Social Cohesion and Community Displacement in Armed Conflict
Appendix

A From Archives to Sample - The Data Collection Procedure

One of the authors contacted the Haganah Archives, during a research visit to Israel. He was invited to come and examine the Village Files and was given permission to take pictures of all the primary source materials. Each Village File was photographed in its original Hebrew. The authors then received a grant to translate the Village Files, a job taken by several native Hebrew speakers. The authors then created a codebook for each variable that was to be created from the information, and the authors, with the help of research assistants, coded the information into tabular form. Because the documents are heavily structured and mostly uniform in the information conveyed, the creation of a codebook for translating the text into tabular form was quite simple and straightforward. Variables were created for each piece of pertinent information (land area, cultivation area, number of clans and families and their political affiliations, number of village leaders, schools, mosques, and so forth). The data collected, in its entirety, will be made available to future researchers for further quantitative analysis of these data. Similarly, for those interested, we have created an online repository of the original (photographed) documents, which will be available upon request.

After collecting all this information and coding the data, the authors made a principled decision on the sampling procedure. Because this article pertains to Arab Palestinian villages which were under direct threat of strategic displacement, our sample consists of all villages in the sub-districts: Acre, al-Ramla, Baysan, Haifa, Jaffa, Nazareth, Safad, and Tiberias. A small number of villages in the sub-districts of Jerusalem, Gaza, Jinin and Tulkarm were excluded from the sample, in places that were behind the front lines of Egyptian, Jordanian, or Iraqi forces which served as a military impediment for Israeli forces. For example, the Jerusalem district also includes the town of Jericho and several smaller villages in the Jordan Valley. These were excluded from the sample because the Jordanian forces quickly overtook the area with the initiation of the inter-state stage of the war, and these villages were not at any point exposed to direct Israeli violence during the war.

To verify the information in the Village Files, specific details such as names of villages, were cross-referenced by the authors with other sources for an additional sub-sample of villages. In his book The Palestinian Exodus from Galilee, 1948, Nazzal documents interviews conducted with refugees about specific

villages and their displacement during the 1948 wars. These conversations, conducted with refugees some thirty years after the events, describe about village life and the circumstances of their evacuation from the village. We use the information provided there to confirm specific details documented in the village files, to validate the veracity of the information provided in the Village Files.

Nazzal, for example, examines the village of al-Ghabisiya, 12 kilometers northeast of Acre. In an interview with a refugee from the village, he reports about the attack by Jewish forces. "they captured Uthman Asad Abdul'al's house, in the southwest corner of the village, and arrested the owner and his son" (Nazzal, Dissertation, pg. 155). The Village Files report the presence of the Abdul'al family, as one of the four clans in the village consisting of roughly 40 individuals. This shows the complementarities between Arab Palestinian sources and the Village Files. The recurrence of family names and similar information about the village remain consistent for the small sample available of villages surveyed by Nazzal in his interviews with refugees. In Al-Zeeb, the al-Saadi and Taha clans appear in both the Village Files and the interviews. In Al-Sumeriya, Nazzal mentions an interview with Ahmad Ibrahim Yusif, presumably the son or grandson of Hussein Yusif, one of the leaders of the Yusif clan in the village, which the Village Files mentions as one of the larger clans in the village (close to 100 members).

B Sample Balance

The Haganah did not survey all villages in Mandate Palestine. The goal of the surveys was to collect important data about each village for strategic purposes, making it unlikely that the villages were randomly selected to be surveyed. To investigate how this non-random selection into the sample may affect our later estimates, we describe how the villages that are in the survey may differ from villages where no survey was conducted. Doing this requires information on villages that were not surveyed. We use the 1945 British Village Census to create a list of villages comprising the total sample, along with relevant covariates. This results in data on 789 villages.

Below, Figure 1 shows which villages received and did not receive a Village File, separating the sample between those in the area later called the West Bank and Gaza, who are irrelevant to our analysis, and those included in our analysis due to their location in proximity to the conflict. Figure 2 illustrues the similarity of distribution on several variables, between villages for which a survey was collected (in blue) and those that weren't (in red). Note that elevation, distance to roads, latitude and longitude, and land area (irrigated and total) all show similar distributions across both samples. As noted in the main text, the main differences regard the ethnic make-up of the villages, which is to be expected.

Figure 1 illustrates the geographic distribution of villages included and excluded from the sample of

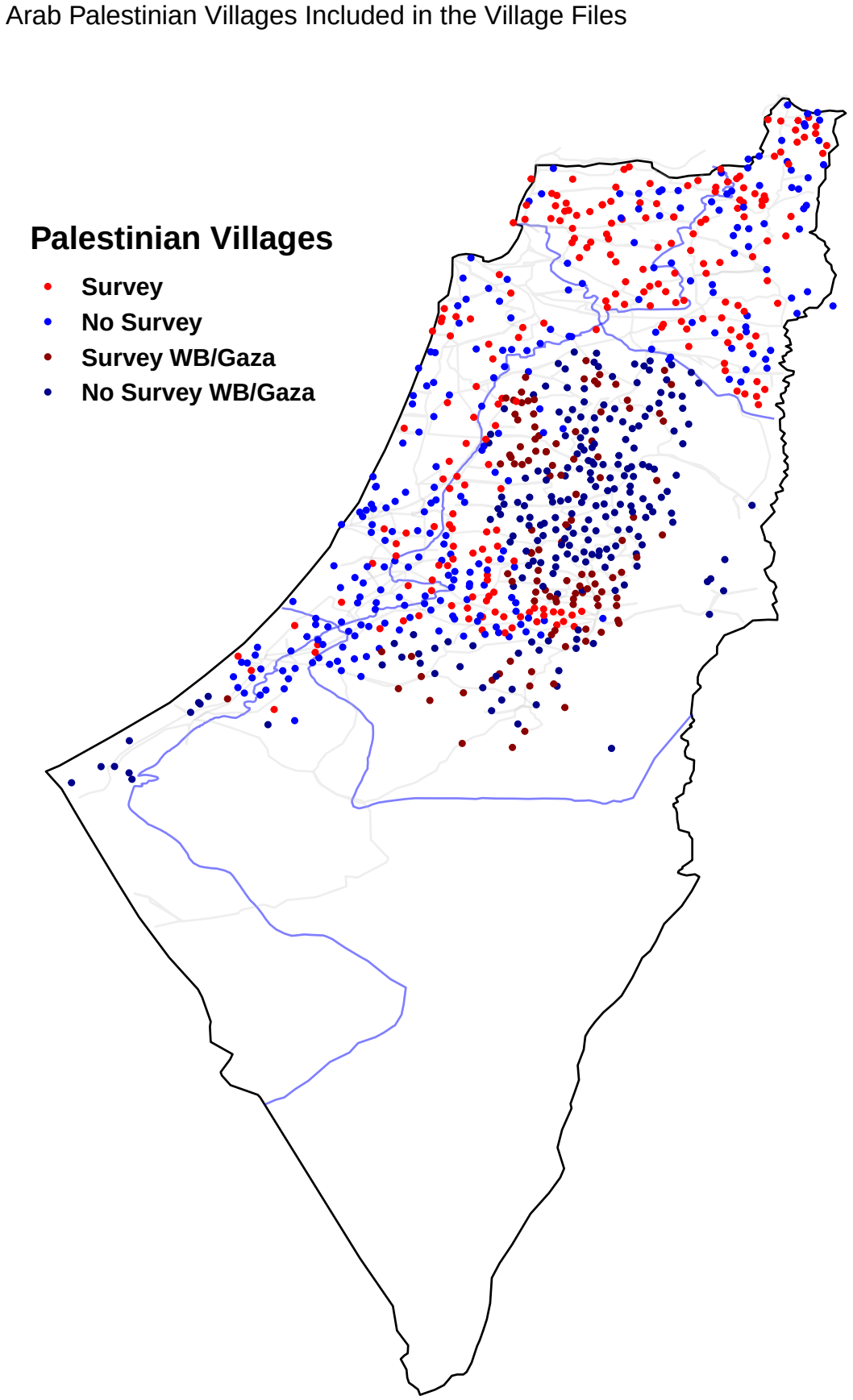
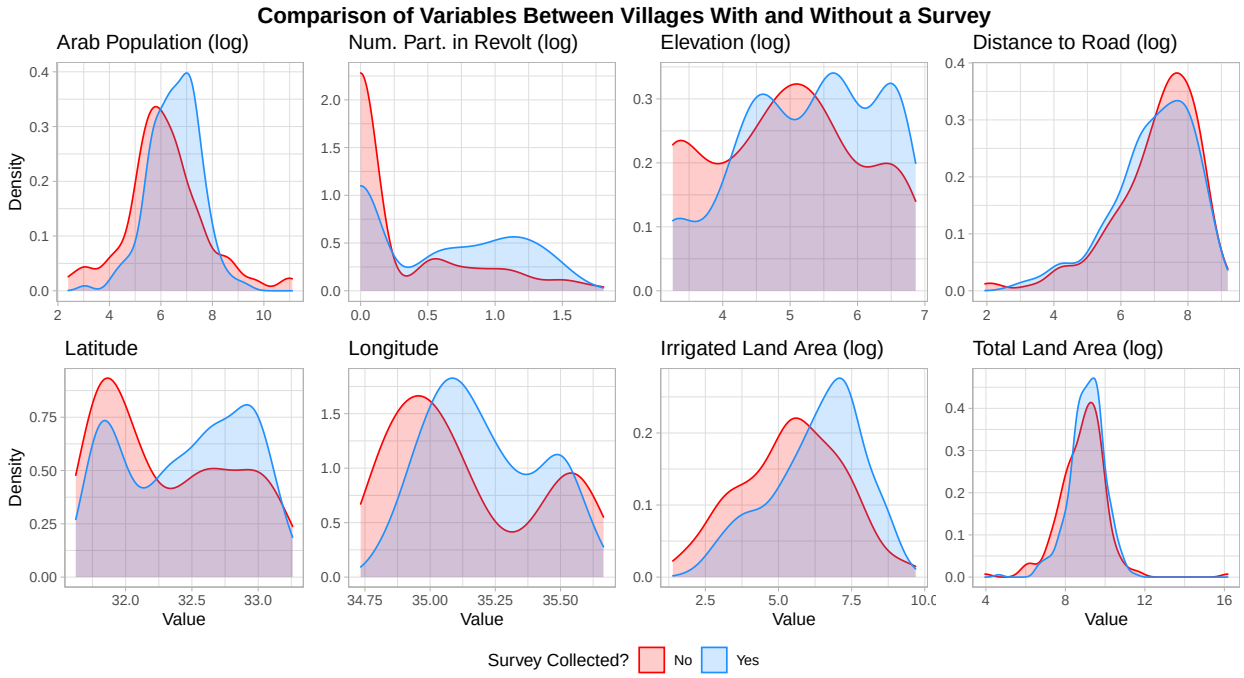


Figure 1: This map shows the location of villages along with whether a Village Survey was conducted, and whether they are located in the current West Bank/Gaza area (excluding all villages in the Jerusalem district).

Figure 2: Balance Plot



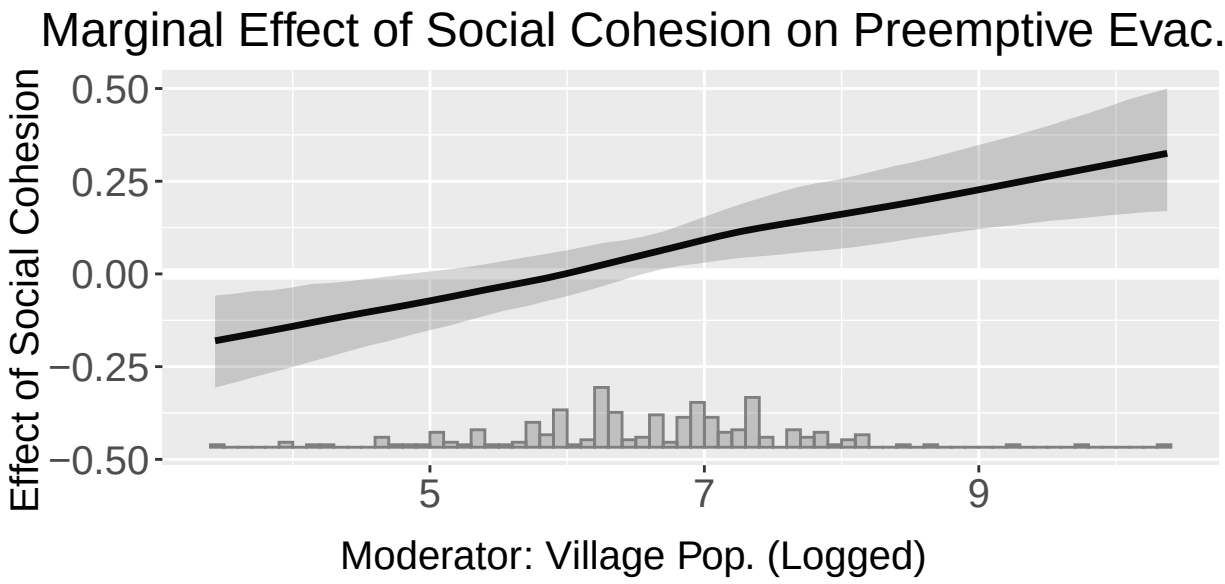
Village Files, and Figure B compares the distribution on key covariates across these groups of villages. As discussed in the paper, for most variables, the density plots are roughly similar across the different types of villages. The main differences are in the demographic variables describing the distribution of Jewish and Arab Palestinian population in the village.

C Research Design and Data Appendix

C.1 Interaction Term Robustness Check

Interaction terms in linear models rely on the assumption that the interaction effect is linear in the parameters. Violations of this assumption can bias the coefficient estimates and substantive interpretation. Here, we use Interflex software to check this assumption (Hainmueller, Mummolo, and Xu 2019). We use the kernel smoothing method in the Interflex software to allow the effect of social cohesion on preemptive evacuation to vary nonlinearly as a function of the village population.

The plot below shows that the interaction term is correctly specified and that the linear assumption is not violated. When using a flexible and nonparametric kernel, the marginal effect of social cohesion increases linearly as village population increases.



C.2 Reactive Evacuation as Dependent Variable

These models replace preemptive evacuation as the dependent variable with reactive evacuation. The effect of social cohesion is null across models. This suggests that, in keeping with our theory, social cohesion only increases the probability of preemptive evacuation and does not increase the probability of reactive evacuation.

Table 1: Social Cohestion and Reactive Depopulation

	Dependent variable:				
	Reactive Evacuation				
	(1)	(2)	(3)	(4)	(5)
Social Cohesion	0.013	0.055	−0.036	0.070	−0.415
	(0.154)	(0.160)	(0.167)	(0.214)	(1.017)
Village Pop					0.024
					(0.263)
Latitude		0.866**		1.006*	0.852**
		(0.393)		(0.565)	(0.396)
Longitude		−2.202***		−1.187	−2.082**
		(0.847)		(1.452)	(0.868)
Military Aged Men				0.110	
				(0.179)	
Total Cultivatable Land				0.251	
				(0.219)	
Mean Elevation			−0.252*	0.065	
			(0.130)	(0.205)	
Land Area			−0.127	−0.446*	
			(0.167)	(0.235)	
Social Cohesion * Village Pop					0.080
					(0.154)
Observations	249	247	238	163	247
Log Likelihood	−157.066	−152.575	−148.489	−98.891	−152.048
Akaike Inf. Crit.	318.132	313.150	304.978	213.781	316.096
Note:					*p<0.1; **p<0.05; ***p<0.01

C.3 Robustness Checks

C.3.1 Matching

To ensure that similar villages are being compared, we use the genetic matching algorithim of (Sekhon 2008) to estimate the effect of social cohesion on depopulation outcomes. Since matching requires a binary variable, we operationalize our measure of social cohesion to take a value of 1 if all *hamulot* are affiliated with the same political party, and 0 if they are not. We are especially interested in comparing villages that are geographically similar. For this reason, we match on the following variables: latitude, longitude, average elevation, and distance to the nearest police office, which also captures proximity to colonial development efforts made by the British.

This exercise results in 56 treated and 56 untreated villages. The estimated effect of our binary measure of social cohesion is 0.177, and the corresponding p-value is 0.0347.

C.3.2 Additional Controls

These models include additional control variables not included in the main text. The results are robust to alternative controls.

Table 2: Additional Controls

	Dependent variable:				
	Preemptive Evacuation				
	(1)	(2)	(3)	(4)	(5)
Social Cohesion	0.406*	0.450*	0.445*	0.980***	0.458*
	(0.247)	(0.244)	(0.233)	(0.330)	(0.265)
Arab Irrigated Land	−0.087				
	(0.127)				
Arab Cereal Land		−0.445**			
		(0.202)			
Arab Cultivatable Land			−0.130		
			(0.221)		
Military Aged Men (log)				0.165	
				(0.278)	
Firearms (Binary)					0.192
					(0.832)
Dist. to Viol. Depop.	0.477	0.131	0.281	0.034	0.034
	(0.294)	(0.285)	(0.272)	(0.372)	(0.307)
Mean Elevation	−0.399**	−0.489***	−0.442**	−0.464**	−0.490**
	(0.181)	(0.188)	(0.176)	(0.222)	(0.202)
Land Area	−0.062	0.735**	0.074	0.834*	−0.099
	(0.236)	(0.323)	(0.260)	(0.436)	(0.274)
Observations	218	226	230	162	181
Log Likelihood	−86.137	−85.083	−93.707	−53.582	−67.808
Akaike Inf. Crit.	184.275	182.166	199.415	119.163	147.616
Note:					*p<0.1; **p<0.05; ***p<0.01

Term	Estimate	Std. Error	p-value
Intercept	2.066	2.010	0.305
Social Cohesion	0.399	0.231	0.086
Mean Elevation	-0.488	0.164	0.003
Land Area	-0.286	0.208	0.173
Military Aged Men	0.183	0.244	0.462

C.3.3 Multiple Imputation Results

As with many types of quantitative data assembled from historical archives, there is missing data in our dataset. To ensure that our results are not driven by patterns of missing data, we perform multiple imputation using the mice R package. This method assumes that the pattern of missing data, conditional on the other covariates, is independent and normally distributed. We note that our missing data problems are almost entirely driven by missingness in the independent variables, as the depopulation data we assembled from the work of historians is exhausitve.

To perform the multiple imputation, we run a logit model with 5 different datasets where the missing data is imputed. The covariates include social cohesion, elevation, land area, and the number of military aged men. The results presented below show that our main findings are robust.

C.3.4 Spatial Regression

This table shows the results of spatial autogressive and spatial error models to estimate the effect of social cohesion on preemptive evacuation.

Table 3:

	Dependent variable:			
	Preemptive Evacuation			
	spatial autoregressive	*spatial error*	*spatial autoregressive*	
	(1)	(2)	(3)	(4)
Social Cohesion	-0.514^{***}	-0.462^{***}	0.043	-0.463^{***}
	(0.153)	(0.146)	(0.028)	(0.162)
Village Pop.	-0.138^{***}	-0.122^{***}	-0.018	-0.118^{***}
	(0.041)	(0.039)	(0.028)	(0.041)
Social Cohesion * Village Pop.	0.083^{***}	0.075^{***}		0.077^{***}
	(0.023)	(0.022)		(0.024)
Mean Elevation			-0.045^{**}	-0.054^{**}
			(0.022)	(0.022)
Land Area			0.012	0.014
			(0.031)	(0.030)
Total Cultivatable Land			0.006	-0.018
			(0.031)	(0.030)
Dist. to Viol. Depop.			0.007	
			(0.033)	
Observations	247	247	231	231
Log Likelihood	-91.697	-92.844	-80.252	-75.357
σ^2	0.119	0.120	0.114	0.109
Akaike Inf. Crit.	195.393	197.689	178.504	168.714
Wald Test (df = 1)	26.252^{***}	23.861^{***}	21.978^{***}	22.179^{***}
LR Test (df = 1)	23.061^{***}	20.766^{***}	18.937^{***}	19.332^{***}
Note:				*p<0.1; **p<0.05; ***p<0.01

C.3.5 Instrumental Variable Regression

As with most forms of observational data, there may be confounding between our independent variable, social cohesion, and our outcome variable, depopulation, even after for controlling for relevant observables. To probe the robustness of our results, we estimate a number of flexible causal inference models (Braumoeller et al. 2018) using repression during the 1936-1939 Palestinian revolt as an instrument for social cohesion. The historiographical literature is quite clear that repression by the British during this revolt devastated Palestinian social and political life in the rural countryside. We have no reason to believe that British repression was completely randomly assigned to villages conditional on some baseline covariates that we observe. Our approach here is not to argue that we identified the causal effect of social cohesion on depopulation outcomes, only to show that our results hold if we assume that repression during the great revolt was as-if randomly assigned.

We use flexible causal inference models as outlined in (Braumoeller et al. 2018). Traditional two-stage least squares assumes that each outcome variable in the two stages is normally distributed. This is not appropriate for our data, since the second stage is a binary variable. We thus specify a normally distributed first stage model and a probit distribution for the second stage. The first stage predicts social cohesion, the second stage predicts preemptive depopulation. We use a range of variables capturing British repression. These include, whether the British demolished a house in a village, whether the British exercised some form of collective punishment in a village, and whether a village was targeted by the Special Night Squad for a raid. The Special Night Squads (SNS) was a joint British-Jewish counterinsurgency unit consisting of about 100 members who targeted Arab villages in the later years of the revolt (Hughes 2015). All three variables are binary variables. We also condition (in both stages) whether a village participated in the revolt, which helps ensure that the variation we are instrument for is variation in repression, not participate in the revolt.

The data on Great Revolt participation and whether there was any collective punishment comes from the Village Files survey we used in the main body of the text. Data from House Demolitions and Special Night Squad raids comes from (Chetrit 2013) which carefully documents these instances from British archival sources, recently released.

Results for three models using different combinations of variables are presented below. As expected, repression in 1936-1939 reduced social cohesion. The estimated effect of social cohesion is higher than in the standard models estimated in the main text, and precisely estimated. This exercise prodives further evidence that social cohesion affected depopulation outcomes in the 1948 war.

Table 4: Two Stage Model Results, Reduced Model

	Estimate	Std. Error	Pr(>|z|)
Second Stage			
Intercept	-2.04380	0.20454	0.00000
Social Cohesion	1.08717	0.10873	0.00000
GR Participation	0.31639	0.22186	0.15385
First Stage			
Intercept	1.61882	0.14054	0.00000
GR Participation	-0.34026	0.16670	0.04124
Collective Punishment	-0.07538	0.11011	0.49363
House Demo.	-0.30466	0.14502	0.03566
SNS Raid	0.14241	0.11369	0.21034

Table 5: Two Stage Model Results, House Demolition as Instrument

	Estimate	Std. Error	Pr(>|z|)
Second Stage			
Intercept	-7.50222	24.63948	0.76076
Social Cohesion	1.20689	0.07228	0.00000
GR Participation	0.25854	0.17450	0.13844
Lat.	-0.05762	0.34281	0.86653
Long	0.14507	0.97983	0.88230
Land Area	0.29265	0.08322	0.00044
Avg. Elevation	-0.03594	0.15485	0.81644
First Stage			
Intercept	-1.69763	10.97895	0.87712
GR Participation	-0.22643	0.14289	0.11305
House Demo.	-0.07355	0.14859	0.62063
Lat.	-0.04619	0.18428	0.80208
Long	0.19177	0.42479	0.65166
Land Area	-0.22576	0.06723	0.00079
Avg. Elevation	-0.02150	0.06471	0.73974

Table 6: Two Stage Model Results, Multiple Repression Variables

	Estimate	Std. Error	Pr(>|z|)
Second Stage			
Intercept	-45.93690	34.59288	0.18420
Social Cohesion	1.18235	0.12032	0.00000
GR Participation	0.37364	0.21406	0.08090
Lat.	-0.55856	0.51559	0.27866
Long	1.70593	1.43853	0.23567
Land Area	0.33939	0.10295	0.00098
Avg. Elevation	-0.25870	0.23475	0.27044
First Stage			
Intercept	-4.02695	12.96410	0.75609
GR Participation	-0.19345	0.16858	0.25116
Collective Punishment	-0.17185	0.14007	0.21989
House Demo.	-0.21978	0.11927	0.06536
SNS Raid	-0.07857	0.13774	0.56841
Lat.	-0.08623	0.21487	0.68821
Long	0.30713	0.50572	0.54364
Land Area	-0.22600	0.07309	0.00199
Avg. Elevation	-0.07082	0.07482	0.34384

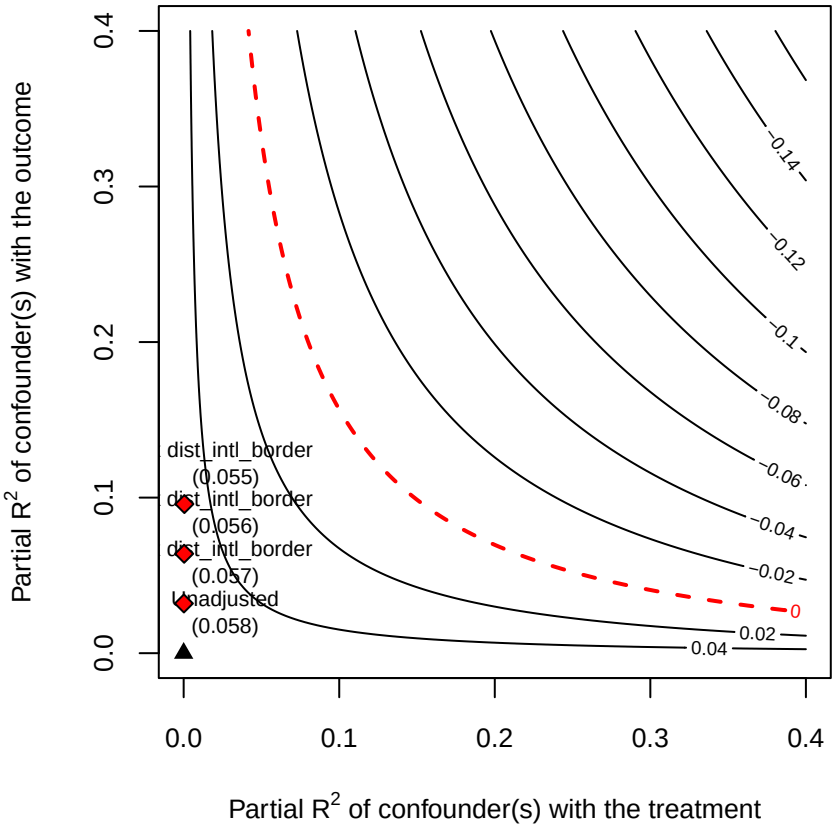
C.3.6 Sensitivity Analysis

This section conducts sensitivity analysis of our main model (Model 3 in Table 1 in the main text) to see how unobserved confounding could affect the effect of social cohesion on preemptive depopulation. To do this, we use the *sensmark* R package. The main exercise is to ask, if there is unobserved confounding, how much would there have to be to eliminate the positive effect of social cohesion on preemptive evacuation? Note that this package and exercise requires the use of OLS as the main model, so we use this model instead of a logit model in our main analysis. Here, the coefficient of social cohesion on preemptive exodus is 0.058, with a t-value of 1.970 and a p-value of 0.05.

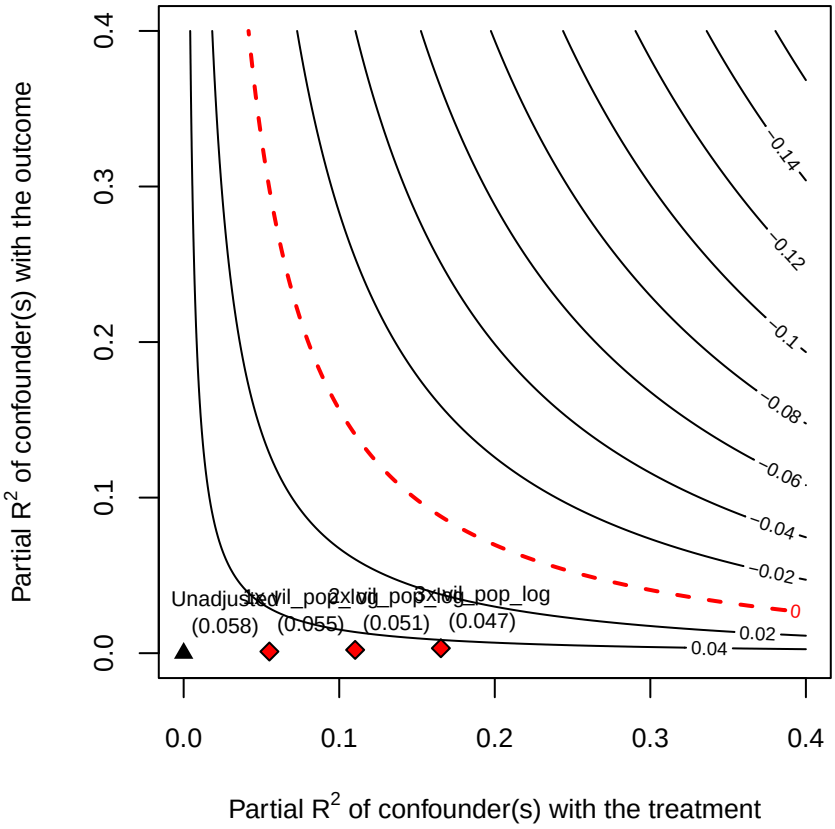
We benchmark the relative level of unobserved confounding by relating that confounding (in other words, setting it in proportion to) the effect of the distance to the international border on preemptive evacuation. We chose this control as a hard test since it is the most precisely estimated control variable (with a p-value in our linear model of 0.00817.)

The contour plot below shows how our estimates change as the level of confounding changes The horizontal axis indicates the partial R^2 of the confounders with the treatment, while the vertical axis indicates the partial R^2 of the confounders with the outcome.

The red diamonds provide estimates of the treatment effect when the level of confounding is proportional to the distance to the international border control. These show that the direction of the effect of social cohesion on preemptive evacuation is robust to confounding that is once, twice, or even three times as strong as the observed effect of the distance to the international border on preemptive evacuation.

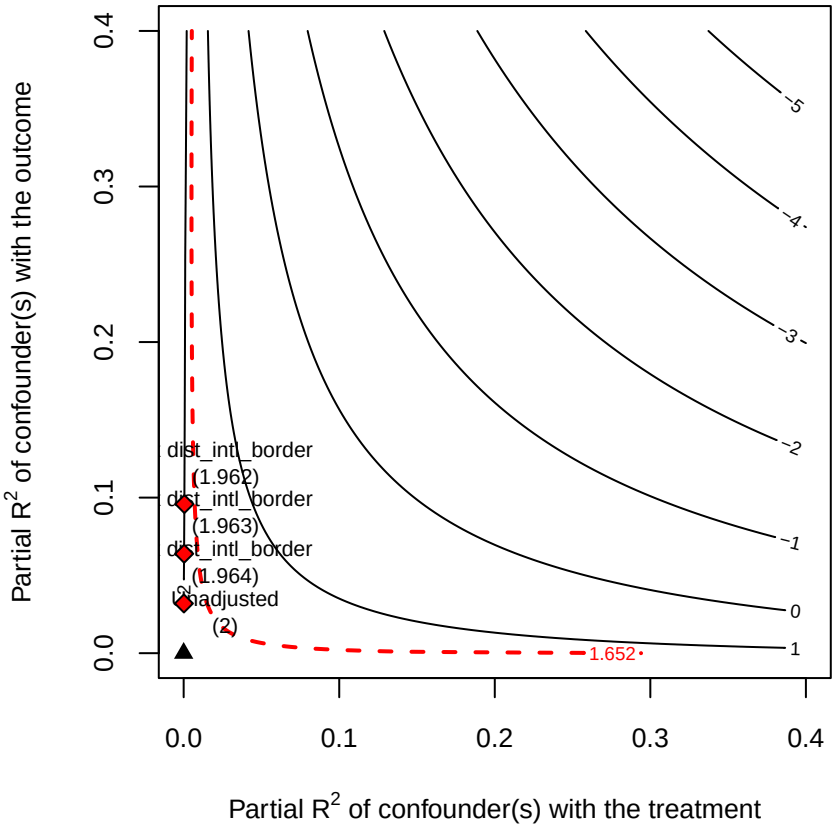


We also perform this exercise with the size of the village population. Population clearly mattered to both Israeli targeting of villages and the proximate cause of a village a village was depopulated by Israeli forces. Here, even if the amount of unobserved confounding is three times the effect of village population on preemptive evacuation, the effect of social cohesion remains positive, although it is attenuated.

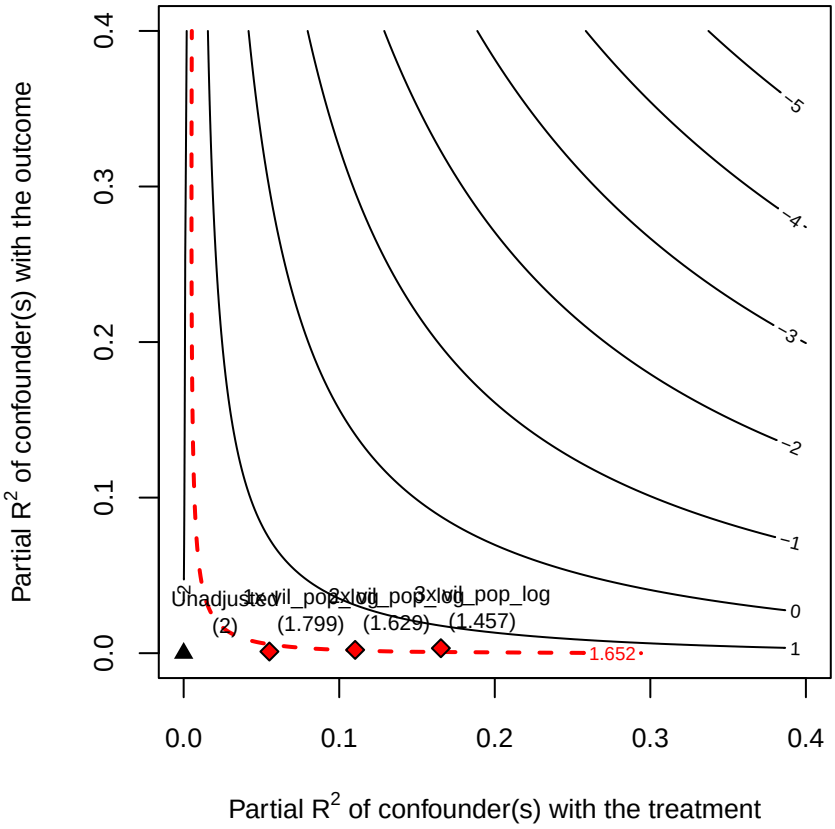


We also explore how unobserved confounding could affect the precision of our estimates of social cohesion on preemptive evacuation. We do this in a similar manner as the previous exercise, benchmarking the unobserved confounding as proportional to the other controls in our model.

First, we benchmark the confounding to the distance to the international border. The t-stat attenuates slightly even if the unobserved confounding is three times the effect of the distance to the international border (and still statistically significant at the $\alpha = 0.10$ level, as indicated by the red-dotted contour line.)



Similar results hold when benchmarking the precision to the effect of village population. Here, if unobserved confounding is 1 or 2 times the effect of village population on preemptive evacuation, social cohesion remains statistically significant. It is only when the unobserved confounding is three times the effect of village population that the effect of social cohesion ceases to be statistically significant. We also note that the t-stat in these linear probability models is lower than the t-stat in our logit models, which we use in the main analysis.


Partial R2 of confounder(s) with the outcome
0.0
0.1
0.2
0.3
0.4
Partial R2 of confounder(s) with the treatment
0.0
0.1
0.2
0.3
0.4
5
4
3
2
1
0
1
Unadjusted
(2)
1x vil_pop_log
(1.799)
2x vil_pop_log
(1.629)
3x vil_pop_log
(1.457)
1.652

C.3.7 Social Cohesion for Different Political Parties

The two main elite political families in Mandate Palestine were the Husseinis and the Nashashibis. Many villages were aligned with either political family. Generally, the Husseini family was more hostile and less accommodating to the Israelis. In our main measure of social cohesion, we treat whether all the clans within a village are aligned with the Husseinis or the Nashashibis. If all the clans are aligned with one family, we add a 1 to the index, and 0 otherwise. There are two potential concerns with this approach, which we address here. It is possible that because the families had different orientations toward the Israelis, that the effect of social cohesion on preemptive evacuation may be conditional on the political alignment of the family. For example, it may be the case that social cohesion may lead to preemptive evacuation when the village clans are affiliated with the Nashashibis, and that social cohesion may lead to violent depopluation when the village clans are affiliated with the Husseinis. This would be consistent with a story where these national families are shaping village level behavior, such as the Husseini family (and the Grand Mufti) told villages to remain and fight the oncoming Israelis. The other concern is that the effect of social cohesion on preemptive evacuation may be driven by the affiliation of all the village clans toward a specific political family, and not be driven by a unified political alignment of the clans in general. This would be the case if the social cohesion index is correlated with preemptive depopulation only when all the village families are affiliated with the Nashashibis, but not the Husseinis (or vice versa).

We present the results of three sets of models below. The first set of models uses a social cohesion index that is identical to the main social cohesion index, but adds in a value of 1 if all the clans in a village are affiliated with the Nashashibi family, and 0 if otherwise (instead of adding a value of 1 if all the clans in a village are affiliated with the Nashashibi *or* the Husseini families). The second set of models adds in a value of 1 if all the clans in a village are affiliated with the Husseini family, and 0 if otherwise. The third set of models examines the effect of this single component of the additive index. These are simple binary variables indicating if all the clans in a village are affiliated with the Husseini family, and another model where the variable indicates if all the clans in the village are affiliated with the Nashashibis.

While not all the coefficients are statistically significant (which is to be expected, since we are throwing out information and collapsing important variation), the estimated effects are consistent across the three measures. Table 4 is especially informative: whether all clans are affiliated with the Husseinis or the Nashashibis has no substantial effect on preemptive depopulation. The important factor is whether all clans are affiliated with the *same* political party, not whether they are affiliated with a *specific* political party.

Table 7: (Nashashibi only) Social Cohesion and Preemptive Evacuation

	Dependent variable:			
	Preemptive Evacuation			
	(1)	(2)	(3)	(4)
Social Cohesion (Nashashibi)	0.354	0.417	0.375	−4.503**
	(0.228)	(0.270)	(0.278)	(2.196)
Village Pop			−0.162	−1.029**
			(0.230)	(0.451)
Dist. to Viol. Depop.		0.274	0.245	0.298
		(0.275)	(0.277)	(0.278)
Mean Elevation		−0.380**	−0.384**	−0.443**
		(0.175)	(0.178)	(0.186)
Land Area		−0.060	−0.008	−0.023
		(0.240)	(0.257)	(0.264)
Total Cultivatable Land		0.065	0.128	−0.043
		(0.231)	(0.254)	(0.286)
Social Cohesion (Nashashibi) * Village Pop				0.742**
				(0.331)
Observations	249	231	231	231
Note:				*p<0.1; **p<0.05; ***p<0.01

Table 8: (Husseini only) Social Cohesion and Preemptive Evacuation

	Dependent variable:			
	Preemptive Evacuation			
	(1)	(2)	(3)	(4)
Social Cohesion (Husseini)	0.314	0.463*	0.432*	−5.222***
	(0.194)	(0.239)	(0.245)	(1.973)
Village Pop			−0.142	−1.329***
			(0.232)	(0.465)
Dist. to Viol. Depop.		0.230	0.208	0.287
		(0.273)	(0.275)	(0.280)
Mean Elevation		−0.396**	−0.397**	−0.474**
		(0.176)	(0.179)	(0.190)
Land Area		−0.069	−0.024	0.004
		(0.242)	(0.258)	(0.272)
Total Cultivatable Land		0.126	0.180	−0.034
		(0.239)	(0.261)	(0.292)
Social Cohesion (Husseini) * Village Pop				0.872***
				(0.301)
Observations	249	231	231	231
Note:				*p<0.1; **p<0.05; ***p<0.01

Table 9: Husseini and Nashashibi affiliation and Preemptive Evacuation

	Dependent variable:	
	Preemptive Evacuation	
	(1)	(2)
All Husseini	0.733*	
	(0.421)	
All Nashashibi		0.967
		(0.737)
Village Pop	−0.202	−0.221
	(0.226)	(0.223)
Dist. to Viol. Depop.	0.162	0.232
	(0.271)	(0.274)
Mean Elevation	−0.433**	−0.413**
	(0.178)	(0.175)
Land Area	−0.055	−0.025
	(0.261)	(0.258)
Total Cultivatable Land	0.152	0.076
	(0.254)	(0.248)
Observations	231	231
Note:	*p<0.1; **p<0.05; ***p<0.01	

C.3.8 Social Cohesion Index

For the measure of clan structure, we adopt the Laakso-Taagepera (L-T) index (Laakso and Taagepera 1979), designed to measure the effective number of political parties in a party system:

$N = \frac{1}{\sum_{i=1}^{n} p_i^2}$;

where N represents the effective number of clans and p_i represents clan i's population as a proportion of the population including all clans. When the effective number of clans as measured by the L-T index is low (close to 1), social cohesion is higher.

References

Braumoeller, Bear F, Giampiero Marra, Rosalba Radice, and Aisha Bradshaw. 2018. "Flexible Causal Inference for Political Science." *Political Analysis* 26(1): 54–71.

Chetrit, Shlomi. 2013. "The Special Night Squads (SNS) During the Arab Revolt – Unit Characteristics and Doctrine According to Military Operations in Kawkab Al- Hawa, Dabburiya and Khirbat Lidd-Al-'Awaddin." PhD thesis.

Hainmueller, Jens, Jonathan Mummolo, and Yiqing Xu. 2019. "How Much Should We Trust Estimates from Multiplicative Interaction Models? Simple Tools to Improve Empirical Practice." *Political Analysis* 27(2): 163–92.

Hughes, Matthew. 2015. "Terror in Galilee: British-Jewish Collaboration and the Special Night Squads in Palestine During the Arab Revolt, 1938–39." *The Journal of Imperial and Commonwealth History* 43(4): 590–610.

Laakso, Markku, and Rein Taagepera. 1979. "'Effective' Number of Parties: A Measure with Application to West Europe." *Comparative political studies* 12(1): 3–27.

Sekhon, Jasjeet S. 2008. "Multivariate and Propensity Score Matching Software with Automated Balance Optimization: The Matching Package for r." *Journal of Statistical Software*.