our Statistical database (StatLine). Most statistics are published annually for the previous calendar year, and often do not appear until a few months into the following year or even years later.
A number of organizations in the Netherlands provide statistics on health and related topics. A well-known institution is the National Institute for Public Health and the Environment (RIVM), whose numbers in this report sometimes are used for comparison.

5. Analyses

With Google Trends interest in a wide variety of health topics can be measured. This section gives merely a few of the possibilities on the basis of several examples. First, the popularity of the category "health" is viewed as a whole. Further, examples are given whether (and how) epidemics such as the outbreak of some infectious diseases are visible in Google Trends. How Google Trends provides insights into Seasonal trends in good intentions in the field of lifestyle is then investigated.
The search behaviour for sexually transmitted diseases is also considered, a subject where we suspect patients generally would prefer the anonymity of an internet search over a face to face discussion and a subject that is unlikely to be accurately answered in a survey.
Finally, an example is given on medical consumerism, namely interest in the alternative care system. These examples are compared where possible with existing data on similar subjects. However, a substantive in-depth analysis of the topics is beyond the scope of this investigation.
In this article, the figures are presented in different ways. Often the trend figure that is directly given by Google Trends is shown, for convenience, but also because for less popular search terms many week numbers are zero (see Section 3.3.3). For displaying seasonal or yearly trends, the numbers had to be downloaded first and edited. Regional differences are displayed through the heat maps as presented by Google Trends.

5.1 Popularity of ‘health’

When no search term is specified but a category is selected, Google Trends shows the relative share of the search behaviour in the selected category in relation to all categories together. Figure 5.1 shows the interest in the category "Health" in the period January 2004 to May 2013. The popularity of health seems to rise slightly (especially) in recent years and there seems to be a seasonal pattern. To gain more insight the (week) figures were downloaded and re-calculated to monthly figures. The 'new' month trend (see figure 5.2) differs slightly from the trend that Google Trends displays in its figure and is somewhat lower in its entirety. The deviation is due to the overlap between weeks and months. Monthly figures are necessarily calculated from the average weekly figure, the weeks are assigned to the month to which the first day of the week belongs. The trend is lower because Google takes the first month to 0 and as a reference point, but in the weekly figures, the first week is taken as the reference point. Figure 5.3 clearly shows a clear seasonal trend. In the summer months and at the end of the year, the interest in health is the lowest. It also appears from figures 5.3 and 5.4 that the interest in health has increased since 2007, with a dip in 2010. This is possibly related to the increased use of the search engine in the elderly (see section 3.4), who may be particularly interested in health.

Figure 5.1 Share of the category ‘health’ in the total number of search terms

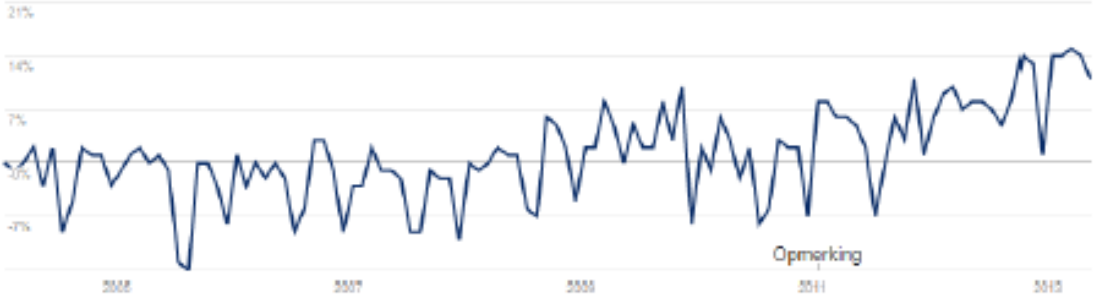


Figure 5.2 Share of the category ‘health’ after editing the downloaded week numbers

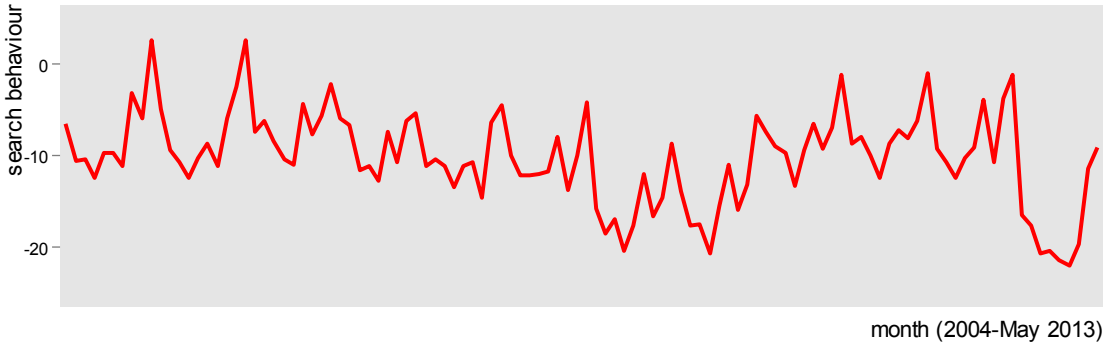


Figure 5.3 Share of the category ‘health’per month

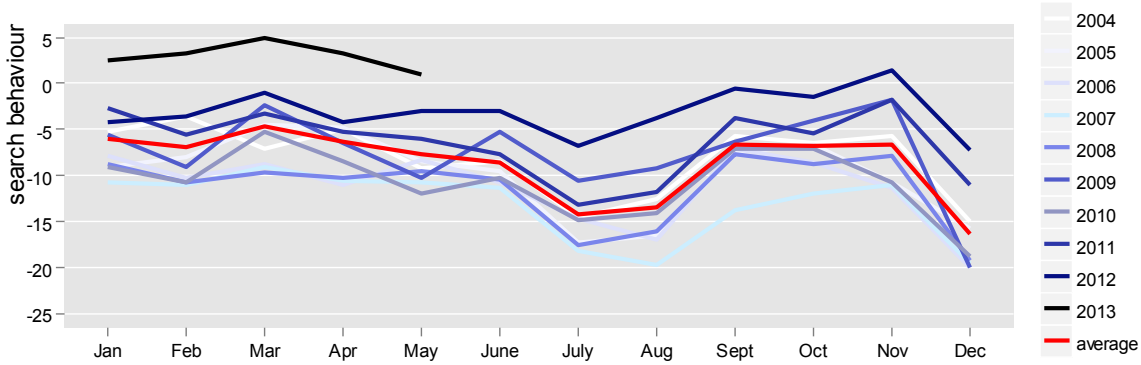
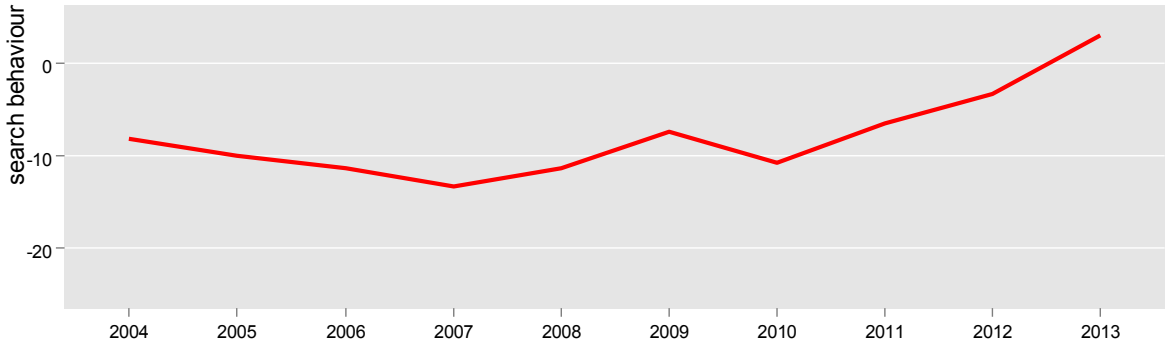


Figure 5.4 Share of the category ‘health’ per year



5.2 Fast and regional data - Epidemics

The correlation between search behaviour and flu epidemics has been described (Ginsberg et al, 2009) and formed the basis for this research. A logical next step is to investigate other epidemics. In this section we have chosen a number of infectious diseases that are included in the National Immunisation Programme and therefore are infrequent, namely measles, mumps and whooping cough. The incidence of these diseases is closely monitored by the RIVM (2013a). This makes it possible to test the Google Trends findings as an ‘incidence measure’ of existing material. Peaks in search behaviour should correspond to peaks in number of reports unless the dark number of (unreported) mild dose infections is high. The reasoning here is to see if Google Trends can be used as a reliable and fast predictor of the outbreak of epidemics for these diseases.

5.2.1 Measles

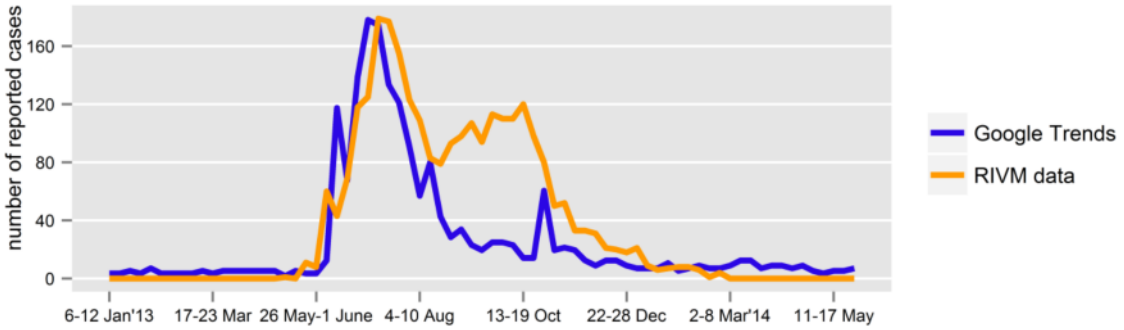
There was an outbreak of the measles in the summer of 2013 in the Dutch 'Bible Belt', an area in which the use of vaccination is lower than average because of the prevailing religious beliefs. The RIVM published the spread over the Netherlands, based on the reports from the municipal health regions. On the 18th of June 2013, the RIVM said on their website *"Given the low vaccination coverage in the Bible Belt, it is likely that the disease will spread among unvaccinated children in this region. From the 1st of May until 13th of June 2013, 64 patients with measles have been reported. The actual number of patients is probably much higher because not all patients go to a doctor. The last measles epidemic in the Netherlands was in 1999/2000."* On July 4, the RIVM wrote *"Since 1st of May 2013 230 cases of measles were reported. Signals from various municipal health regions indicate that this is an underestimate of the actual number of cases."*

The incidence as published by the RIVM is compared with the trend in Google search indices (Figure 1). Google Trends shows an explosion in the term 'measles' from early June 2013, with a peak around mid-July that corresponds well with the first and highest wave of incidence as published by the RIVM. The high incidence in September and October is not reflected in the search behaviour. The search behaviour from 2004 until May 2013 was relatively low, around 2, which corresponds with the absence of measles as the last epidemic was in 1999-2000.

During the heaviest period, the RIVM published the spread of measles very frequent. Figure 5.6 shows four maps that gives the cumulative incidence on 13 June, 27 June and 4 July 2013. These are compared to the relative search behaviour in that period.

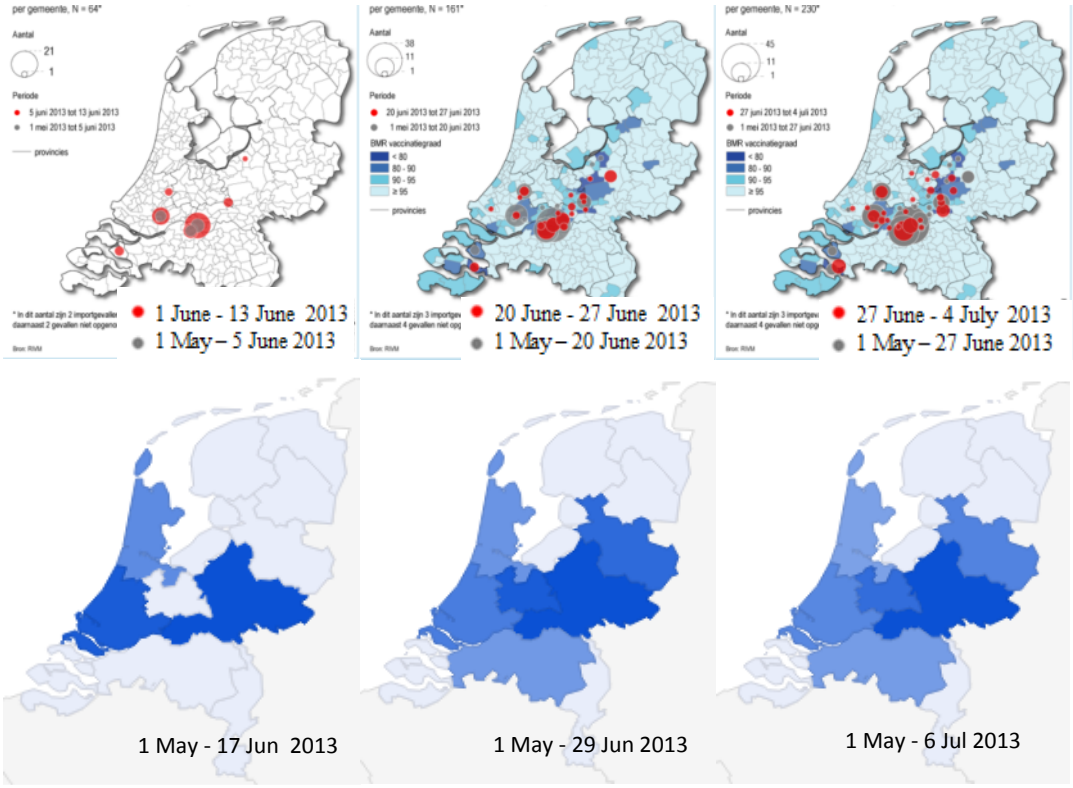
The first heat map (May – 17 June) shows many searches in the central provinces, which correspond to the official figures until the beginning of July. However, at the end of June Google Trends already shows searches in the surrounding provinces while the first official reports from these regions came later on in July. In both figures the epidemic was most serious in the East Central province. Six provinces remained free from measles, which remained the case later in the year. So, although difficult to compare, both sources show a similar distribution in time. Google Trends might even be predictive for the real reports.

Figure 5.5 GT index for 'measles' and reported incidence of measles (RIVM, 2014a).



The GT index is scaled to have its maximum at the same level as the maximum of the RIVM series.

Figure 5.6 Regional distribution 'measles', reports RIVM (top) versus search behaviour (below); May-July 2013



Note 1. A change of colour saturation for a province does not necessarily mean that the search behaviour in that province has changed; the maps only give relative search behaviour within the period that is covered in that map. Note 2. The official figures are published on more detailed regional level than Google Trends (municipalities versus provinces) which makes comparison difficult. Due to the spread over provinces it is not possible to see exactly where people have searched for measles but is clear where no one has searched for this term. Note 3. These Google heat maps cannot be reproduced as exact dates cannot be specified with the period of choice. These heat maps are made on respectively June 18, July 1 and July 7 2013.

5.2.2 Mumps

According to the RIVM there has been an increase in the circulation of the mumps virus in the Netherlands since September 2007. Since January 2009 reporting of cases of mumps has been obligatory in the Netherlands. The actual number of cases of mumps before that time may be higher than published by the RIVM. 359 cases were reported between September 2009 and August 2010, 689 cases between September 2010 and August 2011, and 509 cases between September 2011 and August 2012 (RIVM, 2013).

The search behaviour graph on 'mumps' is quite erratic (figure 5.7) and so the monthly figure is difficult to compare to the seasonal figure of the RIVM. The chart on a seasonal level (figure 5.8) indeed showed an increase in search activity seen from September 2007, which reaches a maximum in 2009-2010. The regional RIVM figures for the 2007-2008 season correspond less well with the local search behaviour in that period than the regional RIVM figures from 2011 to 2012. This could be explained as the report obligation did not exist yet in 2007-2008.

Figure 5.7 Search behaviour for ‘mumps’ (2004-june 2013), monthly figures

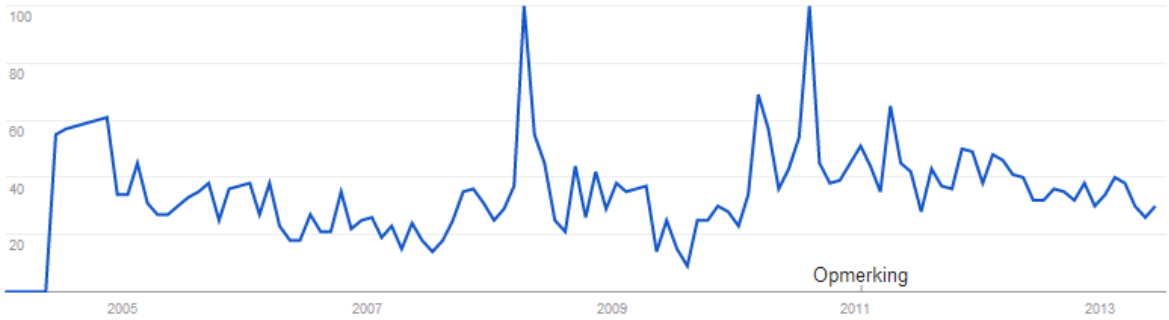


Figure 5.8 Search behaviour for ‘mumps’ (2004-june 2013), seasonal figures

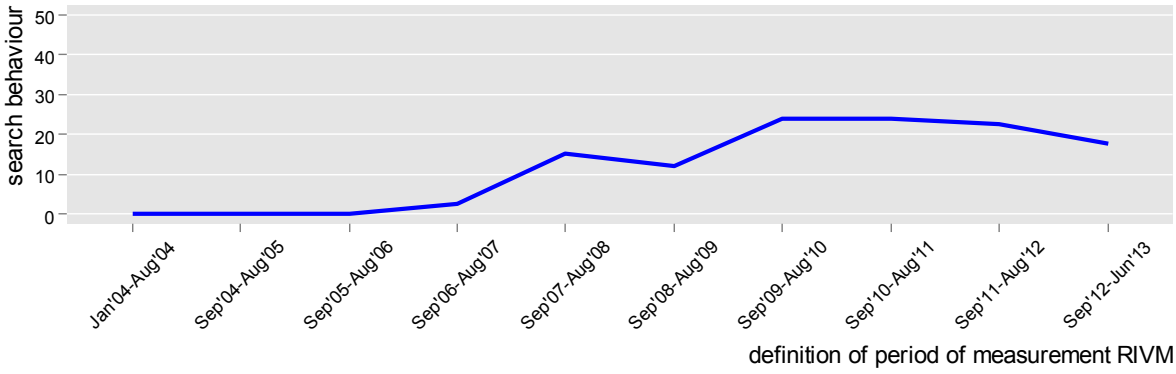
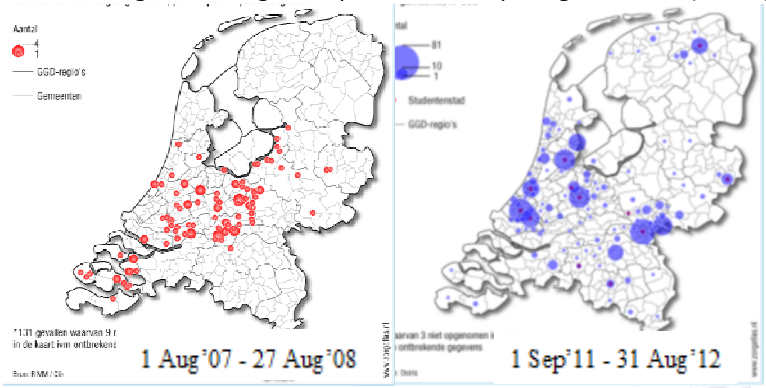
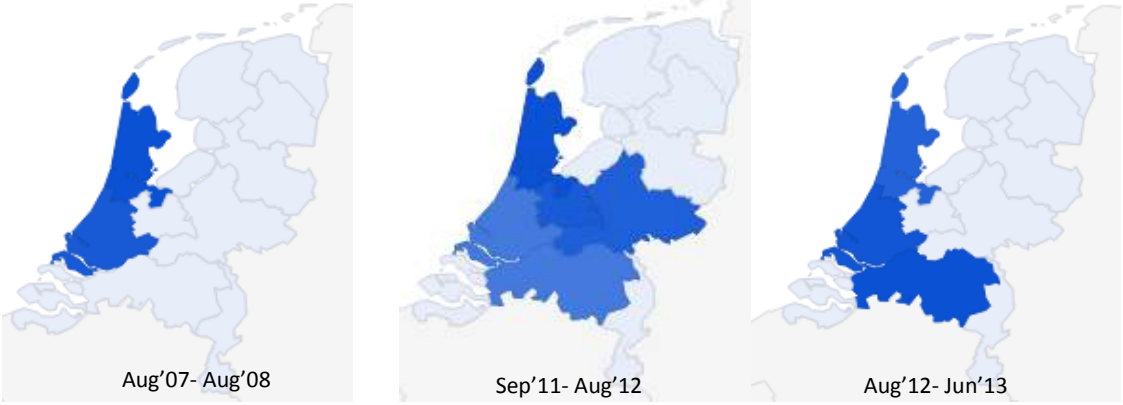


Figure 5.9 Regional spread of ‘mumps’, figures RIVM (above) versus search behaviour (under)





5.2.3 Whooping cough

Google Trends shows increased search activity for whooping cough in 2004 and 2012. In 2004 this was mainly concentrated in one month, namely August, and in 2012 this was relatively spread throughout the whole year. These peak years are also reflected in the figures of the RIVM on the number of reports (figure 5.10)(RIVM, 2012a). According to the RIVM there have been quite a few cases in the period 2007-2009. This cannot really be seen in the search behaviour, in this period the search engine was less used for this disease.
Also the regional variation in search behaviour is very similar to the reports by region according to the RIVM (figure 5.11). Only in province Friesland there was relatively 'often' sought for whooping cough in 2012. Perhaps these were only cases with mild whooping cough symptoms.

Figure 5.9. Search behaviour for 'whooping cough' (2004 – June 2013)

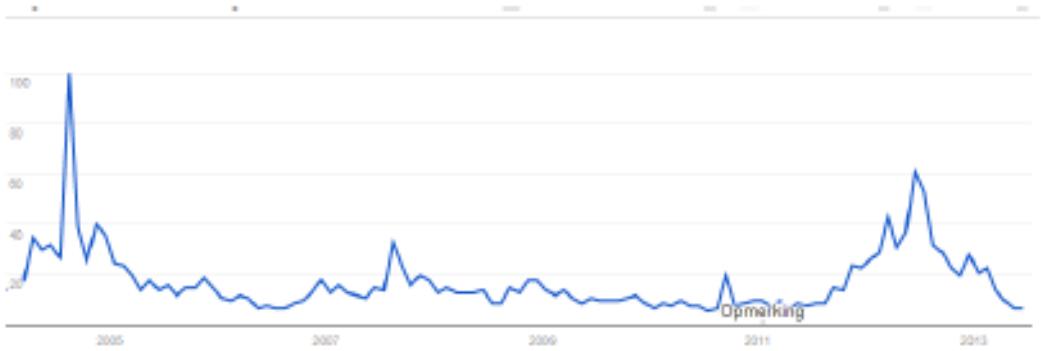


Figure 5.10. Absolute number of whooping cough reports 2004 – Aug 2012, RIVM

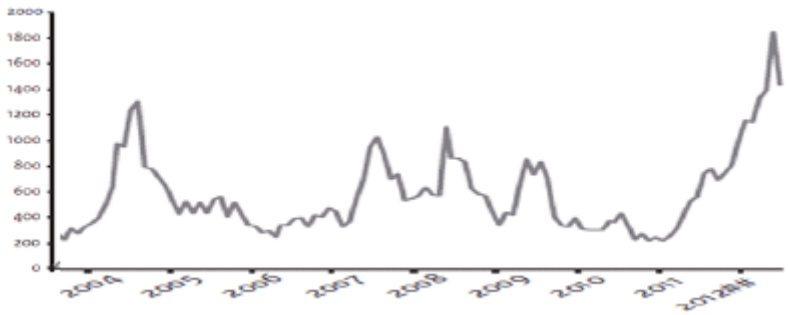
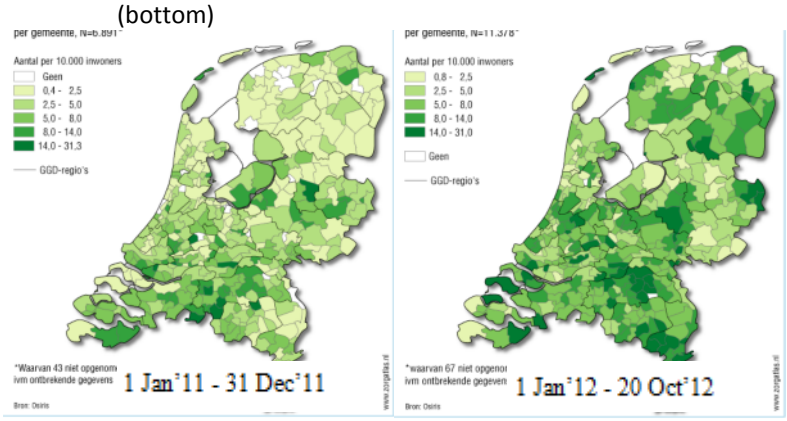
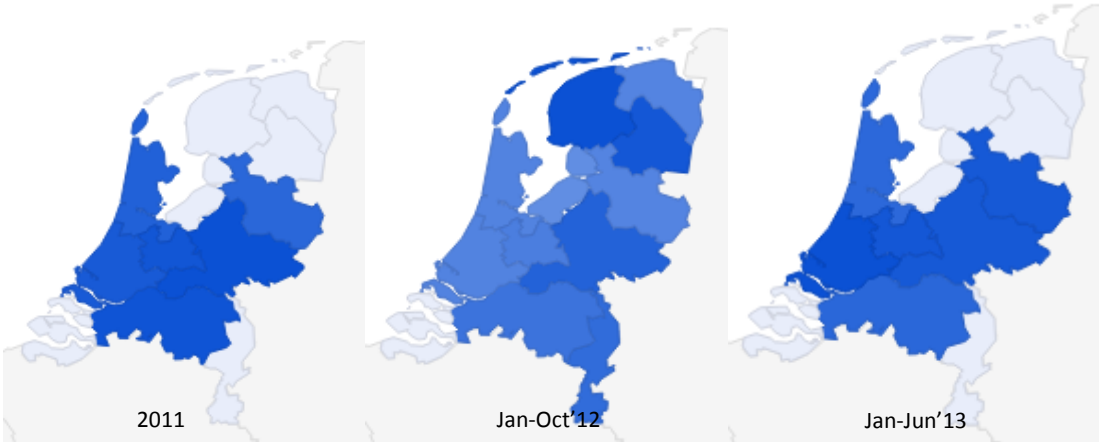


Figure 5.11 Regional spread of ‘whooping cough’, search behaviour (top) versus reports RIVM (bottom)





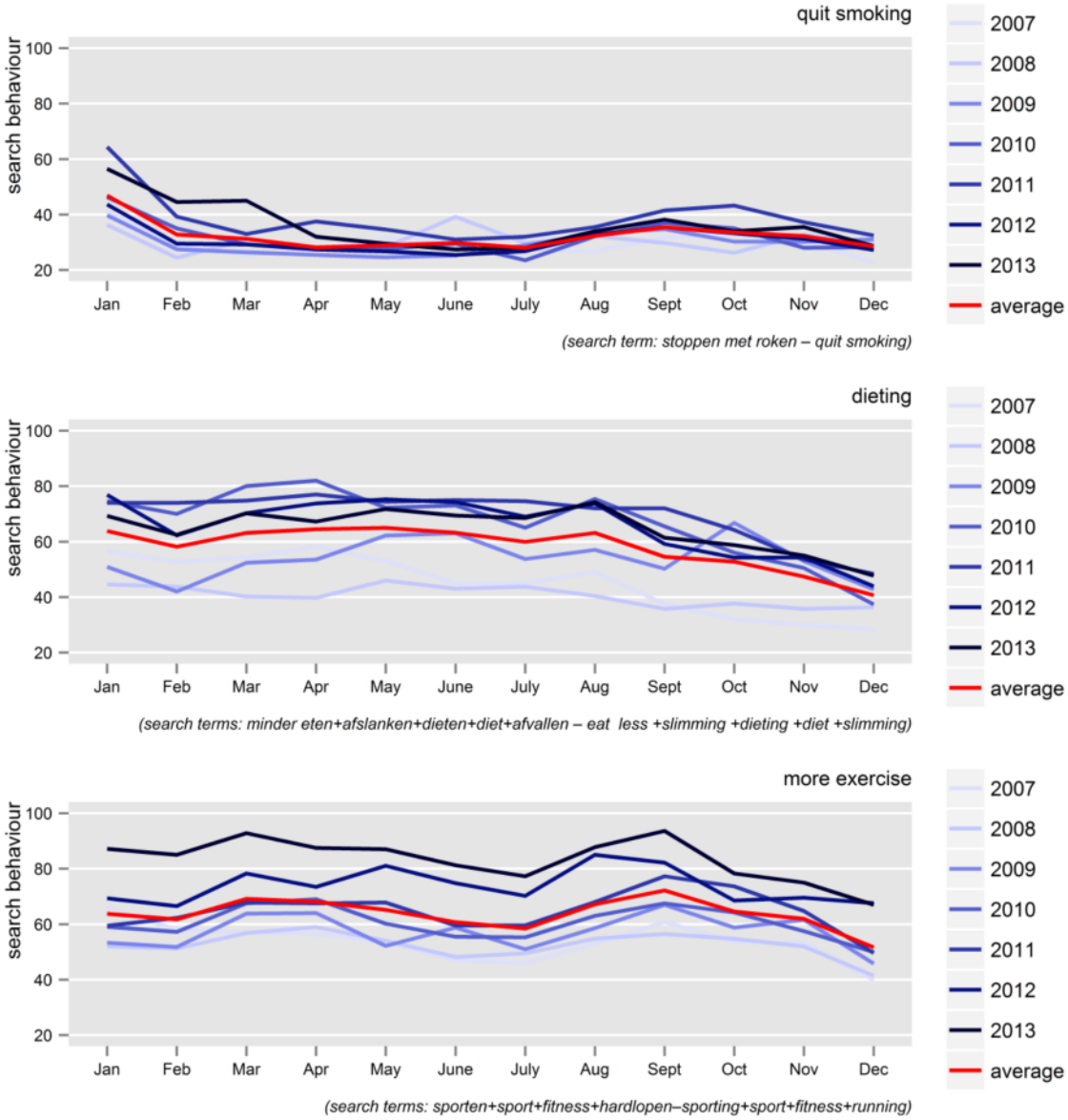
5.3 Seasonal patterns - Lifestyle and preventive behaviour

Most health statistics are published on a yearly basis. However, for some health related topics strong seasonal trends can be expected, for instance the intention to adopt a healthy lifestyle is culturally connected to the beginning of a new year. This section shows the search behaviour for ‘quit smoking’, ‘slimming/dieting’ and ‘more exercise’.

5.3.1 Seasonal patterns

Interest in healthier behaviour shows a clear seasonal trend (figure 5.12). At the end of the year the search behaviour for 'quit smoking' and 'slimming' (dieting) is at rock bottom but then peaks as a new year brings “New Year’s resolutions”. The interest in ‘more exercise’ is more stable during the year, though Google Trends shows a dip in December as well in July, the summer holiday month.

Figure 5.12 Seasonal trends in quitting smoking, slimming/dieting, exercising



These graphs are based on data downloaded from Google Trends. Month figures resemble the averaged weekly data in which the first day of the week determines the month. As the differences between healthy behaviour between December and January is very large, the differences between these months are weakened because the last week of the year is spread over those months. The monthly figures for December and January were therefore recalculated by calibrating the real proportions, read from the trend as displayed by Google, to the February months.

Also another method is investigated to reproduce monthly figures, namely downloading and averaging of the daily figures. If a short period of (less than) 3 months is chosen Google Trends shows the day digits (also in the download file). This is only possible from 2007 onwards. This would thus theoretically make it possible to calculate the month ratios. However, this proved not to be a useful method because many days have such a relatively low search volume that they had been rounded to 0.

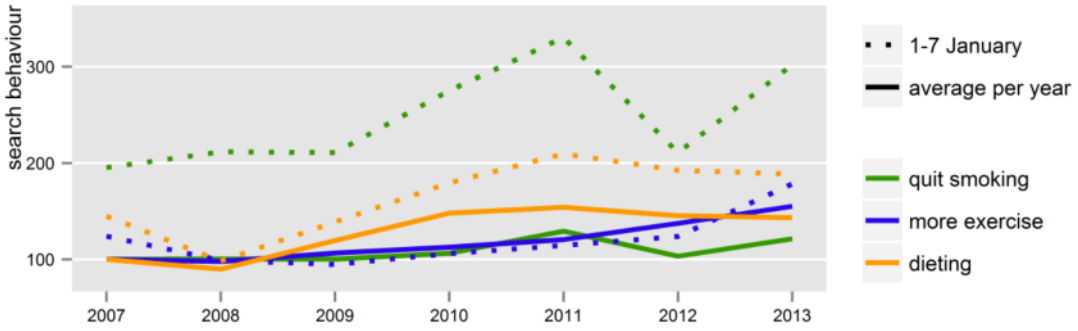
5.3.2 Year trends

The theme page 'Health and Wellbeing' on the website of Statistics Netherlands opens amongst others with a graph that shows the evolution of the percentage of smokers in the Netherlands. An additional trend line that displays the interest in 'quit smoking' might be interesting (Figure 5.13).

Figure 5.13. Trends in smoking behaviour, regular publication Statistics Netherlands (left) and expansion with the trend in search behaviour for 'stop smoking' (right)

Figure 5.14 shows the trends in lifestyle from 2007, in which the first week in January is shown separately. This shows that 'more exercise' slowly but surely gains interest, just as slimming. The desire to quit smoking seems fairly stable, with the exception of the peak in 2011. The large difference in search behaviour between the first week of January and the rest of the year is striking, especially for 'quit smoking'.

Figure 5.14 Trends in search behaviour for healthy life style, per year and per week of the good intentions

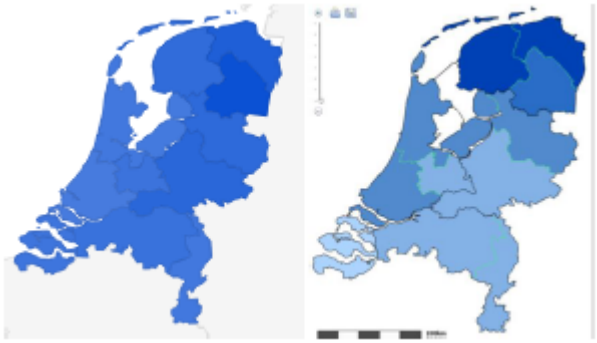


In figure 5.14, the lifestyle aspects are not shown in relationship to one another because this depends too much on the number of search terms per aspect. Therefore the trend for each aspect is calibrated on search behaviour in January 2007.
For determination of the search behaviour in the first weeks of January, the day digits for all January months (for all years separately) have been downloaded. In these January months there were no days with too little search volume and therefore for each year the average number for January 1 to 7 and the average for the entire month of January could be calculated. By applying the proportions of the week versus month numbers on the previously calculated January number, the search behaviour in the first week of January is derived.

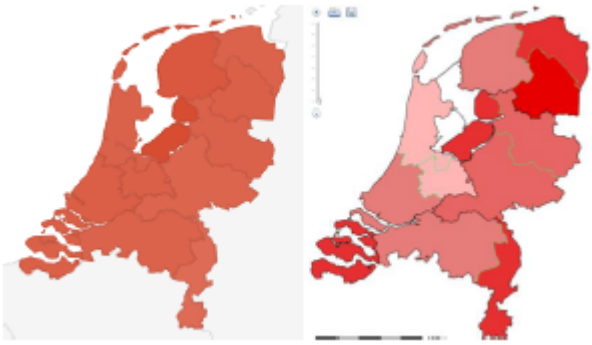
5.3.3 Regional differences

Statistics Netherlands publishes regional differences in lifestyle once every five years. The latest figures from the Health Survey cover the period 2008-2011 (Statistics Netherlands, 2013c). It may be interesting to compare the lifestyle with the desire to improve it on the level of the provinces. In figures 5.15 and 5.16 the search behaviour is compared to the official figures on actual lifestyle, namely the percentage of the Dutch that says it 'sometimes' smokes, the percentage that meets the standard physical activity advice and the percentage that is overweight. Based on the heat maps from Google Trends (figure 5.15) there seems little regional variation in search behaviour for healthier living. However, the magnitude of the difference is similar to the figures of the Health Survey (figure 5.16). The correlation between actual lifestyle and intention to healthy behaviour is not large, and does not have to be there. In provinces where lots of people smoke, a little more often is searched for quitting. Interest in 'slimming' results in less obesity, inhabitants of the middle provinces (Gelderland and Zuid Holland) do not seem to succeed in their intention. And it seems to be that interest in 'more exercise' also results in more exercise, except for in the northern provinces Flevoland and Groningen.

Figure 5.15. Regional spread lifestyle 2008-2011, search behaviour versus Health Survey
Search behaviour for quit smoking (Google Trends) - % that smokes (Health Survey)



Search behaviour for dieting/slimming (Google Trends) - % that is overweight (Health Survey)



Search behaviour for more exercise (Google Trends) - % that meets the recommended standard of exercising (Health Survey)

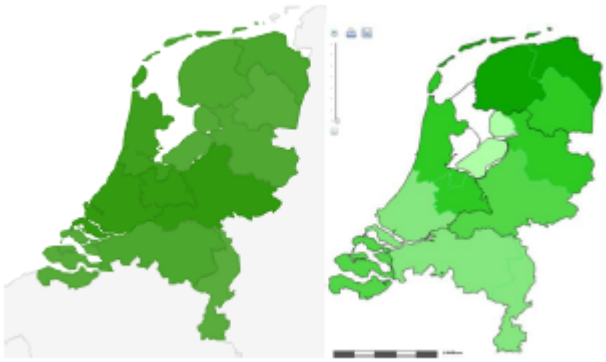
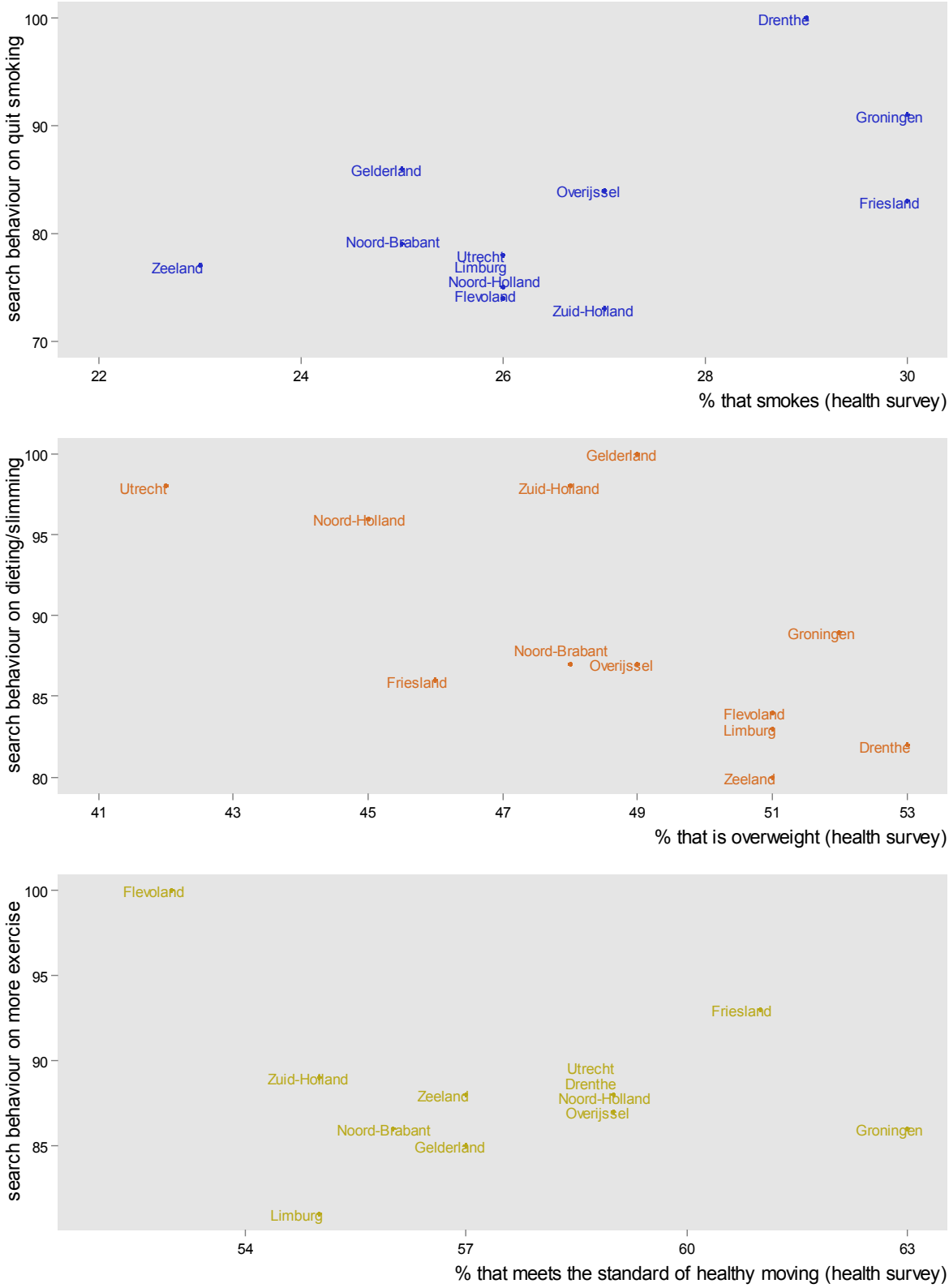


Figure 5.16 Lifestyle per province, search behaviour versus Health Survey



5.4 Sensitive topics - Sexually transmitted diseases

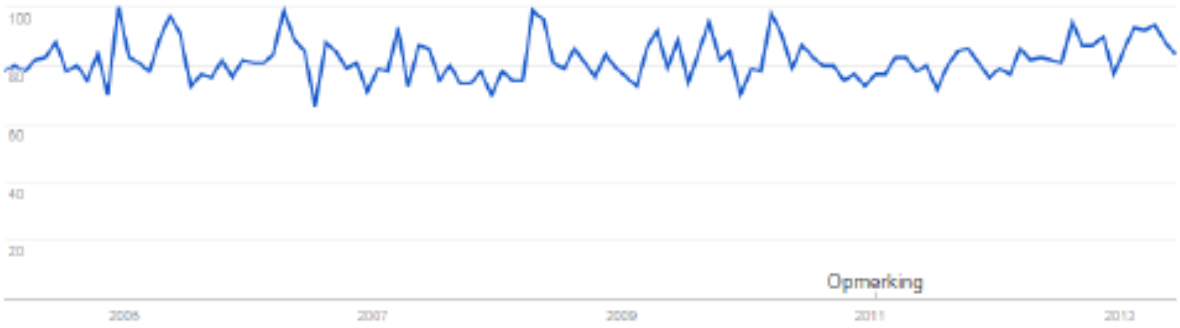
For sensitive or embarrassing diseases the internet will often be a preferred source of information. Sexually transmitted diseases (STDs) are an example of this. In the Netherlands no clear figures are available on the prevalence of STDs (Bruin et al., 2011). The majority of STDs is diagnosed by general practitioners and STD centres, and a significant part is never detected. The Health questionnaire does not ask about such sensitive topics as STD. Only the number of hospitalisations due to STDs is published on the statistical database (Statistics Netherlands, 2013d).

The transfer of STDs is possibly season related. However, this seasonality may in incidence only appear in STDs with a short incubation period and (obvious) symptoms. In this section, the search behaviour of the STDs in total is described and three bacterial STDs are investigated, namely chlamydia, gonorrhoea and syphilis. Regional trends in search behaviour are also presented.

5.4.1 STDs total trend

Figure 5.17 shows the trend in search behaviour of all STD terms that will be discussed in more detail below. Because of the sometimes difficult and unfamiliar concepts here many spelling variations are included in the query analysis. The term AIDS is not included because two extreme peaks in 2004 and 2006 causes the other periods to drop to a relative score of 0. The search behaviour for STDs shows an erratic trend, but seems fairly constant on the whole. There is no clear seasonal trend to be discerned.

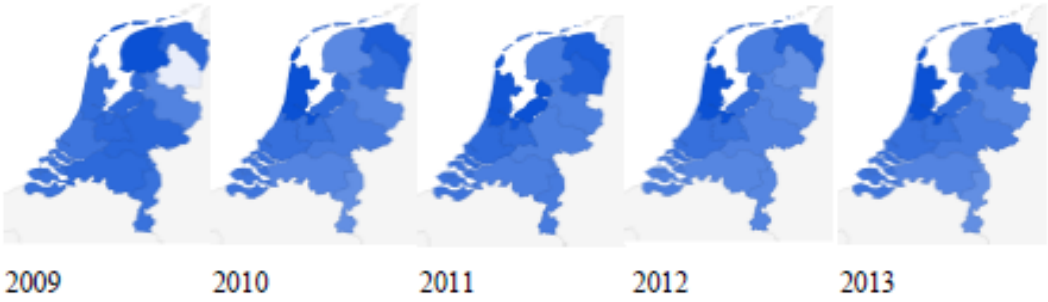
Figure 5.17. Trend in search behaviour of all STDs together, 2004 - June 2013



Search terms:
SOA+geslachtsziekte+gonorroe+gonoroe+druiper+chlamydia+chlamidya+clamidia+clamydia + chlamidia+ chlamydiasis+syfilis+syfillis+syphilis+syphillis+lues+trichomonas+ genitale wratten+ hepatitis B+ hepatitis C+Primaire herpes genitalis+LGV

Relatively most is searched for STDs in the provinces Groningen and Noord-Holland (figure 5.18). Amongst the other provinces there is little difference. It is striking that in 2009 (and 2012) little is sought for STDs in Drenthe, while this province scores relatively high in the other years.

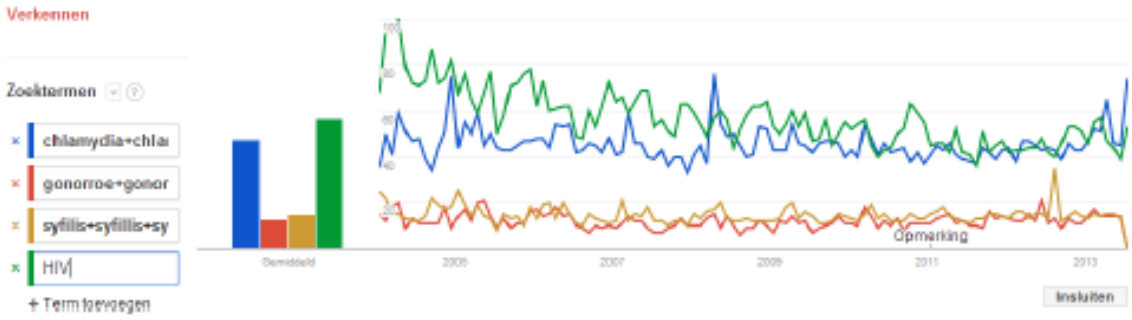
Figure 5.18. Regional trend in search behaviour for all STDs together



A change of colour saturation for a province does not necessarily mean that the search behaviour in that province has changed (see section 3.2)

Search behaviour does not reflect the true incidence ratios of the different kinds of STD. In figure 5.19 the three bacterial STDs are compared with 'HIV', a dreaded but rare viral STD. The most frequently searched term is 'HIV', though there is a downward trend. HIV infections are less common than bacterial STDs. Although good prevalence figures do not exist, it is known that Chlamydia is much more common than Gonorrhea and Syphilis. This indeed appears from the search behaviour. But Gonorrhea on the other hand is more common than Syphilis, and this is not reflected in the search behaviour (unless a major spelling variant is missed here, the terms used in these analyses for each STD are given in next paragraphs).

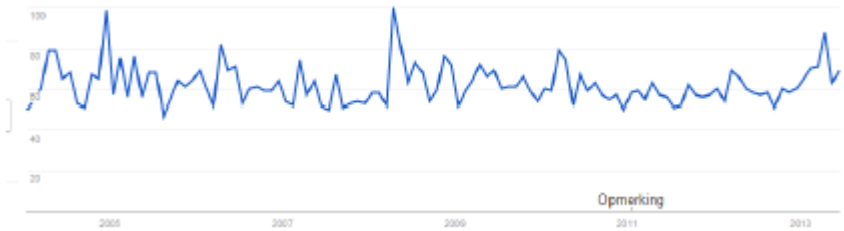
Figure 5.19. Trends in search behaviour for a selection of STDs



5.4.2 Chlamydia

Chlamydia is the most common bacterial STD and occurs mostly in heterosexual women and men. Based on a nationwide screening study completed in 2004, the annual number of new infections is estimated at 60.000 (SOA AIDS Netherlands, 2013), the actual number of people infected is much higher. Chlamydia is usually asymptomatic and is, as a result, spread relatively quickly. The search behaviour for Chlamydia therefore is fairly constant (figure 5.20), except for a few peak months. Probably these peaks reflect media attention. For example, in April 2008 the first large population screening for chlamydia amongst young people has started (NOS, 2008).

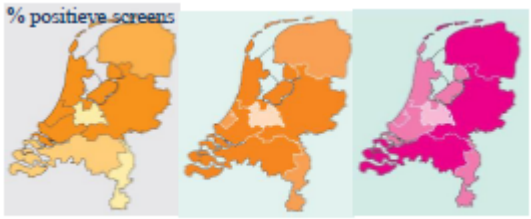
Figure 5.20. Trend in search behaviour for Chlamydia



Search terms: chlamydia+chlamidya+clamidia+clamydia+chlamidia+chlamydiasis

Figure 5-21 shows the regional differences in search behaviour for chlamydia. These are presented alongside the regional distribution of the percentage of positive screenings at the 8 STD centres (RIVM, 2010, 2011, 2012b). These are the only regionally available figures for STDs. Even if there would be more research and coherence between the different surveillance systems, it would be difficult to make regional figures because the screening proceeds anonymously.
The differences between the province's change every year. In 2009 and 2010 Chlamydia was hardly sought in the entire North-Eastern part of the Netherlands, in 2011-2013 there is no difference with most other provinces. The regional differences in search behaviour do not correspond to the differences between the screening results in the 8 STD regions. However, the screening heat maps do not reflect the difference in the incidence. To obtain this the figures have to be corrected for number of people that share the risk groups and the attendance rate at screening. On top of that, most STDs are diagnosed by the GPs.

Figure 5.21. Regional trend chlamydia, screen results (RIVM, above) and search behaviour (below)

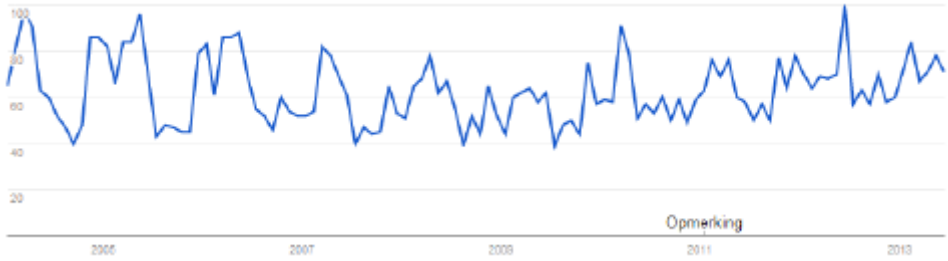


These heat maps (above) do not reflect the actual difference in incidence, but merely the results of the screening (see section 5.4.2).

5.4.3 Gonorrhea

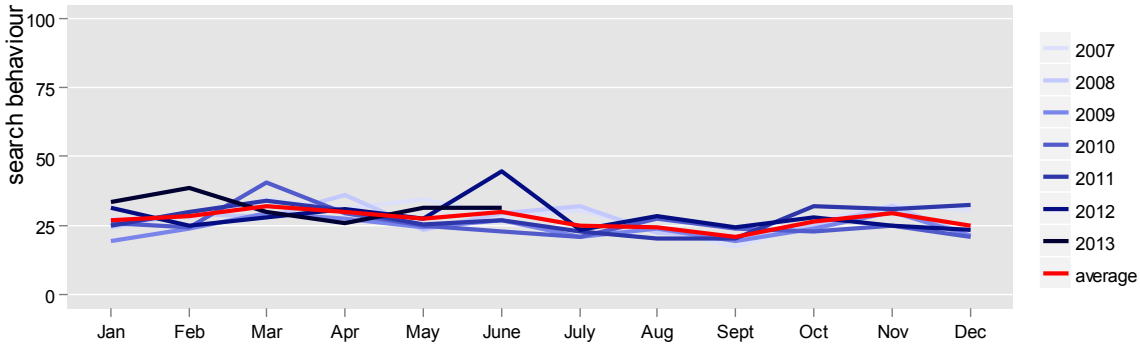
Gonorrhea is also a relatively common STD, with an estimated 5 000 new cases per year (SOA AIDS Netherlands, 2013). Gonorrhea is most common in homosexual men. It has a short incubation period of 2-5 days with clear symptoms. The search behaviour for Gonorrhea should therefore show a seasonal effect if this STD is mainly transmitted in certain periods. The search activity does indeed fluctuate slightly stronger than Chlamydia (figure 5.22). The peaks are caused by media attention (for example, according to Google Trends there was news that an untreatable gonorrhea has surfaced in June 2012). According to the (RIVM, 2013b), the number of diagnoses by general practitioners has decreased slightly since 2007 and increased slightly from 2011. Search behaviour (figure 5.22) also points in that direction. From a display at monthly level (figure 5.23) no clear seasonal trend is shown. In this figure the trend as a whole is much lower than in 5.22, the direct Google presentation. This is because the monthly averages are calculated on weekly figures, in which the first week of June 2012 peaks extremely (see 3.3.3).

Figure 5.22. Trend in search behaviour of Gonorrhea



Search terms: gonorroe+gonoroe+druiper

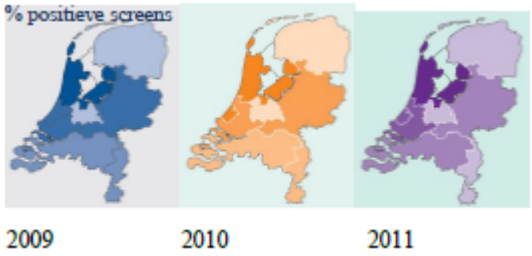
Figure 5.23. Monthly trend in search behaviour of gonorrhoea



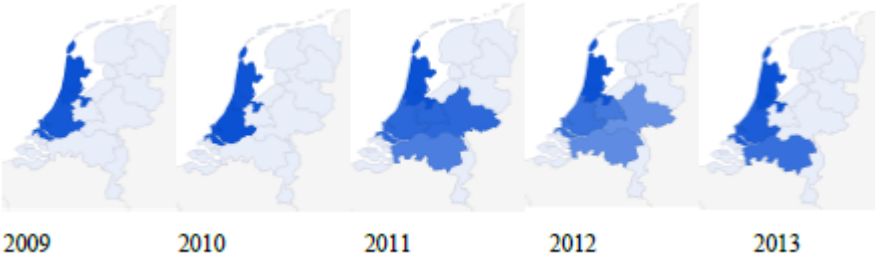
The maximum search behaviour is not 100 because it is based on weekly figures (see 3.3.3).

The regional differences in the incidence of Gonorrhea are quite large and fairly consistent, both regarding search behaviour as the percentage of positive screenings (RIVM, 2010, 2011, 2012b) (figure 5.24). In Noord-, and Zuid Holland in all years relatively much is sought. The search term Gonorrhea has also been used in Brabant, Gelderland and Utrecht since 2011. This might be related to the regional spread or distribution of gay nightlife. We have no figures on this subject.

Figure 5.24. Regional trend gonorrhea, screen results (RIVM, above) versus search behaviour (below)



These heat maps do not reflect the actual difference in incidence, but merely the results of the screening (see section 5.4.2).



5.4.4 Syphilis

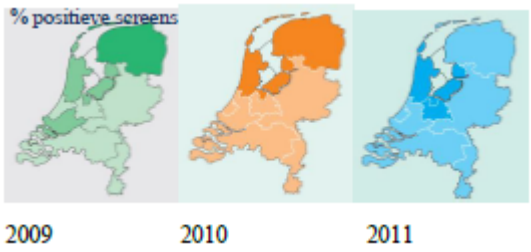
Syphilis is a serious STD if it is discovered too late. It is quite rare, with an estimated 500 new cases per year, the most among homosexual men. The symptoms of the disease occur only months after infection. The search behaviour for syphilis (figure 5.25) is fairly constant except for a peak in August 2012. Google Trends does not show a news item for this peak, and we have been unable to find any explanation for this peak in search behaviour.

Figure 5.25. Trend in search behaviour for syphilis

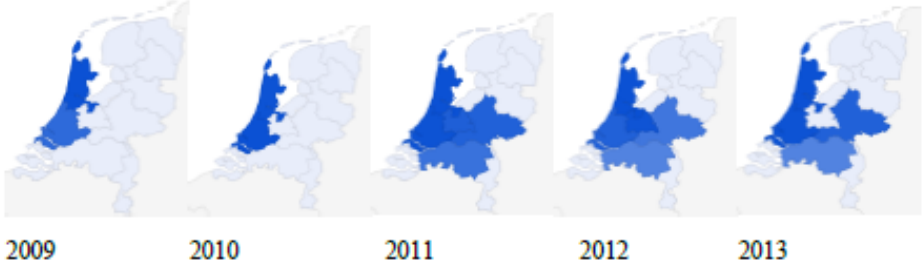
Search terms: syfilis+syfillis+syphilis+syphillis+lues

The trend in regional variation in search behaviour resembles that of Gonorrhea (figure 5.26). In 2009 and 2010 search interest was mainly concentrated in Noord Holland, and Zuid Holland, later spread to other provinces in the middle of the country. Perhaps this has to do with the fact that both Syphilis and Gonorrhea occur primarily among homosexual men.

Figure 5.26. Regional trend syphilis screen results (RIVM, above) versus search behaviour (below)



These heat maps do not reflect the actual difference in incidence, but merely the results of the screening (see Section 5.4.2).



5.5 Trendbreak detection and in-depth analyses - Medical consumption

This section discusses interest in health care. For common mainstream types of health care providers (such as the general practitioner or the physiotherapist) the internet may not be consulted for information on the care itself, but rather for contact information. In the lesser known types of care providers, search behaviour will be a combination of information about the care itself as well as contact information.

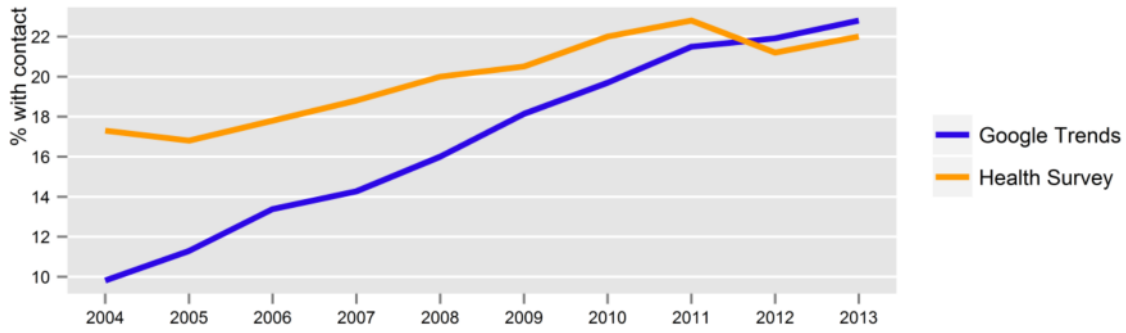
In this section, an example of search behaviour for well-known care is given, namely physiotherapy. In addition, search behaviour for alternative medicine is examined. Statistics Netherlands publishes (trend) data on medical contacts from 1981 (Swinkels, 2011; Statistics Netherlands, 2014b). Here contact with physiotherapists and alternative doctors is included but does not distinguish between different types. This section does not address the relative proportion in search behaviour in physiotherapy and alternative therapy as, with alternative therapy, information about the medicine itself will also be searched for, while a search for for physiotherapy often will only be for contact information.

5.5.1 Physiotherapy

Since 2004 the internet has increasingly been consulted for physiotherapy (figure 5.27). According to Statistics Netherlands, the proportion of Dutch people that have had contact with a physiotherapist has also increased since 2004. Despite the fact that the increase in search behaviour is stronger than the official figures on increase in actual 'contact', Google Trends indicates that the popularity of physical therapy also increased after 2011. This corresponds to the increase in interest in 'moving' (see section 5.3.2), because of which the number of injuries might have risen.

As a side note it is important to realize that the Health Survey results become available only well into the following year, a significant delay compared to the Google Trends figures which are immediately available.

Figure 5.27. Trend physiotherapy 2004-present, contact with physiotherapist (Health Survey) versus search behaviour

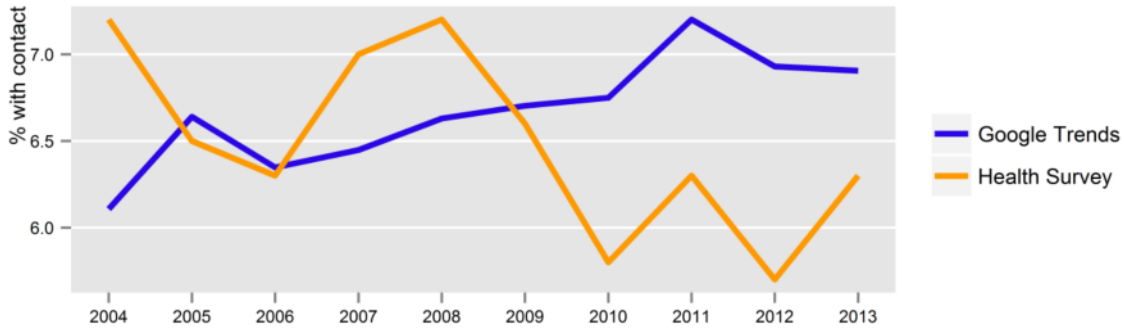


Search terms: fysio+fysiotherapeut+fysiotherapie
The Google Trends index is scaled to have its maximum at the same level as the maximum of the Health Survey series.

5.5.2 Alternative medicine

Since 2004, the Dutch increasingly search for terms that relate to alternative medicine (figure 5.28). In the survey based data from Statistics Netherlands, the trend in contact with practitioners of alternative medicine appears to be erratic. This is partly caused by the redesign of the Health Survey in 2010 in which the questions about contact with practitioners of alternative medicine have been changed (Hupkens, 2011) – specifically to measure the 'new' upcoming therapies. The trend in the Google numbers might therefore give a better picture over this transition period and could even fill in the gaps before the change in the Health Survey on this topic.

Figure 5.28. Trends in alternative medicine 2004-present according to the Health Survey and to GT.

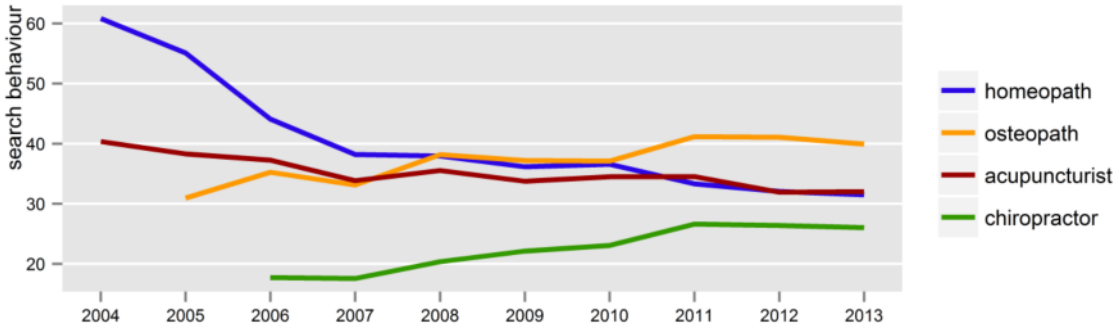


Search Terms: ostheopaat+osteopaat+ostheopathie+osteopathie+Chiropractor+giropractor+ chiropractie+acupuncturist+acupunctuur+Haptonoom+haptonomie+Homeopaat+homeopatie+ magnetiseur+magnesticisme+Antroposoof+antroposofie+Fytotherapeut+Kruidengenezer+ natuurgenezer+Paranormaal genezer+Gebedsgenezer+Religieus genezer
The GT index is scaled to have its maximum at the same level as the maximum of the Health Survey series.

Figure 5.29 and 5.30 show the GT indices for four types of alternative medicine. Interest in the different types is undergoing a pronounced change. In 2004 the Dutch turned for alternative medicine primarily to homeopathy and acupuncture. Especially homeopathy has dropped significantly in popularity since making way for osteopathy and chiropractic. Few official figures are known concerning contact with the various alternative practices. According to the Health Survey in 2002 the homeopaths and acupuncturists were predominantly visited (Statistics Netherlands, 2003; Van Dijk, 2006) – but contact with osteopaths and chiropractors was not offered as possible responses in the questionnaire until 2010 (Statistics Netherlands, 2013a). Recent estimates still show that homeopathy and acupuncture are more popular than chiropractic and osteopathy (Statistics Netherlands, 2014c) .
Figure 5.31 compares the ratio of contact with the different alternative medical disciplines in 2010 to 2012 according to the Health Survey with the ratio in search behaviour in the same years. Striking is the large difference in osteopathy, which is most searched for on the internet, but least used according to the Health Survey. A tentative explanation is the design of the Health Survey questionnaire. The questions about use of alternative medicine start with a filter question to determine whether or not the respondent has made use of alternative medicine. The individual disciplines are presented to the respondent in a follow-up question. In that filter question only the examples homeopathy, acupuncture and chiropractic are mentioned, but not osteopathy (Hupkens, 2011). In the subsequent question, all four are mentioned among others. This analysis suggests a potential bias in the current measurements of osteopath visits as measured by the Health Survey. It may be advisable to list osteopathy as a discipline in the filter question.

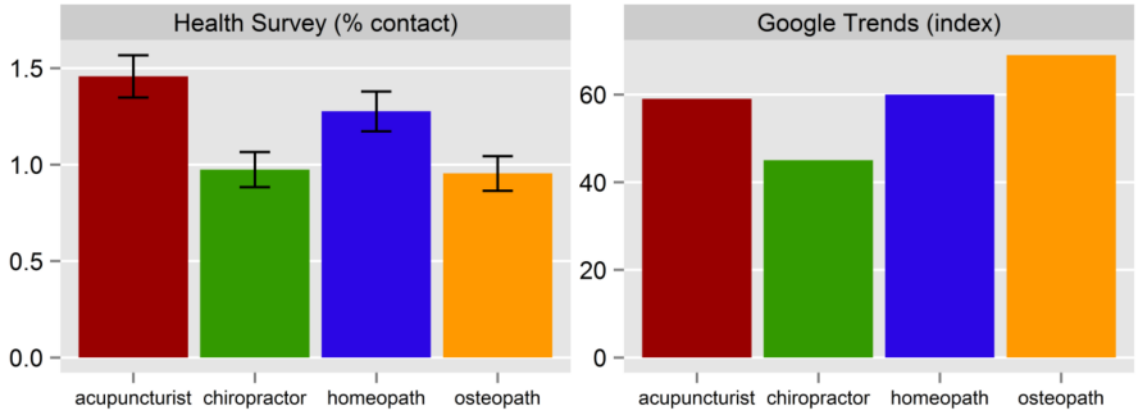
Figure 5.29. Search behaviour for popular kinds of alternative medicine, per month

Figure 5.30. Search behaviour for popular kinds of alternative medicine, per year



Search terms: 'homeopaat+homeopathie+homeopatie', 'ostheopaat+osteopaat+ostheopathie+osteopathie+ osteopatie', 'acupuncturist+acupunctuur+accupunctuur', 'chiropractor+giropractor+chiropractie+ chiropraxie'
The maximum search behaviour is not 100 because it is based on weekly figures (see 3.3.3). Annual averages are not shown if weeks occur where the search behaviour is 0.

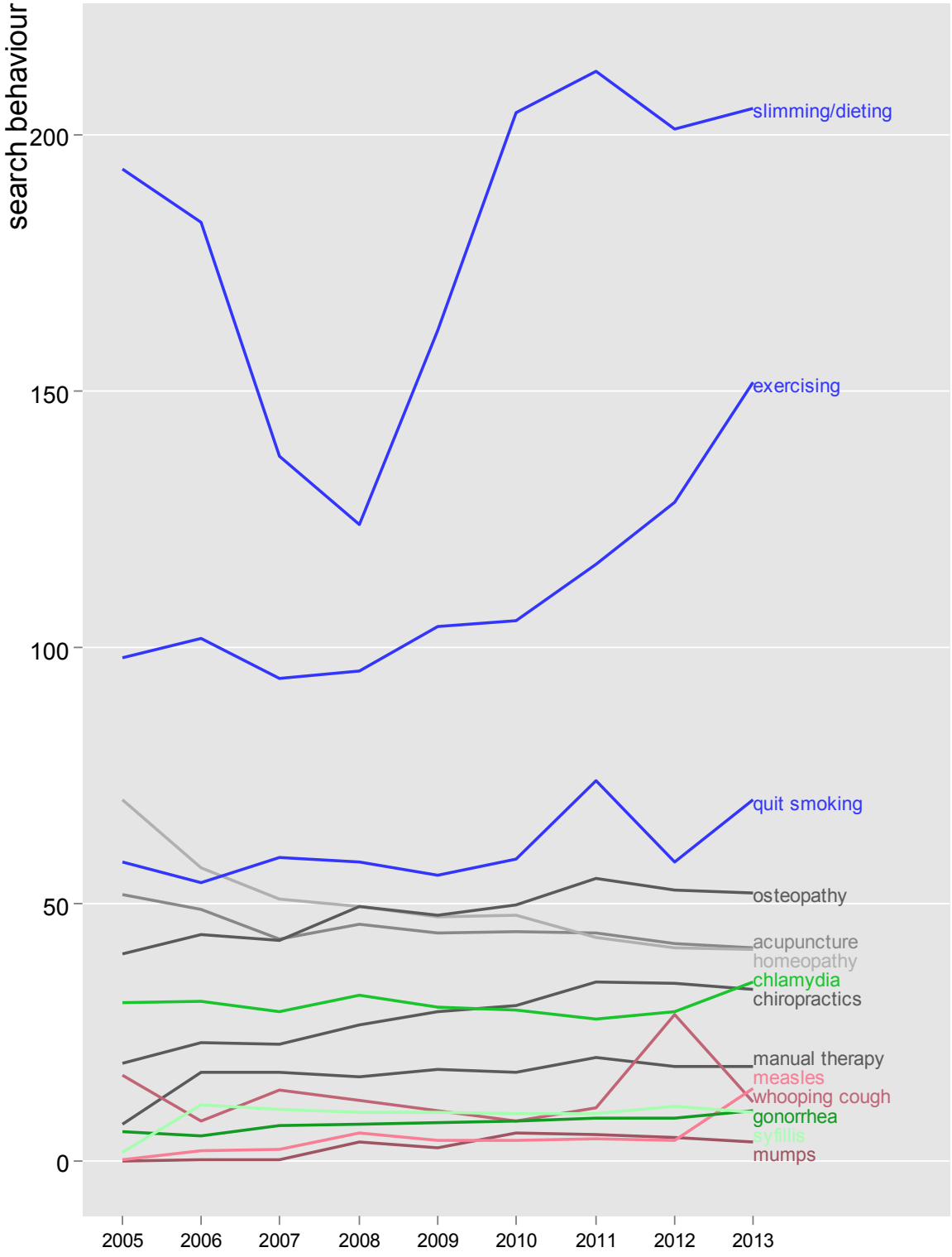
Figure 5.31. Contact with alternative medicine in the past 12 months (Health survey 2010 to 2012), and search behaviour for alternative medicine (2010-2012)



5.6 Examples in perspective

In sections 5.2 to 5.5 a number of examples are given of the use of Google Trends within a health theme. As indicated earlier, it is hard to interpret the relative shares of search behaviour on different topics. It depends very much on the choice of (and the number of) terms by subject. In order to get a general idea of the order of magnitude of the difference in search behaviour between the topics that are discussed in this report they are summarized in a single figure (figure 5.32). This shows that the interest in "healthy behaviour" is the greatest, apart from the search behaviour for the physiotherapy (that rises above this figure). Alternative medicine is also relatively popular. The lowest numbers of searches in Google was for infectious diseases such as STD and the diseases discussed that are included in the national vaccination program.

Figure 5.32. All subjects in perspective



(In this figure all themes are calibrated on the trend line of 'homeopathy', that is why two life styles rise above 100.)

6. Conclusion

Google Trends has some advantages over survey based statistics: the topics are almost unlimited in scope, the data are free of charge, up-to-date and quick to obtain. Google Trends provides seasonal patterns and more temporal and regional detail than many traditional surveys. While survey statistics and internet search indices provide information on the same themes, they are different quantities that cannot be used interchangeably. Rather, they are complementary, with search query data providing a way to enrich existing, classical official statistics. Google Trends cannot give absolute numbers nor does it provide background information about the subpopulation using the Google search engine. In addition Google has full control over the Google Trends methodology which is not fully documented.

This report presents research into the use of Google Trends in relation to health statistics. The results show that Google Trends could be used to monitor the outbreak of an epidemic such as measles, mumps and whooping cough. Google Trends provides data about interest in sensitive topics about which little or no information exists. Often these topics, such as STDs, are not included in health questionnaires because social taboo would cause the information gathered from respondents to be questionable. Google Trends reflects seasonality in interest in a healthier lifestyle. Google Trends can offer extra detail to existing figures that are only published at aggregated level because of small sample sizes. An example is the rise in popularity of different alternative medical treatments. As new or upcoming trends will only get added to traditional surveys after they are fairly well established, Google Trends allows you to extrapolate back to before a survey has caught up with a new topic. An example is the use of osteopathy and chiropractic. In all these cases Google Trends indices provide valuable additional information complementing official statistics on the same themes.

In addition to enriching existing statistics, Google Trends indices could fulfil a more prominent role in official statistics in future. Two possible uses are suggested: using Google Trends during the design of survey questionnaires to avoid subsequent measurement errors and quantifying discontinuities in survey based time series using a correlating Google Trends series. Another possibility is using search behaviour in small area estimation models to improve spatial and temporal detail of survey based official statistics (Bruggink et al., 2013). In these methods, it is important to have correlating auxiliary information from another source than the survey.

It is easy to see how the analyses from this survey can be extended within the area of health, for example on allergies, injuries, and certain medications. Google Trends analyses can also add value to other themes, such as housing, employment, voting behaviour, safety, travelling and leisure.

References

Ayers JW, Althouse BM, Allem JP, et al. (2013). Seasonality in seeking mental health information on Google. American Journal of Preventive Medicine 44: 520-525.

Ayers JW, Ribisl K, Brownstein JS (2011). Using search query surveillance to monitor tax avoidance and smoking cessation following the United States' 2009 “SCHIP” cigarette tax increase. PLoS ONE 6(3): e16777.

Ayoubkhani, D (2012). An investigation into using Google Trends as an administrative data source in ONS. Seminar on New Frontiers for Statistical Data Collection, UNECE Conference of European Statisticians, Geneva.

Bruggink, J.W., Smeets, M., Tennekes, M. (2013). Kleinedomeinschatters bij de gezondheidsenquete. Statistics Netherlands nota, BPA-nr. PPM-2013-01-29-JBRK, Statistics Netherlands Heerlen.

Bruin A, Voorrips L, Ploemacher J, et al (2011). Netherlands Pilot Project on Morbidity Statistics (Eurostat Multi-Beneficiaries Grant 2009, Theme 1.05 - Health and Safety). Statistics Netherlands (CBS), co-ordinator, and National Institute for Public Health and the Environment (RIVM), co-partner. The Hague/Bilthoven, The Netherlands.

Buelens B, Daas P, Burger J, Puts M, Van den Brakel J (2014). Selectivity of big data. Discussion Paper 2014-11, Statistics Netherlands, The Hague/Heerlen, Netherlands.

Carr LJ, Dunsiger SI (2012). Search query data to monitor interest in behavior change: application for public health. PLoS ONE 7(10): e48158.

Choi H, Varian H (2011). Predicting the present with Google Trends. Technical Report available at: http://people.ischool.berkeley.edu/~hal/Papers/2011/ptp.pdf (accessed 3 September, 2014).

Daas, P., van der Loo, M. (2013). Big Data (and official statistics). Paper for the 2013 Meeting on the Management of Statistical Information Systems (MSIS 2013), Paris - Bangkok, France - Thailand.

Dijk, P van. (2006). Omvang alternatieve geneeswijzen in Nederland 21-22 (1): p. 18 (Stichting Tijdschrift voor Integrale Geneeskunde).

Eysenbach, G. (2009). Infodemiology and infoveillance: Framework for an emerging set of public health informatics methods to analyze search, communication and publication behavior on the Internet. Journal of Medical Internet Research, 11(1).

Ginsberg J, Mohebbi MH, Patel RS, et al. (2009). Detecting influenza epidemics using search engine query data. Nature 457:1012-1014.

Google (2012). Google Terms of Service, version 1 March 2012. http://www.google.nl/intl/en/policies/terms/regional.html

Google.org Flu Trends (2013). Explore flu trends around the world. http://www.google.org/flutrends

Google Inc. (2014). Google Trends. Available at: http://www.google.com/trends (accessed 3 September, 2014).

Hupkens C (2011). Revisie van de vragen over medische consumptie in POLS-gezondheid vanaf 2010. Rapport van de adviescommissie Revisie Medische consumptie. Statistics Netherlands (CBS), Den Haag/Heerlen.

iProspect (2014). Nationale search engine monitor. Only in Dutch available at http://www.iprospect.nl/our-world/iprospect-onderzoek/nationale-search-engine-monitor (accessed 3 September, 2014).

Murphy J, Kim A, Hagood H, et al. (2011). Twitter feeds and Google search query surveillance: can they supplement survey data collection? Proceedings of the sixth ASC International Conference, Association for Survey Computing, Bristol.

NOS (Netherlands Broadcasting Foundation) (2008). Groot onderzoek chlamydia bij jongeren. April 2008. Only in Dutch available at: http://nos.nl/artikel/70549-groot-onderzoek-chlamydia-bij-jongeren.html (accessed 3 September 2014).

RIVM (2010). Sexually transmitted infections, including HIV, in the Netherlands in 2009. RIVM report 210261007/2010. Centre for Infectious Disease Control, National Institute for Public Health and the Environment. In collaboration with:STI centres, Stichting HIV Monitoring and HIV treatment centres,ISIS/OSIRIS

RIVM (2011). Sexually transmitted infections, including HIV, in the Netherlands in 2010. RIVM report number: 210261009/2011. Centre for Infectious Disease Control, National Institute for Public Health and the Environment. In collaboration with:STI centres, Stichting HIV Monitoring and HIV treatment centres,ISIS/OSIRIS

RIVM (2012a). Kinkhoest terug van weggeweest. Infectieziekten Bulletin Themanummer vaccinaties oktober 2012.
http://www.rivm.nl/Documenten_en_publicaties/Algemeen_Actueel/Uitgaven/Infectieziekten_Bulletin/Jaargang_23_2012/Oktober_2012

RIVM (2012b). Sexually transmitted infections, including HIV, in the Netherlands in 2011. RIVM report number: 201051001/2012. Centre for Infectious Disease Control, National Institute for Public Health and the Environment. In collaboration with:STI centres, Stichting HIV Monitoring and HIV treatment centres,ISIS/OSIRIS

RIVM (2013a). Infectieziekten. Only in Dutch available at: http://zorgatlas.nl/gezondheid-en-ziekte/ziekten-en-aandoeningen/infectieziekten/ (accessed 18 June 2013, 1 July 2013 and 6 July 2013).

RIVM (2013b). Nationaal Kompas Volksgezondheid Seksueel overdraagbare aandoeningen (soa). Available at http://www.nationaalkompas.nl/gezondheid-en-ziekte/ziekten-en-aandoeningen/infectieziekten-en-parasitaire-ziekten/soa/ (accessed 20 June 2013)

RIVM (2014a). Mazelenepidemie 2013-2014. Only in Dutch available at: www.rivm.nl/Onderwerpen/M/Mazelen/ Mazelenepidemie _2013_ 2014 (accessed 3 September 2014).

RIVM (2014b). Seksueel overdraagbare aandoeningen (soa). Only in Dutch available at http://www.nationaalkompas.nl/ gezondheid-en-ziekte/ziekten-en-aandoeningen/infectieziekten-en-parasitaire-ziekten/soa/ (accessed 3 September 2014).

Soa Aids Nederland (2013). Recente cijfers soa en hiv in Nederland. Expertisecentrum voor HIV/aids en andere SOA. Available at: http://www.soaaids-professionals.nl/epidemiologie

Statistics Netherlands (2003). Vademecum gezondheidsstatistiek Nederland 2003, Voorburg. Only in Dutch available at: http://www.cbs.nl/NR/rdonlyres/77848968-EE71-4C67-83A0-47DEE6A316D4/0/vademecum2003.pdf

Statistics Netherlands (2013a). Korte onderzoeksbeschrijvingen Gezondheid en Welzijn. only in Dutch available at: http://www.cbs.nl/nl-NL/menu/themas/gezondheid-welzijn/methoden/dataverzameling/korte-onderzoeksbeschrijvingen/default.htm

Statistics Netherlands (2013b). Thema's Gezondheid en welzijn. http://www.cbs.nl/nl-NL/menu/themas/gezondheid-welzijn/cijfers/default.htm

Statistics Netherlands (2013c). Gezondheid, leefstijl en medische contacten; regio's, 2008/2011 Database, only in Dutch available at: http://statline.cbs.nl/StatWeb/publication/?DM=SLNL&PA=81976ned&D1=21-30&D2=0,l&D3=25,29&D4=(l-11)-l&VW=T (accessed 1 July 2013).

Statistics Netherlands (2013d). Ziekenhuisopnamen; geslacht, leeftijd en diagnose-indeling VTV. Database, only in Dutch available at: http://statline.cbs.nl/StatWeb/publication/?DM=SLNL&PA=71859ned&D1=1,3,5,7-8&D2=1-2&D3=l&D4=22-26&D5=0,4,9,14,19,24-30&VW=T

Statistics Netherlands (2014a). ICT gebruik van personen naar persoonskenmerken (2005-2013). Database, only in Dutch available at: http://statline.cbs.nl/StatWeb/selection/ ?DM=SLNL&PA=71098NED&VW=T (accessed 3 September 2014).

Statistics Netherlands (2014b). Gebruik medische voorzieningen; t/m 2009/ Medische contacten, ziekenhuisopname, medicijnen; leeftijd, geslacht. Database, only in Dutch available at: http://statline.cbs.nl/StatWeb/publication/?DM=SLNL&PA=7042mc &D1=0-347&D2=0-1,5,10,15,20,25-29&VW=T/ and http://statline.cbs.nl/StatWeb/ publication/?DM=SLNL&PA =81178ned&D1=0,11,22,37,40-41,48,51-52&D2=1-2&D3= a&D4=0&D5=l&VW=T (accessed 3 September, 2014).

Statistics Netherlands (2014c). Bijna 1 miljoen mensen onder behandeling van een alternatieve genezer. Only in Dutch available at: http://www.cbs.nl/nl-NL/menu/themas/gezondheid-welzijn/publicaties/artikelen/archief/2014/2014-4041-wm.htm?RefererType=RSSItem&RSSFeedTitle=Gezondheid%20en%20welzijn (accessed 3 September, 2014).

Swinkels H (2011). Trendcijfers Gezondheidsenquete 1981-2009. Gebruik geneeskundige voorzieningen, gezondheidsindicatoren en leefstijl. Statistics Netherlands (CBS), Den Haag/Heerlen.

Wikipedia (2013). Google Trends. Accessed 12 June 2013. http://en.wikipedia.org/wiki/Google_Trends

Explanation of symbols

.	Data not available
*	Provisional figure
**	Revised provisional figure (but not definite)
x	Publication prohibited (confidential figure)
–	Nil
–	(Between two figures) inclusive
0 (0.0)	Less than half of unit concerned
empty cell	Not applicable
2014–2015	2014 to 2015 inclusive
2014/2015	Average for 2014 to 2015 inclusive
2014/'15	Crop year, financial year, school year, etc., beginning in 2014 and ending in 2015
2012/'13–2014/'15	Crop year, financial year, etc., 2012/'13 to 2014/'15 inclusive

Due to rounding, some totals may not correspond to the sum of the separate figures.

Publisher
Statistics Netherlands
Henri Faasdreef 312, 2492 JP The Hague
www.cbs.nl

Prepress
Studio BCO, Den Haag

Design
Edenspiekermann

Information
Telephone +31 88 570 70 70, fax +31 70 337 59 94
Via contact form: www.cbs.nl/information