



ARTICLE

DOI: 10.1057/s41599-017-0055-7

OPEN

Gluttony and guilt: monthly trends in internet search query data are comparable with national-level energy intake and dieting behavior

Sean Coogan¹, Zhixian Sui¹ & David Raubenheimer¹

ABSTRACT Systems for monitoring population-level diet and nutritional intake have been considered insufficient across many countries. Recently, internet search query data have been used to examine spatial and temporal patterns of public behavior to inform public-health campaigns, policies, and interventions. Seasonal trends in public interest in behavioral change associated with obesity have been documented using such data. However, it has not been validated whether search query data can be related to diet and nutritional intake at the population level. The purpose of this study was to investigate whether trends in search query data related to behavioral changes associated with obesity reflects population nutritional intake and dieting behavior. First, long-term (2004 to 2016) trends in Australian Google search behavior were examined for the terms “weight loss”, “diet”, and “fitness” to establish monthly patterns in relative search volume (RSV). Second, monthly total energy (kJ), macronutrient, and food intake of the Australian population, and the percentage of self-reported dieters, were quantified using data collected as part of a 2011–2012 national-level survey. The two independent data sets were then compared to ascertain similarities in trends. There were distinct patterns in RSV across months, which was significantly higher than the mean during January, and lower during December, for all search terms. The decline in RSV was not linear, however, as there were significantly lower RSVs for terms during May to July, and significantly higher from August to October. Likewise, nutritional data showed a seasonal pattern, with the energy intake of survey participants highest in December and lowest in February, and the percentage of self-reported dieters closely followed monthly patterns in RSV. The proportion of energy from protein was consistent across months examined; however, energy from lipid and carbohydrate + fiber, was variable between months. Likewise, consumption patterns of different food groups was variable across months. Our analysis suggests that search query data can be used to surveil and predict dietary behavior at the population level, which has implications for producing novel and contemporaneous health information and marketing strategies.

¹University of Sydney, Sydney, NSW, Australia. Correspondence and requests for materials should be addressed to S.C. (email: sean.c.p.coogan@gmail.com)

Introduction

There is strong evidence of a global obesity pandemic related to over-nutrition and underactivity (James et al., 2001; Swinburn et al., 2011; Stevens et al., 2012). Incidences of obesity and associated health impacts such as cardiometabolic disease continue to increase, and there have been no reported national-level successes in addressing the problem in over three decades (Ng et al., 2014). There is thus clearly a need to develop effective public-health policies and interventions to combat overweight and obesity. For this, decision makers require contemporaneous information to assess progress and prioritize action relating to overweight and obesity (Ng et al., 2014). Yet, systems for monitoring population-level diet and nutritional intake have been considered insufficient in the majority of countries worldwide (Swinburn et al., 2011).

New approaches to understanding public behavior are emerging in the era of “Big Data”. Internet search query data is one form of Big Data that has been used to provide insight into temporal and biogeographic patterns of public behavior (Brownstein et al., 2009; Nuti et al., 2014). Such data have been used to aid researchers, practitioners, and policy makers in designing spatially and temporally specific public-health campaigns, policies, and behavioral interventions (Carr and Dunsiger, 2012).

One such tool is Google Trends (Google Inc.), a publically available online tool that allows users to access and interact with Internet search data (Google, 2017a). Google Trends analyzes a subset of daily Google search queries and provides data on temporal and geospatial patterns in search volumes for user-inputted terms (Google, 2017b). Google Trends holds considerable potential for deriving meaningful insights from public search behavior (Nuti et al., 2014), and has been used in variety of studies including those related to disease (Ginsberg et al., 2009; Fenichel et al., 2013; Kostkova et al., 2013), tobacco use (Ayers, 2011a, b, 2012), suicide and mental health (McCarthy, 2010; Gunn and Lester, 2013), sexual behavior (Markey and Markey, 2013), and cancer (Glynn et al., 2011; Ayers, 2014).

However, the potential of such data in relation to one of the most pervasive behaviors and important areas of public health, diet and nutrition, remains largely unexplored. An important exception is Carr and Dunsiger (2012), who examined intra-annual patterns in the United States in search queries suggestive of public interest in behavioral change in relation to obesity, diet, physical fitness and smoking. The results showed that searches for the terms “weight”, “diet”, “fitness”, and “smoking” peaked in January and declined through to December, leading the authors to conclude that this type of analysis holds potential for informing interventions, campaigns, and public-health policy around dietary behavior.

The purpose of this study was to extend this approach to the study of diet-related behavior, by testing whether and how relevant Google searches are able to predict actual dietary intake. We first examined long-term trends in Australian Google search behavior to establish monthly patterns in search term usage. We then compared patterns in search term usage with the total energy, food, and macronutrient intakes of the Australian population using data collected as part of a national-level survey.

Methods

Google Trends query data. Google Trends determines the proportion of searches for a given term over a specified geographic region and time period. It then provides users with a graph and optional downloadable output of relative search volume (RSV) for that term scaled from 1–100 (Nuti et al., 2014), which represent search interest relative to the peak popularity for that term, where a value of 100 indicates peak popularity and a score of 0 indicates the term was < 1% as popular as it was during its peak (Google, 2017b). Likewise, a score of 50 indicates the term

was half as popular as its peak during that time period for the specified geographic region.

We followed the recommendations of Nuti et al. (2014) to document our Google Trends search, where we use square brackets to denote the search input. We initially downloaded data for the following search terms independently: [weight]; [diet]; [fitness], and [nutrition]. The terms “weight”, “diet”, and “fitness” were identified as the most commonly searched terms in their respective subject areas (Carr and Dunsiger, 2012); however, after examining the related topics and queries for “weight”, we decided to use the search term [“weight loss”], which had associated queries and topics more relevant to our study. We likewise examined and ruled-out the search term [nutrition] for the same reason. We downloaded data for Australia from 01 January 2004 (the earliest available) to 31 December 2016 (the most current full year of data relative to our study) across “all categories”. The thirteen-year data sets were provided at a monthly resolution. We report on the top three related topics and related queries associated with each of our query terms to help contextualize the search terms in relation to public interest. We accessed and downloaded data for “diet” and “fitness” on 20 February 2017 and for “weight loss” on 22 February 2017. Screenshots of searches can be found in the Supplementary Information Figs. S1–S3.

Google Trends data analysis. We performed time series analysis of Google Trends data using R version 3.3.3 (R Core Team, 2017) to determine monthly trends in RSV. We used the function “decompose” included in the base R package {stats} to perform classical seasonal decomposition using moving averages. The decompose function produces plots with the decomposed time series partitioned into trend (inter-annual), seasonal (intra-annual, monthly), and irregular (error) components (Kendall and Stuart, 1983; R Core Team, 2017). This function first computes the trend component using moving averages and removes it from the time series. The seasonal component is then calculated by averaging for each time unit (month) overall periods (years), and centered. The error component is then calculated by removing the trend and seasonal components from the original time series. We used an additive model for the seasonal component.

We used generalized linear mixed-models (GLMMs) to test statistical differences in monthly RSV using the function “lmer” in the package {lme4} (Bates et al., 2015), where we included a random intercept for “year” to account for inter-annual variation. Month was set as a twelve-level categorical variable with January set as the reference category. We created separate models for each search term. Because RSV was expressed as a proportion ranging from 1 to 100, we logit transformed RSV (“l.RSV”) using the “logit” function in the package {car} (Fox and Weisberg, 2011) prior to GLMM analysis (Warton and Hui, 2011). After creating GLMMs, we performed post-hoc contrasts between least-squares (LS) means predictions for months using the “contrast” function in the package {lsmeans} (Lenth, 2016). We requested “eff” contrasts, which reports on differences between each factor level mean (monthly mean) and the overall mean, to infer whether monthly peaks and troughs in RSV were significantly different from the annual mean. Contrasts were adjusted for multiple inferences using the default “fdr” (a.k.a. “BH”) method (Benjamini and Hochberg, 1995). We assessed the significance of the random component (“year”) in all models using a likelihood ratio test implemented with the “rand” function in the package {lmerTest} (Kuznetsova et al., 2015).

Nutritional survey and data. To examine patterns of Google Trends RSV with population-level nutritional intake, we analyzed data from the 2011–2012 National Nutrition and Physical Activity Survey (NNPAS) performed by the Australian Bureau of Statistics (ABS), full details of which can be found in NNPAS User’s Guide (ABS, 2013). In brief, the purpose of the survey was to collect diet, nutrition, and health information of the Australian population. Dietary intake information was collected using a face-to-face 24-hour recall interview based on an Automated Multiple-pass Method. The 24-hour dietary recall interview collected detailed information on all foods and beverages consumed on the previous day, from midnight to midnight. Total energy and macronutrient intakes of participants were estimated by the NNPAS using the AUSNUT 2011–2013 database, which was developed specifically for the survey (Food Standards Australia New Zealand, 2014a).

A total sample of 12,153 persons (aged 2 years and over) from 9519 private dwellings across Australia were interviewed. The survey started in late May 2011 and lasted until the end of May 2012, with a pause in enumeration between 06 August and 02 October for administrative reasons (ABS, 2013). Demographic and other information of survey participants was collected for a variety of parameters, including age, sex, country of birth, socio-economic status, and lifestyle factors. Age was categorized over twelve levels: 2–3; 4–8; 9–13; 14–18; 19–30; 31–50; 51–70; 70+ years of age. The Socio-Economic Index of Disadvantage for Areas was used to estimate socio-economic status quintiles, where the first quintile indicates the least advantaged (ABS, 2014). Lifestyle factors, including whether or not survey participants were on a weight loss or other diet, were self-reported.

For the purpose of our study, we restricted our analysis to those aged from 15 to 65 years old based on the internet user’s statistics presented in ABS (2016), and that completed the first NNPAS 24-h interview. We excluded people that declared they did not speak English at home, because they would be unlikely to enter our English search terms. This left us with data from a total of 7744 respondents, with the

results weighted to reflect the Australian population following ABS (2013, 2014). The weighting procedure was provided by the ABS to ensure that the survey estimates conformed to the independently estimated distribution of the population by age, sex, and area of usual residence, rather than to the distribution within the sample itself (ABS, 2013). Limited NNPAS data was available for May 2011, thus was excluded from our analysis. Likewise, August and September were not included due to the aforementioned pause in enumeration. Full demographic information for respondents in our study is given in Supplementary Information Table S1.

We examined four dietary factors from the NNPAS survey in relation to Google Trends RSV: (1) average monthly total energy intake (kJ); (2) average monthly proportional macronutrient intake; (3) average monthly food group intake (grams); and (4) monthly percentage of self-reported dieters. Total energy included energy from protein, total fat, total carbohydrate, dietary fiber, alcohol, and organic acids (ABS, 2013). Macronutrients included protein, total fat, and available carbohydrate (including sugar alcohols) plus dietary fiber. For our study, we expressed the macronutrient intake of the survey respondents on a percent energy basis, where macronutrients were expressed as a percentage of total macronutrient plus fiber derived energy (100% = protein (%) + lipid (%) + (carbohydrate % + fiber %)). We estimated carbohydrate + fiber energy by subtracting protein, total fat, and alcohol energy from total energy. We assessed food intake of survey participants across ten categories based on the Australian Dietary Guidelines (Australian Government, 2015): (1) meat, poultry, and fish (including red meat, offal/organ, poultry, fish/seafood, and processed meat); (2) meat alternatives (eggs, legumes/beans, nuts/seeds); (3) non-starchy vegetables (excluding legumes and beans); (4) starchy vegetables (potato, sweet potato, taro, cassava, and corn); (5) grains and cereals (bread, cereals, rice, pasta); (6) dairy products (milk, cheese, yoghurt and alternatives); (7) fruit; (8) alcoholic drinks (beer, wine, spirits, and mixed drinks with ethanol); (9) discretionary foods (cakes, biscuits, pastries, deep-fried fast foods, crisps, ice cream, confectionary, butter, sweet spreads; and (10) discretionary beverages (soft drinks, fruit drinks, energy drinks, sports drinks). To assess consumption of all foods within a food group, all individually recorded food items and foods as part of a mixed dish were included. The individual food components from a mixed dish were estimated using the AUSNUT 2011–13 recipe file (Food Standards Australia New Zealand, 2014b). For example, the amounts of poultry and non-starchy vegetable components in a chicken stir fry were disaggregated and classified accordingly. Composite drinks, for example tea or coffee with milk, were disaggregated into dairy products (or alternatives) and drinking water.

Statistical analysis of nutritional data. Statistical analysis of nutritional data was performed in SPSS for Windows 22.0 (SPSS Inc., Chicago, IL, USA). Descriptive statistics were used to report the proportions of respondents and average intakes. We tested for changes in monthly total energy, individual macronutrient, and food group intakes using analysis of covariates procedure (ANCOVA, two-sided) with month of the year as the fixed factor. Total energy, macronutrient, and food intakes were associated with survey respondent's age, sex, and socio-economic status, and were thus adjusted for all tests. Country of birth and physical activity level were not associated with energy intake, but were previously proven to be associated with food choice (Sui et al., 2016, 2017), hence, were included as covariates in ANCOVA. We used Bonferroni post-hoc tests for multiple inferences. For all tests, a p -value of ≤ 0.05 was considered statistically significant.

We used nutritional geometry to examine and visualize patterns in monthly proportional macronutrient intake using right-angled mixture triangles (RMT). RMTs allowed us to plot the three-dimensional macronutrient plus fiber energy proportions in an x - y plot format, and have an established framework for supporting interpretation of data patterns (Raubenheimer, 2011; Raubenheimer et al., 2016). We plotted protein on the x -axis and lipid on the y -axis, with the values of the remaining proportion of carbohydrate + fiber energy being read along the diagonal implicit z -axis, which is inversely related to the origin of the plot (Raubenheimer, 2011). For example, values plotted on the hypotenuse of the RMT would have a value of 0% carbohydrate + fiber, and, conversely, values at the plot origin would have a value of 100%.

Data congruence. To assess the congruence between Google Trends RSV and the NNPAS survey data, we directly compared monthly total energy intake, and the self-reported dieting behavior of those on a "low calorie or weight loss diet", with RSV for the same time period as the NNPAS survey. We did not apply correlation statistics because of the complex nature of the NNPAS survey data. Demographic information of dieters is given in Supplemental Information Table S2.

Results

Google Trends. Associated search topics and queries given by Google Trends for each of our search terms suggested that "weight loss", "diet" and "fitness" were generally associated with interest in behavioral change associated with obesity. For "weight loss", the top three associated queries were: "diet"; "weight loss diet"; and "best weight loss". For "diet", the top three associated queries were: "diet plan"; "diet food"; and "paleo". For "fitness", the top three related queries were: "fitness first"; "anytime fitness"; and

"fitness australia". Full tables of related search topics and queries are given in the Supplementary Information Tables S3–S12.

Decomposition of RSV suggested a monthly pattern for all search terms with similar patterns observed for each ("weight loss", Fig. 1; "diet", Fig. 2; "fitness", Fig. 3) (Table 1). LS means contrasts of GLMMs (Supplementary Information Table S13) suggested that there were significantly different monthly peaks and troughs in I.RSV predictions, with January highest, and December lowest for all search terms (Table 2; Fig. 4). The term "weight loss" showed significant peaks in January, September, and October, and significant troughs in March, May, June, July, and December. The term "diet" showed significant peaks in January, August, and September, and significant troughs in December. The term "fitness" showed significant peaks in January and February, and significant troughs in May, June, November, and December. The random component ("year") was significant for all three

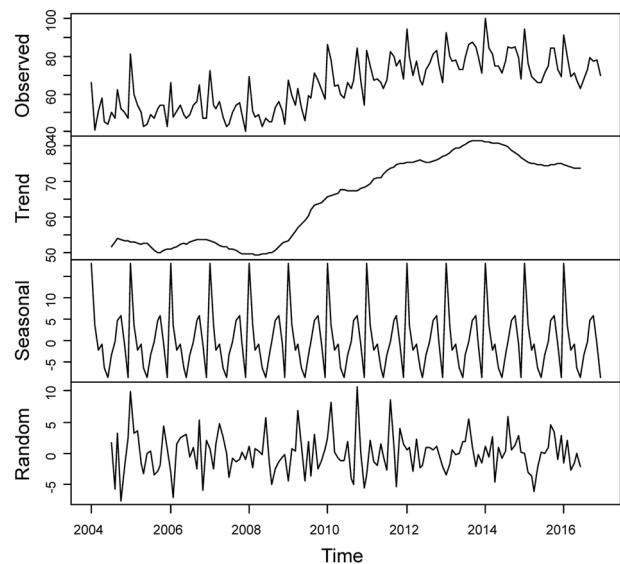


Fig. 1 Decomposed Google Trends RSV for the term "weight loss." The seasonal (i.e., monthly) component was calculated by averaging for each time unit (month) overall periods (year), and then centered

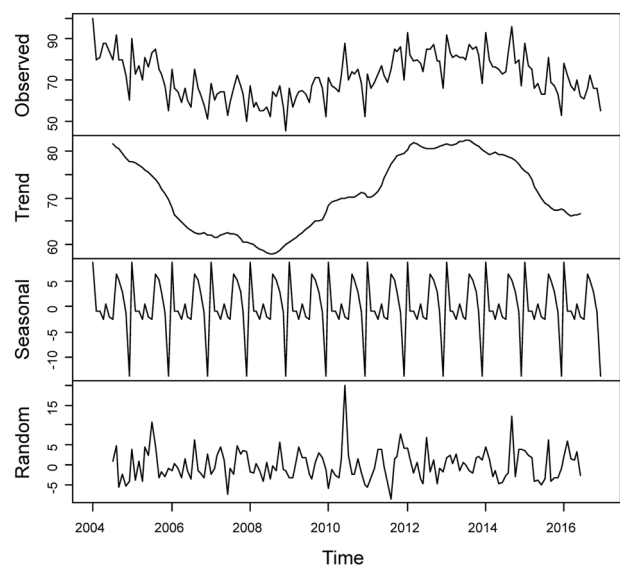


Fig. 2 Decomposed Google Trends RSV for the term "diet." The seasonal (i.e., monthly) component was calculated by averaging for each time unit (month) overall periods (year), and then centered

search terms (all $p < 2e-16$), suggesting that overall RSV differed across years while retaining monthly trends.

2011/2012 Energy intake and Google Trends RSV. We found that energy intake from June 2011 to May 2012 varied between months, and that the trends in energy intake closely matched the pattern in Google Trends RSV, especially from October through May (Table 3; Fig. 5). We focussed on testing between-month differences in total energy intake for two periods: (1) June and July 2011; and (2) November 2011 to February 2012 (Table 3). Energy intake declined significantly from June to July ($p < 0.001$; Table 3). This is consistent with the increase in RSV for “weight loss” from June to July, but inconsistent with the decrease in RSV for “diet” and “fitness”. However, the long-term trends in I.RSV for “weight loss” during June and July, and during June for “fitness”, were lower than average during this time (Tables 1–2), which is consistent with the overall relatively high-energy intake during these months.

Monthly average energy intake rose significantly from November 2011 to reach highest in point in December (Table 3). This is consistent with the trough in RSV for “weight loss” and “diet”, while RSV for “fitness” was relatively stable, yet low. The long-term monthly trend in I.RSV was significantly lower than average during December for three search terms, and the lowest of all months (Tables 1 and 2; Figs. 1, 2, 3, and 4). Energy intake decreased from December to January, yet not significantly (Table 3; Fig. 5). Energy intake in January 2012 was still significantly higher than during November 2011. January 2012 also saw the highest peak in RSV for all three terms during the survey period, which was consistent with the multi-year trend (Tables 1 and 2;

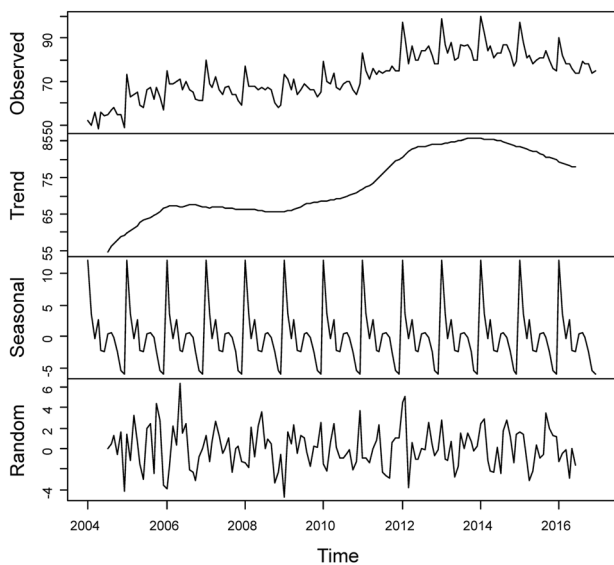


Fig. 3 Decomposed Google Trends RSV for the term “fitness.” The seasonal (i.e., monthly) component was calculated by averaging for each time unit (month) overall periods (year), and then centered

Figs. 1, 2, 3, and 4). Energy intake was lowest during February, and this was significantly lower than in December and January, but not November (Table 3; Fig. 5). RSV declined for all terms from January to February 2012, yet the RSV for “fitness” was still relatively high, which is consistent with the multi-year trend being significantly higher than average (Tables 1 and 2; Fig. 4).

Self-reported dieters and Google Trends RSV. The percentage of self-reported survey participants on a weight loss or low calorie diet varied monthly from June 2011 to May 2012, and this trend closely followed monthly patterns in RSV for our search terms, especially from December to May (Fig. 6). The number of dieters was relatively high in June and declined to July, which is inconsistent with the multi-year trend in I.RSV being lower during this time, and with the decrease in energy intake. However, the percent of dieters and RSV very closely matched the pattern of intake from December 2011 to May 2012. December saw low troughs in RSV and self-reported dieters, and both rose sharply in January 2012. Self-reported dieters reached its highest in February while search volume dropped for all terms. March saw a secondary trough in both self-reported dieters and RSV for all search terms, which is consistent with the multi-year trend in I.RSV being significantly lower for “weight loss”, while the long-term trend for the other search terms were not significantly different than average during this time. There was a secondary peak in dieting behavior in April consistent with RSV; however, energy intake was relatively stable in April and May (Fig. 5).

Monthly macronutrient intake and food consumption. The percent energy from protein was not significantly different across all months examined (Table 3; Fig. 7), while the energy from fat

Table 2 Contrast p-values of monthly GLMM estimates using least-squares (LS) means predictions in logit transformed Relative Search Volume (I.RSV) for the search terms “weight loss”, “diet”, and “fitness”

Month	“Weight loss”	“Diet”	“Fitness”
1. January	<0.0001 (+)	<0.0001 (+)	<0.0001 (+)
2. February	0.1224	0.5942	0.0049 (+)
3. March	0.0124 (-)	0.5942	0.3538
4. April	0.4339	0.2767	0.0719
5. May	<0.0001 (-)	0.5942	0.0137 (-)
6. June	<0.0001 (-)	0.2767	0.0077 (-)
7. July	0.0016 (-)	0.0949	0.9337
8. August	0.9745	0.0003 (+)	0.3538
9. September	<0.0001 (+)	0.0008 (+)	0.6912
10. October	<0.0001 (+)	0.2744	0.0761
11. November	0.9745	0.2767	<0.0001 (-)
12. December	<0.0001 (-)	<0.0001 (-)	<0.0001 (-)

Contrasts examined the differences between each monthly mean compared to the overall mean, where significance was inferred at $\alpha \leq 0.05$ (in bold). Plus (+) indicates significantly > the mean, and minus (-) indicates significantly < the mean. Contrasts were adjusted for multiple inferences using the “BH” method (Benjamini and Hochberg, 1995)

Table 1 Decomposed seasonal (monthly) trend in Google Trends Relative Search Volume (RSV) for the search term queries “weight loss”, “diet”, and “fitness”

Term	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Weight loss	17.9	3.6	-2.3	-1.0	-6.5	-8.7	-3.4	-0.3	4.7	5.8	-1.0	-8.7
Diet	8.6	-0.8	-0.8	-2.4	0.5	-2.0	-2.4	6.3	5.2	2.8	-1.1	-13.7
Fitness	11.9	3.6	-0.4	2.6	-2.1	-2.4	0.3	0.5	-0.2	-2.4	-5.4	-6.0

The monthly component was calculated by averaging for each time unit (month) overall periods (year), and then centered

and carbohydrate + fiber varied depending on the month. From June to July, the percent energy derived from lipid was not significantly different; however, the proportion of energy from carbohydrate + fiber was significantly higher in July (Table 3; Fig. 7a). Regarding food categories, the amount of alcoholic drinks, discretionary beverages, and dairy products all increased in July relative to June, while the amount of meat alternatives decreased (Table 3).

Despite the significant difference in total energy intake, the proportion of macronutrients consumed between November and December was not significantly different (Table 3; Fig. 7b). The amount of alcohol consumed increased significantly from November to December, the energy from which was not included in the proportional analysis of macronutrient intake. Other foods that increased from November to December include: meat/

poultry/fish; starchy vegetables; dairy products; discretionary foods; and discretionary beverages. Meat alternatives decreased from November to December. There was significantly less fat and more carbohydrate + fiber consumed in January than during November and December (Table 3; Fig. 7b). Alcoholic drinks and discretionary foods were consumed less in January than both November and December, while grains and cereals were consumed more. There were significantly less starchy vegetables and dairy consumed in January than December. The proportion of fat consumed was significantly less in February than in November and December, and the proportion of carbohydrate + fiber consumed was less than in January (Table 3; Fig. 7b). The proportion of protein consumed was highest in February, although not statistically different. Consumption of meat/poultry/fish and alcoholic drinks was significantly lower, and meat alternatives significantly higher, in February than the three other months.

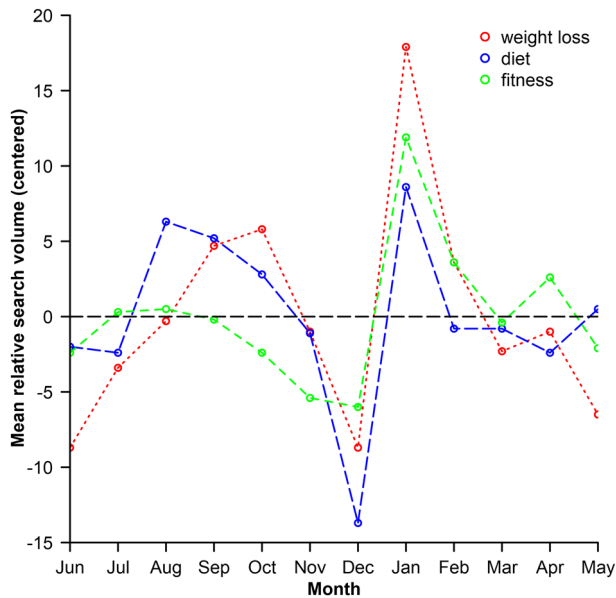


Fig. 4 The decomposed monthly trends (2004–2016) in RSV for the three search terms “weight loss”, “diet”, and “fitness”

Discussion

In this paper, we examined the digital mapping of online behavior with dietary habits to show that monthly patterns in public interest in behavior change as measured using search query data can be associated with population-level patterns in energy, macronutrient, and food intake. By demonstrating congruence between these two disparate sources of data, each of which contains nominally independent information, we show that food and nutritional intake joins the growing list of public behaviors assessable through analysis of Google Trends data.

Google Trends studies have been used to examine a variety of health care topics and inform associated public-health strategies (Nutti et al., 2014). For example, such studies have been particularly successful in correlating search query data with disease surveillance data (Pelat et al., 2009), such as for influenza (Kang et al., 2013), norovirus (Desai et al., 2012), and chickenpox (Valdivia and Monge-Corrella, 2010), and have ultimately improved the monitoring and prediction of disease outbreaks (Zhou et al., 2011). Thus, an implication of our study is that freely available search query data can be added to the list of approaches used to tackle the complex issue of obesity by informing public-health nutrition and dieting strategies.

Table 3 Summary of mean monthly energy intakes including alcohol (kJ), percent macronutrient (protein, lipid, and carbohydrate + fiber) energy, and food group intake (g) of survey participants of the 2011–2012 NNPAS survey for the months of June and July 2011, and November 2011 to February 2012

	Jun	Jul	p-value	Nov	Dec	Jan	Feb
Energy (kJ)	9018.2	8870.5	<0.001	8743.6	9065.8 ^a	8927.7 ^a	8599.9 ^{b,c}
Macronutrients (% kJ)							
Protein	19.2	19.1		19.2	19.3	19.3	19.5
Fat	32.6	32.0		33.0	33.0	32.1 ^{a,b}	32.6 ^{a,b}
Carbohydrate + fiber	48.2	48.9	0.01	47.8	47.7	48.6 ^{a,b}	47.9 ^c
Food intake (g)							
Meat/poultry/fish	158.7	155.0		157.0	163.2 ^a	162.2 ^a	151.0 ^{a,b,c}
Meat alternatives	41.2	34.7	<0.001	32.7	30.0 ^a	31.2 ^a	35.1 ^{a,b,c}
Non-starchy vegetables	171.1	167.9		158.7	162.7	164.9 ^a	165.3 ^a
Starchy vegetables	49.3	52.8		42.2	49.8 ^a	44.9 ^b	40.7 ^{b,c}
Grains/cereals	245.5	246.2		220.5	216.3	230.3 ^{a,b}	212.8 ^c
Dairy products	270.6	297.7	<0.001	262.5	278.8 ^a	266.4 ^b	285.8 ^{a,c}
Fruit	184.8	180.8		209.0	204.6	223.5	220.4
Alcoholic drinks	236.0	261.7	<0.001	298.4	311.4 ^a	276.3 ^{a,b}	252.1 ^{a,b,c}
Discretionary foods	201.0	193.3		194.5	209.3 ^a	187.1 ^{a,b}	182.7 ^{a,b}
Discretionary beverages	253.3	276.2	<0.001	280.1	370.4 ^a	313.3 ^{a,b}	297.3 ^{a,b,c}

Statistical tests for Jun–Jul and Nov–Feb were performed separately. Only significant p-values are given
^aSignificant difference (p < 0.01) from November 2011 by Bonferroni post-hoc test
^bSignificant difference (p < 0.01) from December 2011 by Bonferroni post-hoc test
^cSignificant difference (p < 0.01) from January 2012 by Bonferroni post-hoc test

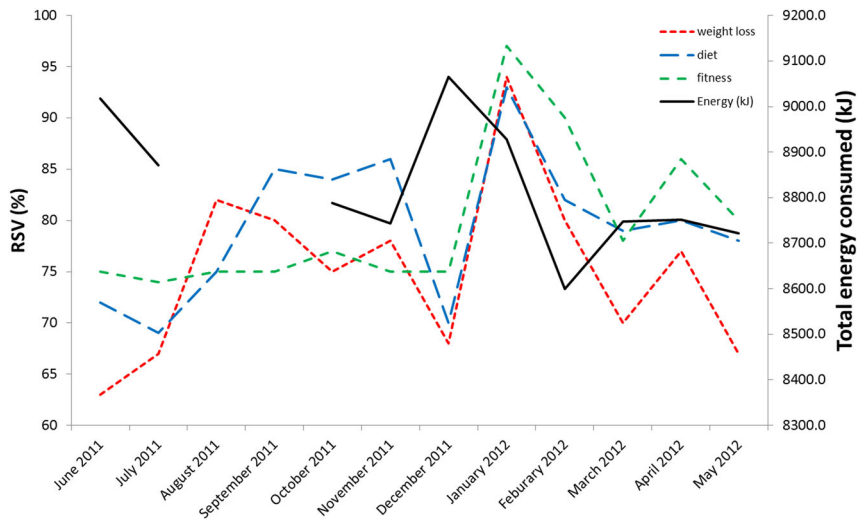


Fig. 5 Monthly total energy intake (kJ) of 2011–2012 NNPAS survey participants (secondary y-axis) plotted alongside Google Trends RSV in Australia (primary y-axis) during the same period. Note that NNPAS survey data collection was paused during August and September 2011 and thus was not included in our analysis

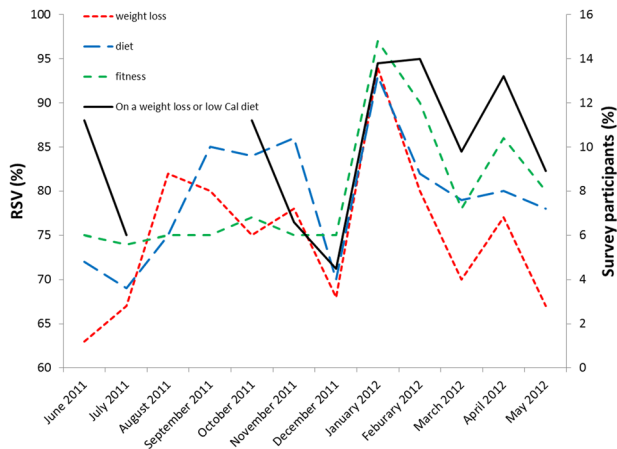


Fig. 6 The monthly percentage of 2011–2012 NNPAS survey participants self-reported as being on a weight loss or low-calorie diet (secondary y-axis) plotted alongside Google Trends RSV for the terms “weight loss”, “diet”, and “fitness” over the same period (primary y-axis)

One application could be to use RSV to ascertain the effectiveness of nutrition, fitness, and obesity campaigns. Surveillance studies, for instance, have been used to determine the real-time effectiveness of public-health campaigns, such as the effectiveness of annual breast cancer campaigns (Glynn et al., 2011), the influence of media coverage on public and professional online information needs (Kostkova et al., 2013), and to gauge public interest in a particular treatment, campaign, or policy (Liu et al., 2012; Davis et al., 2012; Harsha et al., 2014). Another useful application of Google Trends data might be to track popular diet trends (e.g., “paleo” was identified in our related search terms) to help to inform public nutritional patterns and anticipate their possible implications. While our study focused at the national level, identifying spatial patterns within a country is another potential application of Google trends data, such as have been identified for HIV incidence (Jena et al., 2013), kidney stones (Willard and Nguyen, 2013), and sexually transmitted infections (Johnson and Mehta, 2014).

Nutritional search terms might also help identify what information might be lacking for those seeking to lose weight.

Research suggests that peoples’ willingness to lose weight is strong (Timperio et al., 2000), and findings from an English health survey suggested that most adults had an understanding of what comprised a healthy diet, and eating less at mealtimes was the most common strategy of weight control (Roberts and Marvin, 2011). In that survey, eating sufficient vegetables and fruit was the most frequently cited component of a healthy diet, followed by limiting fat intake, which is consistent with what our study showed for Australians during January and February. However, lack of knowledge around the health consequences of poor weight management, healthy eating, and safe exercise have been identified as barriers to weight management in other studies, such as a study of pregnant women who paradoxically had access to freely provided weight loss information (Sui et al., 2013; Dodd et al., 2014), suggesting a possible mismatch between what they required and what was supplied.

Another challenge in combatting obesity is that weight management intervention studies and public-health champions frequently report low compliance (Sui et al., 2012; Kristensen et al., 2017). Even for those that have started weight management, long-term maintenance can be poor (de Vos et al., 2016). Our results may partially explain the low compliance with interventions and public-health messages by further demonstrating that people’s interest in weight management and dietary behavior is not consistent. Thus, public-health interventions might explore ways in which to market campaigns in keeping with public interest, for example, by supplying the public with large amounts of accurate and accessible information during the months in which interest is high, and providing motivation and healthy behavior maintenance strategies when interest is low.

In December, for instance, the low RSV, high-energy intake, patterns of food consumption, and low percentage of self-reported dieters is consistent with the Australian holiday season, which being in the Southern Hemisphere includes both Christmas and summer holidays. The spike in RSV, decrease in energy intake, and high percentage of dieters in January is consistent with the post-holiday New Year’s resolution period (Kassirer and Angell, 1998). January, however, also includes the “Australia Day” holiday (26 January), which may have contributed to the overall high monthly energy intake. Energy intake was lowest in February and percentage of dieters was highest, while RSV declined, suggesting that people had obtained their online dieting

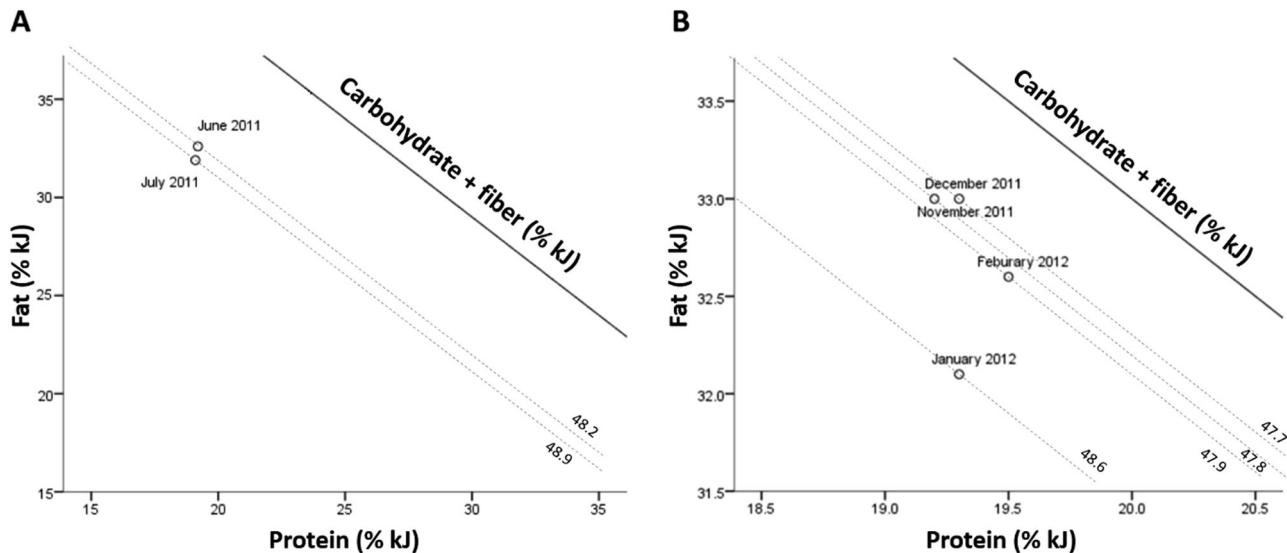


Fig. 7 Right-angled mixture triangle (RMT) showing average monthly percent energy derived from protein, fat, and carbohydrates + fiber of 2011–2012 NNPAS survey participants for: **a** June and July 2011; and **b** November 2011 to February 2012. The RMT has been zoomed-in on the data points to aid interpretation. The percent energy from carbohydrate + fiber is shown on the implicit z-axes, which are depicted as negatively sloped isolines. Any point along a particular z-axis isoline has the same value of carbohydrate + fiber energy, with co-varying amounts of protein and fat

information in the previous month and had been fully on their diet in February.

That RSV declined, and energy intake increased, following February, is consistent with failing New Year's resolutions (Kassirer and Angell, 1998). However, in contrast to the United States (Carr and Dunsiger, 2012), the pattern of RSV did not decline steadily through the year; there were spikes in RSV in April 2012, and from August to October 2011. Unfortunately, the missing August and September survey data precluded assessing energy intake or self-reported dieting during this time. The secondary peaks in RSV during Austral late-winter/spring (i.e., August to October) may be related to the reversal of seasons between the hemispheres, and the prevalent Australian “beach culture” (James, 2000; Edwards et al., 2003), where public interest in weight loss and dieting may be indicative of people trying to “get in shape” for the summer season.

There have been a number of studies considering seasonal variation in energy and macronutrient intake, with inconsistent results. For example, Bernstein et al. (2016) found that intake of energy, macronutrients, and food groups by healthy adults in metropolitan Washington D.C. did not vary between seasons. Yet, seasonal differences in protein (Mori et al., 1981), lipid (Mori et al., 1981; Owaki et al., 1996), and dietary fiber (Owaki et al., 1996; Tokudome et al., 2002) intake were observed in Japanese studies. Similarly, energy and protein tended to be higher in the autumn and winter in those over 65 in Great Britain (Doyle et al., 1999); daily energy, carbohydrate, and lipid intake varied seasonally in an American study (Ma et al., 2006); lipid intake increased in winter in a study of male Israeli factory workers (Shahar et al., 1999); and summer carbohydrate intake was higher, and fat intake lower, in a Brazilian study (Rossato et al., 2010).

Our study found that in Australia, total energy, lipid, and carbohydrate + fiber changed monthly, yet protein remained relatively stable. The consistent protein intake is in keeping with evidence suggesting that humans more strongly regulate their intake of protein over non-protein macronutrients (Simpson and Raubenheimer, 2005; Martinez-Cordero et al., 2012; Gosby et al., 2014; Raubenheimer et al., 2016). A variety of animals have been shown to have separate appetite systems for protein, fat, and

carbohydrates (Raubenheimer and Simpson, 1997; Berthoud and Seeley, 2000). For those animals, such as humans, in which protein is more strongly regulated than the other macronutrients, low-percent protein diets can drive excess energy intake through increased consumption of surplus non-protein energy to gain limiting protein (the “protein leverage hypothesis”; Simpson and Raubenheimer, 2005).

Changes in energy intake from November to December, in which the proportion of macronutrients were not significantly different, appeared to be driven by increased food and alcohol intake. In January and February, the population on average adjusted the types of foods consumed and reduced fat intake in keeping with weight loss strategies. Although not significant, the proportion of protein in the diet during February was highest, which is in keeping with popular high-protein diets and reduced energy intakes (Raubenheimer et al., 2015). The increase in proportional carbohydrate plus fiber energy may be a passive result of decreasing fat intake while holding protein stable. The types of carbohydrates and fats consumed also changed, for example, there were less starchy vegetables, discretionary foods, grains and cereals and meat consumed, and more meat alternatives and dairy, during February which had the highest percentage of self-reported dieters.

While the patterns of RSV and survey data were in general agreement around and after the summer holiday season, the patterns in June and July were somewhat harder to interpret due partly to the missing monthly data before and after the period. The spike in energy intake in June 2011 was consistent with low RSV for our search terms; however, there was a discrepancy with self-reported dieters, which remained high during this time. The diets of survey participants, however, seemed to become poorer in July despite decreased energy intake, as evidenced by the increase in alcohol and discretionary beverages, and decrease in meat substitutes consumed. Dairy may have increased due to increased consumption of hot beverages, to which milk is commonly added (e.g., Dupas et al., 2006), during the cooler months. Depression and boredom have been cited as factors contributing to emotional triggers to eating (Roberts and Marvin, 2011). Internet search query data has been used to identify trends in suicide and depression (McCarthy, 2010; Gunn and Lester, 2013). In

Australia, seasonal affective disorder during the winter/autumn months has been characterized by increased appetite and carbohydrate cravings (Boyce and Parker, 1988). This may be consistent with our data showing the relatively high-proportion of carbohydrate + fiber energy in the diet, and relatively high total energy consumed, in June and July overall.

One limitation of our study was the lack of correlation statistics due to the complex nature of the NNPAS survey data. Another obvious limitation of the survey data was the missing data from August to September. Furthermore, although the respondent sample of the NNPAS survey was weighted to reflect the whole population, we cannot be sure the people reported on a low calorie or weight loss diet (64.1% women; Supplementary Information Table S2) were representative of the dieting population. While we were only able to investigate one incomplete year of survey data for validation purposes, the repeating trends observed using Google Trends suggests that there is a repeating pattern in population dietary behavior. We recognize that all weight loss search terms possibly entered by the public might not be captured using only a few terms; however, Pelat et al. (2009) found that one well-chosen search term related to infectious diseases was enough to produce a time series highly correlated with incidence. Thus, effective Google Trend analysis requires careful search term selection (Nuti et al., 2014).

The limitations of using Google Trends have been well documented (e.g., Nuti et al., 2014), including, for example, not being able to ascertain user characteristics, capturing only the behavior of those using the Google Search engine, and the lack of information on the methodology employed by Google to generate and analyze search data. According to one source, however, Google's search engine accounted for 92.8% of the Australian market share in March 2015 (payerclick, 2015), which suggests that the use of Google Trends was a suitable source of search query data for Australia. Yet, despite the limitations there are several advantages to using online search query data (Ginsberg et al., 2009; Nuti et al., 2014), which is a freely accessible tool that provides high volume population-level behavioral data that can be used to generate meaningful insights and public-health strategies.

Conclusions

The key insight of this paper is that spatial and temporal patterns in public interest in behavior change associated with obesity can be observed using online search query data, and that such data can be related to nutritional intake and dieting behavior at the national level. Furthermore, we documented distinct seasonal patterns in food, energy, and macronutrient intake in the Australian population. Novel and innovative approaches to combatting obesity and other nutritional issues in a timely and spatially appropriate manner can likely be developed through incorporating such technology. Our study, therefore, has implications for integrating web-based data within future national-level surveys.

Received: 13 May 2017 Accepted: 27 November 2017

Published online: 09 January 2018

References

- Australian Bureau of Statistics (2013) Australian health survey: user's guide, 2011-2013. Australian Government Publishing Service, Australian Bureau of Statistics, Canberra, Australia. <http://www.abs.gov.au/AUSSTATS/abs@.nsf/Lookup/4363.0.55.001Main+Features12011-13?OpenDocument>. Accessed Mar 2017
- Australian Bureau of Statistics (2014) Australian health survey: nutrition first results-foods and nutrients, 2011-2012. Australia Bureau of Statistics, Canberra, Australia
- Australian Bureau of Statistics (2016) 8146.0—Household use of information technology, Australia, 2014-15. <http://www.abs.gov.au/ausstats/abs@.nsf/mf/8146.0#>. Accessed Feb 2017
- Australian Government (2015) Australian dietary guidelines. <https://www.eatforhealth.gov.au/>. Accessed Mar 2017
- Ayers JW, Ribisl KM, Brownstein JS (2011a) Tracking the rise in popularity of electronic nicotine delivery systems (electronic cigarettes) using search query surveillance. *Am J Prev Med* 40:448-453
- Ayers JW, Ribisl K, Brownstein JS (2011b) Using search query surveillance to monitor tax avoidance and smoking cessation following the United States' 2009 "SCHIP" cigarette tax increase. *PLoS ONE* 6:e16777
- Ayers JW, Althouse BM, Allem JP et al. (2012) A novel evaluation of world no tobacco day in Latin America. *J Med Internet Res* 14:288-298
- Ayers JW, Althouse BM, Noar SM et al. (2014) Do celebrity cancer diagnoses promote primary cancer prevention? *Prev Med* 58:81-84
- Bates D, Maechler M, Bolker B et al. (2015) Fitting linear mixed-effects models using lme4. *J Stat Softw* 67:1-48
- Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc Ser B* 57:289-300
- Bernstein S, Zambell K, Amar MJ et al. (2016) Dietary intake patterns are consistent across seasons in a cohort of healthy adults in a metropolitan population. *J Acad Nutr Diet* 116:38-45
- Berthoud H, Seeley R (2000) Neural and metabolic control of macronutrient intake. CRC Press, Boca Raton, USA
- Boyce P, Parker G (1988) Seasonal affective disorder in the Southern Hemisphere. *Am J Psychiatry* 145:96-99
- Brownstein JS, Freifeld CC, Madoff LC (2009) Digital disease detection — harnessing the Web for public health surveillance. *New Engl J Med* 360:2153-2157
- Carr LJ, Dunsiger SI (2012) Search query data to monitor interest in behavior change: application for public health. *PLoS ONE* 7:e48158
- Davis NF, Smyth LG, Flood HD (2012) Detecting internet activity for erectile dysfunction using search engine query data in the Republic of Ireland. *Bju Int* 110:E939-E942
- de Vos BC, Runhaar J, van Middelkoop M et al. (2016) Long-term effects of a randomized, controlled, tailor-made weight-loss intervention in primary care on the health and lifestyle of overweight and obese women. *Am J Clin Nutr* 104:33-40
- Desai R, Hall AJ, Lopman BA et al. (2012) Norovirus disease surveillance using Google internet query share data. *Clin Infect Dis* 55:E75-E78
- Dodd JM, Cramp C, Sui Z et al. (2014) The effects of antenatal dietary and lifestyle advice for women who are overweight or obese on maternal diet and physical activity: the LIMIT randomized trial. *BMC Med* 12:161
- Doyle W, Crawley H, Robert H et al. (1999) Iron deficiency in older people: interactions between food and nutrient intakes with biochemical measures of iron; further analysis of the National Diet and Nutrition Survey of people aged 65 years and over. *Eur J Clin Nutr* 53:552-559
- Dupas CJ, Marsset-Baglieri AC, Ordonaud CS et al. (2006) Coffee antioxidant properties: effects of milk addition and processing conditions. *J Food Sci* 71: S253-S258
- Edwards A, Skinner J, Gilbert K (2003) Some like it hot: the beach as a cultural dimension. vol 3. Meyer & Meyer Verlag, Oxford, UK
- Fenichel EP, Kuminoff NV, Chowell G (2013) Skip the trip: air travelers' behavioral responses to pandemic influenza. *PLoS ONE* 8:e58249
- Food Standards Australia New Zealand (2014a) Australian food, supplement and nutrient database (Ausnut). Food Standards Australia New Zealand, Canberra, Australia
- Food Standards Australia New Zealand (2014b) Australian Food Supplement and Nutrient Database: Food Recipe File 2011/13 [24/05/2016]. <http://www.foodstandards.gov.au/science/monitoringnutrients/ausnut/ausnutdatafiles/Pages/foodrecipe.aspx>. Accessed March 2017
- Fox J, Weisberg S (2011) An {R} companion to applied regression, second edition. Sage, Thousand Oaks, USA
- Ginsberg J, Mohebbi MH, Patel RS et al. (2009) Detecting influenza epidemics using search engine query data. *Nature* 457:1012-1014
- Glynn RW, Kelly JC, Coffey N et al. (2011) The effect of breast cancer awareness month on internet search activity - a comparison with awareness campaigns for lung and prostate cancer. *BMC Cancer* 11:442
- Google Trends (2017a) Google Trends. <https://www.google.com.au/trends/>. Accessed Mar 2017
- Google Trends (2017b) Welcome to the Google Trends help center. <https://support.google.com/trends/#topic>. Accessed Mar 2017
- Gosby AK, Conigrave AD, Raubenheimer D et al. (2014) Protein leverage and energy intake. *Obes Rev* 15:183-191
- Gunn JF, Lester D (2013) Using google searches on the internet to monitor suicidal behavior. *J Affect Disord* 148:411-412
- Harsha AK, Schmitt JE, Stavropoulos SW (2014) Know your market: use of online query tools to quantify trends in patient information-seeking behavior for varicose vein treatment. *J Vasc Interv Radiol* 25:53-57

- James RJ (2000) From beaches to beach environments: linking the ecology, human-use and management of beaches in Australia. *Ocean Coast Manag* 43:495–514
- James PT, Leach R, Kalamara E et al. (2001) The worldwide obesity epidemic. *Obes Res* 9:228S–233S
- Jena AB, Karaca-Mandic P, Weaver L et al. (2013) Predicting new diagnoses of HIV infection using internet search engine data. *Clin Infect Dis* 56:1352–1353
- Johnson AK, Mehta SD (2014) A comparison of internet search trends and sexually transmitted infection rates using Google trends. *Sex Transm Dis* 41:61–63
- Kang M, Zhong HJ, He JF et al. (2013) Using Google Trends for influenza surveillance in South China. *PLoS ONE* 8:e55205
- Kassirer JP, Angell M (1998) Losing weight — an ill-fated New Year's resolution. *N Engl J Med* 338:52–54
- Kendall M, Stuart A (1983) *The advanced theory of statistics*, vol 3. Griffin, London, UK
- Kostkova P, Fowler D, Wiseman S et al. (2013) Major infection events over 5 years: how is media coverage influencing online information needs of health care professionals and the public? *J Med Internet Res* 15:167–190
- Kristensen M, Pelletier X, Ross AB et al. (2017) A high rate of non-compliance confounds the study of whole grains and weight maintenance in a randomised intervention trial — the case for greater use of dietary biomarkers in nutrition intervention studies. *Nutrients* 9:55
- Kuznetsova A, Brockhoff PB, Christensen RHB (2015) lmerTest: tests in linear mixed effects models. R package version 2.0-33. <https://CRAN.R-project.org/package=lmerTest>
- Lenth RV (2016) Least-squares means: the R package lsmeans. *J Stat Softw* 69:1–33
- Liu R, Garcia PS, Fleisher LA (2012) Interest in anesthesia as reflected by keyword searches using common search engines. *J Anesth Clin Res* 3:1000187
- Ma Y, Olendzki BC, Li W et al. (2006) Seasonal variation in food intake, physical activity, and body weight in a predominantly overweight population. *Eur J Clin Nutr* 60:519–528
- Markey PM, Markey CN (2013) Seasonal variation in internet keyword searches: a proxy assessment of sex mating behaviors. *Arch Sex Behav* 42:515–521
- Martinez-Cordero C, Kuzawa CW, Sloboda DM et al. (2012) Testing the Protein Leverage Hypothesis in a free-living human population. *Appetite* 59:312–315
- McCarthy MJ (2010) Internet monitoring of suicide risk in the population. *J Affect Disord* 122:277–279
- Mori S, Saito K, Wakasa Y (1981) Studies on annual fluctuation of food intake in female college students. *Jpn J Nutr Diet* 39:243–257
- National Health and Medical Research Council (2013) *Eat for health - Providing the scientific evidence for healthier Australian diets*. National Health and Medical Research Council, Canberra, Australia
- Ng M, Fleming T, Robinson M et al. (2014) Global, regional, and national prevalence of overweight and obesity in children and adults during 1980–2013: a systematic analysis for the Global Burden of Disease Study 2013. *Lancet* 384:766–781
- Nuti SV, Wayda B, Ranasinghe I et al. (2014) The use of Google Trends in health care research: a systematic review. *PLoS ONE* 9:e109583
- Owaki A, Takatsuka N, Kawakami N et al. (1996) Seasonal variations of nutrient intake assessed by 24 h recall method. *Jpn J Nutr Diet* 54:11–18
- paperclick (2015) Australian Internet Search Activities. <http://www.paperclick.net.au/australian-internet-statistics-2015/>. Accessed Mar 2017
- Pelat C, Turbelin C, Bar-Hen A et al. (2009) More disease tracked using Google Trends. *Emerg Infect Dis* 15:1327–1328
- R Core Team (2017) *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria, <http://www.R-project.org/>
- Raubenheimer D (2011) Toward a quantitative nutritional ecology: the right-angled mixture triangle *Ecol Monogr* 81(3):407–427
- Raubenheimer D, Simpson SJ (1997) Integrative models of nutrient balancing: application to insects and vertebrates. *Nutr Res Rev* 10:151–179
- Raubenheimer D, Gosby AK, Simpson SJ (2015) Integrating nutrients, foods, diets, and appetites with obesity and cardiometabolic health. *Obesity* 23:1741–1742
- Raubenheimer D, Simpson SJ, Le Couteur DG et al. (2016) Nutritional ecology and the evolution of aging. *Exp Gerontol* 86:50–61
- Roberts K, Marvin K (2011) Knowledge and attitudes towards healthy eating and physical activity: what the data tells us. National Obesity Observatory, Oxford, UK. http://www.noo.org.uk/uploads/doc/vid_11171_Attitudes.pdf. Accessed Mar 2017
- Rossato SL, Olinto MT, Henn RL et al. (2010) Seasonal effect on nutrient intake in adults living in Southern Brazil. *Cad Saude Publica* 26:2177–2187
- Shahar D, Froom P, Harari G et al. (1999) Changes in dietary intake account for seasonal changes in cardiovascular disease risk factors. *Eur J Clin Nutr* 53:395–400
- Simpson SJ, Raubenheimer D (2005) Obesity: the protein leverage hypothesis. *Obes Rev* 6:133–142
- Stevens GA, Singh GM, Lu Y et al. (2012) National, regional, and global trends in adult overweight and obesity prevalences. *Popul Health Metr* 10:22
- Subar AF, Frey CM, Harlan LC et al. (1994) Differences in reported food frequency by season of questionnaire administration: The 1987 National Health Interview Survey. *Epidemiology* 5:226–233
- Sui Z, Grivell RM, Dodd JM (2012) Antenatal exercise to improve outcomes in overweight or obese women: A systematic review. *Acta Obstet Gynecol Scand* 91:538–545
- Sui Z, Turnbull D, Dodd J (2013) Enablers of and barriers to making healthy change during pregnancy in overweight and obese women. *Australas Med J* 6:565–577
- Sui Z, Wong WK, Louie JC et al. (2017) Discretionary food and beverage consumption and its association with demographic characteristics, weight status, and fruit and vegetable intakes in Australian adults. *Public Health Nutr* 20:274–281
- Sui Z, Zheng M, Zhang M et al. (2016) Water and Beverage Consumption: Analysis of the Australian 2011–2012 National Nutrition and Physical Activity Survey. *Nutrients* 8:678
- Swinburn BA, Sacks G, Hall KD et al. (2011) The global obesity pandemic: shaped by global drivers and local environments. *Lancet* 378:804–814
- Timperio A, Cameron-Smith D, Burns C et al. (2000) The public's response to the obesity epidemic in Australia: weight concerns and weight control practices of men and women. *Public Health Nutr* 3:417–424
- Tokudome Y, Imaeda N, Nagaya T et al. (2002) Daily, weekly, seasonal, within- and between-individual variation in nutrient intake according to four season consecutive 7 day weighed diet records in Japanese female dietitians. *J Epidemiol* 12:85–92
- Valdivia A, Monge-Corrella S (2010) Diseases tracked by using Google Trends, Spain. *Emerg Infect Dis* 16:168
- Warton DI, Hui FKC (2011) The arcsine is asinine: the analysis of proportions in ecology. *Ecology* 92:3–10
- Willard SD, Nguyen MM (2013) Internet search trends analysis tools can provide real-time data on kidney stone disease in the United States. *Urology* 81:37–42
- Zhou XC, Ye JP, Feng YJ (2011) Tuberculosis surveillance by analyzing Google Trends. *IEEE Trans Biomed Eng* 58:2247–2254

Data availability

Google Trends data are freely available online (<https://trends.google.ca/trends/>). The specific Google Trends data sets used in our analysis can be obtained from the corresponding author upon reasonable request. Data from the 2011–2012 National Nutrition and Physical Activity Survey (NNPAS) are available upon request from the Australian Bureau of Statistics (ABS; <http://www.abs.gov.au/websitedbs/D3310114.nsf/home/Expected+and+available+Microdata>).

Additional information

Supplementary information: accompanies this paper at <https://doi.org/10.1057/s41599-017-0055-7>.

Competing interests: The authors declare no competing financial interests.

Reprints and permission information is available online at <http://www.nature.com/reprints>

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2018