Annual variation in Internet keyword searches: Linking dieting interest to obesity and negative health outcomes
Patrick M Markey and Charlotte N Markey
*J Health Psychol* 2013 18: 875 originally published online 19 September 2012
DOI: 10.1177/1359105312445080

The online version of this article can be found at:
http://hpq.sagepub.com/content/18/7/875
Individuals who are overweight are at a substantially higher risk than peers of healthy weight for various negative health outcomes and can therefore benefit from effective and long-term weight loss strategies (Markey and Markey, 2005, 2011a). Although healthy dieting can lead to weight loss, it is questionable if most individuals who engage in dieting maintain their weight loss across time. A recent review of studies examining the long-term benefits of dieting found little evidence that diet programs lead to long-term weight loss (Mann et al., 2007). Instead, it appears that the weight loss experienced due to dieting tends to be temporary, and after about 2 years, most individuals gain back any weight they had lost.

Weight cycling, the repeated loss and regaining of weight, is often observed among dieters and obese individuals (Brownell and Rodin, 1994; Field et al., 1999; Kroke et al., 2002; Stice et al., 1999). Weight cycling appears to have a negative impact on individuals’ general mental health (e.g. depression and self-esteem; Brownell and Rodin, 1994; Friedman and Brownell, 1995;
Lowe, 1993; Rosen et al., 1987, 1990). Studies even suggest that weight cycling is linked to increased physical health risks, such as heart disease, diabetes, and stroke (French et al., 1997; Hamm et al., 1989). Taken together, past research on dieting suggests that, for most individuals, diets have little to no positive effects on long-term weight loss or health. Instead, it appears that it is extremely common for dieters to gain more weight than they lost during a diet. What is perhaps most concerning is that such weight cycling appears to have negative effects on health and longevity. These negative effects of weight cycling led Mann et al. (2007), in their seminal review of the dieting literature, to stress the importance of continued research examining the potential outcomes of weight cycling.

In order to extend and complement previous research examining dieting, the current research employs a novel methodology—Internet keyword searches—to examine both the annual cycle of dieting interests and the potentially negative effects of these cycles on obesity and health at a national level. Given that about 78 percent of the population in the United States has access to the Internet (World Bank, 2011), Internet keyword searches have become a valuable source of information about what issues, concerns, and desires individuals residing within the United States are thinking about at any given moment (Markey and Markey, in press). By simply typing a few words into an Internet search engine (e.g. Google), individuals are able to obtain information on any topic of interest. For example, a person might type the word “diet” or “Nutrisystem” into Google’s search engine when attempting to find information about dieting. Past researchers have successfully used Internet keyword searches to examine seasonal affect disorder (Yang et al., 2010), to track H1N1 outbreaks (Ginsberg et al., 2009), suicide (McCarthy, 2010), and even pornography searches (Markey and Markey, 2010, 2011b, in press).

Within the United States, 80 percent of the adult Internet users (about 93 million individuals) have used the Internet to search for health-related topics, and almost half of these searches relate to dieting (Fox, 2005; Fox and Rainie, 2002). Although individuals start dieting during various times of the year, people often begin dieting as a form of a New Year’s resolution. In 2011, losing weight was the second most popular New Year’s resolution right behind quitting smoking (Marist College Institute for Public Opinion, 2010). Due to the lack of success in dieting, it has been argued that most will eventually give up these New Year’s resolutions and few will ever maintain any weight loss (Kassirer and Angell, 1998; Norcross et al., 1989; Polivy and Herman, 2002). It seems probable that many individuals who start a diet as a result of a New Year’s resolution will slowly lose interest in dieting after the New Year only to regain interest in dieting the following January.

The current research will empirically determine whether or not there is an annual dieting cycle that peaks around New Year and then slowly declines and whether such an annual dieting surge is related to obesity and negative health outcomes. It is predicted (Prediction 1) that a 12-month linear model will fit Internet searches for dieting-related keywords at the national level (i.e. Internet searches for dieting will occur most frequently in January and then will linearly decrease until surging again in the following January; see Figure 1). Additionally, if a 12-month linear model does fit dieting-related keyword searches, the size of the dieting-related keyword surge can also be computed (see Figure 1). The dieting-related keyword surge indicates the percentage increase in searches for dieting-related keywords that tends to occur between December and January and helps quantify the size of the dieting cycle. Given the link between dieting and the negative consequences of weight cycling (French et al., 1997; Hamm et al., 1989), the presence and size of a dieting-related keyword surge might serve as an important predictor of negative health outcomes.

In order to examine the potential negative association between dieting-related keyword surges and health outcomes, data will also be examined at the state level. Specifically, for each state and the District of Columbia, a dieting-related
keyword surge will be computed along with an overall indicator of how frequently individuals within a given location search for dieting-related keywords using the Internet (called the overall dieting-related keyword search). These two variables (dieting-related keyword surge and overall dieting-related keyword search) will then be used to predict state-level obesity rates and negative health outcomes. Given the relationship between weight cycling and obesity (c.f. Field et al., 1999; Kroke et al., 2002) and obesity and dieting (i.e. individuals who are overweight tend to diet the most; c.f. Hill, 2004), it is predicted (Prediction 2) that both a dieting-related keyword surge and overall dieting-related keyword search will be uniquely related to obesity. In other words, it is expected that individuals located in states with high levels of obesity will tend to search for dieting-related keywords more frequently, and will tend to show a greater interest in these searches between December and January than individuals located in states with low levels of obesity.

Finally, because past studies suggest that weight cycling is linked to increased risk of heart disease, diabetes, and stroke (French et al., 1997; Hamm et al., 1989), it is predicted (Prediction 3) that mortality rates for these diseases will be higher in states with greater dieting-related keyword surges. In other words, it is expected that individuals located in states where there was a high dieting-related keyword surge between December and January, suggesting higher incidences of weight cycling in these states, will be more likely to die from diabetes, heart disease, and stroke than individuals located in states where the dieting-related keyword surge is not as high.

Of course, some caution is always necessary when using aggregated keyword search data to predict the behaviors and interests of individuals.

Figure 1. Expected monthly changes in dieting-related keyword searches. This figure displays the predicted 12-month linear model and the dieting-related keyword surge that is expected to occur between December and January.
(c.f. Markey and Markey, 2010, 2011b, in press; Wagner, 1982). This methodology also does not allow for causal associations to be determined. However, even with these limitations, there are tremendous benefits associated with this methodology. Data yielded from keyword searches are not reliant on self-reports of dieting interest or behaviors and therefore avoids many of the issues and limitations associated with the use of self-report data (Funder, 1999; John and Robins, 1993). In short, individuals searching for dieting-related keywords are doing so because they have some interest in this topic. This methodology also provides an extremely large database derived from billions of keyword searches that can be examined at different time points and different geographic locations. Therefore, results yielded from Internet keyword searches have the potential to extend and complement past research examining links between dieting and obesity as well as weight cycling and deleterious health consequences, which have often been conducted using either self-report data (c.f. Hamm et al., 1989; Lowe, 1993; Rosen et al., 1987, 1990) or relatively small samples from a single geographical location (c.f. Anderson et al., 1999; Foster et al., 1996; Lantz et al., 2003; Pekkarinen and Mustajoki, 1997; Walsh and Flynn, 1995).

Method

Participants

Participants for this study were individuals residing in the United States who entered select keywords into the Google search engine between January 2005 and March 2011.

Procedure

Dieting-related keyword surge. In order to examine the proposed 12-month linear model and compute dieting-related keyword surges (see Figure 1) at the state and national levels, Google Trends was utilized to determine how often individuals searched for dieting-related keywords between January 2005 and March 2011. The dieting-related keywords included “diet” and “dieting” along with popular diet program names (e.g. “Weight Watchers,” “Nutrisystem,” “Jenny Craig,” “Atkins,” etc.). Google Trends examines Google web searches to determine how many searches for the given set of keywords had been conducted in a given month relative to the average number of searches on Google for those keywords over the entire observed time period. Although Google Trends does not supply the raw number of searches that occurred, the information it provides allows for the computation of a monthly relative search volume index (RSVI). This value represents the monthly percent increase or decrease of searches for a set of keywords relative to the norm (Google, 2009; Markey and Markey, 2010, 2011b). The interpretation of an RSVI is fairly straightforward. An RSVI score of 0 percent indicates that the monthly search volume for a given keyword set was equivalent to the norm for the given keyword set. A monthly RSVI score of 2 percent indicates that the search volume for a given keyword set was 2 percent higher than the norm and an RSVI score of −2 percent indicates that the search volume for a keyword set was 2 percent lower than the norm in a given time period. The resulting state and national RSVI scores between January 2005 and March 2011 were used to examine whether or not a 12-month linear trend in dieting-related keyword searches tended to occur and the size of dieting-related keyword surge at the state and national levels.

Overall dieting-related keyword searches. Because the RSVI value computed to examine dieting-related keyword surges provides information about the percentage change in dieting-related keyword searches relative to the norm of dieting-related keyword searches, it does not provide information about the overall number of dieting searches conducted in a given state. In other words, although RSVI scores are useful to examine changes in keyword searches, they do not provide information about the total number of searches conducted in a given area. However,
Google Insights (a service similar to Google Trends) can be used to compute overall dieting-related keyword searches for each state (Google, 2011). To do this, Google Insights examines Google web searches to determine how many total searches for a given set of keywords had been conducted in each state during a given time period relative to the overall number of keyword searches on any topic within that state. The resulting coefficient controls for differences in population and Internet penetration within various states by considering the number of dieting-related keyword searches relative to all keyword searches in a given geographic area. Next, Google Insights standardizes these state coefficients by dividing each by the greatest state-level coefficient and multiplying by 100 (Google, 2011). This resulting value can range between 0 and 100. In this range, a value of 100 would indicate that the given state had the greatest overall dieting-related keyword searches between January 2005 and March 2011. A state with half of the dieting-related keyword searches as the highest state would receive a value of 50, and so forth (Google, 2011). Figure 2 displays the overall dieting-related keyword searches for each state and the District of Columbia. As can be seen in Figure 2, there was a considerable amount of variability in overall dieting-related keyword searches across the states and the District of Columbia. The states that produced the highest overall dieting-related keyword searches included Tennessee (100), Florida (100), and Mississippi (98), and the states with the lowest overall dieting-related keyword searches were Virginia (59), Vermont (62), and California (69).

State obesity rates and health status. Consistent with the guidelines provided by the Centers for Disease Control and Prevention (CDC), this study

![Overall Dieting-Related Keyword Searches](image_url)
operationalized obesity rates as the percent of adults who had a body mass index (BMI) greater than or equal to 30 within a given state. Each state’s obesity rate was obtained from the CDC (2010). In order to operationalize various health outcomes of each state, the CDC provided death rate information for each of the causes examined in this study (heart disease, stroke, and diabetes). Specifically, for each state, the number of people who died, per 100,000, due to a given cause was computed (Xu et al., 2010).

**Results**

*Prediction 1: detecting a dieting-related keyword surge at the national level*

In the following analyses, the focus was examining annual changes that occurred in dieting-related keyword searches. Therefore, all the national RSVI data were de-trended by removing any variance attributable to linear trends. This allows the annual analysis to be conducted and removes any artifacts in the results that might be due to the general increase in popularity of Google. This was accomplished by regressing time on monthly RSVI values and conducting all subsequent analyses on these residuals (Warner, 1998).

Figure 3 displays the monthly changes in searches related to dieting each month for the period between January 2005 and March 2011. A visual inspection of these data suggests a relatively consistent dieting-related keyword surge occurring in January followed by a linear decrease in searches until the next year. To formally test Prediction 1, each month was coded in a linear manner starting at January (coded 11) to December (coded 0). To make the interpretation of this analysis easier, these monthly codes were divided by 11 (e.g. January = 1, June = .55, December = 0, etc.). The fit of a 12-month linear model was then tested by regressing the RSVI value on these monthly codes via a time-series analysis using ordinary least

![Figure 3](image-url)
Markey and Markey

881

squares (OLS) regression methods (Warner, 1998). The fit of a 12-month linear model to these data can then be estimated using a multiple R with a high multiple R value indicating that the model accounts for a large amount of the observed variance. Additionally, because January was coded 1 and December was coded 0, the slope of this analysis estimates the average size of the dieting-related keyword surge.

Results indicated that a 12-month linear model accounted for 77 percent of the variance in dieting searches (Multiple R = .88, p < .01). Additionally, searches for dieting-related keywords surged an average of 29 percent between December and January. This coefficient (i.e. the dieting-related keyword surge), suggests that, on average, there was a 29 percent increase in searches for dieting-related keywords between December and January. The lagged autocorrelations were not statistically significant, suggesting that the residuals from this analysis are essentially noise and the significance tests that can be interpreted with a degree of certainty (Warner, 1998). Figure 3 graphically displays the resulting 12-month linear model for this analysis. Taken together, and consistent with Prediction 1, these results indicate that searches for dieting tend to peak in January and then show a linear decrease until the next year.

Predictions 2 and 3: relating states’ dieting-related keyword surges and overall dieting-related keyword searches to obesity rates and health indicators

As was done at the national level, a 12-month linear model was also fitted to the RSVI scores for each state and the District of Columbia. The results from these analyses indicated that a 12-month linear model adequately fit all the states (mean Multiple R = .65; standard deviation (SD) = .13). Additionally, although all the states displayed a 12-month linear trend, the size of the dieting-related keyword surge varied by state (see Figure 4). For example, the states

![Diagram](image_url)

**Figure 4.** Dieting-related keyword surge for each state. Scores represent the average percentage increase in searches for dieting-related keywords in a given state between December and January.
with the highest dieting-related keyword surges included South Dakota (54%), Tennessee (50%), and North Dakota (46%). The states with the lowest dieting-related keyword surges included Vermont (18%), Hawaii (22%), and California (27%). This indicates that for some states (e.g. South Dakota, Tennessee, etc.) there tends to be a fairly large increase in dieting-related keyword searches between December and January, whereas for other states (e.g. Vermont, Hawaii, etc.), the increase for dieting-related keyword searches during this time period is more modest.

In order to examine Prediction 2, state obesity rates were regressed on states’ dieting-related keyword surges and states’ overall dieting-related keyword searches. As predicted, both of these variables were uniquely related to state obesity rates (see Table 1). Such a finding suggests that individuals located in states with high levels of obesity tend to search for dieting-related keywords more frequently and tend to show a greater increase in these searches between December and January than individuals located in states with low levels of obesity. In a similar manner, to test Prediction 3, each of the health variables was regressed on dieting-related keyword surges and overall dieting-related keyword searches. The results from this analysis found that, as expected, dieting-related keyword surges consistently and significantly predicted these various health outcomes. The effect sizes yielded from these analyses were moderate (Cohen, 1992) and have practical significance because they predicted mortality (Ferguson, 2009). Specifically, individuals located in states where there was a high dieting-related keyword surge were more likely to die from diabetes, heart disease, and stroke than individuals located in states where the dieting-related keyword surge was not as high. Perhaps most interesting is that a dieting-related keyword surge was a significant predictor of these negative health outcomes even after controlling for overall dieting-related keyword searches. Such a finding is consistent with the notion that diet cycling is a more important predictor of negative health outcomes than simply dieting in general.

Table 1. Multiple regression analyses predicting states’ obesity and causes of death from states’ overall dieting-related searches and states’ dieting-related keyword surge between December and January.

<table>
<thead>
<tr>
<th></th>
<th>B</th>
<th>SE B</th>
<th>Beta</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Obesity (Multiple R = .58</strong>)**</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Overall dieting-related keyword searches</td>
<td>.14</td>
<td>.04</td>
<td>.39** (.13 to .60)</td>
</tr>
<tr>
<td>Dieting-related keyword surge</td>
<td>.18</td>
<td>.06</td>
<td>.35** (.08 to .57)</td>
</tr>
<tr>
<td><em><em>Diabetes (Multiple R = .38</em>)</em>*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Overall dieting-related keyword searches</td>
<td>−.01</td>
<td>.08</td>
<td>−.02 (−.29 to .26)</td>
</tr>
<tr>
<td>Dieting-related keyword surge</td>
<td>.36</td>
<td>.12</td>
<td>.38** (.12 to .60)</td>
</tr>
<tr>
<td><strong>Heart disease (Multiple R = .45</strong>)**</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Overall dieting-related keyword searches</td>
<td>.86</td>
<td>.43</td>
<td>.26 (−.01 to .49)</td>
</tr>
<tr>
<td>Dieting-related keyword surge</td>
<td>1.54</td>
<td>.65</td>
<td>.31* (.04 to .54)</td>
</tr>
<tr>
<td><em><em>Stroke (Multiple R = .37</em>)</em>*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Overall dieting-related keyword searches</td>
<td>.10</td>
<td>.12</td>
<td>.11 (−.17 to .37)</td>
</tr>
<tr>
<td>Dieting-related keyword surge</td>
<td>.47</td>
<td>.19</td>
<td>.33* (.06 to .56)</td>
</tr>
</tbody>
</table>

SE: standard error.

* N = 51 states and the District of Columbia. Values in parentheses represent the 95 percent confidence interval for a given beta.

*p < .05; **p < .01.
Discussion

This study extends past research investigating links between dieting and obesity as well as weight cycling and negative health outcomes by examining Internet keyword searches between January 2005 and March 2011. As expected, it was found that searches related to dieting peaked in January and then linearly decreased until December, when they then, on average, surged 29 percent in January (see Figure 3). The observed dieting-related keyword surge between December and January is consistent with the notion that many individuals start dieting as a form of a New Year’s resolution and then slowly lose interest in dieting until the next New Year (Kassirer and Angell, 1998; Norcross et al., 1989; Polivy and Herman, 2002). The occurrence and size of such a keyword surge is of interest given previous research suggesting that weight cycling is related to various negative health outcomes.

This study also found that the size of the dieting-related keyword surge and the overall number of dieting searches each uniquely predicted obesity at the state level. The results indicated that individuals located in states that had high levels of obesity tended to search for dieting-related keywords more frequently than individuals located in states with low levels of obesity. Such a result is consistent with previous research suggesting that individuals who are overweight tend to diet the most (Hill, 2004). Although this finding may not be particularly surprising, it is interesting that state dieting-related keyword surge (i.e. the size of a state’s dieting cycle that occurs between December and January) predicted state obesity even when the overall dieting-related keyword search was controlled. Such a finding complements past research that has suggested that weight cycling is related to obesity (Field et al., 1999; Kroke et al., 2002).

Not only did state dieting-related keyword surge uniquely predict state obesity but it was also related to various negative health outcomes. Specifically, individuals located in states that had high dieting-related keyword surges, suggesting higher incidences of weight cycling, expressed higher mortality rates due to diabetes, heart disease, and stroke than individuals located in states with low dieting-related keyword surges. Such findings are consistent with studies that have found that weight cycling tends to have negative effects on longevity and might be particularly predictive of increased risks of heart disease, diabetes, and stroke (French et al., 1997; Hamm et al., 1989). It is interesting to note that, in this study, overall, dieting-related keyword search was not a significant predictor of these negative health outcomes. In other words, when predicting mortality at the state level, it is how much individuals located in a state tend to cycle in their interest in dieting between December and January, not the overall interest in dieting expressed by these individuals, that is the best predictor of adverse health outcomes.

Limitations and conclusions

Although the results of this study extend previous research examining dieting behaviors through the use of an extremely large data set (created by examining billions of Internet searches from various geographic locations during the past 7 years), there are several limitations that should be noted. Other than location, Google Trends does not supply user data, such as gender. Although past research suggests that weight cycling affects both men and women, it is unclear whether or not the observed relations in this study are equivalent for men and women. Additionally, because the data used in this study are based on aggregated data, some caution is warranted when using these results to predict the behaviors of individuals. However, given that these findings are consistent with results examining dieting behaviors and negative outcomes (c.f. French et al., 1997; Hamm et al., 1989), it is unlikely that the use of aggregate data are a confound in this study.
Finally, results from this study are somewhat limited because Google does not provide the raw number of times a term was searched for in a given month. Such information would be useful in order to better understand exactly the size of the dieting-related keyword surges observed in this study. However, given the number of Internet searches done each month using Google (especially for diet information; Fox, 2005; Fox and Rainie, 2002), it seems likely that even modest dieting-related keyword surges could reflect thousands of additional searches. Therefore, the national diet surge found in this study (29%) likely reflects thousands (if not millions) of additional searches that occur annually during December and January for dieting information.

In order to examine the relation between interest in dieting and various negative health outcomes at the national level, this study employed a novel methodology—Internet keyword searches. Given the prevalence of Internet use, this methodology allows for a large-scale sampling of various topics researchers might be interested in examining. Past studies have already employed this methodology to examine topics as diverse as seasonal affect disorder (Yang et al., 2010), the challenge hypothesis (Markey and Markey, 2010, 2011b), changes in pornography use (Markey, in press), flu outbreaks (Ginsberg et al., 2009), and suicide (McCarthy, 2010). In a similar manner, this study demonstrated how researchers can use this methodology to examine topics of interest in various areas of health psychology. Researchers utilizing this methodology will soon realize that it provides a quick and inexpensive means of examining an extremely large set of data from various geographic locations across many years.

An examination of weight cycling and its correlates at a national level is critical in the current era of obesity. Health-care professionals and their patients need to have a clear understanding of the potentially deleterious consequences of pursuing weight loss without a long-term, lifelong approach in mind. This study provides further support for the strong possibility that a “New Year’s Resolution diet” may do more harm than good in the long term.

**Funding**

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

**References**


