UNDER EMBARGO UNTIL JANUARY 7, 2014, 12:01 AM ET

Population Health Concerns During the United States' Great Recession

Benjamin M. Althouse, PhD, ScM, Jon-Patrick Allem, MA, Matthew A. Childers, PhD, MPP, Mark Dredze, PhD, John W. Ayers, PhD, MA

Background: Associations between economic conditions and health are usually derived from cost-intensive surveys that are intermittently collected with nonspecific measures (i.e., self-rated health).

Purpose: This study identified how precise health concerns changed during the U.S. Great Recession analyzing Google search queries to identify the concern by the query content and their prevalence by the query volume.

Methods: Excess health concerns were estimated during the Great Recession (December 2008 through 2011) by comparing the cumulative difference between observed and expected (based on linear projections from pre-existing trends) query volume for hundreds of individual terms. As performed in 2013, the 100 queries with the greatest excess were ranked and then clustered into themes based on query content.

Results: The specific queries with the greatest relative excess were *stomach ulcer symptoms* and *headache symptoms*, respectively, 228% (95% CI=35, 363) and 193% (95% CI=60, 275) greater than expected. Queries typically involved symptomology (i.e., *gas symptoms*) and diagnostics (i.e., *heart monitor*) naturally coalescing into themes. Among top themes, headache queries were 41% (95% CI=3, 148); hernia 37% (95% CI=16, 142); chest pain 35% (95% CI=6, 313); and arrhythmia 32% (95% CI=3, 149) greater than expected. Pain was common with back, gastric, joint, and tooth foci, with the latter 19% (95% CI=4, 46) higher. Among just the top 100, there were roughly 205 million excess health concern queries during the Great Recession.

Conclusions: Google queries indicate that the Great Recession coincided with substantial increases in health concerns, hinting at how population health specifically changed during that time.
(Am J Prev Med 2014;46(2):166–170)

Introduction

Studies have linked recessions with poorer health using surveys,[1–3] but rising costs[4] mean fewer questions are included with preference for nonspecific items such as self-rated health.[2,5] Moreover, the time associated with collecting and developing data sharing protocols means data sets are not publicly available for years.[6] Given these limitations, to understand the potential health effects of the December 2008 through 2011 U.S. Great Recession, investigators turned to expert opinion,[7,8] community studies,[9,10] or convenient samples.[11]

Web data are emerging to fill these surveillance gaps.[12–17] For instance, it was previously found that in the U.S., Google queries for psychological distress increased 16% (95% CI=9, 24) for each 1% increase in home foreclosures the prior month.[18] Herein, that approach is expanded by assessing the hypothesis that health concerns, primarily of psychosocial etiology, also increased during the Great Recession.[2] By monitoring hundreds of systematically selected Google query trends, this novel approach takes the traditional self-rated health questionnaire to the next level by identifying precise health concerns by the query content and their prevalence by the query volume.

From the Santa Fe Institute (Althouse), Santa Fe, New Mexico; Keck School of Medicine (Allem), University of Southern California, Los Angeles; Graduate School of Public Health (Ayers), San Diego State University, San Diego, California; School of Public and International Affairs (Childers), University of Georgia, Athens, Georgia; and the Human Language Technology Center of Excellence (Dredze), Johns Hopkins University Baltimore, Maryland

Address correspondence to: John W. Ayers, PhD, MA, Graduate School of Public Health, San Diego State University, 9245 Sky Park Court, Suite 230, San Diego CA 92123. Email: ayers.john.w@gmail.com.

0749-3797/$36.00
http://dx.doi.org/10.1016/j.amepre.2013.10.008

Methods

Search trends in the U.S. were downloaded from Google Trends (google.com/trends), Google's public database. Trends automatically returns a weekly relative search volume (RSV) time series, reflecting the proportion for queries inputted by the investigator relative to all queries each week, then normalized on a 0 to 100 scale to the highest observed search proportion (e.g., RSV=50 is 50% of the highest search proportion). This corrects for increases in absolute searches over time due to changes in Internet access or disposable time.

Root terms were used for the most common health concerns of psychosomatic origin[19] including with economic etiology[20–24]: *chest*, *headache*, *heart*, *pain*, and *stomach*. Additional terms were added using Trends' related-terms utility, which identifies associated terms by either content (e.g., contains similar language) or users search behavior (e.g., a search session for *headache* might involve *anxiety*). The 50 most-related terms for the five roots and the 10 most-related terms for these queries were considered, yielding 2755 candidate queries. Duplicates and unrelated (i.e., *tool chest*) or unclear/nonspecific (i.e., *symptoms*) terms were removed, yielding 1044 terms. Queries that had more than 5% of weeks with very low volume (RSV < 1) were purged, yielding 343 terms.

The Great Recession was defined as December 2008 through 2011.[25] The cumulative weekly difference between weekly observed RSV and counterfactual RSV (derived from a projection of the best-fitting line from January 2006 through November 2008) was divided by the mean counterfactual value, representing the percentage increase over expected RSV during the Great Recession.

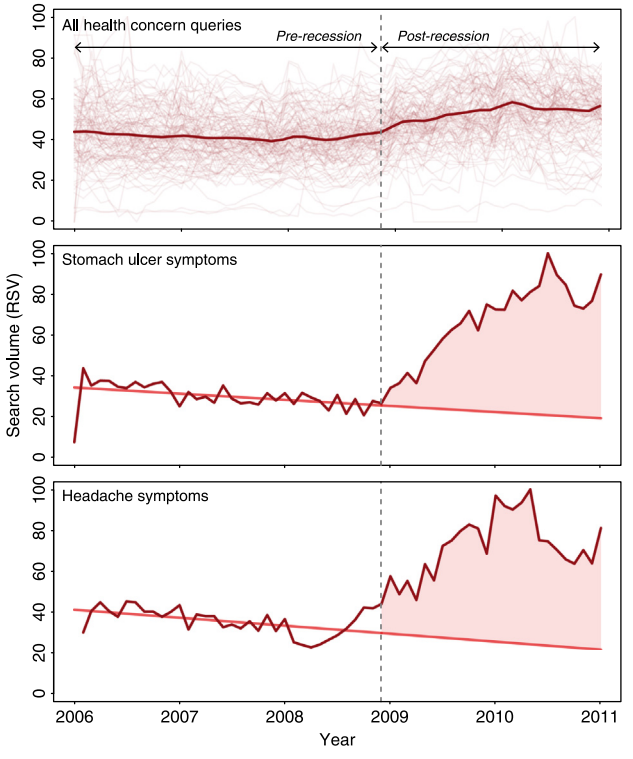

Figure 1. Time trends for all health concern queries around the Great Recession

Note: Figure shows time trends before and during the Great Recession (shaded) for all 383 health problem–related queries analyzed. Each line indicates a specific query trend, with the bold line indicating the mean trend (loess) across all queries.

Ninety-five percent CIs were derived from bootstrapped estimates of these ratios.[26] The alpha was based on typical standards, assuming each query represented a unique statistical test.[27] This strategy was applied to each query, ranking the top 100 by greatest excess, and later to the mean trend for groups of similar queries, grouped according to our opinion and ICD-10 codes.[28]

This modified interrupted time series[29,30] is useful for assessing overall changes in health concerns agnostic to their potential motivation. This approach is also immune to cyclical trends that bias pure pre–post comparisons and yields comparable results across all queries regardless of the potential shape of the observed RSV trends during the Great Recession. The use of linear projections was evaluated using the Akaike information criterion (AIC), assuming a good pre–Great Recession fit indicated that the Great Recession projection was valid.[31] The model fits were typically better for a linear model than many nonlinear alternatives (e.g., quadratics). A better AIC was not associated with higher estimates for excess query volume during the Great Recession.

To demonstrate practical significance, raw search volumes were downloaded from Google Adwords (google.com/adwords). Adwords' monthly volume was transformed by scaling to the time length of the Great Recession and then multiplying by the RSV model estimates described above.

Results

Figure 1 displays a spaghetti plot and mean trend for excess searches during the U.S. Great Recession. All queries had a relative excess of 26% (95% CI=3, 138) during the Great Recession. Figure 1 also shows example trends for *stomach ulcer symptoms* and *headache symptoms* queries; each with pronounced increases during the Great Recession. Searches were expected to decline five RSV each year for both; however, they increased rapidly during the Great Recession, even reaching their peak proportion (RSV=100).

Figure 2 shows rankings by RSV excess during the Great Recession for the top 100 significantly increasing queries. Leading the top 100, *stomach ulcer symptoms* had 228% (95% CI=35, 363) and *headache symptoms* 193% (95% CI=60, 275) greater relative volume during the Great Recession, accounting for roughly 1,480,000 and 1,520,000 excess queries, respectively. Across the top 100, about 205,000,000 excess health concern searches occurred during the Great Recession. Queries typically involved symptomology (i.e., *anxiety symptoms* [57%; 95% CI=12, 86]) or diagnostics (i.e., *heart rate monitor* [31%; 95% CI=3, 69] or 1,890,000).

Figure 3 displays rankings for thematic groups of queries. Headache queries had the greatest excess relative volume over the Great Recession (41%; 95% CI=3, 48), followed by hernia queries (37%; 95% CI=16, 142); chest pain queries (35%; 95% CI=6, 313); and arrhythmia queries (32%; 95% CI=3, 149). A common topic across themes was pain, with gastric (19%; 95% CI=5, 45); back (19%; 95% CI=4, 112); and joint pain (11%; 95%

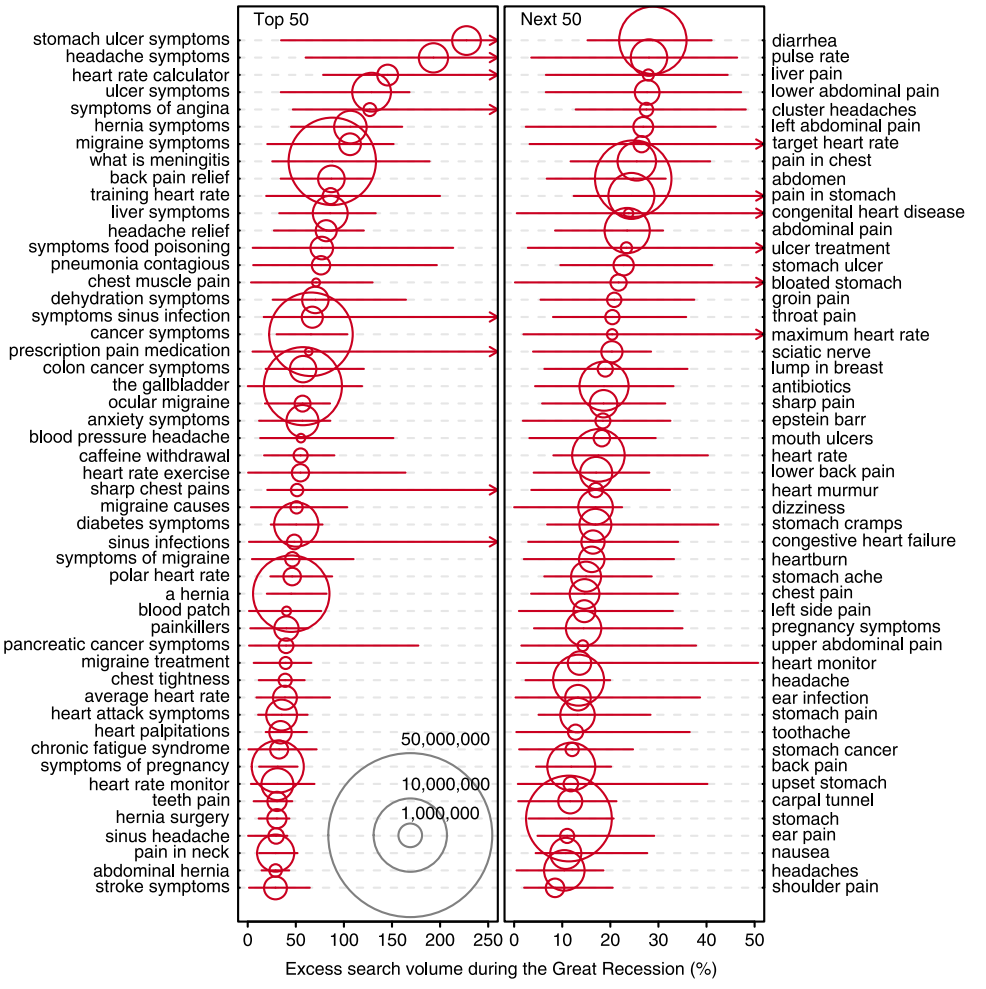

Figure 2. Ranking of excess health concern queries during the Great Recession

Note: Each circle represents the mean excess search volume during the Great Recession, and each line the 95% CI for that excess, as estimated from an interrupted time series, comparing the ratio of observed relative search volume (RSV) during the Great Recession versus expected volume (as estimated from a historic linear projection). Circles are sized according to the estimated increase in raw search volume during the Great Recession. Queries are ranked by the mean percentage increase, with only the top 100 significant associations shown.

CI=1, 21) queries higher than expected. Toothache-related queries were 19% (95% CI=4, 46) greater than expected during the Great Recession, including *teeth pain* (31%; 95% CI=6, 46) and *toothache* (13%; 95% CI=1, 37). Themes also involved cancer, potential respiratory infections, and reproductive health, with cancer searches up 32% (95% CI=5, 105); congestion up 26% (95% CI=1, 201); and pregnancy up 22% (95% CI=6, 49).

Discussion

The U.S. Great Recession was associated with increases in a range of health concerns, potentially indicative of worsening population health. Changes comprised arrhythmia, back pain, cancer, congestion, chest pain, gastric pain, headaches, hernia, joint pain, pregnancy, toothache, and ulcer concerns.

Implications for Improved Surveillance

The study herein provides a new methodological lens for one of the most studied phenomenon in public health. Most studies rely on self-reported survey responses, where those affected by a recession report generally poorer health[1] or poorer physical functioning.[32] Internet search queries, however, are potentially a more-specific measure, identifying the precise health concern under investigation and potentially any relating health problem.

This approach can be especially useful for hypothesis generation and empirically identifying outcomes for inclusion in surveys. For instance, investigators may use query trends to understand how individuals describe their symptoms to inform specific question wording. As health concerns around recessions vary over time/geography, query archives provide the potential to measure unique time/geographic changes for the provision of investigator resources.

Implications for Health and Health Advocacy

Query-based sentinels may inform clinical practice, where clinicians use health concern query trends to alter their screening. The stigma surrounding some concerns[33] or limited access to healthcare delivery[34] may prevent patients from engaging with clinical care. The Internet can potentially be a low-stigma and low-cost venue to reach patients who search for but do not otherwise receive screening or treatment for their concerns. Sponsored search results appearing on the first page of links for concern queries may direct searchers to webpages with information, professional screening, referrals to clinicians, and online treatments.

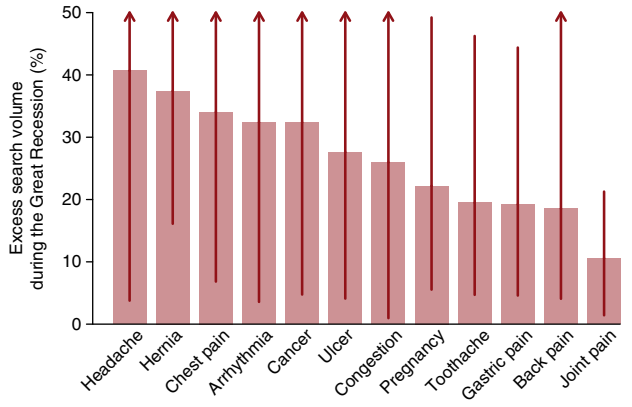

Figure 3. Health concern themes during the Great Recession
Note: Each bar represents the mean excess search volume during the Great Recession, and each line the 95% CI for that excess, as estimated from an interrupted time series, comparing the ratio of observed relative search volume (RSV) during the Great Recession versus expected volume (as estimated from a historic linear projection). Query themes are ranked by the mean percentage increase. Each bar is indicated by a group of similar queries, with the specific queries included in each category described in the text and included in Figure 2. Arrows indicate values of the CI outside of the axis range. Themes with fewer than two search terms were omitted.

Limitations

Health concerns manifested on Google search may differ from the population as a whole for many reasons, including multiple queries emerging from one household, demographic profiles of Internet searchers versus nonsearchers (including non–English speakers) and queries being unrelated to a true concern. Nevertheless, query trends are representative of population trends for many health[35] and nonhealth[36] outcomes.

The selection of root terms affected the outcomes studied, however, many of the subsequent terms identified share no direct commonality with the roots, for example, *pregnancy test*, suggesting minimal bias. The etiology of some concerns was unclear, for example, *heart rate monitor* may be indicative of a concern over sinus tachycardia or an interest in monitoring rates for exercise. However, the majority of queries could be linked to a specific health problem (e.g., ulcer). Only increases during the Great Recession were monitored (as strongly influenced by prior expectations[2]), however, a series of placebo tests find query patterns for other outcomes that should have declined indeed do, for example, *Washington Mutual*. The true counterfactual cannot be observed, but no obvious biases were identified in this linear projection.

The Great Recession has diverse impacts on health, including some positive ones (i.e., fewer automobile-related fatalities).[37] However, many positive effects relate to resource constraints, in contrast to the psychosocial mechanistic focus in this study that suggests recessions are bad for health. It may be possible to explore positive recession impacts using queries, but this awaits future study.

Conclusion

This study represents a novel use of web data, suggesting query trends may provide a sensitive, specific, timely and cost-effective account of how precise health concerns in the U.S. changed during the Great Recession. Extensions of this approach are needed to measure similarly detailed accounts of health concerns and their relationship with the socio-political environment. Regardless, this study outlines the initial steps toward fulfilling the promises of big data for preventive medicine.

This research was supported by a Google.org grant to JWA. A National Science Foundation (NSF) Graduate Research Fellowship and the Omidyar Foundation supported BMA. Google.org and the NSF had no role in the design and conduct of the study; in the collection, management, analysis, and interpretation of the data; or in the preparation, review, or approval of the manuscript. This report was improved by comments from a colloquium at Children's Hospital Boston, Stanford Prevention Research Center, University of North Carolina, and University of Southern California Keck School of Medicine. We thank Sanjay Basu, MD, PhD, Carl Latkin, PhD, Kurt Ribisl, PhD, and J. Niels Rosenquist, MD, PhD, for their helpful feedback.

JWA and BMA share an equity stake in a consultancy, Directing Medicine LLC, that helps other investigators implement some of the ideas embodied in this work. The data generation procedures, however, are not proprietary. There are no other conflicts of interest relevant to this study.

References

1. Burgard SA, Brand JE, House JS. Toward a better estimation of the effect of job loss on health. J Health Soc Behav 2007;48:369–84.
2. Catalano R, Goldman-Mellor S, Saxton K, et al. The health effects of economic decline. Annu Rev Public Health 2011;32:431–50.
3. McKee-Ryan F, Song Z, Wanberg CR, Kinicki AJ. Psychological and physical well-being during unemployment: a meta-analytic study. J Appl Psychol 2005;90:53–76.
4. Boland M, Sweeney MR, Scallan E, Harrington M, Staines A. Emerging advantages and drawbacks of telephone surveying in public health research in Ireland and the U.K. BMC Public Health 2006;6:208.
5. Cooper B. Economic recession and mental health: an overview. Neuropsychiatry 2011;25:113–7.
6. King G. Ensuring the data-rich future of the social sciences. Science 2011;331:719–21.
7. Bennett GG, Scharoun-Lee M, Tucker-Seeley R. Will the public's health fall victim to the home foreclosure epidemic? PLoS Med 2009;6:e1000087.
8. Catalano R. Health, medical care, and economic crisis. N Engl J Med 2009;360:749.
9. Pollack CE, Lynch J. Health status of people undergoing foreclosure in the Philadelphia region. Am J Public Health 2009;99:1833–9.

10. Pollack CE, Kurd SK, Livshits A, Weiner M, Lynch J. A case–control study of home foreclosure, health conditions, and health care utilization. J Urban Health 2011;88:469–78.
11. Alley DE, Lloyd J, Pagán JA, Pollack CE, Shardell M, Cannuscio C. Mortgage delinquency and changes in access to health resources and depressive symptoms in a nationally representative cohort of Americans older than 50 years. Am J Public Health 2011;101:2293–8.
12. Ayers JW, Ribisl K, Brownstein JS. Using search query surveillance to monitor tax avoidance and smoking cessation following the United States' 2009 "SCHIP" cigarette tax increase. PLoS One 2011;6:e16777.
13. Ayers JW, Ribisl KM, Brownstein JS. Tracking the rise in popularity of electronic nicotine delivery systems (electronic cigarettes) using search query surveillance. Am J Prev Med 2011;40:448–53.
14. Ayers JW, Althouse BM, Allem JP, Ford DE, Ribisl KM, Cohen JE. A novel evaluation of World No Tobacco Day in Latin America. J Med Internet Res 2012;14:e77.
15. Ayers JW, Althouse BM, Allem JP, Rosenquist JN, Ford DE. Seasonality in seeking mental health information on Google. Am J Prev Med 2013;44:520–5.
16. Ayers JW, Althouse BM, Johnson M, Cohent JE. Circaseptan (weekly) rhythms in smoking cessation contemplations. JAMA Int Med 2013; doi: http://dx.doi.org/10.1001/jamainternmed.2013.11933.
17. Paul MJ, Dredze M. You are what you tweet: analyzing Twitter for public health. Fifth International AAAI Conference on Weblogs and Social Media (ICWSM 2011); Jul 2011; Barcelona, Spain.
18. Ayers JW, Althouse BM, Allem JP, et al. Novel surveillance of psychological distress during the Great Recession. J Affect Disord 2012;142:323–30.
19. Katon W, Sullivan M, Walker E. Medical symptoms without identified pathology: relationship to psychiatric disorders, childhood and adult trauma, and personality traits. Ann Intern Med 2001;134:917.
20. Kuhn A, Lalive R, Zweimüller J. The public health costs of job loss. J Health Econ 2009;28:1099–115.
21. Lihm HS, Park SH, Gong EH, Choi JS, Kim JW. Relationship between occupational stress and gastric disease in male workers. Korean J Fam Med 2012;33:311–9.
22. Rios R, Zautra AJ. Socioeconomic disparities in pain: the role of economic hardship and daily financial worry. Health Psychol 2011;30:58–66.
23. Schulz AJ, Kannan S, Dvonch JT, et al. Social and physical environments and disparities in risk for cardiovascular disease: the healthy environments partnership conceptual model. Environ Health Perspect 2005;113:1817–25.
24. Gallo WT, Teng HM, Falba TA, Kasl SV, Krumholz HM, Bradley EH. The impact of late career job loss on myocardial infarction and stroke: a 10-year follow up using the health and retirement survey. Occup Environ Med 2006;63:683–7.
25. Business Cycle Dating Committee, National Bureau of Economic Research; www.nber.org/cycles/dec2008.html. 2008.
26. King G, Tomz M, Wittenberg J. Making the most of statistical analyses: improving interpretation and presentation. Am J Pol Sci 2000;44:341–55.
27. Rothman KJ. No adjustments are needed for multiple comparisons. Epidemiology 1990;1:43–6.
28. World Health Organization. ICD-10: International statistical classification of diseases and related health problems. Geneva: World Health Organization, 2004.
29. Reeves A, Stuckler D, McKee M, Gunnell D, Chang SS, Basu S. Increase in state suicide rates in the USA during economic recession. Lancet 2012;380:1813–4.
30. Ramsay CR, Matowe L, Grilli R, Grimshaw JM, Thomas RE. Interrupted time series designs in health technology assessment: lessons from two systematic reviews of behavior change strategies. Int J Technol Assess Health Care 2003;19:613–23.
31. Posada D, Buckley TR. Model selection and model averaging in phylogenetics: advantages of Akaike information criterion and Bayesian approaches over likelihood ratio tests. Syst Biol 2004;53:793–808.
32. Gallo WT, Bradley EH, Siegel M, Kasl SV. Health effects of involuntary job loss among older workers: findings from the health and retirement survey. J Gerontol B Psychol Sci Soc Sci 2000;55:S131–S140.
33. Link BG, Phelan JC. Stigma and its public health implications. Lancet 2006;367:528–9.
34. Hoffman C, Paradise J. Health insurance and access to health care in the U.S. Ann N Y Acad Sci 2008;1136:149–60.
35. Althouse BM, Ng YY, Cummings DA. Prediction of dengue incidence using search query surveillance. PLoS Negl Trop Dis 2011;5:e1258.
36. Goel S, Hofman JM, Lahaie S, Pennock DM, Watts DJ. Predicting consumer behavior with Web search. Proc Natl Acad Sci U S A 2010;107:17486–90.
37. Ruhm CJ. Are recessions good for your health? Q J Econ 2000;115:617–50.