Assignment 2: Hurricane Katrina Evacuees

Andrew Heiss (working with Jessica Lewis, Michelle Welch, Ben McCartney, and Jennifer Shen)

PubPol 604 • September 30, 2013

All the code and logs for this assignment are available at https://github.com/andrewheiss/Causal-inference-code.

1. Evacuee Outcomes in March 2006

In 2006, people who were evacuated from the New Orleans area during Hurricane Katrina had varying employment and health outcomes. On average, evacuees worked 33.6 weeks during the previous year (with a 95% confidence interval (CI) of 31.3–35.8 weeks). However, as seen in Figure 1, the distribution is partially bimodal, with around a third of evacuees working between 0–10 weeks, and two-thirds working between 40–50 weeks, suggesting that the impact of the evacuation is uneven across the sample of evacuees. Evacuees earned an average of $22,586 (95% CI [$20,009, $25,164]) and collected an average of $129 (95% CI [$63, $194]) in unemployment benefits during the previous year. Both of these distributions are unimodal and skewed to the right. Finally, over 80% of evacuees reported to be in good health or better, with only 5% reporting poor health (see Figure 1).

However, beyond noticing the bimodal distribution of weeks worked, it is difficult to make any sort of inference regarding the employment or health outcomes of Hurricane Katrina evacuees without any sort of comparison group or counterfactual.

**Figure 1: Evacuee Outcomes, March 2006**

2. Comparison of 2005 and 2006 Outcomes for Evacuees

Better information about the impact of the evacuation can be found by comparing employment and health outcomes of evacuees before and after the hurricane (see Figure 2). Evacuees appear to have worked fewer weeks on average (36.1 weeks in 2005, 33.6 weeks in 2006), but the difference is statistically insignificant (t = 1.38, p = 0.169). Similarly, the average annual income of evacuees did not change significantly before and after the hurricane evacuation ($23,609 in 2005, $22,586 in 2006, t = 0.40, p = 0.686). Health outcomes tracked each other closely between the two years and are also not significantly different (χ^2 = 2.74, p = 0.602).

However, the amount of unemployment insurance compensation is significantly impacted by hurricane evacuation. On average, those who were destined for evacuation received around $37 (95% CI [$0, $74]) in unemployment benefits in the year preceding Katrina. That compensation increased to an average of $129 (95% CI [$63.50, $194]) in the year following the hurricane—a significant increase of $92 (t = -2.39, p = 0.017). It tentatively seems that being evacuated is correlated with job loss and extended unemployment.

Figure 2: Evacuee Outcomes in 2005 and 2006
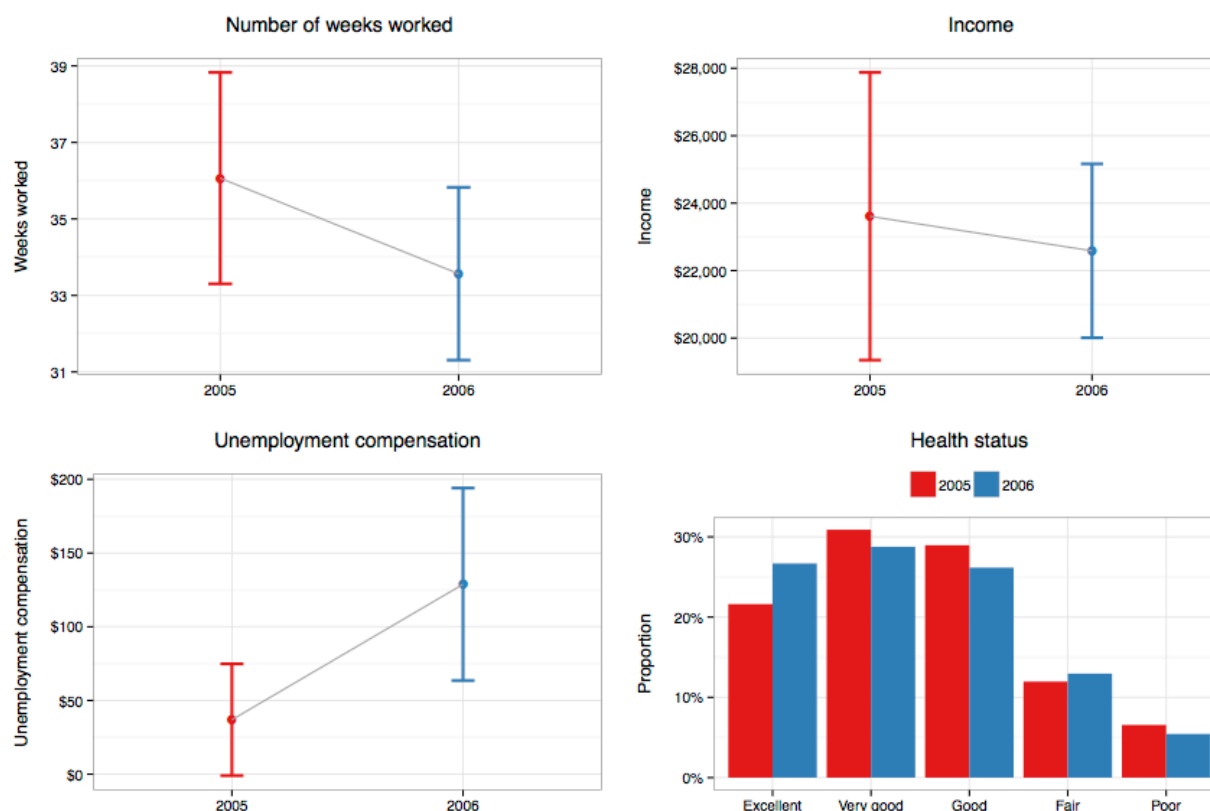
We can control for factors such as age, race, sex, and education to ensure that these characteristics are not driving the association between evacuation and employment and health outcomes. Table 1 (in the appendix) presents the OLS and ordered logistic regression results for weeks worked, income, unemployment compensation, and health status. When controlling for other evacuee characteristics, the effect of the evacuation (as measured by the "year" coefficient)

continues to be significant only for unemployment income, indicating a nearly $70 increase in compensation between 2005 and 2006 (t = 2.26, p = 0.0245). Though age and race are significantly associated with the number of weeks worked and health outcomes, and education is associated with changes in the number of weeks worked, health outcomes, and income, the actual evacuation has no significant impact on any other dependent variable.

The strength of these models is perhaps most easily understood visually. Figure 3 shows the predictions of the main economic and health outcomes, following 1,000 Monte Carlo draws from the model coefficient estimates given a hypothetical modal evacuee: a 21-year-old black male with a high school diploma. The simulated OLS results show violin plots of the results, where the distribution of the number of weeks worked or income is rotated counterclockwise and mirrored. The plot of health outcomes shows the predicted probabilities of the hypothetical respondent reporting different levels of health.

Figure 3: Simulated 2005 and 2006 Evacuee Outcomes for a hypothetical 21-year-old black male with a high school diploma (300 draws for health status, 1,000 draws for all others)
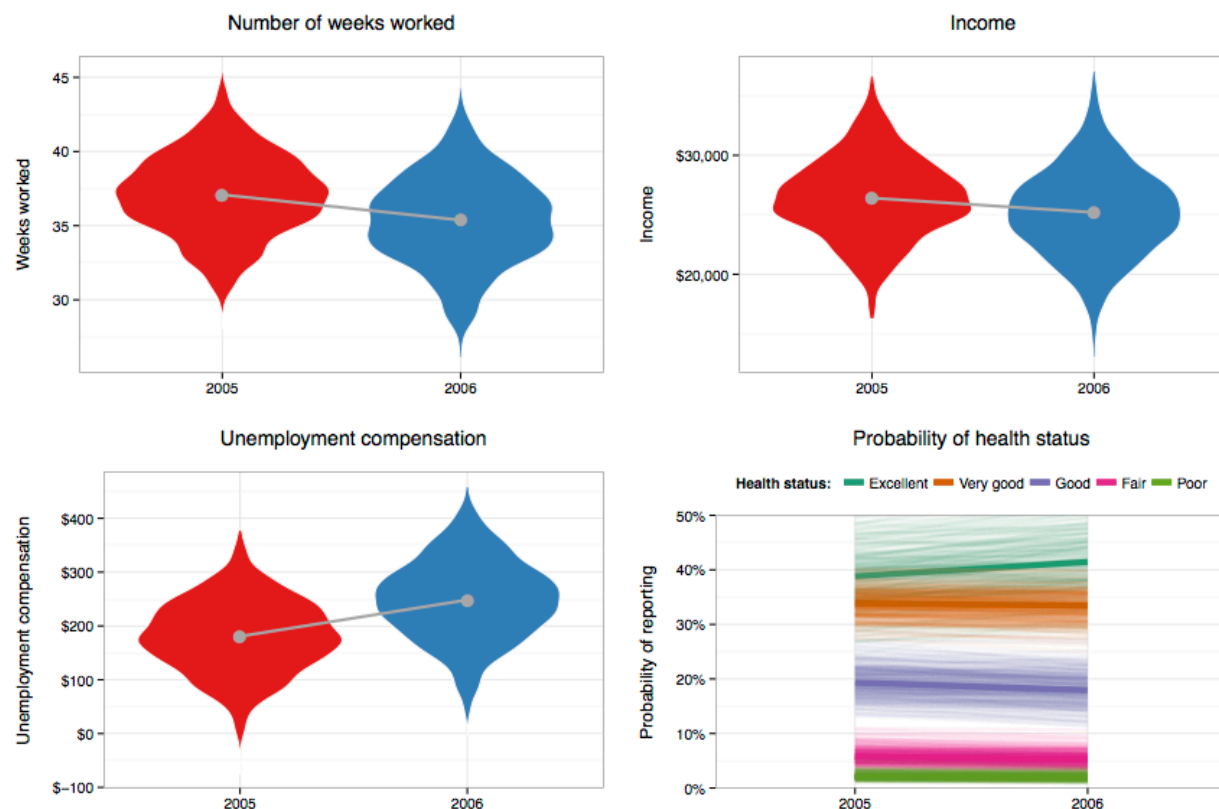
As seen in the plots, there is no visual difference in the number of weeks worked or income earned before and after the evacuation—the bulk of each of the distributions clearly overlap, making the effect indistinguishable from zero. Similarly, the probability of reporting any given health status is virtually the same in both 2005 and 2006, with "excellent" and "very good" most likely and equiprobable, and "fair" and "poor" highly improbable. Most interesting, however, is the difference in unemployment compensation. Though we have seen that the evacuation ostensibly has a significant impact on the amount of unemployment insurance evacuees

received, the two simulated distributions still have substantial overlap, meaning that the effect—though significant—may be overstated by simply looking at a regression table.

Thus, looking only at the 2005 and 2006 outcomes of evacuees, we can safely say that the evacuation had little impact on the number of weeks worked, income, and health, and some small significant impact on the amount of unemployment income received.

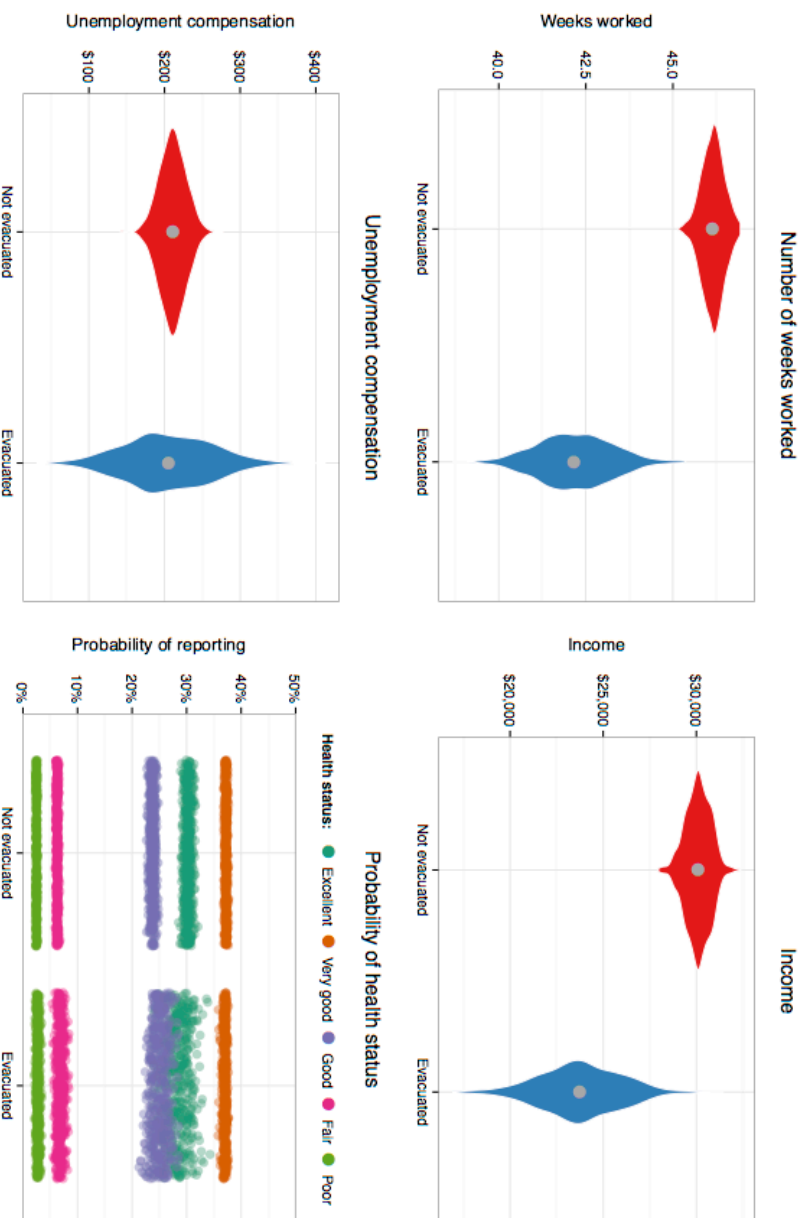## 3. Comparison of 2006 Outcomes for Evacuees and Non-Evacuees

Rather than compare evacuee outcomes between 2005 and 2006, we can compare post-evacuation outcomes for non-evacuees and evacuees (see Figure 4). In contrast with 2005 and 2006 evacuee outcomes, evacuees and non-evacuees have clearly different employment outcomes. Non-evacuees worked an average of nearly 39 weeks in 2006 (95% CI[38.8, 39.1]), while evacuees only worked 33.6 weeks (95% CI [31.3, 35.8]), resulting in a significant difference of 5.4 weeks, or more than a month of forgone work (95% CI [3.1, 7.7], t = 4.68, p < 0.001). Similarly, evacuees had substantially lower incomes than non-evacuees. Evacuees earned $22,586 (95% CI [$20,009, $25,164]) following the hurricane, nearly $10,000 less ($9,781, 95% CI [$7,182, $12,379], t = 7.4, p < 0.001) than the $32,366 (95% CI [$32,038, $32,694]) their non-evacuating counterparts earned during the same time. Evacuees also reported lower health outcomes in 2006, with significantly higher proportions of "fair" and "poor" responses (χ^2 = 10.33; χ^2 = 3.52) and lower proportions of "very good" responses (χ^2 = 3.33; χ^2 = 17.65 for the whole health variable). Only unemployment compensation appears to be unaffected by the evacuation—as is readily apparent in Figure 4, there is no significant difference (difference = 0.02, t < 0.001, p = 0.99) in average insurance income between evacuees and non-evacuees.

Figure 4: Non-Evacuee and Evacuee Outcomes in 2006
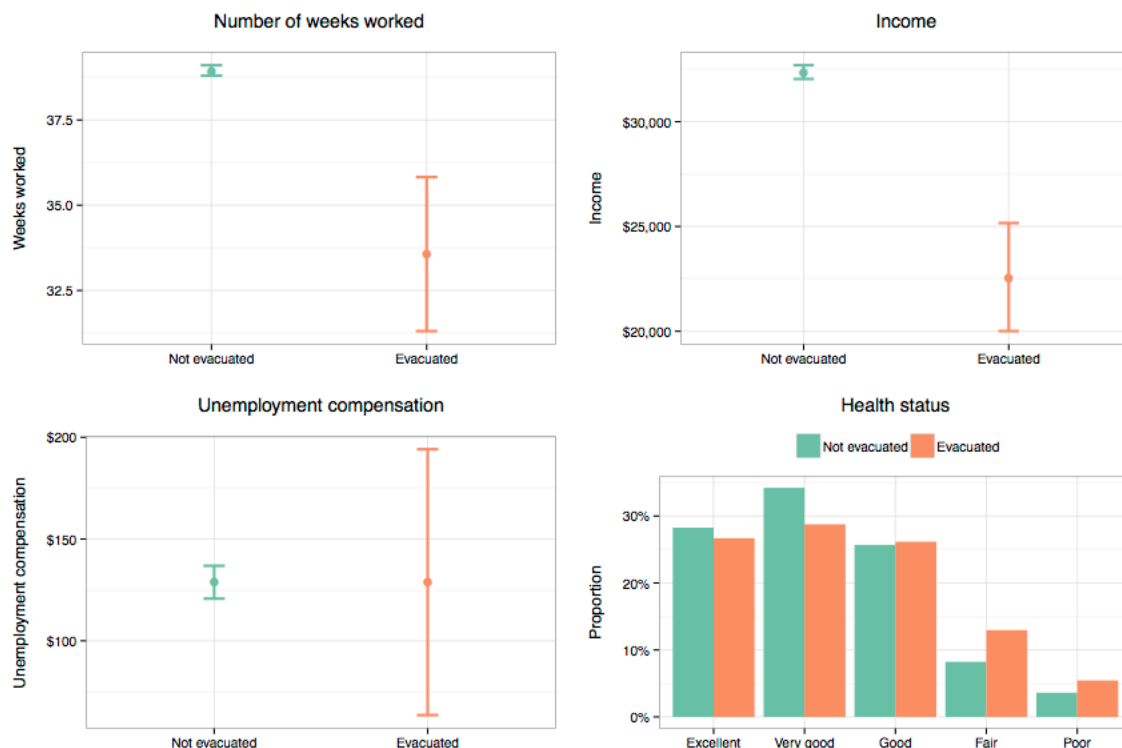
Though it appears that evacuation had a sizable impact on these employment and health outcomes, factors such as age, race, gender, and education also have an effect. As seen in Table 2 (in the appendix), race, gender, and education are all significantly associated with changes in outcomes—in general, black females work and earn less than white males, and those with more education work and earn more. Even after controlling for other factors, though, the hurricane evacuation is still significantly associated with fewer weeks worked and lower annual income. Evacuees worked 4 fewer weeks (t = -3.9, p < 0.001) and earned nearly $6,400 less (t = -3.01, p = 0.003) than those who were not evacuated. The evacuation is not associated with any significant difference in health status or unemployment compensation. Monte Carlo simulation confirms these findings—Figure 4 shows a clear visual difference between non-evacuee and evacuee income and employment. Average unemployment compensation is unchanged between the two groups, and the only change in health status is that hypothetical evacuees are equally likely to report "excellent" or "good" health.

Figure 5: Simulated Non-Evacuees and Evacuees in 2006 for a hypothetical 21-year-old black male with a high school diploma (500 draws for health status, 1,000 draws for all others)

Consequently, by looking at the 2006 outcomes of non-evacuees and evacuees, we can potentially infer that evacuating from Hurricane Katrina had a negative impact on employment and income.

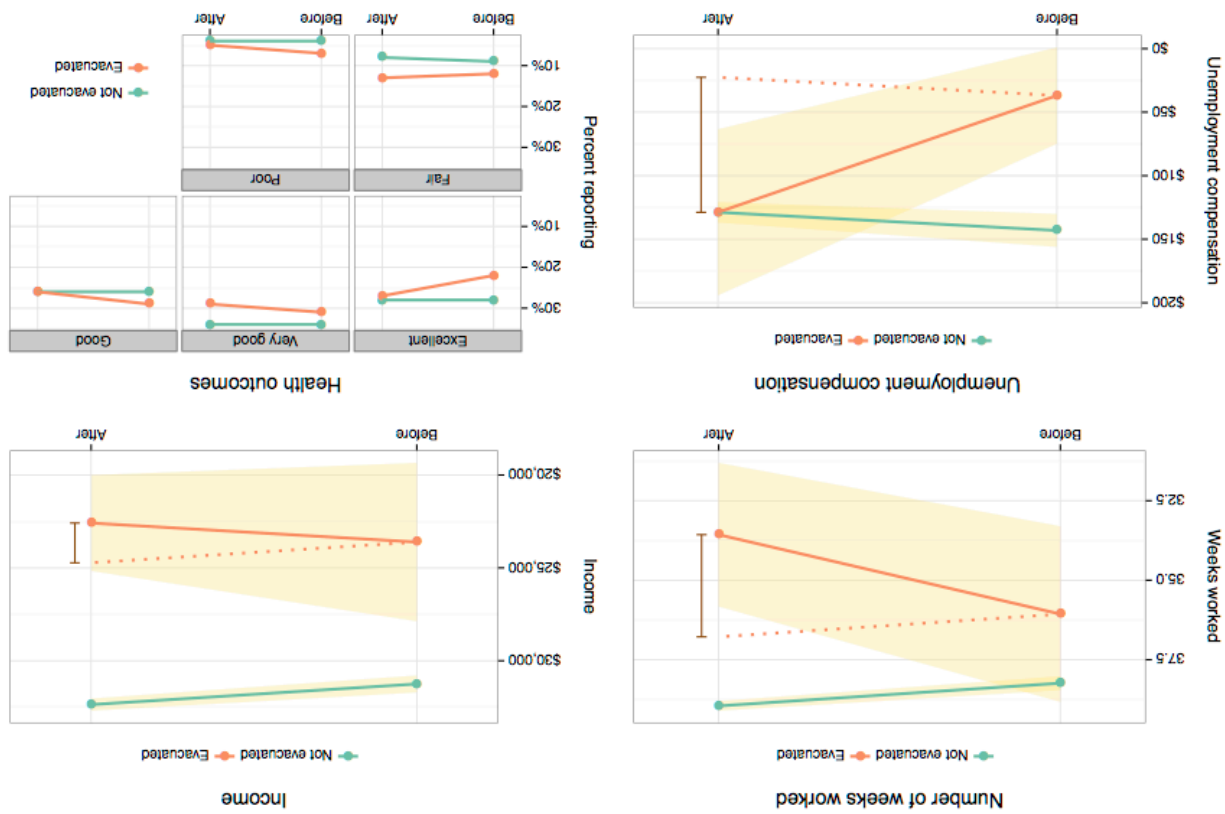# 4. Comparison of 2005 and 2006 Outcomes for Evacuees and Non-Evacuees

So far, we have looked at the potential impact of evacuation by comparing the 2005 and 2006 outcomes of evacuees and by comparing the 2006 outcomes of evacuees and non-evacuees, both of which suggested different types of statistically significant relationships between employment and health outcomes. We can combine these analyses and more precisely estimate the impact of evacuation by comparing the outcomes of non-evacuees and evacuees in both 2005 and 2006 with a difference-in-difference style estimate. This allows us decompose the data and remove (1) the secular time trend and (2) the inherent differences between evacuees and non-evacuees, thus isolating the precise effect of the evacuation on hurricane evacuees. The difference-in-difference estimation was done with the following model, the results of which are presented in Table 3 (in the appendix):

$$\text{Outcome} = \beta_0 + \beta_1 \text{After storm} + \beta_2 \text{Evacuated} + \beta_3 \text{After storm} \times \text{Evacuated} + \text{Controls}$$

The effect of the evacuation is most easily seen on a graph. Figure 6 shows the outcomes for both 2005 and 2006 for evacuees and non-evacuees. The treatment effect, or β_3 in the model, is the vertical distance between the actual level of the outcome for evacuees and the hypothetical level in the absence of the evacuation (represented by the dotted orange line) assuming the evacuees followed a trend parallel to the non-evacuees. If the treatment effect is statistically significant, it indicates that the evacuation itself had a clearly distinguishable impact on the difference in outcome.

Though Figure 6 does show a potentially large treatment effect for the number of weeks worked and unemployment compensation, only unemployment income is significant at a 95% confidence level (t = 2.72, p = 0.007), while weeks of employment is only significant at a 90% confidence level (t = -1.77, p = 0.077), indicating that the evacuation only had a strong causal impact on unemployment compensation, increasing claims by $106 on average. The evacuation had no impact on income or health.

When controlling for age, race, gender, and education, the relationship generally persists—the treatment effect remains statistically significant for unemployment compensation, resulting in an average increase in \$114 ($t = 2.92$, $p = 0.003$). However, the effect of the evacuation disappears for the number of weeks worked and is only significant at an 85% confidence level ($t = -1.48$, $p = 0.14$). Despite the fact that the evacuation effect is significant for one outcome, almost every control variable explains the variation in outcomes better than the treatment effect. This is visible in Figure 7, a Monte Carlo simulation using the same modal hypothetical person: the treatment effect tends to be within the distribution of simulated outcomes, thus demonstrating the relative weakness of the significance.

Figure 6: Non-Evacuee and Evacuee Outcomes in 2005 and 2006

Figure 7: Simulated Non-Evacuee and Evacuee Outcomes in 2005 and 2006 for a hypothetical 21-year-old black male with a high school diploma (500 draws for health status, 1,000 draws for all others)
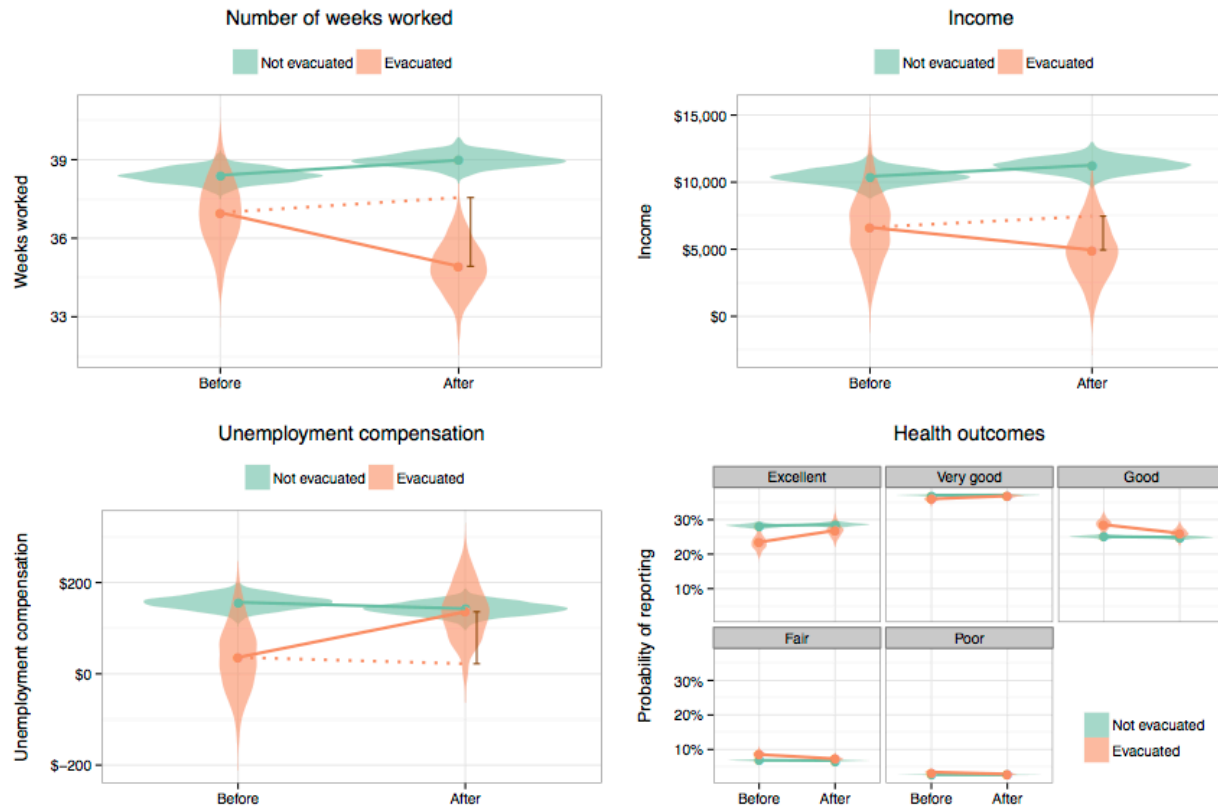
Therefore, by using difference-in-difference regression to compare the before- and after-storm outcomes of non-evacuees and evacuees, we can safely infer that evacuation had a positive impact on the amount of unemployment compensation people requested following the storm (meaning they received more), but had no impact on the amount of time spent working, income, or health.

5. Intent-to-Treat and Treatment-on-the-Treated Effects

Part A: Effect on noncompliers

Before declaring that the evacuation had a clear causal increase in the amount of unemployment compensation received, we first need to determine if the treatment, or actual evacuation, only had an impact on those who complied with the treatment (i.e. those who evacuated and have not yet returned to their houses in 2006) and not on noncompliers (i.e. those who have returned home or who deny ever having evacuated). To capture the noncomplier effect we can decompose the difference-in-difference interaction term used previously into two parts: complier x after the storm and noncomplier x after the storm. The coefficient of these new terms indicates the effect of the treatment on each of these groups.

As seen in Table 4 (in the appendix), this new interaction term is insignificant for weeks worked, income, and health, indicating that the change due to the evacuation (however limited it might be, considering that the evacuation had no significant impact on these variables anyway) is limited to those who complied with the treatment, or remained long-term evacuees. The complier interaction term is significant for unemployment compensation, though, meaning that despite the fact that the evacuation was tentatively causally linked to an increase in unemployment income, it is impossible to determine if that change was due to the evacuation itself or due to remaining evacuated in the long-term.

Part B: Treatment-on-the-Treated Effect

Given the discovery that the evacuation had different effects on compliers and noncompliers, we cannot estimate the effect on the evacuation on long-term evacuees directly. Instead, the results in question 4 represent intention-to-treat (ITT) effects, demonstrating the impact of the evacuation in general. Instead of looking at ITT effects, we can calculate the treatment-on-the-treated (TOT) effect to determine the precise impact of the evacuation on long-term evacuees.

The TOT effect is calculated by dividing the change in treatment outcomes by the change in probability of complying with the treatment, or

$$\text{TOT} = \frac{\text{Outcome}_\text{treatment} - \text{Outcome}_\text{control}}{P(\text{Compliance})_\text{treatment} - P(\text{Compliance})_\text{control}}.$$

The numerator of the equation is equal to the interaction coefficients calculated in question 4, or the ITT. The denominator of the equation is based on the probability of compliance in the different experimental groups. Of the 386 people in the sample who were evacuated, 97 remained evacuees in 2006, implying that there is a 25% chance that evacuees become long-term evacuees, or a 25% chance of complying with the treatment. Because of the design of the natural experiment, there is no chance that non-evacuees became long-term evacuees in 2006. The TOT effect on for long-term evacuees is thus

$$\text{TOT} = \frac{\text{ITT}}{0.2513}.$$

The TOT effects for the four main outcomes are presented in Table 5 (though only unemployment compensation is causally impacted by the evacuation). By taking noncompliance into account, we can thus see that the evacuation increases unemployment income by an average of $455 *for long-term evacuees*, or those who receive the actual treatment, which is a much larger effect than what was found in question 4.

Table 5: ITT and TOT effects for each outcome

Outcome	ITT	TOT
(Weeks worked)	-2.569	-10.222
(Income)	-2,554.218	-10,164.210
Unemployment compensation	114.396	455.226
(Health status)	-0.165	-0.655

Appendix

Table 1: 2005 and 2006 outcomes for evacuees

	Dependent variable:			
	Weeks worked	Earnings	Unemployment income	Health
	OLS	OLS	OLS	Ordered logistic
	(1)	(2)	(3)	(4)
Year (2006)	−1.652	−1,261.200	69.778[*]	−0.110
	(1.223)	(1,679.517)	(27.271)	(0.104)
Age	−0.157[*]	−146.807	−2.333	0.049[***]
	(0.075)	(95.730)	(1.568)	(0.006)
Race (Black)	−4.492[*]	−2,397.509	89.074	0.482[**]
	(1.998)	(1,732.053)	(62.868)	(0.163)
Sex (Female)	−7.727[***]	−13,940.650[***]	2.141	0.189
	(1.726)	(2,281.671)	(36.128)	(0.146)
High school	6.616[*]	4,934.234[*]	44.035	−0.341
	(2.886)	(1,970.417)	(49.687)	(0.235)
Some college	6.426[**]	6,362.530[**]	75.841	0.122
	(2.125)	(2,189.071)	(57.718)	(0.180)
Bachelor's degree	−0.807	6,001.959	−28.704	−0.943[***]
	(2.717)	(3,121.825)	(102.675)	(0.247)
Post-graduate work	4.507	28,001.180[**]	−86.681	−0.131
	(3.243)	(8,608.237)	(85.260)	(0.319)
Constant	46.075[***]	40,902.670[***]	95.270	
	(4.923)	(7,016.240)	(109.544)	
Observations	645	645	645	645
R^2	0.105	0.218	0.022	
Adjusted R^2	0.094	0.209	0.009	
Residual Std. Error (df = 636)	21.501	26,443.790	540.698	
F Statistic (df = 8; 636)	9.354[***]	22.218[***]	1.757	

[*]p<0.05; [**]p<0.01; [***]p<0.001; OLS models use robust standard errors

Table 2: 2006 outcomes for evacuees and non-evacuees

	Dependent variable:			
	Weeks worked	Earnings	Unemployment income	Health
	OLS	*OLS*	*OLS*	*ordered logistic*
	(1)	(2)	(3)	(4)
Evacuated	−3.996***	−6,387.017***	−4.453	0.086
	(1.098)	(1,198.395)	(33.441)	(0.095)
Age	−0.244***	3.500	−0.171	0.036***
	(0.007)	(12.597)	(0.324)	(0.001)
Race (Black)	−1.347***	−2,895.401***	31.001*	0.441***
	(0.272)	(383.704)	(13.976)	(0.023)
Sex (Female)	−7.787***	−19,407.870***	−62.915***	0.103***
	(0.147)	(311.532)	(8.339)	(0.014)
High school	8.642***	9,912.908***	20.048	−0.609***
	(0.294)	(298.188)	(12.860)	(0.024)
Some college	2.613***	7,060.417***	6.016	−0.335***
	(0.197)	(290.304)	(11.462)	(0.018)
Bachelor's degree	1.875***	15,197.090***	−48.776***	−0.498***
	(0.203)	(489.667)	(11.748)	(0.020)
Post-graduate work	1.615***	19,452.690***	−25.665	−0.232***
	(0.247)	(985.681)	(14.866)	(0.026)
Constant	51.758***	42,447.640***	224.980***	
	(0.441)	(715.215)	(19.991)	
Observations	73,343	73,343	73,343	73,343
R^2	0.088	0.151	0.001	
Adjusted R^2	0.088	0.151	0.001	
Residual Std. Error (df = 73334)	20.020	41,576.650	1,108.324	
F Statistic (df = 8; 73334)	888.450***	1,629.430***	12.552***	

*p<0.05; **p<0.01; ***p<0.001; OLS models use robust standard errors

Table 3: 2005 and 2006 outcomes for evacuees and non-evacuees

	Dependent variable:							
	Weeks worked		Earnings		Unemployment income		Health	
	OLS		*OLS*		*OLS*		*ordered logistic*	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
2006	0.716***	0.574***	1,116.547***	827.497**	−14.347	−14.232	−0.032**	−0.016
	(0.136)	(0.130)	(286.996)	(265.085)	(7.798)	(7.782)	(0.012)	(0.012)
Evacuated	−2.172	−1.400	−7,640.662***	−3,799.238	−106.216***	−119.123***	0.391***	0.249*
	(1.405)	(1.341)	(2,174.313)	(1,952.132)	(20.327)	(20.914)	(0.113)	(0.113)
Treatment (2006 * evacuated)	−3.214	−2.569	−2,139.839	−2,554.218	106.236**	114.396**	−0.129	−0.165
	(1.815)	(1.732)	(2,543.577)	(2,286.768)	(39.123)	(39.155)	(0.147)	(0.148)
Age		−0.273***		−21.378*		−0.187		0.037***
		(0.006)		(10.224)		(0.303)		(0.001)
Race (Black)		−1.453***		−3,224.639***		36.243*		0.446***
		(0.224)		(305.350)		(15.191)		(0.019)
Sex (Female)		−7.723***		−19,242.020***		−73.006***		0.098***
		(0.121)		(253.225)		(7.184)		(0.011)
High school		8.692***		9,887.709***		17.138		−0.654***
		(0.238)		(246.068)		(12.888)		(0.019)
Some college		2.741***		6,925.844***		14.291		−0.323***
		(0.161)		(237.463)		(9.644)		(0.014)
Bachelor's degree		1.693***		14,909.540***		−54.539***		−0.493***
		(0.168)		(398.505)		(9.857)		(0.016)
Post-graduate work		1.601***		19,277.150***		−23.612*		−0.240***
		(0.205)		(807.102)		(11.852)		(0.022)
Constant	38.238***	52.335***	31,250.160***	42,705.840***	143.143***	254.082***		
	(0.112)	(0.370)	(233.176)	(608.262)	(6.625)	(20.299)		
Observations	110,011	110,011	110,011	110,011	110,011	110,011	110,011	110,011
R^2	0.0005	0.092	0.0004	0.149	0.00005	0.002		
Adjusted R^2	0.0005	0.092	0.0003	0.149	0.00002	0.001		
Residual Std. Error	21.111 (df = 110007)	20.121 (df = 110000)	44,887.240 (df = 110007)	41,416.940 (df = 110000)	1,161.520 (df = 110007)	1,160.667 (df = 110000)		
F Statistic	18.274*** (df = 3; 110007)	1,115.710*** (df = 10; 110000)	13.625*** (df = 3; 110007)	1,926.219*** (df = 10; 110000)	1.833 (df = 3; 110007)	17.424*** (df = 10; 110000)		

*p<0.05; **p<0.01; ***p<0.001; OLS models use robust standard errors

Table 4: Effect of evacuation on compliers and noncompliers

	Dependent variable:			
	Weeks worked	Earnings	Unemployment income	Health
	OLS	OLS	OLS	ordered logistic
	(1)	(2)	(3)	(4)
Noncomplier*2006	−0.183	−902.126	95.143[*]	−0.244
	(1.828)	(2,416.017)	(37.582)	(0.157)
Complier*2006	−9.675[***]	−7,476.259[**]	171.756	0.097
	(2.533)	(2,804.015)	(95.535)	(0.226)
Treatment (2006 * evacuated)	52.346[***]	42,713.280[***]	253.995[***]	
	(0.370)	(608.269)	(20.302)	
2006	0.574[***]	827.468[**]	−14.231	−0.016
	(0.130)	(265.086)	(7.782)	(0.012)
Ever evacuated	−1.402	−3,800.670	−119.106[***]	0.249[*]
	(1.341)	(1,952.130)	(20.914)	(0.113)
Age	−0.273[***]	−21.514[*]	−0.186	0.037[***]
	(0.006)	(10.224)	(0.303)	(0.001)
Race (Black)	−1.443[***]	−3,217.866[***]	36.164[*]	0.446[***]
	(0.224)	(305.386)	(15.189)	(0.019)
Sex (Female)	−7.722[***]	−19,241.660[***]	−73.010[***]	0.098[***]
	(0.121)	(253.224)	(7.184)	(0.011)
High school	8.686[***]	9,883.548[***]	17.186	−0.654[***]
	(0.238)	(246.075)	(12.890)	(0.019)
Some college	2.746[***]	6,928.722[***]	14.257	−0.323[***]
	(0.161)	(237.463)	(9.645)	(0.014)
Bachelor's degree	1.689[***]	14,907.080[***]	−54.511[***]	−0.493[***]
	(0.168)	(398.517)	(9.858)	(0.016)
Post-graduate work	1.604[***]	19,279.580[***]	−23.640[*]	−0.240[***]
	(0.205)	(807.113)	(11.853)	(0.022)
Observations	110,011	110,011	110,011	110,011
R^2	0.092	0.149	0.002	
Adjusted R^2	0.092	0.149	0.001	
Residual Std. Error (df = 109999)	20.119	41,416.780	1,160.671	
F Statistic (df = 11; 109999)	1,015.891[***]	1,751.288[***]	15.869[***]	

[*] p<0.05; [**] p<0.01; [***] p<0.001; OLS models use robust standard errors