Published in final edited form as:

*Urology*. 2011 August ; 78(2): 267–271. doi:10.1016/j.urology.2011.01.010.

# Use of Google Insights for Search to track seasonal and geographic kidney stone incidence in the United States

**Benjamin N. Breyer**,
Department of Urology, University of California, San Francisco

**Saunak Sen**,
Department of Epidemiology and Biostatistics, University of California San Francisco

**David S. Aaronson**,
Department of Urology, University of California, San Francisco

**Marshall L. Stoller**,
Department of Urology, University of California, San Francisco

**Bradley A. Erickson**, and
Department of Urology, University of California, San Francisco

**Michael L. Eisenberg**
Department of Urology, University of California, San Francisco

## Abstract

**Objective—**To determine if internet search volume for kidney stones has seasonal and geographic distributions similar to known kidney stone incidence.

**Materials and Methods—**Google Insights for Search analyzes a portion of Google web searches from all Google domains to compute how many searches are performed for a given term relative to the total number of searches done over a specific time interval and geographic region. Selected terms related to kidney stones were examined to determine which most closely tracked kidney stone incidence. Google Insights for Search data was correlated with hospital admissions for the emergent treatment of nephrolithiasis found through the Nationwide Inpatient Sample. Ambient temperature in Seattle and New York were compared to search volume for these regions to display qualitative relationships.

**Results—**The term "kidney stones" had the highest seasonal correlation of terms examined (r=0.81, p=0.0014). Google Insights for Search output and National Inpatient Sample admissions also correlated when regions were compared (r=.90, p=0.005). Qualitative relationships between ambient temperatures and kidney stone search volume do exist.

Address for Correspondence: Benjamin N. Breyer, MD, 400 Parnassus St, Ste A-660, Campus Box 0738, Department of Urology, University of California San Francisco, San Francisco, CA, 94143 U.S.A., telephone: 415-353-2200, bbreyer@urology.ucsf.edu.



None of the authors have any financial conflicts or disclosures related to this manuscript.

**Conclusion—**Internet search volume activity for kidney stones correlates with temporal and regional kidney stone insurance claims data. In the future, with improved modeling of search detection algorithms and increased internet usage, search volume has the potential to serve as a surrogate for kidney stone incidence.

### Keywords

Nephrolithiasis; internet; seasons

## Introduction

Kidney stones (KS) often cause significant pain and suffering, can lead to visits to the emergency room, life threatening infections, and renal compromise.[1] In the United States, it is a common medical condition affecting 13% of men and 7% of women in their lifetime.[2] Nephrolithiasis accounted for $2.1 billion of health care expenditure in 2000 in the United States, a 50% increase from 1994.[1] Once diagnosed with KS, 50% of patients will develop a recurrence by 5 years.[3] In addition, KS incidence appears to be increasing.[2]

KS are known to display distinct geographic and seasonal variations whereby stone disease is more prevalent in warmer climates and during hotter times of the year. A "stone belt" has been described in the Southeast United States where an increased number of symptomatic stone disease is known to occur.[2,4,5] Other researchers have also established relationships between regional and temperature variations and stone volume.[6,7] However, estimating the burden of disease can be difficult given the varying locations of presentation (outpatient, inpatient, emergency department) and the different types of stone treatment (medical, surgical, dietary).[1]

Google's search engine (Google Inc, Mountain View, CA, U.S.A; http://www.google.com) has emerged as a ubiquitous tool of worldwide internet users. In 2009, Google commanded a 65% share of the roughly 130 billion online searches performed in the United States.[8] In August 2008, Google introduced Insights for Search (GIS) (http://www.google.com/insights/search), which provides normalized, relative search volume for selected terms over specific time ranges and geographic regions.

Our group recently reported the ability of GIS to track the incidence of diseases such as diabetes, myocardial infarction and blood pressure.[9] Our present objective is to determine if search queries in GIS may correlate with KS incidence and adequately represent seasonal or geographic variation. We hypothesize that GIS search volume will correlate with known KS distributions across time and geography, allowing for estimation of disease burden thus making it a novel tool for future studies of nephrolithiasis.

## Materials and Methods

### Google Insights for Search

Institutional Review Board approval was not required for this secondary analysis of a deidentified national data set. GIS (http://www.google.com/insights/search) (Google, Mountain View, CA, USA) was queried to generate seasonal and geographic specific information pertaining to nephrolithiasis-related search phrases (accessed May 10, 2010). GIS analyzes a portion of Google web searches from all Google domains to compute how many searches have been done for the entered term relative to the total number of searches done over a specific time interval and geographic region. GIS provides a normalized and relative search volume of entered queries on a scale of 0–100 and does not provide absolute search numbers. Search data is normalized to the total search volume for a specific region. The relative search volume may be interpreted as the probability that a random user searched for a particular search term from a specific location and time.

GIS uses Internet Protocol (IP) address information from Google search users to generate estimates regarding research query location. In order to protect privacy, a minimum threshold of searches must be entered for analysis inclusion. When query term data is insufficient, 0 is shown for the search volume.

## Nationwide Inpatient Sample

In order to verify that GIS search volume correlates with nephrolithiasis incidence, we compared monthly search volume to data of emergent hospital admissions caused by nephrolithiasis. We utilized the Nationwide Inpatient Sample (NIS) from 2006 and 2007.[10] NIS is a database of inpatient discharge abstracts collected via federal-state partnerships as part of the Agency for Healthcare Research and Quality's Healthcare Cost and Utilization Project. The NIS contains population adjusted discharge data from U.S, non-federal hospitals located in 40 states. This approximates a nationally representative 20% stratified sample of US non-federal hospitals (representing 39,541,948 total discharges).

NIS data includes patient discharge r.ecords, including the primary reason for admission (coded by International Classification of Disease-9) and whether it was an emergent or elective admission. We identified cases using ICD-9-CM codes listed under the primary diagnosis for each emergent admission. Nephrolithiasis was identified by the following codes: (592) Calculus of kidney and ureter; (592.0) Calculus of kidney, Nephrolithiasis NOS, Renal calculus or stone, Staghorn calculus, Stone in kidney, (592.1) Calculus of ureter, Ureteric stone, Ureterolithiasis; (592.9) Urinary calculus, unspecified; (788.0) Renal colic. ICD-9 codes were further separated into acute (788.0, 592.1) versus non-acute stone events (592.0).

## Term Screening

GIS was queried with various sets of phrases to determine which most strongly correlated with admissions for nephrolithiasis identified in the NIS. The following terms were used to search GIS: “kidney stones”, “flank pain”, “kidney stone pain”, “kidney stone symptoms”, “kidney”, “passing kidney stone”. “California”, “textbook” and “hernia” were entered as non-related terms to determine if correlations existed with non-related terms. The term “kidney stones” was also searched with a parameter to display only “health” related searches. Multiple permutations and combinations of search terms were grouped to determine which would best correlate with the NIS monthly data.

### Study population

**National—**We analyzed GIS output on a national, regional and metropolitan level. For national data, a time interval (January 2006 to December 2007) and global region (United States) were selected *a priori*. This time interval was selected to match available data from the Nationwide Inpatient Sample. This produced 104 weeks of normalized search volume that we divided into monthly intervals.

**Regional—**States within the contiguous United States were compared for relative search activity. GIS was queried with the term "kidney stones" to create a relative rank of states, with each assigned a relative search volume between 0–100. Years 2006 and 2007 were averaged together. States were then grouped based on region as previously described.[4] This regional GIS data was then compared to NIS regional emergent KS-related admissions. Regions were defined as follows: Northwest (Idaho, Oregon, Washington); Midwest (Illinois, Indiana, Iowa, Kentucky, Michigan, Minnesota, Missouri, Ohio, Wisconsin); Southeast (Florida, Georgia, North Carolina, South Carolina, Tennessee); South Central (Arkansas, New Mexico, Oklahoma, Texas); Southwest (Arizona, California, Nevada, Utah); Northeast (Delaware, Connecticut, New Hampshire, Maine, Maryland, Massachusetts, New Jersey, New York, Rhode Island, Vermont, Virginia, West Virginia, Washington D.C.); North Central (Colorado, Kansas, Nebraska, North Dakota, South Dakota). Wyoming, Delaware, and North Dakota were excluded from analysis because the search volume did not reach the minimum threshold needed to register a value. The following states were not included because NIS data was not available: Alabama, Idaho, Louisiana, Montana, Mississippi, Missouri, and Pennsylvania.

### Metropolitan

To obtain a qualitative representation of the relationship of weather patterns and search volume we examined 2 major metropolitan areas with a high density of internet users, New York City and Seattle. A GIS query of each metropolitan area from January 2005 to January 2010 was completed, yielding 260 weeks of search volume for the selected term, which was converted into monthly totals. Average daily temperature data was obtained from the National Weather Service for both cities and converted to monthly averages and compared to search volume.

### Statistical analysis

Pearson correlation coefficients were used to assess correlations between search phrases and NIS output. A heat map was generated from GIS to demonstrate the national distribution of relative search volume for each state from January 1, 2005 to December 31, 2009. Search data was exported from GIS and analyzed using STATA 11 (Statacorp, College Station, TX, USA).

## Results

### Term-selection and Correlation of GIS and NIS

The term "kidney stones" held the highest correlation with NIS monthly KS incidence (r=0.81, p=0.0014) (Figure 1) (Table 1). GIS output correlated more closely with "acute"

ICD-9 codes than chronic (r=0.84, p=0.0006 vs. r=0.78, p=0.0026, respectively). Interestingly, when the term “kidney stones” was restricted to health related searches by GIS, the correlation coefficient went down to 0.67, p=0.025.

GIS data was also correlated with regional NIS data (r=.90, p=0.005). GIS and NIS relative ranks were similar. For GIS, the Southeast ranked highest (Southeast (1), South Central (2), Midwest (3), Southwest (4), Northeast (5), North Central (6) and Northwest (7)). The rankings for NIS were the same except Northwest was 6 and North Central 7. A heat map of relative search volumes for the United States from January 1, 2005 to December 31, 2009 is displayed in Figure 2.

### Metropolitan Temperature and GIS output

Plots of mean monthly ambient temperature and GIS search volume for the term “kidney stones” are shown in Figure 3 for (A) New York and Seattle (B). The weather data has a predictable cyclic pattern. GIS data does not follow a pattern as clearly. Coordinated up strokes and down strokes can be seen for GIS activity and the ambient temperature. Spikes of activity can also be seen in cold weather months with unclear implications. Areas where search volume did not reach threshold values can be seen as zero readings in the Figure 3.

## Discussion

In the present study, we determined which Google search terms best correlated with known kidney stone incidence. GIS output for the term “kidney stone” replicated regional and temporal patterns of nephrolithiasis incidence in the United States. The appeal and promise of using GIS to study human disease comes from the ability to harvest the data entered by millions of people in real time. Potentially, minute to minute data on disease search patterns could be used to help improve understanding of disease pathophysiology and guide treatment.

The website Google Flu Trends monitors search data of flu related terms to localize outbreaks 7–10 days earlier than traditional systems.[11,12] Other infectious disease diagnoses such as avian influenza and West Nile virus have demonstrated high correlation with Google search volume.[11,12] We previously demonstrated that GIS can track non-infectious disease incidence for diseases such as stroke, myocardial infarction and diabetes. [9] The current study establishes that internet search volume may be used to estimate KS incidence.

The pathophysiology of nephrolithiasis remains poorly understood. Evidence exists that exposure to warm temperatures increases stone formation risk likely through dehydration, acidosis and mineral supersaturation.[13] Likewise, the amount of light exposure may increase the risk of stone formation via vitamin D production and calcium absorption.[2,5] Thus, longer daylight hours and warmer temperatures may explain why nephrolithiasis is more common in the summer months and in the Southeastern region of the United States.[4,5]

We investigated the qualitative relationship between average ambient temperature and search volume. While patterns emerged, the correlation was not perfect. This is a reflection of the complexity of the relationship between KS and ambient weather conditions. While weather plays a role in stone disease, it is one of many factors contributing to nephrolithiasis. In addition, genetics, diet, sunlight exposure, body mass index play a role in the pathophysiology of stone disease. [6,14–17] In addition, while we studied average temperature, it is likely that daily temperature highs and lows also impact the pathophysiologic response leading to stone formation. That is to say that it is possible that dramatic shifts in temperature result in acute dehydration that trigger bouts of kidney stones. A potential strength of GIS is its immediacy which could provide finer detail of minute to minute, hour to hour disease burden. At present, we were only able to obtain weekly data. In addition, a single search term (such as “kidney stones”) is very likely to be inadequate to effectively model nephrolithiasis incidence. It is likely that groups of terms strung together would have the highest correlation. Unfortunately, without the aid of Google programmers we do not have access to search data to perform such a query. When Google produced Google Flu to track flu breaks with internet searches, they devised a search monitoring strategy that incorporated 45 different terms. [11,12] This dramatically increased their ability to track flu outbreaks. Our methods could be vastly improved if we had the ability to build a comprehensive and tailored model.

Just as the biologic processes leading to stone formation are complex, so too are the factors that lead to internet search traffic. Holiday shopping or celebrity scandals can alter normal internet traffic and alter normal internet search patterns. [9] Physicians, patient’s entering multiple queries, patient family members and other lay people likely all contribute to internet search volume for a specific disease. It is unclear how this will impact internet search as a surrogate for disease incidence. A further limitation of the study is that Google does not provide details of their methods for generating the relative search volume. While ubiquitous, use of Google requires reading skills and computer access. This may produce a sample over-representative of people from a higher socioeconomic background who can read and have internet access. Potentially individuals who are in more sedentary work are in a higher socioeconomic and have improved internet access and are over-represented contributing to the distribution of stone disease. We correlated the GIS output to aggregated administrative claims data, which may introduce inaccuracies in actual kidney stone incidence. We limited our study to one national inpatient administrative claims data source. Our methods could be improved by using multiple different sources of kidney stone incidence data, such as claims from emergency room and outpatient visits, as well specific patient health data from national repositories.

Further collaboration with the search engine directly may allow more precision in the tracking of nephrolithiais by modeling fluctuations in search volume, GIS output versus humidity, daily sunlight exposure, and daily high-low temperature differentials. While this study is specific to kidney stones, this type of technology could be applied more broadly to other disease processes to improve our understanding of disease incidence and etiology. Expanding our understanding of disease burden by measuring disease with internet search volume could lead to breakthroughs in disease pathophysiology and treatment.

## Conclusions

Internet search volume activity for kidney stones correlates with temporal and regional kidney stone insurance claims data. In the future, with improved modeling of search detection algorithms and increased internet usage, search volume has the potential to serve as a surrogate for kidney stone incidence.

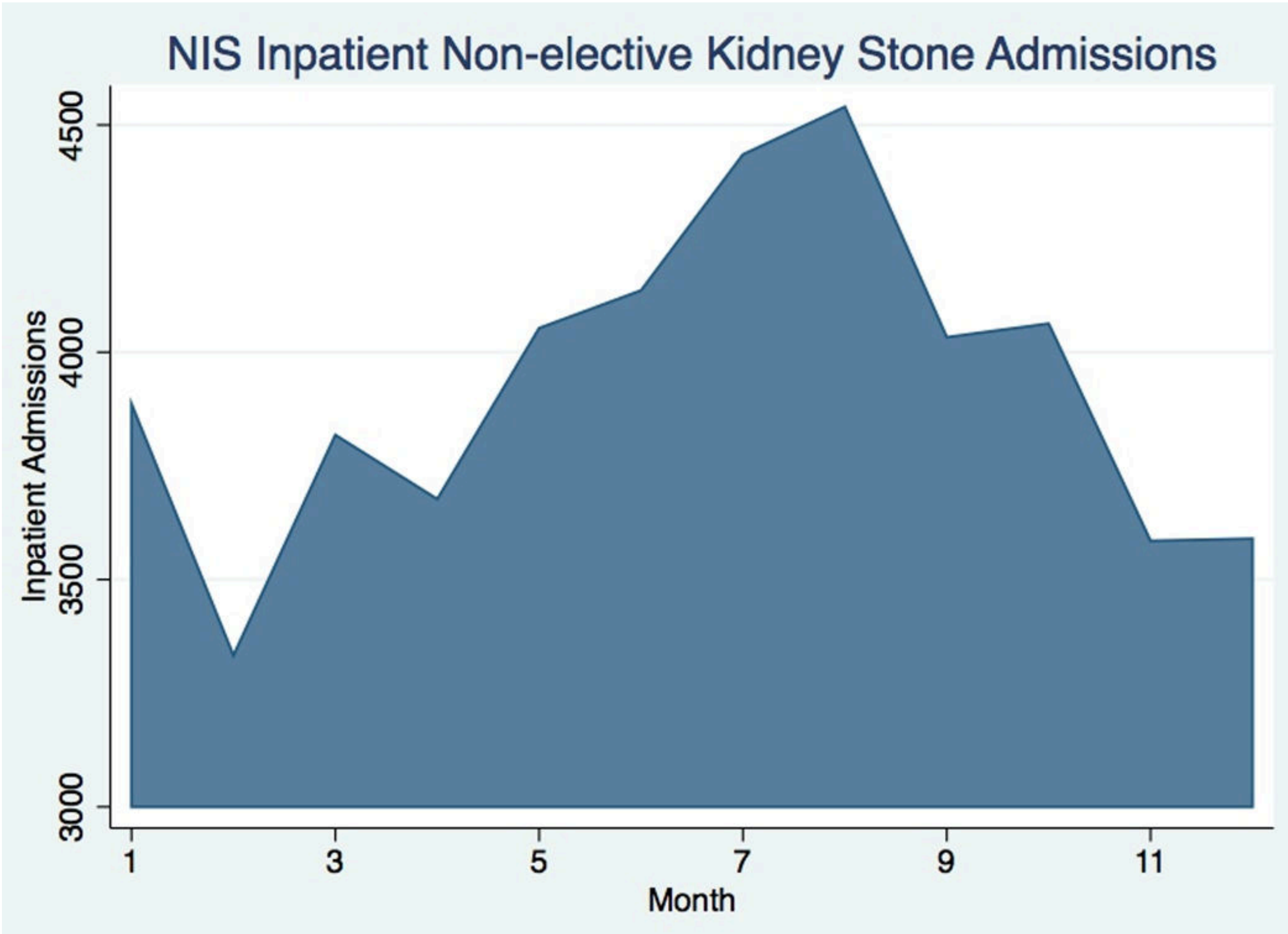
## Acknowledgement

BNB is supported by NIH grant K12DK083021.

## References

1. Pearle MS, Calhoun EA, Curhan GC. Urologic diseases in America project: urolithiasis. J Urol. 2005; 173:848–857. [PubMed: 15711292]
2. Brikowski TH, Lotan Y, Pearle MS. Climate-related increase in the prevalence of urolithiasis in the United States. Proc Natl Acad Sci U S A. 2008; 105:9841–9846. [PubMed: 18626008]
3. Glowacki LS, Beecroft ML, Cook RJ, Pahl D, Churchill DN. The natural history of asymptomatic urolithiasis. J Urol. 1992; 147:319–321. [PubMed: 1732583]
4. Soucie JM, Coates RJ, McClellan W, Austin H, Thun M. Relation between geographic variability in kidney stones prevalence and risk factors for stones. Am J Epidemiol. 1996; 143:487–495. [PubMed: 8610664]
5. Stamatelou KK, Francis ME, Jones CA, Nyberg LM, Curhan GC. Time trends in reported prevalence of kidney stones in the United States: 1976–1994. Kidney Int. 2003; 63:1817–1823. [PubMed: 12675858]
6. Chen YK, Lin HC, Chen CS, Yeh SD. Seasonal variations in urinary calculi attacks and their association with climate: a population based study. J Urol. 2008; 179:564–569. [PubMed: 18082222]
7. al-Hadramy MS. Seasonal variations of urinary stone colic in Arabia. J Pak Med Assoc. 1997; 47:281–284. [PubMed: 9510632]
8. [Accessed on March 7, 2010] Monthly U.S. Search Engine Rankings. 2009. Available at http://www.comscore.com/Press_Events/Press_Releases?year=1218&keywords=&location=0&searchBtn=Search
9. Breyer BN, Eisenberg ML. Use of Google in study of noninfectious medical conditions. Epidemiology. 21:584–585. [PubMed: 20539114]
10. Edited by AFHRA Quality. , editor. H. N. I. S.. Introduction to the HCUP Nationwide Inpatient Sample (NIS). Rockville, MD: 2006–2007.
11. Carneiro HA, Mylonakis E. Google trends: a web-based tool for real-time surveillance of disease outbreaks. Clin Infect Dis. 2009; 49:1557–1564. [PubMed: 19845471]
12. Ginsberg J, Mohebbi MH, Patel RS, Brammer L, Smolinski MS, Brilliant L. Detecting influenza epidemics using search engine query data. Nature. 2009; 457:1012–1014. [PubMed: 19020500]
13. Atan L, Andreoni C, Ortiz V, Silva EK, Pitta R, Atan F, Srougi M. High kidney stone risk in men working in steel industry at hot temperatures. Urology. 2005; 65:858–861. [PubMed: 15882711]
14. Eisner BH, Eisenberg ML, Stoller ML. Relationship between body mass index and quantitative 24-hour urine chemistries in patients with nephrolithiasis. Urology. 75:1289–1293. [PubMed: 20018350]
15. Eisner BH, Eisenberg ML, Stoller ML. Influence of body mass index on quantitative 24-hour urine chemistry studies in children with nephrolithiasis. J Urol. 2009; 182:1142–1145. [PubMed: 19625057]
16. Borghi L, Schianchi T, Meschi T, Guerra A, Allegri F, Maggiore U, Novarini A. Comparison of two diets for the prevention of recurrent stones in idiopathic hypercalciuria. N Engl J Med. 2002; 346:77–84. [PubMed: 11784873]

17. Khan SR, Canales BK. Genetic basis of renal cellular dysfunction and the formation of kidney stones. Urol Res. 2009; 37:169–180. [PubMed: 19517103]


NIS Inpatient Non-elective Kidney Stone Admissions
Inpatient Admissions
4500
4000
3500
3000
1
3
5
7
9
11
Month

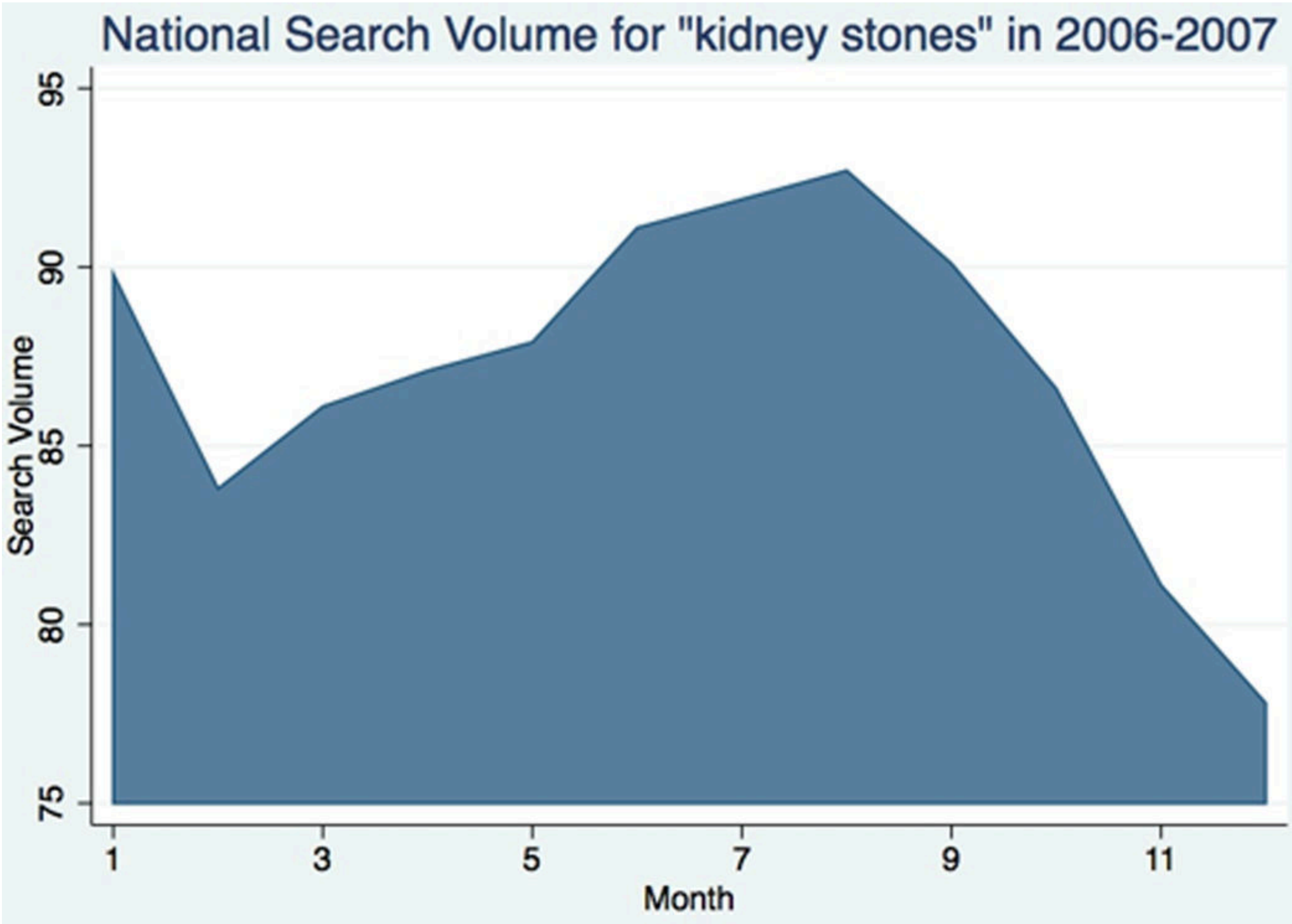



**Figure 1.**

Comparison of the (A) Nationwide Inpatient Sample non-elective kidney stone related hospital admissions and the (B) Google Insights for Search output for search term "kidney stones" over the same period (2006–2007).

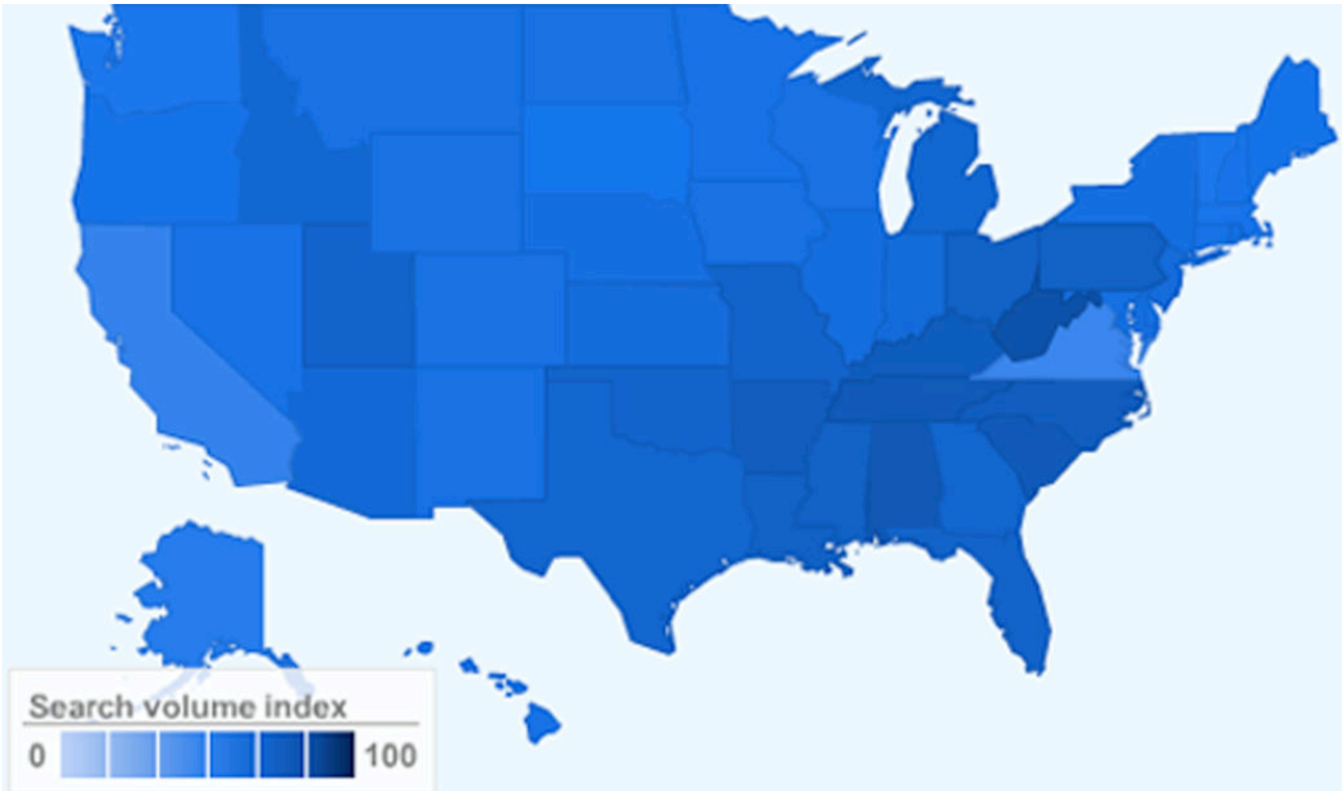
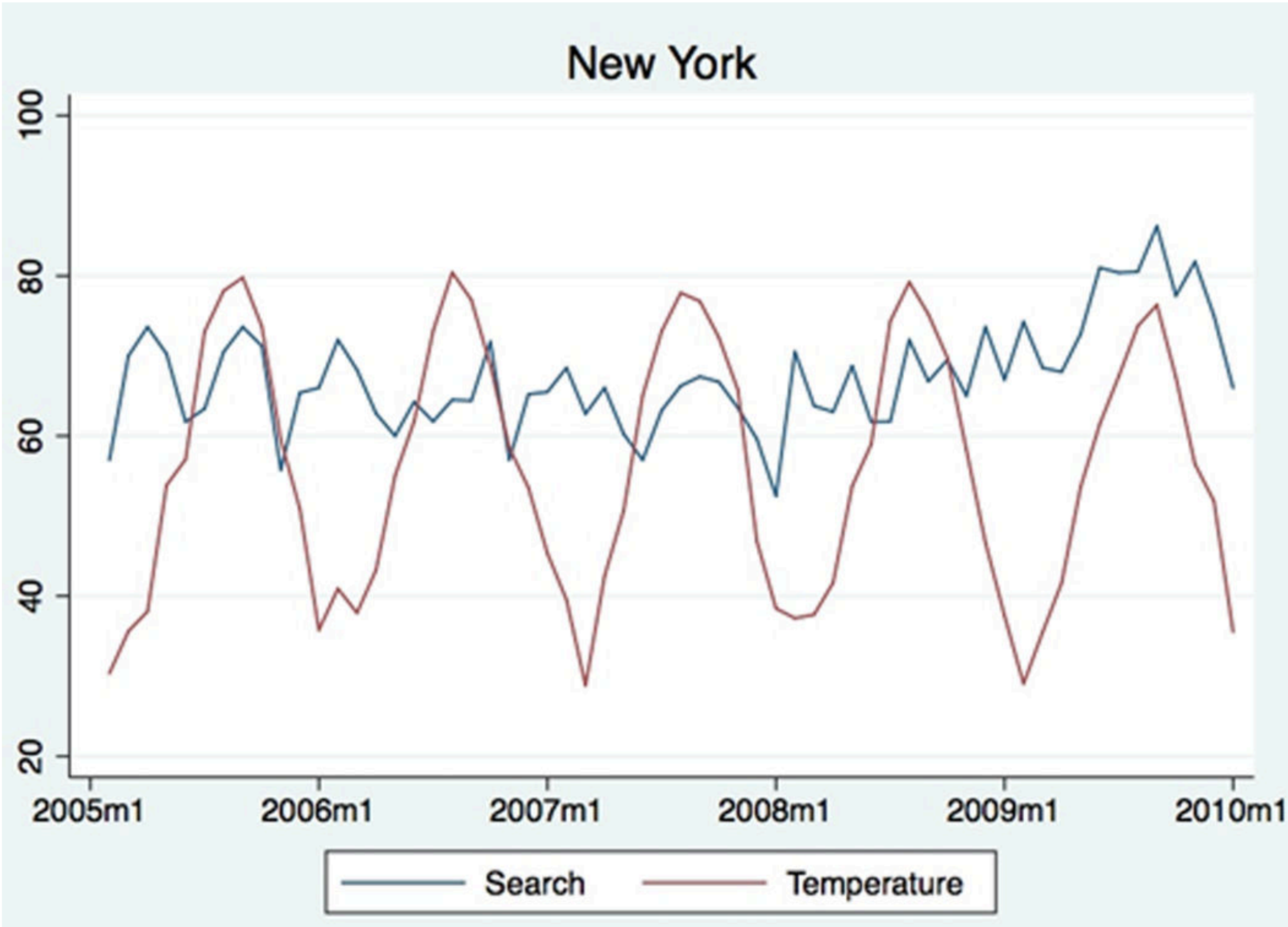



**Figure 2.**

Heat map generated from Google Insights for Search depicting relative kidney stone search volume in the United States during January 1, 2005 to December 31, 2009.


New York
100
80
60
40
20
2005m1
2006m1
2007m1
2008m1
2009m1
2010m1
Search
Temperature

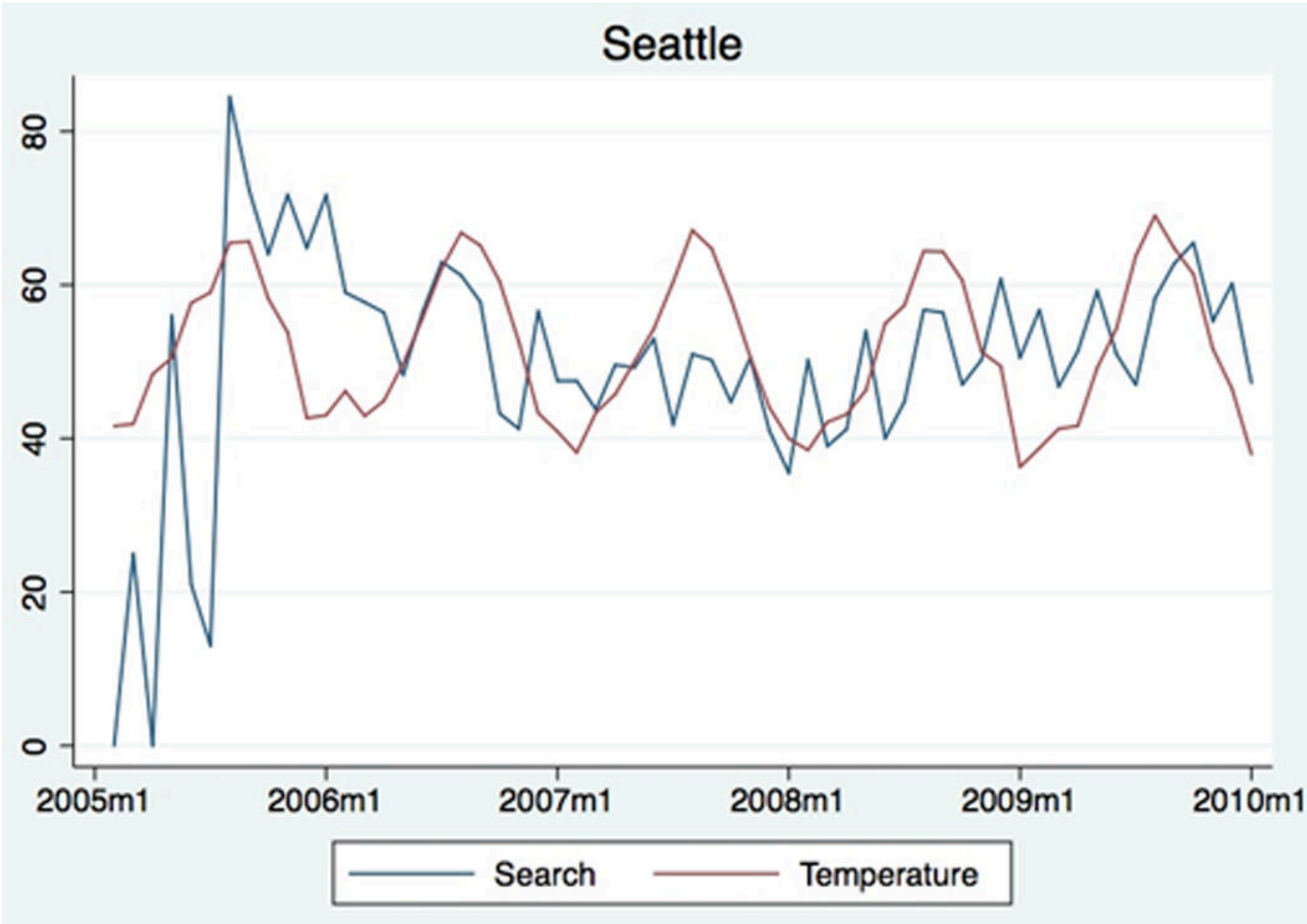



**Figure 3.**

Monthly Google Insights for Search volume (blue) for search term "kidney stones" and daily temperature averages (red) over time (2005–2009). 3A. New York City. 3B. Seattle.

**Table 1**

Pearson correlation of various search terms and the Nationwide Inpatient Sample monthly inpatient kidney stone related admissions from 2006 and 2007 in the United States

| | Search term | r | p |
|---|---|---|---|
| 1. | "kidney stones" | 0.81 | 0.0014 |
| 2. | "flank pain" | 0.28 | 0.362 |
| 3. | "kidney stone pain" | −0.06 | 0.854 |
| 4. | "kidney stone symptoms" | 0.16 | 0.624 |
| 5. | "kidney" | 0.15 | 0.639 |
| 6. | "passing kidney stone" | −0.32 | 0.308 |
| 7. | "California" | 0.39 | 0.209 |
| 8. | "textbook" | 0.17 | 0.593 |
| 9. | "hernia" | 0.41 | 0.191 |
| 10. | "kidney stones" restricted to health | 0.67 | 0.025 |