Time-series analyses

We proceeded through the following steps. First, we estimated an equation predicting monthly suicide incidence in the absence of any covariates. Second, we used Box-Jenkins methods to identify and model autocorrelation in residual (i.e., unexpected) values of the equation estimated in Step 1 (1). Third, we added all independent variables to the equation (suicide methodology search share, unemployment rate, and suicide-related searches in the news) and estimated their coefficients. Fourth, we inspected the residuals from the full equation to ensure they exhibited no temporal patterns. Finally, we repeated steps 1 through 4 but substituted the independent variable of suicide methodology search share with depressive symptoms search share, and subsequently, suicide help search share. Given that we examined six distinct suicide-related search terms, we conducted six separate time-series tests.

The steps above required that we estimate the following time-series equation:

$$Y_t = \left(\omega_0 B^0 + \omega_1 B^1\right) S_t + \omega_2 N_t + \left(\omega_3 B^0 + \omega_4 B^1\right) U_t + \frac{\left(1 - \theta B^q\right)}{\left(1 - \phi B^p\right)} a_t$$

Where:

Y_t is the incidence of suicide for England in month t.

S_t is the standardized volume of suicide-related searches on the internet, in England in month t. A share is the number of queries for a particular search term or set of terms in a month divided by the total number of queries in that region-month.

B^n is the value of the variable at month t-n.

ω_0 to ω_1 are the estimated parameters for the suicide-related search variable, lagged at 0 and 1 month before the incidence of suicide. These are the parameters of interest.

N_t is the standardized volume of news-related suicide searches in England in month t.

ω_2 is the estimated parameter for the news-related search variable.

U_t is the unemployment rate in England in month t.

ω_3 and ω_4 are the estimated parameters for the unemployment rate variable lagged at 0 and 1 month before the incidence of suicide.

θ is the moving average parameter.

ϕ is the autoregressive parameter.

B^p and B^q are backshift operators that yield the value of a at year t-p for autoregressive and t-q for moving average patterns respectively.

a_t is the error term at month t.

We performed six separate time-series tests, one for each of the following terms or set of terms (S_t):

1. “suicide”

2. “suicide and methods”

3. “suicide and depression”

4. “depression”

5. “depression and help”

6. “suicide and help”

As a robustness check, we performed outlier detection routines to identify and remove outliers in suicide that may distort estimated parameters. Outliers in the time series could artificially inflate standard errors and lead to a type II error (i.e., false acceptance of the null). We used software from Scientific Computing Associates (2) to apply the outlier detection and correction method, proposed by Chang, Tiao, and Chen (3), to the regression model. This method iteratively adds binary variables for each month to find any that, if added to the equation, would have coefficients with *t* values greater than 3.5. The method also adjusts the ARIMA parameters as outliers are added. This routine detected no outliers in all tests.

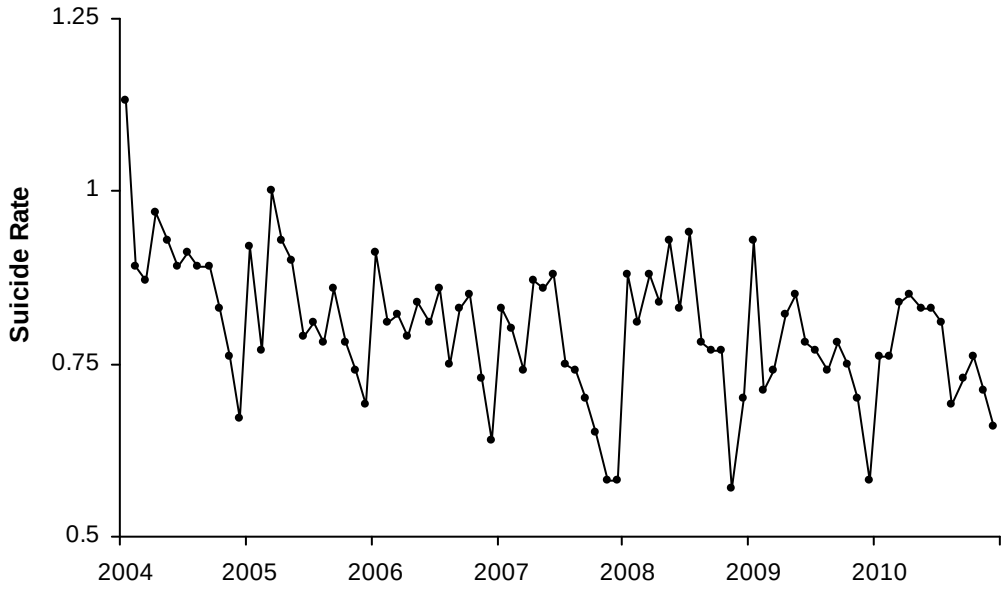

Figure 1. Monthly incidence of suicide in England (per 100,000 population) for 84 months spanning January 2004 to December 2010.

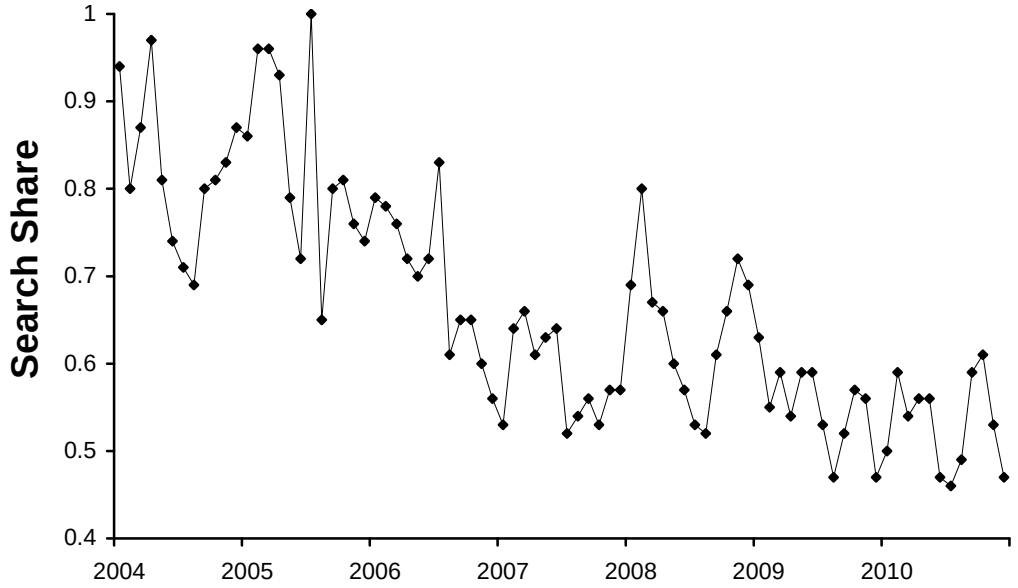

Figure 2. Monthly, standardized Google search share for "*suicide*" in England for 84 months spanning January 2004 to December 2010. (Search share for other five suicide-related searches not shown.)

Table 1. Time series coefficients predicting the incidence of suicide in England (n = 84 months beginning January 2004).

	Search: "Suicide"		**Search: "Suicide & Methods"**		**Search: "Suicide & Depression"**		**Search: "Depression"**		**Search: "Depression & Help"**		**Search: "Suicide & Help"**	
	Coef.	**SE**	**Coef.**	**SE**	**Coef.**	**SE**	**Coef.**	**SE**	**Coef.**	**SE**	**Coef.**	**SE**
Constant	.55	.175***	.53	.25*	.60	.15***	.43	.17***	.43	.29***	.62	.24**
Google Search lagged at:												
0 months	.0013	.0010	.0005	.0007	.0012	.0006*	.0045	.0014**	.0016	.0007*	.0010	.0006†
1 month	.0004	.0009	.0008	.0007	-.0005	.0005	-.0009	.0014	-.0007	.0007	.0009	.0006
News search, suicide Lagged at 0 months	-.0006	.0006	-.0005	.0006	-.0002	.0006	-.0003	.0006	-.00004	.0006	.0005	.0008
Unemployment Rate Lagged at:												
0 months	.046	.067	.046	.067	.065	.067	.063	.065	.048	.064	.014	.063
1 month	-.041	.066	-.041	.066	-.058	.066	-.054	.064	-.045	.065	-.010	.062
Autocorrelation												
MA at 12 months	.81	.098***	.85	.094***	.86	.096***	.94	.092***	.87	.096***	.83	.10***
AR at 1 month	.26	.12*	.25	.12*	.33	.12*	.27	.12*	.32	.11*	none	none
AR at 12 months	.91	.04***	.93	.04***	.87	.04***	.92	.03***	.94	.04***	.94	.04***

† p<0.10; *p<0.05; **p<0.01; ***p<0.001; all tests are 2-tailed.

References

1. Box G, Jenkins G, Reinsel G: Time series analysis: forecasting and control. London, Prentice Hall, 1994

2. Liu L-M, Hudak G: Forecasting and Time Series Analysis Using the SCA Statistical System, Vol. 1. Chicago, Scientific Computing Associates Corp., 1994

3. Chang I, Tiao G, Chen C: Estimation of time series parameters in the presence of outliers. Technometrics 30:193–204, 1988