

**Environmental Health Risks and Housing Values:  
Evidence from 1600 Toxic Plant Openings and Closings  
Supplementary Online Appendix**

Janet Currie

Lucas Davis

Michael Greenstone

Reed Walker<sup>1</sup>

Alternative Timing of Assigning Plant Activity to Birth Outcomes

Below we describe alternative methods of assigning birth outcomes to plant events due to the ambiguity within a year for when a plant opens/closes. In the body of the paper, we restrict our birth outcomes to births between November and December of year  $t-1$  and January and February of year  $t$ . This sample restriction reflects the fact that we do not observe *when* during the year a plant opens or closes. Census reporting occurs in March so we have a 12-month window during which each opening/closing could have occurred. By focusing on births towards the end of the year (or the beginning of the following year), we increase the likelihood of that birth being exposed to these events. The idea is that a baby born January 1, 2002 has not been exposed to any of the toxic plant activity for calendar 2002, but was exposed to toxic emissions in 9 out of 12 months of 2001. Similarly, a baby born in November 2001 was exposed to toxic emissions for 9 out of 12 months of 2001. In contrast, timing is much more difficult for births in the middle of the year. A June birth, for example, may or may not have been exposed to a change in plant operations depending on *when* during the year the change occurred. Restricting our sample to these four months also has the additional advantage of limiting the extent to which seasonality in plant activity or birth outcomes affects our findings.

We nonetheless investigated the sensitivity of our estimates to additional sample selection criteria in two separate ways. First, we created a dataset of plant-month-year observations for all births around a given plant in a given month-year. We then calculated the expected exposure of a birth to a plant opening or closing assuming that plant openings/closings are uniformly distributed by month throughout a year. This allows us to calculate the expected fraction of in utero exposure to a plant opening/closing based on month of birth. We use this expected fraction as a continuous measure of treatment for the birth-months immediately surrounding the plant. For birth-months long before/after the event, they are unambiguously treated or not-treated given their “distance” from the ambiguously timed event. A summary of this

---

<sup>1</sup> Currie: Department of Economics, Princeton University. Email: jcurrie@princeton.edu. Davis: Haas School of Business, University of California, Berkeley. Email: ldavis@haas.berkeley.edu. Greenstone: Department of Economics, University of Chicago. Email: mgreenst@uchicago.edu. Walker: Robert Wood Johnson Scholar of Health Policy, University of California, Berkeley. Email: rwalker@haas.berkeley.edu. Currie thanks the John D. and Catherine T. MacArthur Foundation and the Environmental Protection Agency (RE: 83479301-0) for support. We thank seminar participants at University of Delaware, Princeton, MIT, Stanford, UC Davis, RAND, and the University of Oregon as well as Trudy Cameron for helpful comments. The research in this paper was conducted while the authors were Special Sworn Status researchers of the U.S. Census Bureau with generous guidance from Jim Davis, Jonathan Fisher, and Maggie Levenstein. Research results and conclusions expressed are those of the author and do not necessarily reflect the views of the Census Bureau, the MacArthur Foundation, or the EPA. This paper has been screened to insure that no confidential data are revealed. Previous versions of this paper were circulated as “Toxic Pollutants and Infant Health: Evidence from Plant Openings and Closings” and “Do Housing Prices Reflect Environmental Health Risks? Evidence from More than 1600 Toxic Plant Openings and Closings.”

assignment mechanism is described in the tables below, where the last line of each table represents the expected fraction of gestation exposed to a plant opening/closing conditional on month of birth. This expected fraction makes up our treatment variable.

		Month of Birth for Year 2007 Opening				
		Apr-06	May-06	...	Oct-07	Nov-07
Month of Opening for Year 2007 Opening	Apr-06	1	2	...	9	9
	May-06	0	1	...	9	9
	Jun-06	0	0	...	9	9
	Jul-06	0	0	...	9	9
	Aug-06	0	0	...	9	9
	Sep-06	0	0	...	9	9
	Oct-06	0	0	...	9	9
	Nov-06	0	0	...	9	9
	Dec-06	0	0	...	9	9
	Jan-07	0	0	...	9	9
	Feb-07	0	0	...	9	9
	Mar-07	0	0	...	8	9
	Expected Number of Months Gestational Exposure	0.08	0.25	...	8.92	9.00
	Expected Fraction of Gestation	0.01	0.03	...	0.99	1.00

		Month of Birth for Year 2007 Closing				
--	--	--------------------------------------	--	--	--	--

		Apr-06	May-06	...	Oct-07	Nov-07
Month of Closing for Year 2007 Closing	Apr-06	9	8	...	0	0
	May-06	9	9	...	0	0
	Jun-06	9	9	...	0	0
	Jul-06	9	9	...	0	0
	Aug-06	9	9	...	0	0
	Sep-06	9	9	...	0	0
	Oct-06	9	9	...	0	0
	Nov-06	9	9	...	0	0
	Dec-06	9	9	...	0	0
	Jan-07	9	9	...	0	0
	Feb-07	9	9	...	1	0
	Mar-07	9	9	...	2	1
	Expected Number of Months Gestational Exposure	9	8.917	...	0.250	0.083
	Expected Fraction of Gestation	1	0.991	...	0.028	0.009

We also explored models that exclude ambiguous birth months altogether and simply look at pre-post outcomes for births/gestation clearly on either side of openings and closings. This has the added benefit of being somewhat agnostic about the prenatal production technology, whereas the continuous assignment mechanism above imposes some structure in the form of additive linearity across months in utero. In both this specification and in the “fractional apportionment” specification, we restricted the sample to an “event window” that limits the sample to 2 years before and after an event (i.e. as in columns (7) and (8) of the baseline table to avoid putting too much weight on observations far before and after the event. Results from both exercises are presented in Appendix Table A7.

#### A Note on Geocoding

Our housing transaction data provides information on the street address of the housing parcel which we use to geocode our data. We used Yahoo Geocoding API to match street address data with latitude and longitude via a Python routine. The Yahoo API provides much better geocoding capabilities in the case of

minor errors in street addresses relative to some more commonly used geocoding routines (e.g. ArcGIS). For example, if one were to type in a single street address without a city identifier into Yahoo, and that address was fairly unique (e.g. 2342 Bristol Lane, KY), Yahoo would more than likely be able to figure out where the exact location of the house was without city information. Using the Yahoo Geocoding API, we were able to match over 98% of our transactions to latitude and longitude. We geocoded the birth data using ArcGIS v9.3 geocoding engine combined with the US Census Bureau's Tiger/Line reference layer. For each state, we audited the match using external web services such as Google Maps, and we found that the ArcGIS match was accurate for nearly all of the randomly sampled birth addresses.

Table A1: Effect of Plant Operating Status on Ambient Air Pollution: Pollutant Specific Coefficients

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Benzene	Cumene	Dichloromethane	Lead (TSP)	Methylene Chloride	Nickel	Styrene	Toluene
Plant Open×Dist	-9.1502*** (2.3810)	-37.4841*** (8.5595)	-0.8801 (0.7433)	-0.6627 (6.0252)	0.2579 (1.3544)	-10.6628*** (0.8827)	-21.0906 (14.0195)	-1.9633 (4.9933)
Plant Open×Dist <sup>2</sup>	7.5063*** (2.2812)	28.2352*** (7.3815)	0.9816 (0.8171)	2.2785 (5.6212)	0.2332 (1.5556)	7.8313*** (0.9525)	20.9231 (13.3794)	0.5489 (4.2769)
Plant Open×Dist <sup>3</sup>	-2.5056*** (0.8857)	-8.7304*** (2.7271)	-0.3856 (0.3201)	-1.5783 (2.1001)	-0.2477 (0.6738)	-2.4739*** (0.3816)	-7.8927 (4.8879)	0.1807 (1.3857)
Plant Open×Dist <sup>4</sup>	0.2944** (0.1194)	0.9455** (0.3630)	0.0492 (0.0408)	0.2873 (0.2700)	0.0469 (0.0951)	0.2837*** (0.0509)	0.9772 (0.5901)	-0.0600 (0.1524)
Monitor Count	49	23	51	32	21	11	51	67
N	1106	358	1077	532	315	324	1000	1970

NOTES: This table reports regression coefficients from 8 separate regressions. The unit of observation is the monitor-plant pair and the dependent variable in all regressions is ambient hazardous pollution, standardized so each pollutant has mean 0 and standard deviation is 1. All regressions include monitor-pair fixed effect and an indicator variable for whether each plant is operating 1(Plant Operating). Standard errors are two-way clustered on plant and monitor.

Table A2: The Effect of Toxic Plants on Housing Values: Model Sensitivity to Alternative Controls

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	<b>Baseline</b>					<b>First Difference</b>			
1(Plant Operating) $\times$ 1(<0.5 Miles)	-0.027*** (0.008)	-0.032*** (0.007)	-0.030*** (0.007)	-0.025*** (0.007)	-0.025*** (0.008)	-0.034*** (0.008)	-0.020* (0.010)	-0.020** (0.010)	-0.014** (0.007)
N	34736	34736	34736	34736	34736	1114248	1114248	1114248	1114248
Plant $\times$ Distance-Bin FE	X	X	X	X	X				
Census Tract Quadratic Trends		X	X	X	X		X	X	X
State $\times$ Year FE			X					X	
County $\times$ Year FE				X					X
Plant $\times$ Year FE					X				

NOTES: This table reports regression coefficients from 9 separate regressions. The dependent variable in all regressions is housing values (in logs). In columns (1)-(5), the data have been aggregated to plant by distance by year cells, and regressions are weighted by the group-level cell size. In columns (6)-(9), we estimate the model on the set of houses for which we observe 2+ sales in our sample period. Each regression reports estimates of the effect of plant operating status on local housing values, where 1(Plant Operating) is a indicator variable equal to one for plants that have opened and/or have not yet closed. Standard errors two-way clustered by plant and year are in parentheses.

Table A3: The Effect of Toxic Plants on Local Housing Values and Low Birthweight: Using 2-4 Mile Radius Comparison Group

	0-0.5 Miles		0.5-1 Miles		0-1 Miles		0-1 Miles (+/- 2 years)	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
<b>Panel A: Estimated Effect of Plant Operation on Housing Values</b>								
1(Plant Operating) $\times$ 1(< 1 Mile)	-0.032*** (0.009)	-0.026*** (0.008)	-0.013** (0.007)	-0.017*** (0.006)	-0.018*** (0.007)	-0.019*** (0.006)	-0.011* (0.006)	-0.012*** (0.005)
N	34736	34736	34736	34736	34736	34736	30492	30492
PlantCount	2171	2171	2171	2171	2171	2171	2171	2171
<b>Panel B: Estimated Effect of Plant Operation on Low Birthweight</b>								
1(Plant Operating) $\times$ 1(< 1 Mile)	0.0001 (0.0011)	-0.0000 (0.0011)	0.0009* (0.0006)	0.0011** (0.0005)	0.0007 (0.0006)	0.0008 (0.0005)	0.0017** (0.0008)	0.0020*** (0.0008)
N	88922	88922	88922	88922	88922	88922	63301	62984
PlantCount	3438	3438	3438	3438	3438	3438	3438	3438
Plant $\times$ Distance-Bin FE	X	X	X	X	X	X	X	X
State $\times$ Year FE	X		X		X		X	
Plant $\times$ Year FE		X		X		X		X

NOTES: This table reports regression estimates from specifications that are identical to Table 2 (Panel A) and Table 4 (Panel A) except the comparison group is observations between 2 and 4 miles from a plant.

Table A4: The Effect of Toxic Plants on Local Housing Values and Low Birthweight, Distance Radius Specification

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	0-0.5	0-0.5	0.5-1	0.5-1	1-1.5	1-1.5	1.5-2	1.5-2	0-1	0-1	1-2	1-2
<b>Panel A: Estimated Effect of Plant Operation on Housing Values</b>												
1(Plant Operating)	-0.018** (0.009)	-0.018** (0.008)	-0.011 (0.007)	-0.013* (0.007)	-0.002 (0.006)	-0.006 (0.007)	0.004 (0.006)	-0.000 (0.006)	-0.015** (0.007)	-0.016** (0.007)	0.000 (0.006)	-0.003 (0.006)
N	17336	17336	17368	17368	17368	17368	15976	15976	17368	17368	17368	17368
PlantCount	2167	2167	2171	2171	2171	2171	1997	1997	2171	2171	2171	2171
<b>Panel B: Estimated Effect of Plant Operation on Low Birthweight</b>												
1(Plant Operating)	0.0004 (0.0011)	0.0003 (0.0012)	0.0012** (0.0005)	0.0012** (0.0005)	-0.0012* (0.0007)	-0.0012* (0.0007)	-0.0003 (0.0004)	-0.0002 (0.0005)	0.0009 (0.0006)	0.0009 (0.0006)	-0.0003 (0.0003)	-0.0003 (0.0003)
N	42077	42077	42140	42140	42114	42114	40556	40556	74890	74890	75002	75002
PlantCount	3254	3254	3256	3256	3256	3256	3136	3136	5787	5787	5794	5794
Plant×Distance-Bin FE	X	X	X	X	X	X	X	X	X	X	X	X
Year FE	X		X		X		X		X		X	
State×Year FE		X		X		X		X		X		X

NOTES: This table reports regression estimates from specifications that are identical to Table 2 (Panel A) and Table 4 (Panel A) except that each regression is estimated using only observations from the distance range indicated in the column headings. For example, the regressions described in columns (1) and (2) include only observations from within 0.5 miles of a plant.



Table A5: The Effect of Toxic Plants on Low Birthweight: Model Sensitivity to Alternative Controls

	(1)	(2)	(3)	(4)	(5)
<b>Estimated Effect of Plant Operation</b>					
1(Plant Operating) $\times$ 1(< 1 Mile)	0.0016*** (0.0006)	0.0016*** (0.0006)	0.0014** (0.0006)	0.0013** (0.0006)	0.0014** (0.0007)
N	88958	88958	88958	88958	88958
PlantCount	3438	3438	3438	3438	3438
Plant $\times$ Distance-Bin FE	X	X	X	X	X
Maternal Characteristics		X	X	X	X
Census Tract Quadratic Trends			X	X	X
State $\times$ Year FE				X	
Plant $\times$ Year FE					X

NOTES: This table reports regression coefficients from 9 separate regressions aimed at evaluating the sensitivity of the birth outcome estimates to alternative specifications. The dependent variable in all columns is the mean incidence of low birthweight where the data have been aggregated to plant by distance by year cells. Cell level averages have been adjusted for maternal characteristics including age, education, race, and smoking behavior, as well as for month of birth, birth order, and gender of child. See text for details. The mean incidence of low birthweight in our sample is 0.07. The comparison group in all columns is births between 1 and 2 miles from a plant. All columns control for census tract characteristics (interacted with quadratic trends) and regressions are weighted by the group-level cell size. Multiple births are dropped from regressions. Standard errors are two-way clustered by plant and year.

Table A6: The Effect of Toxic Plants on Maternal Characteristics

	(1) African American	(2) African American	(3) Hispanic	(4) Hispanic	(5) Mother's Education	(6) Mother's Education	(7) Teenage Mother	(8) Teenage Mother
1(Plant Operating) $\times$ 1(<1 Mile)	-0.007*** (0.002)	-0.006*** (0.002)	-0.002 (0.002)	-0.001 (0.002)	0.008 (0.010)	0.001 (0.009)	-0.001 (0.001)	-0.000 (0.001)
N	88062	88062	87438	87438	82608	82608	89302	89302
PlantCount	3438	3438	3438	3438	3438	3438	3438	3438
Mean	0.201	0.201	0.219	0.219	12.555	12.555	0.137	0.137
	(9) Smoker	(10) Smoker	(11) White College	(12) White College	(13) Predictive Index	(14) Predictive Index	(15) Fertility	(16) Fertility
1(Plant Operating) $\times$ 1(<1 Mile)	-0.001 (0.001)	0.002 (0.001)	0.002** (0.001)	0.001 (0.001)	0.001*** (0.000)	0.001*** (0.000)	0.002 (0.005)	-0.000 (0.004)
N	88077	88077	86093	86093	89388	89388	89388	89388
PlantCount	3438	3438	3438	3438	3438	3438	3438	3438
Mean	0.149	0.149	0.151	0.151	8.084	8.084	4.233	4.233
Plant $\times$ Distance-Bin FE	X	X	X	X	X	X	X	X
State $\times$ Year FE	X		X		X		X	
Plant $\times$ Year FE		X		X		X		X

NOTES: This table reports regression coefficients from 16 separate regressions, 8 per panel. The dependent variable is listed in the column heading, and data have been aggregated to plant by distance by year cells. Cell level averages have been adjusted for maternal characteristics. See text for details. The comparison group in all columns is births between 1 and 2 miles from a plant. The dependent variable "Predictive Index" is created by first running a regression of birthweight on observable characteristics of the mother. The predicted values from this regression are used as a summary index of demographic changes. Fertility is measured as the log number of births in a cell. Regressions are weighted by the group-level cell size (with the exception of "Fertility"). All regressions control for tract characteristics (excluding the dependent variable of interest), interacted with quadratic trends. Multiple births are dropped from regressions. Regressions are weighted by the group-level cell size. Standard errors two-way clustered by plant and year.

Table A7: The Effect of Toxic Plants on Low Birthweight: Alternative Timing Assumptions

	(1)	(2)	(3)	(4)
	Fractional Assignment		Outer Donut	
	0-1 (+/- 2 years)	0-1 (+/- 2 years)	0-1 (+/- 2 years)	0-1 (+/- 2 years)
<b>Panel A: Estimated Effect of Plant Operation</b>				
1(Plant Operating) $\times$ 1(< 1 Mile)	0.0010 (0.0007)	0.0011* (0.0006)	0.0005 (0.0007)	0.0010 (0.0007)
N	753329	753329	695179	695179
PlantCount	3438	3438	3438	3438
<b>Panel B: Estimated Effect of Plant Openings and Closings</b>				
1(Plant Opened) $\times$ 1(< 1 Mile)	0.0017** (0.0009)	0.0022*** (0.0008)	0.0018* (0.0010)	0.0025** (0.0011)
1(Plant Closed) $\times$ 1(< 1 Mile)	0.0000 (0.0008)	-0.0003 (0.0008)	0.0001 (0.0009)	-0.0003 (0.0009)
$H_0$ : Opening=-Closing (p-value)	0.087	0.056	0.103	0.097
N	753329	753329	695179	695179
Plant Count	3438	3438	3438	3438
Plant $\times$ Distance-Bin FE	X	X	X	X
State $\times$ Year FE	X		X	
Plant $\times$ Year FE		X		X

NOTES: This table reports regression coefficients from 8 separate regressions, 4 per panel. The dependent variable in all regressions is the mean incidence of low birthweight where the data have been aggregated to plant by distance by month-year cells. Cell level averages have been adjusted for maternal characteristics including age, education, race, and smoking behavior, as well as for month of birth, birth order, and gender of child. Columns (1) and (2) assign plant operating status “fractionally” based on the expected number of months a plant was open during the child’s gestational period  $\in [0, 1]$ . Columns (3) and (4) focus only on births in which there is no ambiguity as to whether the plant was opened or closed (i.e. dropping births just around the plant opening/closing event). In all columns we focus on the sample +/- 2 years from the plant opening event. Panel A estimates the effect of plant operating status on local birth outcomes, where 1(Plant Operating) is a indicator variable equal to one for plants that have opened and/or have not yet closed. Panel B estimates the asymmetric effect of plant openings/closings. Panel B reports p-values from tests that the two coefficients are equal in magnitude but of opposite sign. All columns control for census tract characteristics (interacted with quadratic trends) and regressions are weighted by the group-level cell size. Multiple births are dropped from regressions. Standard errors are two-way clustered by plant and year.

Table A8: The Effect of Toxic Plants on Birthweight, Additional Evidence

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	$bw \leq 1000$	$bw > 1000$ $bw \leq 1500$	$bw > 1500$ $bw \leq 2000$	$bw > 2000$ $bw \leq 2500$	$bw > 2500$ $bw \leq 3000$	$bw > 3000$ $bw \leq 3500$	$bw > 3500$ $bw \leq 4000$	$bw > 4000$ $bw \leq 4500$	$bw > 4500$ $bw \leq 5000$
<b>Panel A: Estimated Effect of Plant Operation</b>									
1(Plant Operating) $\times$ 1(< 1 Mile)	-0.0003 (0.0002)	0.0005** (0.0002)	0.0005** (0.0002)	0.0007 (0.0006)	-0.0019 (0.0011)	-0.0003 (0.0011)	0.0017 (0.0013)	-0.0008 (0.0006)	-0.0001 (0.0004)
Mean	0.005	0.006	0.013	0.047	0.181	0.374	0.277	0.082	0.015
<b>Panel B: Estimated Effect of Plant Openings and Closings</b>									
1(Plant Opening) $\times$ 1(< 1 Mile)	-0.0001 (0.0004)	0.0003 (0.0003)	0.0011*** (0.0004)	0.0013 (0.0008)	-0.0006 (0.0025)	0.0007 (0.0023)	0.0001 (0.0026)	-0.0016 (0.0011)	-0.0011** (0.0005)
1(Plant Closing) $\times$ 1(< 1 Mile)	0.0004 (0.0003)	-0.0006** (0.0003)	-0.0001 (0.0003)	-0.0004 (0.0007)	0.0025* (0.0013)	0.0008 (0.0011)	-0.0026* (0.0014)	0.0004 (0.0005)	-0.0004 (0.0005)
$H_0$ : Opening=-Closing (p-value)	0.616	0.463	0.080	0.310	0.510	0.555	0.398	0.280	0.004
Mean	0.005	0.006	0.013	0.047	0.181	0.374	0.277	0.082	0.015

NOTES: This table reports regression coefficients from 18 separate regressions, 9 per panel, on a regression of 88958 plant-distance-year cells and 3438 plants. The dependent variable in each regression is an indicator variable for whether a birth falls in a particular birthweight range as indicated by column headings, and the data have been aggregated to plant by distance by year cells. The regression sample changes as one moves across the columns, indicated by the column headings. For example, the specification in column (1) examines the relative likelihood of a birth being below 1000 grams within 1 mile of a plant responds to plant operating status, relative to the comparison group. The comparison group in all columns is births between 1 and 2 miles from a plant. Panel A estimates the effect of plant operating status on local birth outcomes, where 1(Plant Operating) is a indicator variable equal to one for plants that have opened and/or have not yet closed. Panel B estimates the asymmetric effect of plant openings/closings. Panel B reports p-values from tests that the two coefficients are equal in magnitude but of opposite sign. All columns control for plant $\times$ distance-bin and plant $\times$ year fixed effects, census tract characteristics (interacted with quadratic trends), and regressions are weighted by the group-level cell size. Multiple births are dropped from regressions. Standard errors are two-way clustered by plant and year.