Supplementary Information for "Violence is Like a Contagious Disease: The Spread of Violence through Social Networks"

Table S1 contains summary statistics for the sample including the variables we use in our analyses, measured for both the participant and the participant's friend.

Table S2 contains the correlations between the three violence measures for the overall sample, and for males and females separately. We analyzed each violent behavior separately to see if we would obtain converging results across the three measures. However, the different violence measures were significantly correlated. Although we analyze the three measures separately, we interpret the qualitatively similarity in results across the three measures as further evidence that they assess a common construct (i.e., violence).

Details on the Permutation Method

We generated 1,000 networks in which we retained the network structure and the overall prevalence of the violent behavior, but randomly reassigned the violent behavior measures among the participants. That is, for all participants we kept their network structure exactly as the participants reported. We then randomly re-ordered the assignment of the violent behavior variable in the full set of participants and re-assigned it to each participant. In this way, each participant is still connected to the same individuals, but the behavior of the individuals is random. In doing so, we are able to test whether the observed correlation is a product of the structure of the network (i.e., do we see high correlation in behavior between connected individuals even when the behavior is randomly assigned) or is different than chance (i.e., do we observe a high correlation in behavior only in the observed network). This method has been widely used in similar studies on obesity[1], smoking[2], happiness[3], and others[4].

We compared the observed association between the participant's behavior and the friend's behavior to this random baseline. In doing so, we were able to construct a null distribution of values from which we estimated standard errors for confidence intervals[5]. To construct 95% confidence intervals, we selected the 25th (< 2.5%) and 975th (>2.5%) values from the ordered set of values from the randomly generated networks as the 95% confidence interval of the null distribution of the association of behavior between friends.

Details on the Regression Models

In the tables below we show the regression results that underlie the models described in the main text. Each table shows the regression results for all participants, male participants, female participants, siblings, and all participants from a single school. We chose the school because it was the largest school in which there was variation in all of our model variables. This is the second largest school (*n*=783) in the Add Health study (the largest school did not contain variation across all of the racial groups). We also tried regressions on other, smaller schools. Although the statistical significance of the results varied considerably among these smaller schools, the direction of the results was largely consistent with our main findings.

An association in the behavior of friends in a social network may be a result of at least three processes: (1) influence, in which the behavior in one individual induces change in the behavior of another; (2) homophilic selection, in which an individual's choice of friends is a result of behavioral similarity and (3) confounding, in which individuals in the same social network are exposed to similar stimuli at the same time (e.g., the presence versus absence of anti-bullying campaigns or security guards at the school)[5]. Through the use of repeated measures of violent behavior, repeated measures of network relationships, and measurements of the directionality of the ties (e.g., who nominated whom as a friend), researchers are better able to distinguish between these explanations[6]. We were interested in measuring the effects of influence of friends above and beyond the effects of homophilic selection and confounding.

By including the violent behavior of nominated friends in Wave I we are able to identify associations between the participant's violent behavior and that of the participant's social contact net of the association exhibited in Wave I. The relationship identified between the participant and the friend in Wave II is more likely to be related to social influence than to social selection had we not included the lagged variables for both the participant and the participant's social contact. Inclusion of the friend's violent behavior in the previous wave has been shown to control for the likelihood that a participant selects a friend based on similarity in the characteristic of interest[7]. In each model, the coefficient at Wave II (e.g., "friends' violent behavior") reflected the effect of social influence controlling for other confounding variables in the model.

The key coefficient in these models that measures the extent to which a friend's violent behavior is influential on the participant's violent behavior is on the variable for the friend's contemporaneous violent behavior. We estimated linear regression models where we considered various versions of violent behavior in the participant as the outcome variable using generalized estimating equation (GEE) procedures, which helps account for multiple observations of the same participant across participant-friend pairings, and we assumed an independent working correlation structure for the clusters8. The GEE regression models in the tables provide results for linear regression of the focal participant's violent behavior on prior violent behavior, friend's violent behavior, and friend's prior violent behavior, and other covariates.

We estimated logit models in which we dichotomize the dependent variables and normal models in which we consider a continuous version of the dependent variables. The GEE regression models presented in the tables in the supplementary information provide parameter estimates. The results reported in the text and figures have been transformed into risk ratios for ease of interpretation. The mean effect sizes and 95% confidence intervals of the risk ratio estimates were calculated through simulating the first difference in the participant's Wave II violent behavior (changing from not engaging in violent behavior to engaging in violent behavior) using 1,000 randomly drawn sets of estimates from the coefficient covariance matrix and assuming all other variables are held at their means[9].

As noted in the main text, for the model of pulling a knife or a gun using a dichotomous version of the dependent variable among female participants we omit two control variables, the mother's education and the participant's alcohol usage. Together, these variables perfectly predict the outcome variable among the small number of female participants (n=34) who had pulled a knife or a gun. Although we present the results for female participants who pulled a knife or a gun on someone (see Table S5), these results might not be reliable because of the small number of female participants who reported doing this. In both the model for females, we found non-significant relationships between participants pulling a knife or a gun and friends doing the same. Although this may represent the true relationship, with such a small sample of female participants who engaged in this behavior we report these results with caution.

To test whether our results were robust to school context, we conducted analyses on a sinlge-large school (see single school models in each of Tables S3-S5). These analyses show that most of our main results are either significant, or that the coefficient estimated for the single-school regressions is within the confidence interval of the estimate from the overall sample. These results suggest that the results we find in the overall sample are not driven by school context.

Modeling Contagion

The analytical model we used relies on data with certain characteristics and makes various assumptions about how to model contagion in observational data. First, we note that the model we use relies on both repeated measures of the behavioral variable of interest (fighting, hurting someone badly, and pulling a weapon) for both sides of a social relationship and for at least two points in time. In this way, our sample is limited to participants who maintain relationships over time. It is possible that some relationships are formed or dissolved due, at least in part, to the violent behavior of one of the participants. As such, by limiting to only relationships that are maintained over time we limit the potential for tie formation to be influenced by the behavior we are studying.

A considerable concern when modeling contagion using observational data has been the ability to adequately control for social selection (homophily) as a competing source of clustering in the network. Using a lagged dependent variable measured on the alter (friend) has been used to control for social selection. This has been shown to be effective for at least partially controlling for social selection in observational data both analytically[10] and through simulation[11]. Using this model we are never certain that we have effectively controlled for social selection. However, we have controlled for social selection as effectively as possible given the data at hand.

Although the analytic strategy that we employed to identify peer effects has been widely used, its ability to separate social network effects from social selection or confounding is an area of scholarly debate. Some researchers have questioned whether this strategy is appropriate by showing that other study attributes seem unlikely to transfer socially (e.g., height and acne[12]), and also by showing evidence of contagion[13]. However, research has also shown that these concerns may result from the ways in which the model is specified[14], and that the results on implausible health outcomes are not robust when subjected to sensitivity analyses and may be the result of some small aspect of social contagion (e.g., social transfer of attitudes about reporting health outcomes[15]).

More broadly, researchers have suggested that there is a generic problem in assessing social contagion in observational data[16]. Although this is a serious concern, we have attempted to provide the best evidence available given that we have observational data. Like most observational studies, it is possible that a variable that has not adequately been accounted for is biasing our results. However, we believe that this research contributes to this literature and should be read with an understanding that homophily or confounding are not fully accounted for in the model we use.

Another suggestion that is commonly made when data contain nested structure such as the data we employ here (students are nested in schools) is to use fixed effects. However, the use of fixed effects with panel data in which there are many fixed effects (in this study, 153 schools) and few waves of data observation (in this study, 2 waves), severely biases coefficients toward zero.[17,18]

Our goal in this study is not to prove that the analytic strategy we used is correct. Rather, this research shows that aggressive behavior appears to transfer socially among adolescents given the model we use. Although there has been considerable work to better understand this model, including concern over its ability to identify peer effects separate from social selection and confounding, we present the study results as showing evidence of contagion, but also acknowledge that future work should attempt to use research designs that better allow researchers to separate social influence from other factors.

Sensitivity Analysis

Previous research has shown that a method for understanding the extent to which an estimate from a contagion model such as that we present here is sensitive to an unaccounted for variable.[19] This analysis allows researchers to vary two factors: (1) the extent to which the unaccounted for variable is associated with homophily, or the tendency for individuals to form relationships based on the dependent variable, and (2) how much the unaccounted for variable is associated with the outcome of interest – in our case violent behavior. Using this framework, we tested how unaccounted for variables with varying relationships to both social selection and violent behavior would have affected our analysis.

We conducted sensitivity analysis for all statistically significant results that we report (i.e., $p<0.05$) from the dichotomous versions of our model in which sensitivity analyses are more easily interpreted. The results of these analyses are shown in Tables S6-S9. These results show varying degrees of sensitivity to unobserved variables that are related to homophily or confounds. Table S6 shows the sensitivity for the association between a participant hurting someone badly and a friend hurting someone badly. These results show that for the estimate of 55% to be completely explained by the unobserved variable, the unobserved variable would need to have a fairly strong effect on the participant's likelihood of hurting someone badly. In this instance, the unobserved variable would need to have a 2.5-fold increase on the likelihood that the participant hurt someone badly and have a prevalence of 0.8 in friendships in which the friend had hurt someone badly and only a prevalence of 0.2 in friendships in which the friend had not hurt someone badly. That said, the unobserved variable would need to only be modestly related to the likelihood that the participant had hurt someone badly for the confidence interval to include 0%.

Table S7 shows the sensitivity for the association between a participant hurting someone badly and a friend hurting someone badly among male participants. These results show that for the estimate of 82% to completely be explained by the unobserved variable, the unobserved variable would need to have a fairly strong effect on the participant's likelihood of hurting someone badly. In this instance, the unobserved variable would need to have a 3-fold increase on the likelihood that the participant hurt someone badly and have a prevalence of 0.8 in friendships in which the friend had hurt someone badly and only a prevalence of 0.2 in friendships in which the friend had not hurt someone badly. Because the association between the participant hurting someone badly and a friend hurting someone badly is stronger among males, the relationship between the unobserved variable would need to be fairly strong and/or highly clustered (related to homophily) in the network for the confidence interval to include 0%.

Table S8 shows the sensitivity for the association between a participant having been in a fight and a sibling having been in a fight among all participants. These results show that for the estimate of 38% to completely be explained by the unobserved variable, the unobserved variable would need to have a fairly strong effect on the participant's likelihood of having been in a fight. In this instance, the unobserved variable would need to have a 2-fold increase on the likelihood that the participant had been in a fight and have a prevalence of 0.8 in friendships in which the friend had been in a fight and only a prevalence of 0.2 in friendships in which the friend had not been in a fight. In this case, the unobserved variable would need to only be modestly related to the likelihood that the participant had been in a fight for the confidence interval to include 0%.

Table S9 shows the sensitivity for the association between a participant hurting someone badly and a sibling hurting someone badly among all participants. These results show that for the estimate of 78% to completely be explained by the unobserved variable, the unobserved variable would need to have a fairly strong effect on the participant's likelihood of hurting someone badly. In this instance, the unobserved variable would need to have a 3-fold increase on the likelihood that the participant hurt someone badly and have a prevalence of 0.8 in friendships in which the friend had hurt someone badly and only a prevalence of 0.2 in friendships in which the friend had not hurt someone badly. Because the association between the participant hurting someone badly and a sibling hurting someone badly is stronger relatively strong, the relationship between the unobserved variable would need to be fairly strong and/or highly clustered (related to homophily) in the network for the confidence interval to include 0%.

These sensitivity analyses show that the significant results we find are robust to the inclusion of other variables we have not included that could be related to either homophily, violent behavior, or both. Of course, these analyses do not show that such variables could not possibly explain our results, but do show how prevalent and strong they would have to be to impact our findings. Further, the degree of sensitivity is similar to that found in other studies of social influence on obesity and smoking[17], as well as divorce[20].

Figure S1. Influence of friends and siblings on fighting in Wave II of the Addhealth data. *Note*. Effects are estimated using generalized estimating equation (GEE) logit models of serious fighting on several different subsamples of the Add Health social network. All models control for the participant's lagged behavior from Wave I, the friend or sibling's behavior in Wave II and Wave I, and covariates from the participant (see supplementary information). Circles denote means. Horizontal bars denote 95% CI.

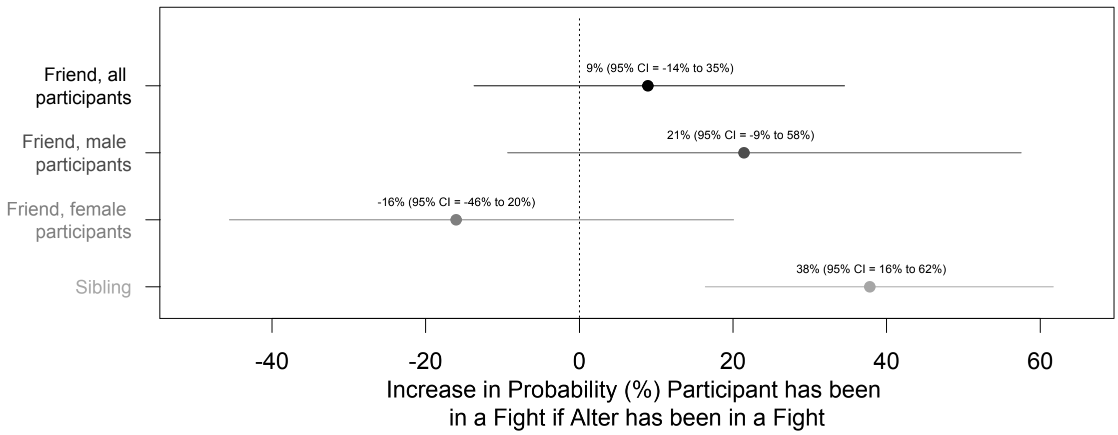

Figure S2. Influence of friends and siblings on pulling a weapon in Wave II of the Addhealth data. *Note*. Effects are estimated using generalized estimating equation (GEE) logit models of pulling a knife or gun on several different subsamples of the Add Health social network. All models control for the participant's lagged behavior from Wave I, the friend or sibling's behavior in Wave II and Wave I, and covariates from the participant (see supplementary information). Circles denote means. Horizontal bars denote 95% CI.

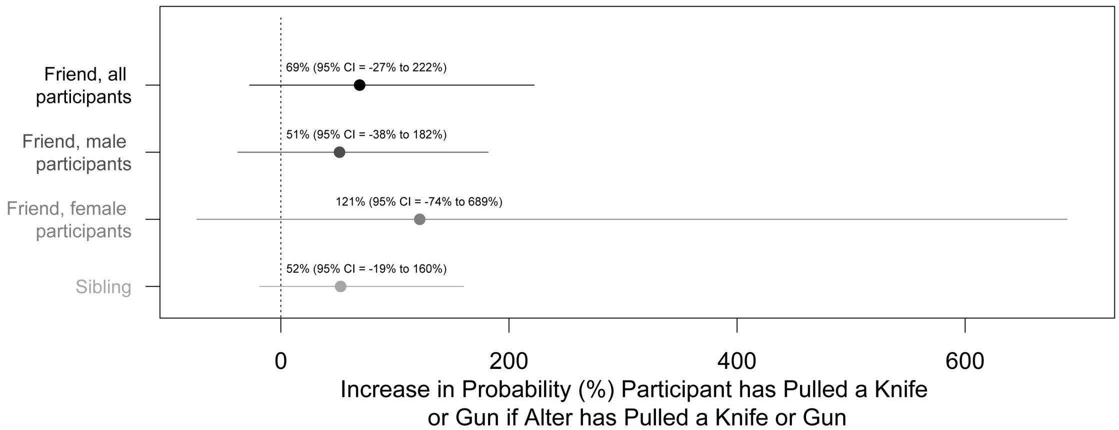

Table S1. Summary statistics of the sample. Note: Parent's education is a 10 item scale (0 = never went to school; 1 = 8th grade or less; 2 = 8th grade, but did not graduate from high school; 3 = went to a business trade, or vocational school instead of high school; 4 = high school graduate; 5 = completed a GED; 6 = went to a business, trade or vocational school after high school; 7 = went to college, but did not graduate; 8 = graduated from a college or university; 9 = professional training beyond a 4-year college or university).

	Wave I				Wave II			
	M	*SD*	Min	Max	*M*	*SD*	Min	Max
Participant has been in serious fight	0.46	0.79	0	3	0.24	0.54	0	3
Friend has been in serious fight	0.36	0.70	0	3	0.35	0.68	0	3
Participant has seriously hurt someone	0.26	0.61	0	3	0.10	0.39	0	3
Friend has seriously hurt someone	0.19	0.52	0	3	0.19	0.52	0	3
Participant has pulled a knife or a gun	0.06	0.30	0	2	0.06	0.29	0	2
Friend has pulled a knife or a gun	0.05	0.27	0	2	0.05	0.28	0	2
Number of times nominated as a friend	0.71	1.49	0	15	0.72	1.50	0	15
Total number of social contacts	2.19	2.13	1	18	2.19	2.13	1	18
Eigenvector centrality	0.03	0.09	0	0.71	0.03	0.09	0	0.71
Participant female	0.51	0.50	0	1				
Participant age	15.81	1.59	11	21				
Household income (1000s of dollars)	46.06	52.21	0	999				
Mother's education	5.45	2.40	0	9				
Hispanic	0.17	0.38	0	1				
Black	0.23	0.42	0	1				
Asian	0.07	0.26	0	1				

Table S2. Correlations among violence measures for all participants, and for males and females separately.

	All participants					Male participants					Female participants				
	serious fight, Wave I	serious fight, Wave II	seriously hurt someone, Wave I	seriously hurt someone, Wave II	pulled knife or gun, Wave I	serious fight, Wave I	serious fight, Wave II	seriously hurt someone, Wave I	seriously hurt someone, Wave II	pulled knife or gun, Wave I	serious fight, Wave I	serious fight, Wave II	seriously hurt someone, Wave I	seriously hurt someone, Wave II	pulled knife or gun, Wave I
serious fight, Wave II	.38					.35					.36				
seriously hurt someone, Wave I	.57	.31				.57	.29				.53	.28			
seriously hurt someone, Wave II	.31	.63	.36			.30	.66	.34			.27	.55	.35		
pulled knife or gun, Wave I	.30	.22	.36	.25		.31	.22	.36	.25		.25	.16	.29	.19	
pulled knife or gun, Wave II	.24	.36	.26	.43	.35	.23	.37	.25	.43	.35	.21	.31	.24	.40	.30

Table S3. Results from a generalized estimating equation (GEE) regression of "Participant has been in a serious fight, Wave 2" on "Friend has been in a serious fight, Wave 1 and Wave 2," controlling for whether the participant been in a serious fight, Wave 1, participant sex, age, race, ethnicity, household income, mother's education, the participant's usage of alcohol, and the mother's usage of alcohol. Variables for "has been in a serious fight" have been dichotomized ("*never*" = 0, "*1 or 2 times*" or "*3 or 4 times*" or "*5 or more times" = 1*).

	All			Male			Female			Sibling			Single School		
	Estimate	SE	p	Estimate	SE	p	Estimate	SE	p	Estimate	SE	p	Estimate	SE	p
Friend has been in serious fight, Wave II	0.09	0.10	0.38	0.23	0.14	0.10	-0.21	0.19	0.29				0.51	0.29	0.08
Sibling has been in a serious fight, Wave II										0.39	0.10	<0.01			
Participant has been in serious fight, Wave I	1.72	0.11	<0.01	1.62	0.14	<0.01	1.89	0.18	<0.01	1.42	0.09	<0.01	1.51	0.31	<0.01
Friend has been in serious fight, Wave I	0.12	0.09	0.22	0.14	0.12	0.21	0.18	0.18	0.31				0.12	0.27	0.66
Sibling has been in serious fight, Wave I										0.17	0.09	0.06			
Participant female	-0.65	0.11	<0.01							-0.55	0.09	<0.01	-0.96	0.33	<0.01
Participant age	-0.13	0.04	<0.01	-0.11	0.05	0.03	-0.15	0.06	0.01	-0.13	0.03	<0.01	-0.36	0.20	0.07
Household income	>-0.01	<0.01	<0.01	>-0.01	<0.01	0.02	>-0.01	<0.01	0.06	>-0.01	<0.01	0.33	0.01	0.01	0.04
Mother's education	-0.03	0.02	0.18	-0.04	0.03	0.15	-0.03	0.04	0.39	0.01	0.02	0.72	-0.13	0.07	0.09
Hispanic	0.16	0.15	0.29	0.12	0.19	0.52	0.20	0.25	0.41	0.34	0.12	<0.01	0.08	0.58	0.89
Black	0.31	0.14	0.03	0.18	0.18	0.34	0.48	0.21	0.03	0.25	0.10	0.02	0.90	0.55	0.10
Asian	0.14	0.21	0.50	-0.08	0.25	0.74	0.51	0.35	0.15	-0.10	0.20	0.60	0.56	0.57	0.32
Participant alcohol,	0.33	0.03	<0.01	0.30	0.04	<0.01	0.38	0.05	<0.01	0.28	0.03	<0.01	0.34	0.10	<0.01

Wave II															
Mother alcohol consumption	-0.01	0.01	0.23	-0.01	0.01	0.25	-0.01	0.02	0.74	0.02	0.01	0.01	-0.05	0.02	0.03
Constant	-0.15	0.65	0.81	-0.31	0.85	0.72	-0.59	0.98	0.55	-0.18	0.48	0.71	3.95	3.72	0.29
Deviance	646.94			408.73			235.51			207.17			88.95		
Null Deviance	795.67			492.97			280.54			236.98			119.34		
N	5913			2827			3086			4904			783		

Table S4. Results from a generalized estimating equation (GEE) regression of "Participant has seriously hurt someone, Wave 2" on "Friend has seriously hurt someone, Wave 1 and Wave 2," controlling for whether the participant seriously hurt someone, Wave 1, participant sex, age, race, ethnicity, household income, mother's education, the participant's usage of alcohol, and the mother's usage of alcohol. Variables for "has seriously hurt someone" have been dichotomized ("*never*" = 0, "*1 or 2 times*" or "*3 or 4 times*" or "*5 or more times" = 1*).

	All			Male			Female			Sibling			Single School		
	Estimate	SE	p	Estimate	SE	p	Estimate	SE	p	Estimate	SE	p	Estimate	SE	p
Friend has seriously hurt someone, Wave II	0.44	0.19	0.02	0.63	0.21	<0.01	-0.86	0.89	0.34				0.49	0.77	0.52
Sibling has seriously hurt someone, Wave II										0.60	0.18	<0.01			
Participant has seriously hurt someone, Wave I	1.80	0.16	<0.01	1.70	0.18	<0.01	2.09	0.32	<0.01	1.77	0.13	<0.01	2.69	0.47	<0.01
Friend has seriously hurt someone, Wave I	0.45	0.14	<0.01	0.35	0.18	0.05	0.78	0.43	0.07				-0.20	0.58	0.73
Sibling has seriously hurt someone, Wave I										0.25	0.14	0.08			
Participant female	-0.79	0.18	<0.01							-0.64	0.13	<0.01	-1.29	0.57	0.02
Participant age	-0.16	0.05	<0.01	-0.13	0.06	0.03	-0.20	0.11	0.08	-0.10	0.04	0.01	0.16	0.32	0.60
Household income	-0.01	<0.01	0.02	-0.01	<0.01	0.02	>-0.01	<0.01	0.38	>-0.01	<0.01	0.29	0.01	0.01	0.51
Mother's education	>-0.01	0.04	0.90	0.01	0.04	0.77	-0.04	0.06	0.49	0.02	0.03	0.40	-0.07	0.12	0.55
Hispanic	0.27	0.21	0.20	0.26	0.26	0.32	0.31	0.38	0.41	0.45	0.17	0.01	1.13	1.05	0.28

Black	0.25	0.21	0.23	0.07	0.24	0.79	0.60	0.37	0.10	0.33	0.15	0.02	2.24	1.03	0.03
Asian	-0.05	0.28	0.87	0.14	0.30	0.64	-1.04	1.12	0.35	-0.30	0.33	0.37	1.49	1.00	0.14
Participant alcohol, Wave II	0.41	0.04	<0.01	0.39	0.05	<0.01	0.47	0.08	<0.01	0.29	0.04	<0.01	0.50	0.15	<0.01
Mother alcohol consumption	0.01	0.01	0.18	0.01	0.01	0.34	0.02	0.02	0.22	0.01	0.01	0.08	-0.01	0.02	0.53
Constant	-0.98	0.94	0.30	-1.40	1.06	0.19	-1.22	2.01	0.54	-1.82	0.71	0.01	-8.26	6.25	0.19
Deviance	294.88			211.18			80.81			328.75			38.60		
Null Deviance	346.93			249.30			90.20			384.02			57.43		
N	5913			2827			3086			4904			783		

Table S5. Results from a generalized estimating equation (GEE) regression of "Participant has pulled a knife or gun on someone, Wave 2" on "Friend has pulled a knife or gun on someone, Wave 1 and Wave 2," controlling for whether the participant has pulled a knife or gun on someone, Wave 1, participant sex, age, race, ethnicity, household income, mother's education, the participant's usage of alcohol, and the mother's usage of alcohol. Variables for "has pulled a knife or gun" have been dichotomized ("*never*" = 0, "*1 or 2 times*" or "*3 or 4 times*" or "*5 or more times" = 1*).

	All			Male			Female			Sibling			Single School		
	Estimate	SE	p	Estimate	SE	p	Estimate	SE	p	Estimate	SE	p	Estimate	SE	p
Friend has pulled a knife or gun, Wave II	0.47	0.37	0.21	0.36	0.38	0.35	0.43	1.09	0.69				0.99	0.70	0.16
Sibling has pulled a knife or gun, Wave II										0.40	0.30	0.19			
Participant has pulled a knife or gun, Wave I	2.68	0.35	<0.01	2.56	0.37	<0.01	3.57	0.70	<0.01	2.06	0.21	<0.01	2.24	1.01	0.03
Friend has pulled a knife or gun, Wave I	0.08	0.53	0.87	0.08	0.47	0.87	-7.81	340.49	0.98				-0.03	1.11	0.98
Sibling has pulled a knife or gun, Wave I										0.44	0.28	0.12			
Participant female	-1.25	0.26	<0.01							-0.79	0.17	<0.01	-1.23	0.61	0.05
Participant age	-0.27	0.09	<0.01	-0.22	0.10	0.03	-0.19	0.22	0.39	-0.13	0.05	0.01	-0.26	0.37	0.49
Household income	<0.01	<0.01	0.39	<0.01	<0.01	0.28	<0.01	<0.01	0.84	-0.01	0.00	0.01	0.01	0.01	0.33
Mother's education	0.02	0.06	0.75	-0.01	0.06	0.87				-0.01	0.04	0.78	-0.23	0.17	0.18
Hispanic	0.73	0.43	0.09	0.35	0.39	0.37	1.66	0.71	0.02	0.18	0.23	0.45	0.53	1.62	0.74
Black	1.19	0.33	<0.01	0.84	0.34	0.01	1.57	0.82	0.06	0.50	0.19	0.01	2.31	1.64	0.16
Asian	0.37	0.52	0.47	0.59	0.54	0.28	-40.35	0.60	<0.01	0.05	0.41	0.91	2.08	1.51	0.17
Participant alcohol,	0.45	0.08	<0.01	0.03	0.01	<0.01				0.38	0.05	<0.01	0.52	0.22	0.02

Wave II															
Mother alcohol consumption	0.02	0.01	<0.01	0.36	0.09	<0.01	-0.10	0.13	0.42	0.01	0.01	0.72	0.13	0.04	<0.01
Constant	-0.39	1.41	0.78	-0.68	1.50	0.65	-2.43	3.48	0.49	-1.26	0.98	0.20	-0.92	9.64	0.92
Deviance	137.15			105.53			27.64			207.17			18.62		
Null Deviance	159.07			122.93			33.63			236.98			32.80		
N	5913			2827			3086			4904			783		

Table S6. Sensitivity analysis for association between friend's hurting someone badly and participant hurting someone badly among all participants. The extent of homophily or confounding simulates the prevalence (π_1) of the unobserved variable among participants who have hurt someone badly. We assume that for participants that did not hurt someone badly the prevalence is $\pi_0 = 1 - \pi_1$. When $\pi_1 = 0.5$ we assume the unobserved variable has no relationship to homophily or confounding and thus has no effect on the values we observed. When we assume that $\pi_1 = 1$, we assume a maximum amount of homophily or confounding, producing larger changes to what we observed.

	Effect of unobserved variable on probability of hurting someone badly (risk ratio)				
Extent of homophily or confounding in unobserved variable	1	1.5	2	2.5	3
0.5	55 (5 to 124)	55 (5 to 124)	55 (5 to 124)	55 (5 to 124)	55 (5 to 124)
0.6	55 (5 to 124)	43 (-3 to 107)	36 (-8 to 96)	31 (-12 to 89)	27 (-14 to 83)
0.7	55 (5 to 124)	32 (-11 to 91)	19 (-20 to 71)	10 (-26 to 58)	3 (-30 to 49)
0.8	55 (5 to 124)	22 (-18 to 76)	3 (-30 to 49)	-8 (-38 to 32)	-17 (-43 to 21)
0.9	55 (5 to 124)	12 (-24 to 62)	-10 (-39 to 30)	-24 (-49 to 10)	-34 (-55 to -4)
1	55 (5 to 124)	3 (-30 to 49)	-23 (-48 to 12)	-38 (-58 to -10)	-48 (-65 to -25)

Table S7. Sensitivity analysis for association between friend's hurting someone badly and participant hurting someone badly among male participants. The extent of homophily or confounding simulates the prevalence (π_1) of the unobserved variable among participants who have hurt someone badly. We assume that for participants that did not hurt someone badly the prevalence is $\pi_0 = 1 - \pi_1$. When $\pi_1 = 0.5$ we assume the unobserved variable has no relationship to homophily or confounding and thus has no effect on the values we observed. When we assume that $\pi_1 = 1$, we assume a maximum amount of homophily or confounding, producing larger changes to what we observed.

	Effect of unobserved variable on probability of hurting someone badly (risk ratio)				
Extent of homophily or confounding in unobserved variable	1	1.5	2	2.5	3
0.5	82 (17 to 65)	82 (17 to 65)	82 (17 to 65)	82 (17 to 65)	82 (17 to 65)
0.6	82 (17 to 65)	68 (8 to 52)	59 (2 to 44)	53 (-1 to 39)	49 (-4 to 35)
0.7	82 (17 to 65)	55 (0 to 41)	39 (-11 to 26)	29 (-17 to 17)	21 (-22 to 10)
0.8	82 (17 to 65)	43 (-8 to 30)	21 (-22 to 10)	8 (-31 to -3)	-2 (-37 to -11)
0.9	82 (17 to 65)	32 (-15 to 19)	5 (-32 to -4)	-11 (-43 to -19)	-22 (-50 to -29)
1	82 (17 to 65)	21 (-22 to 10)	-9 (-42 to -18)	-27 (-53 to -34)	-39 (-61 to -45)

Table S8. Sensitivity analysis for association between sibling's fighting and participant fighting among all participants. The extent of homophily or confounding simulates the prevalence (π_1) of the unobserved variable among participants who have been in a fight. We assume that for participants that have not been in a fight the prevalence is $\pi_0 = 1 - \pi_1$. When $\pi_1 = 0.5$ we assume the unobserved variable has no relationship to homophily or confounding and thus has no effect on the values we observed. When we assume that $\pi_1 = 1$, we assume a maximum amount of homophily or confounding, producing larger changes to what we observed.

	Effect of unobserved variable on probability of fighting (risk ratio)				
Extent of homophily or confounding in unobserved variable	1	1.5	2	2.5	3
0.5	38 (16 to 62)	38 (16 to 62)	38 (16 to 62)	38 (16 to 62)	38 (16 to 62)
0.6	38 (16 to 62)	27 (7 to 50)	21 (1 to 42)	16 (-2 to 36)	13 (-5 to 33)
0.7	38 (16 to 62)	18 (-1 to 38)	6 (-11 to 24)	-2 (-18 to 15)	-8 (-23 to 8)
0.8	38 (16 to 62)	8 (-9 to 27)	-8 (-23 to 8)	-18 (-31 to -4)	-26 (-38 to -13)
0.9	38 (16 to 62)	0 (-16 to 17)	-20 (-33 to -6)	-32 (-43 to -21)	-41 (-50 to -31)
1	38 (16 to 62)	-8 (-23 to 8)	-31 (-42 to -19)	-45 (-54 to -35)	-54 (-61 to -46)

Table S9. Sensitivity analysis for association between sibling's hurting someone badly and participant hurting someone badly among all participants. The extent of homophily or confounding simulates the prevalence (π_1) of the unobserved variable among participants who have hurt someone badly. We assume that for participants that did not hurt someone badly the prevalence is $\pi_0 = 1 - \pi_1$. When $\pi_1 = 0.5$ we assume the unobserved variable has no relationship to homophily or confounding and thus has no effect on the values we observed. When we assume that $\pi_1 = 1$, we assume a maximum amount of homophily or confounding, producing larger changes to what we observed.

	Effect of unobserved variable on probability of hurting someone badly (risk ratio)				
Extent of homophily or confounding in unobserved variable	1	1.5	2	2.5	3
0.5	78 (24 to 145)	78 (24 to 145)	78 (24 to 145)	78 (24 to 145)	78 (24 to 145)
0.6	78 (24 to 145)	64 (14 to 126)	56 (8 to 114)	50 (4 to 106)	46 (1 to 100)
0.7	78 (24 to 145)	52 (6 to 109)	36 (-5 to 87)	26 (-12 to 73)	19 (-17 to 63)
0.8	78 (24 to 145)	40 (-3 to 93)	19 (-17 to 63)	5 (-27 to 45)	-4 (-33 to 32)
0.9	78 (24 to 145)	29 (-10 to 77)	3 (-28 to 42)	-13 (-39 to 20)	-24 (-47 to 5)
1	78 (24 to 145)	19 (-17 to 63)	-11 (-38 to 23)	-29 (-50 to -2)	-41 (-59 to -18)

[1] Christakis, N.A. & Fowler, J.H. (2007). The Spread of Obesity in a Large Social Network over 32 Years. *The New England Journal of Medicine*, 357, 370-379.
[2] Christakis, N.A. & Fowler, J.H. (2008). The collective dynamics of smoking in a large social network. *The New England Journal of Medicine*, 358, 2249-2258.
[3] Fowler, J.H. & Christakis, N.A. (2008). Dynamic spread of happiness in a large social network: longitudinal analysis over 20 years in the Framingham Heart Study. *British Medical Journal*, 337:a2338.
[4] Christakis, N.A. & Fowler, J.H. (2013). Social contagion theory: Examining dynamic social networks and human behavior. Statistics in Medicine, 32(4), 556-577. doi:10.1002/sim.5408
[5] Christakis, N.A. & Fowler, J.H. (2013). Social contagion theory: Examining dynamic social networks and human behavior. *Statistics in Medicine, 32*(4)*,* 556-577. doi:10.1002/sim.5408
[6] Liang, K.Y. & Zeger, S.L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika*, *73*(1)*,* 13-22. DOI:10.1093/biomet/73.1.13
[7] Fowler, J. H., & Christakis, N. A. (2008). Estimating peer effects on health in social networks: A response to Cohen-Cole and Fletcher; and Trogdon, Nonnemaker, and Pais. *Journal of Health Economics*, *27*(5), 1400-1405.
[8] Bollen, K. & Stine, R. (1990). Direct and Indirect Effects: Classical and bootstrap estimates of variability. In C. C. Clogg (Ed.) *Sociological methodology* (pp. 115-140). Oxford, England: Blackwell.
[9] Carrington, P.J., Scott, J., & Wasserman, S. (2005). *Models and methods in social network analysis.* Cambridge, United Kingdom: Cambridge University Press.
[10] Carrington, P.J., Scott, J., & Wasserman, S. (2005). *Models and methods in social network analysis.* Cambridge, United Kingdom: Cambridge University Press.
[11] Fowler, J., Heaney, M.T., Nickerson, D.W., Padgett, J.F. & Sinclair, B. Causality in Political Networks. *American Politics Research 39*, 437-480.
[12] Cohen-Cole, E. & Fletcher, J.M. (2008). Detecting implausible social network effects in acne, height, and headaches: longitudinal analysis. *British Medical Journal, 337:a2533.*
[13] Cohen-Cole, E. & Fletcher, J.M. (2008). Is obesity contagious? Social networks vs. environmental factors in the obesity epidemic. *Journal of Health Economics, 27* (5), 1382-1387.
[14] Fowler, J.H. & Christakis, N.A. (2008). Estimating Peer Effects on Health in Social Networks. *Journal of Health Economics, 27*(5), 1400-1405.
[15] Christakis, N.A. & Fowler, J.H. (2013). Social Contagion Theory: Examining Dynamic Social Networks and Human Behavior. *Statistics in Medicine, 32*(4), 556-577.
[16] Shalizi, C.R. & Thomas, A.C. (2011). Homophily and contagion are generically confounded in observational social network studies. *Sociological Methods & Research, 40*, 211-239.
[17] Nerlove, M. (1971). Further evidence on the estimation of dynamic economic relations from a time series of cross sections. *Econometrica 39*, 359-387.
[18] Nickell, S. (1981). Biases in dynamic models with fixed effects. *Econometrica 49*, 1417-1426.

[19] Vanderweele, T. (2011). Sensitivity Analysis for Congation Effects in Social Networks. *Sociological Methods and Research 40*, 240-255.
[20] McDermott, R., Fowler, J.H., Christakis, N.A. (2013). Breaking Up Is Hard to Do, Unless Everyone Else Is Doing It Too: Social Network Effects on Divorce in a Longitudinal Sample. *Social Forces 92*(2), 491-519.