Social Interactions Most Influential in Preventing and Treating Depression

Jay Kim

December 3, 2014

Abstract:

The study focused on identifying and determining the impact of variables influential in affecting the depressive symptoms of an individual. The variables used were presence of support, adequacy of support, number of close friends, family monthly poverty index, age, gender, type of work, race, Spanish speaking, other language speaking, hours worked, and an interaction between gender and family monthly poverty index. To examine the data and fulfill the study's goals, sub-models based on characteristics of the variables chosen were constructed and tested with a full model. Through performing multiple linear regression and bivariate analyses, it was determined that variables under the social sub-model had the greatest predictive value and the largest impact on depressive symptoms. The discussion and conclusion focused on the impact each variable and sub-model had on depressive symptoms and explored its implications on depression treatment strategies today.

Introduction:

It is estimated that over 9 million Americans suffer from depression. As mental illnesses are slowly gaining acceptance within mainstream society, it is becoming more important to identify factors contributing to depression and the extent these factors affect depression.

Based on the available data, this study focused on factors of depression related to the social interactions of an individual, an individual's economic status, and an individual's cultural heritage. In particular, this study considered the support system of the participant, the participant's family poverty index, work habits or lack thereof, age, gender, interaction between family poverty index and gender, racial identity, and language spoken.

A support system, defined for the purposes of this study as the presence and adequacy of support for an individual as well as the number of close friends, is critical to both the prevention and coping of depression. It should be noted that a person's social interactions and thus support system may depend on other factors such as age and gender. Onset of depression is also highly dependent on age and gender, as it is noted that those over 32 and females have higher incidence of depression.

Similarly, the income of an individual's family affects the number and severity of stressors an individual may face. It was believed that an interaction term existed between gender and a family's poverty level index as poverty affects genders in different manners, most notably due to gender roles and expectations. The study sought to account for work habits such as the hours worked at a job and employment status as well since they may also affect the stressors of an individual's economic condition. Employment status was thus explored as well. For the purposes of this study, unemployment will refer to a lack of job.

It is known that immigrants, or those that identify strongly with their cultural identity, face greater stressors due to either assimilation or microaggressions. To account for this, race was chosen as a variable. However, it was recognized that race was insufficient to account for cultural identity as certain individuals may not associate themselves strongly with their race. To test an individual's association with their race, language spoken at home was factored into the model as those that speak their native language are more likely to identify strongly with other aspects of his/her culture.

Methods:

To obtain the model, the program R 3.1.1 was used with RStudio 0.98.1049. Data was gathered from the 2007-2008 NHANES survey at http://wwwn.cdc.gov/nchs/nhanes/search/nhanes_continuous.aspx from the Depression Screener, Acculturation, Demographics Variables and Sample Weights, Income, Social Support, and Occupation datasets. The criteria for selecting participants for the NHANES survey was a 4-stage sample design, aimed with gathering a sample proportional to the distribution found in the national population, with oversampling to the Mexican-American,black, low income, adolescent, and elderly populations. For the data gathered in this study, 2968 individuals were used in the study. All individuals are 40 years or older. To account for missing data, an assumption was made that data was missing at random and a complete case analysis was used.

Due to limited data, a depression index was constructed to measure the state of depression of an individual relative to the sample. To do so, 7 different depression indicators were standardized and added.

The model was created with the depression index as the outcome variable and predictor variables Presence of Support, Adequacy of Support, Number of Close Friends, Family Monthly Poverty Index, Gender, Gender*Family Monthly Poverty Index, Age, Hours Worked, Type of Work, Race, and Spanish Speaking.

$$Depression\,Index = \beta_0 + \beta_1 Presence\,of\,Support + \beta_2 Adequacy\,of\,Support + \beta_3 Number\,of\,Close\,Friends +$$
$$\beta_4 Family\,Monthly\,Poverty\,Index + \beta_5 Gender + \beta_6 Gender * Family\,Monthly\,Poverty\,Index + \beta_7 Age +$$
$$\beta_8 Hours\,Worked + \beta_9 Unemployed\,but\,Looking\,for\,Work + \beta_{10} Unemployed\,and\,Not\,Looking\,for\,Work +$$
$$\beta_{11} Hispanic(not\,Mexican - American) + \beta_{12} White(non - Hispanic) + \beta_{13} Black(non - Hispanic) +$$
$$\beta_{14} Other\,Races + \beta_{15} Spanish\,Speaking + \beta_{16} Other\,Languages$$

To create the full model, it was necessary to check the assumptions of a multiple regression model. To do so, a residual plot was created and checked for the absence of patterns to suggest homoscedasticity. After observing any possible patterns within the residual plot, a histogram of the depression index was constructed. This plot was to check the overall shape of the data as well as note the existence of any possible influential points for future reference. The linearity assumption was tested through constructing a plot of each variable against the depression index. Distribution of the data of the variable was also tested through univariate analysis.

Adjusted variable plots and simple linear regressions of each term were then created to visualize and quantify the contribution of each term to the model's accuracy. Using backward stepwise elimination, variables with limited practical importance and negative adjusted R-squared values and high p-values were eliminated. To detect any multicollinearity, a scatterplot matrix was created and VIF was calculated. To verify any influential points, plots of standardized residuals that displayed the leverage of each point were created (code shown in source file). Any influential points will be discarded due to their influence on the data and the model's accuracy after assessing whether the point is a fluke or has meaningful implications. Distributions of each variable was also plotted to observe any interesting trends in the raw data itself. Distributions of the full model and the sub-models' fitted values will be compared against the depression index to observe any possible differences.

To create sub-models representing social, economic, and cultural identities, variables were chosen based on their implications on the individual. After determining the variables necessary for each model, these sub-models were created and a F-test was conducted against the full model to see the statistical significance of these variables.

Results:

As seen in Table 2, the social model had the greatest predictive ability amongst the sub-models according to adjusted R-Squared value with a value of .079. The economic model had an adjusted R-Squared value of .055 and the cultural model had an adjusted R-Squared value of .002. The full model had an adjusted R-Squared value of .144. All sub-models were statistically significant with a p-value of 2.2e-16 under a F-test.

The presence of support, the adequacy of support, number of friends, family monthly poverty index, age, gender, interaction term of gender and family monthly poverty index, unemployed and not looking for work, Hispanic but not Mexican-American, and non-Hispanic white were all statistically significant variables. Under the social model, all variables were statistically significant. Under the economic model, family monthly poverty index, the interaction term, and unemployed and not looking for work are all statistically significant. Under the cultural model, only Hispanics that are not Mexican-Americans and those that speak a language others than Spanish and English were statistically significant.

As shown in Table 1, under a bivariate analysis with all the variables against the depression index, only unemployed but looking for work, race (parent categorical variable), Hispanic but not Mexican-American, non-Hispanic white, non-Hispanic black, other races, English speaking, Spanish speaking and other language speaking variables were not statistically significant. Under an α value of .15, unemployed but looking for work, Spanish speaking, and other language speaking were statistically significant. All other variables were statistically significant.

Presence of support, number of close friends, family monthly poverty index, age, gender, english speaking, hours worked, and the interaction term all had negative coefficients. The largest negative coefficient was found with gender and presence of support. Adequacy of support, unemployed and both looking for and not looking for work, Hispanic (non-Mexican-American), non-Hispanic white, non-Hispanic black, other races, Spanish speaking, and other language speaking variables all had positive coefficients. Spanish speaking, other language speaking, and adequcy of support had the largest positive coefficients. In the full model, as shown in Table 2, the only noticable difference in the direction of coefficients existed with the other language speaking variable, which became positive. In the sub-models, only the interaction term and hours worked changed directions. All other trends appeared to be unchanged.

Under univariate analysis, only age and gender had a uniform distribution. Hours had somewhat of a symmetric distribution, excluding values of 0. All other variables faced skewed left or skewed right. The other language speaking and Spanish speaking variables faced extreme skewness. Under bivariate analysis, linearity was generally shown except in the cases of family monthly poverty index and age. Bimodal distribution appeared within family monthly poverty index while a sinusoidal distribution appeared in the age variable. As seen in Figure 1, the distribution of the depression index was skewed right with almost an exponential pattern. Under both the full and social models, the distribution was approximately symmetric. Under the economic model, the distribution was skewed left. Distribution for the cultural model was difficult to determine but appears symmetric.

There were no observations within the data that could be classified as 'unusual' with a Cook's Distance value greater than 1 and a standardized residual greater than 2. There were multicollinearity issues with race and English speaking, which was expected due to the relation between both. No other variables had multicollinearity issues.

There were interesting findings with the residuals. There was a negatively linear cutoff of residuals where past the line, no residuals existed. This was present within all models. For the full model, residuals were centered around 0 and became more dispersed the greater the residual and predicted values were. For the social model, residuals followed a semi-bowtie pattern (semi due to the linear cutoff). For the economic model, residuals began to fan out as fitted values increased. For the cultural model, residuals were highly concentrated over a large range where fitted values were approximately 0. As fitted values moved away from 0, residuals became highly dispersed over a limited range.

Discussion:

The p-value of all the models were sufficiently low enough to reject the null hypothesis that the regression modelss' accuracy was due to chance alone. The relatively low adjusted R-Squared values must be considered in the context of this study. It should be noted that adjusted R-Squared values of approximately .20 are considered to be adequate within social science studies. Furthermore, this study aims to examine the impact of these variables rather than create a model to predict depression among a certain demographic.

Interpretation of Results:

One of the most interesting findings from this study is that while the existence of support can greatly decrease the symptoms of depression, if the support system is deemed poor or inadequate, depression symptoms are greater than those without a support system. This held to be true consistently across both the full and social models. Although the age and family monthly index coefficients seem to be relatively small,these variables have a rather large additive effect, as age was measured in years and the poverty index can be maxed at a value of 5. In fact, upon usage on the data available, since the youngest individual was 40 years old, the smallest effect of age on these individuals is -2.84 which exceeds the value of any other coefficient. Social variables on average had the largest effect on the depression index in both the relevant sub-model compared to other sub-models and full model, with 3 out of 5 of its variables having a coefficient greater than .90. Out of the 3 sub-models, the social sub-model showed the greatest predictive ability, with an adjusted R-Squared value of .07939.

Hours worked however only had any significant implications in the simple linear regression model. Once other variables are accounted for, the association becomes less clear as neither the full model nor the economic model showed hours worked as a statistically significant variable. The effect of gender on depressive symptoms, as supported by literature, seems to affect females more negatively than males. The interaction term presented an interesting implication. Males that were poor tended to face less depressive symptoms than those that are wealthy under the full model. However, under the economic model, there is the opposite trend. The economic model's implications should be treated with higher regard however, due to a large difference in statistical significance, where the interaction term had a p-value of .05155 compared to a p-value of 8.75e-06. Thus, it is seen that those that are poor face more depressive symptoms than those that are wealthy. However, females are still seen as being more negatively affected by depressive symptoms, even in light of economic circumstance. It was not surprising to learn that unemployment increased depressive symptoms, particularly to those that have stopped searching for jobs. What was striking though was the difference found between those searching for and those that stopped searching for jobs. The difference in the coefficient associated with those searching for and those that stopped searching for jobs suggests that it is only the mentality of not being able to gain a job that greatly increases the depression index. The actual state of being unemployed did not greatly increase the depression index, which was surprising, particularly since it's more likely that these individuals were more recently laid off. Economic variables had the least impact on depressive symptoms under both the relevant sub-model compared to other sub-models and the full model. It also had the least predictive power, with the smallest adjusted R-Squared value.

The implications of racial identity on depressive symptoms however was unexpected. Although I had initially believed that whites would face the least amount of depressive symptoms, it appears that being white increases depressive symptoms. However, under the cultural model and simple linear regression, being white was not statistically significant. The language that the individual spoke however had a rather large effect on depressive symptoms. Although this variable similarly suffers from a lack of statistical significance, it shows that those that speak Spanish may face greater stressors due to their association with their culture. Speaking another language seemed to reduce depressive symptoms of an individual, although this is likely to be attributed to the wide-encompassing definition of 'other language spoken.' Cultural variables were shown to have limited effect on depressive symptoms. Although the Spanish speaking variable had the largest coefficient value out of all the models, the variable was not statistically significant, bring in doubt to the variable's validity. The other variables of the cultural model were shown only to have a moderate effect.

Model Limitations:

One of the largest limitations of this study was the usage of a depression index. The depression index was constructed from 7 symptoms of depression, ranging from feeling hopeless to thoughts you were better off dead. While the index may be useful in assessing depressive symptoms, it is not necessarily indicative of depression itself. The most serious issue with the index was that these questions were asked about the past 2 weeks of an individual's behavior. Depression is known to occur periodically and triggered through certain events, rather than a constant state of being. As a result, without direct knowledge of whether the individual was diagnosed depressed, many individuals that were not experiencing a depressive episode would not be reported in the data. Those that are medicated specifically for depression may also report average or near average values, creating unreliable identification as to who may have depression.

Furthermore, some of these variables, while being highly indicative of depression such as poor appetite and energy, may be observed in other mental or physical illnesses. While it may be argued that eliminating such unreliable variables would be a better course of action, it was determined that there were too few indicators that could reliably assess depression and so additional variables were needed. By adding more variables, it was believed that it would limit or offset the impact of mental or physical illnesses on the index's reliability to assess depression.

The method used by the NHANES to gather data also introduced biases to the study. Only individuals 40 years or older were elgible for use in the social support dataset. As a result, the study is limited to a certain demographic. This limits the goals of this model, particularly as average depression onset is 32 years old. As seen in the univariate analyses, the distribution of variables were generally skewed, a possible result of gathering data from datasets with different purposes and undergoing complete case analysis. For example, the NHANES was focused on oversampling Mexican-Americans and low income populations. In addition, the data available categorized various ethnic groups and languages under the umbrella term 'other', creating problems from the inherent differences of these groups and languages.

Model building was also restricted in this study. For both family monthly poverty index and age against depression index, a curvilinear relationship was exhibited, although for the sakes of this study's goals, this was not explored as other variables expressed some forms of linearity.

There also appeared to be a residual error within the study. Within all the models, there appeared to be some form of a linear cutoff, after which no residuals existed past. The extraordinarily colinear nature of this cutoff however suggested that this error was systematic and would not greatly affect the implications of this study.

Conclusion:

The model shows that being male has the single largest effect on limiting an individual's depressive symptoms. Age and poverty may on certain conditions also have the largest effects on decreasing depressive symptoms as well. These factors, all of which were determined to be 'social' variables, suggest that it is an individual's ability to interact with others that makes the greatest impact on depressive symptoms rather than the environment one is exposed to. Economic and cultural variables, both of which are highly important in the types of social interactions one is exposed to, were seen to have only a relatively small effect on one's depressive symptoms. Seeing as many of the social variables had negative coefficients, perhaps it is preventive strategies focused on forming meaningful social interactions that have a greater effect on reducing depression incidence than strategies that rely on other factors such as economic condition and cultural assimilation.

Figures and Tables:

Table 1: Bivariate Analysis of Variables and Depression Index

	Variables	Coefficients	P-Values	Adjusted R-Squared Values
1	Presence of Support	-1.1982	1.93E-06	0.007279
2	Adequacy of Support	2.1548	2.20E-16	0.04277
3	Number of Close Friends	-0.046784	7.56E-07	0.007882
4	Family Monthly Poverty Index	-0.5448	2.20E-16	0.04461
5	Age	-0.04102	8.33E-12	0.01529
6	Gender	-1.0239	2.84E-12	0.01599
7	Type of Work	*	4.23E-13	0.01837
8	Unemployed but looking for work	1.0028	1.25E-01	**
9	Unemployed and not looking for work	1.1144	5.30E-14	**
10	Race	*	3.05E-01	0.0002807
11	Hispanic(non-Mexican-American)	0.56071	6.29E-02	**
12	Non-Hispanic White	0.06247	7.74E-01	**
13	Non-Hispanic Black	0.06027	8.12E-01	**
14	Other Races	-0.17505	7.04E-01	**
15	English Speaking	-0.07953	6.33E-01	-0.0002603
16	Spanish Speaking	4.20793	1.38E-01	0.0004056
17	Other Language Speaking	4.20793	1.38E-01	0.0004056
18	Hours Worked	-0.022974	1.08E-12	0.01662
19	Gender*Family Monthly Poverty Index	-0.42682	2.20E-16	0.03191

$*$ Categorical variable with multiple factors

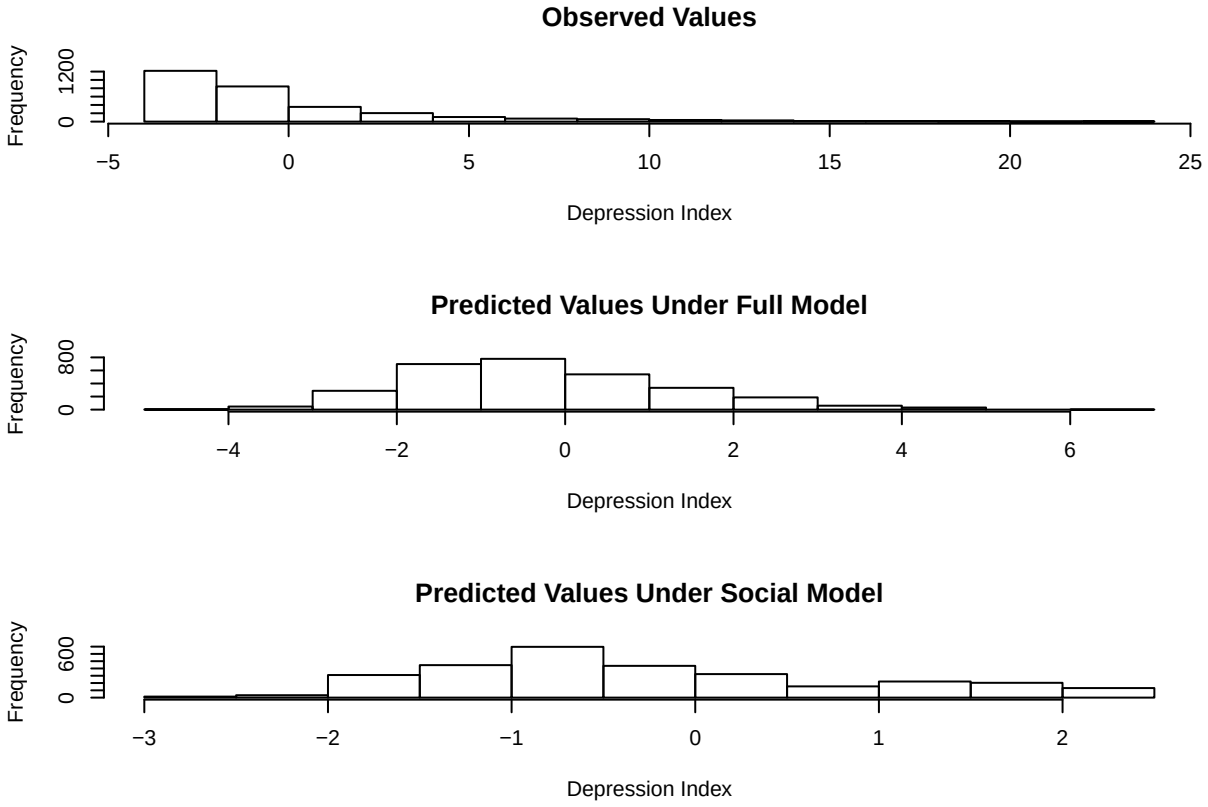
** Categorical variable under the parent variable

Table 2: Comparison of Coefficients, Standard Error, and Statistical Significance of Full and Sub-Models

	Full Model	Social	Economic	Cultural
	(1)	(2)	(3)	(4)
support	−1.347***	−1.597***		
	(0.244)	(0.246)		
moresupport	1.911***	2.085***		
	(0.184)	(0.188)		
friends	−0.025***	−0.023**		
	(0.009)	(0.009)		
poverty	−0.442***		−0.360***	
	(0.066)		(0.056)	
age	−0.071***	−0.028***		
	(0.007)	(0.006)		
gender	−1.127***	−0.902***		
	(0.253)	(0.142)		
inter	0.171*		−0.225***	
	(0.088)		(0.051)	
hours	−0.003		0.007	
	(0.007)		(0.007)	
typework3	0.286		0.522	
	(0.681)		(0.712)	
typework4	1.505***		0.919***	
	(0.328)		(0.340)	
race2	0.617**			0.561*
	(0.280)			(0.301)
race3	0.882***			0.079
	(0.210)			(0.218)
race4	0.391			0.072
	(0.238)			(0.253)
race5	0.479			0.158
	(0.468)			(0.503)
span	2.446			4.673
	(2.642)			(2.844)
other	−0.698			−0.982*
	(0.485)			(0.521)
Constant	5.088***	3.108***	0.274	−0.337*
	(0.566)	(0.429)	(0.345)	(0.192)
Adjusted R^2	0.144	0.079	0.055	0.002
Residual Std. Error	3.708 (df = 2951)	3.846 (df = 2962)	3.895 (df = 2962)	4.005 (df = 2961)

Note: *p<0.1; **p<0.05; ***p<0.01

- Additional note: Top number is coefficient and bottom is standard error



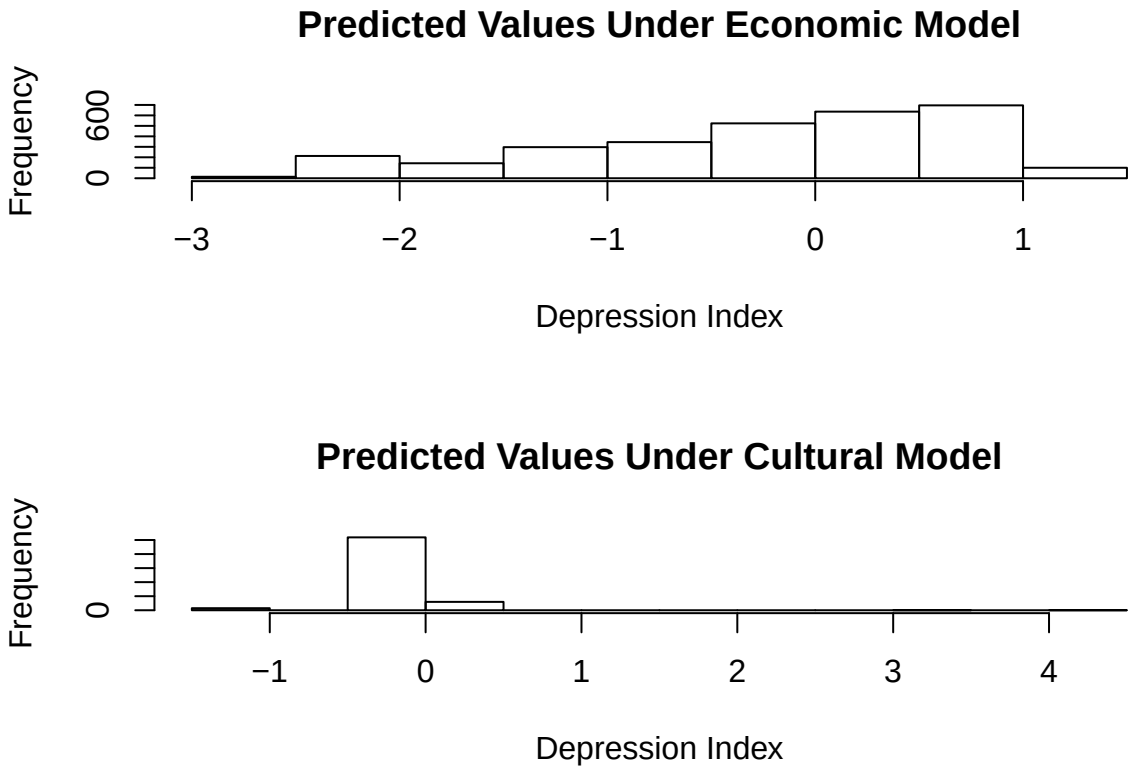


Figure 1. Comparing the Distribution of Fitted Values to the Depression Index

References:

Belle, Deborah, and Joanne Doucet. "Poverty, Inequality, and Discrimination as Sources of Depression Among U.S. Women." Psychology of Women Quarterly 27.2 (2003): 101-13. Web. 10 Dec. 2014.

Bhugra, D. "Depression in Migrants and Ethnic Minorities." Advances in Psychiatric Treatment 10.1 (2004): 13-17. Web. 9 Nov. 2014.

CDC. NHANES 2007-2008 Survey. Sept.-Oct. 2009. Raw data. National, n.p.

Virtanen, M., J. E. Ferrie, A. Singh-Manoux, M. J. Shipley, S. A. Stansfeld, M. G. Marmot, K. Ahola, J. Vahtera, and M. Kivimäki. "Long Working Hours and Symptoms of Anxiety and Depression: A 5-year Follow-up of the Whitehall II Study." Psychological Medicine 41.12 (2011): 2485-494. NCBI. Web. 9 Nov. 2014.

"What Is Depression?" - Depression Research Clinic. Stanford School of Medicine, n.d. Web. 10 Dec. 2014.

Yen, Irene H., and George A. Kaplan. "Poverty Area Residence and Changes in Depression and Perceived Health Status: Evidence from Alameda County Study." International Journal of Epidemiology 28 (1999): 90-94. Oxford Journals. 13 July 1998. Web. 9 Nov. 2014.

Zimet, George D., Nancy W. Dahlem, Sara G. Zimet, and Gordon K. Farley. "The Multidimensional Scale of Perceived Social Support." Journal of Personality Assessment 52.1 (1988): 30-41. 16 Dec. 1986. Web. 9 Nov. 2014.