

University of Toronto  
Department of Economics



Working Paper 643

Military Training Exercises, Pollution, and their Consequences  
for Health

By Gustavo J. Bobonis, Mark Stabile and Leonardo Tovar

August 14, 2019

# Military Training Exercises, Pollution, and their Consequences for Health<sup>\*</sup>

Gustavo J. Bobonis<sup>†</sup>, Mark Stabile<sup>‡</sup>, and Leonardo Tovar<sup>§</sup>

August 2019

**Abstract:** Militaries around the world perform training exercises in preparation for war. We study the relationship between *in utero* exposure to military exercises (bombing) and early-life health outcomes, combining data on naval bombing exercises in Vieques, Puerto Rico, and the universe of births from 1990-2003. Using a differences-in-differences design, we find that the sudden end of bombing practices is associated with a 56-79 percent decrease in the incidence of congenital anomalies. The evidence is generally consistent with the channel of environmental pollution; increases in arsenic levels in waters surrounding the live impact area.

**Keywords:** Infant health; military activity; environmental pollution; maternal stress.

**JEL Codes** I15, I14, O1

---

<sup>\*</sup> This is a significant revision and extension of an earlier paper, circulating under the title “Bombs and Babies: U.S. Navy Bombing Activity and Infant Health in Vieques, Puerto Rico” (NBER Working Paper #2909). We thank Francesco Amodio, Michael Baker, Loren Brandt, Esther Duflo, Fred Finan, Jeanne Lafortune, Elaine Liu, Aprajit Mahajan, Ted Miguel, Teresa Molina, Ben Olken, Rohini Pande, José Tessada, Tom Vogl, and seminar and conference participants at UC Berkeley, Toronto, Catholic University of Chile-Economic History and Cliometrics Lab Second Annual Conference 2014, the 11th Annual Workshop of the Households in Conflict Network, the CIREQ Montreal Applied Economics Conference 2016, and the NBER Environment and Energy Economics Spring 2017 Meeting, the editor (Manisha Shah), and various referees for helpful comments and suggestions that improved the paper substantially. We would also like to thank Aly Somani and Jordan Scantlebury for excellent research assistance, the Institute of Statistics of Puerto Rico, Dr. Heriberto Marín, and Dr. Arturo Massol Deyá for sharing administrative data and for their general support throughout, as well as the Canadian Institute for Advanced Research and the CONICYT/Programa de Investigación Asociativa (Project SOC 1102) for financial support. We are responsible for any errors that may remain.

<sup>†</sup> Corresponding author: Department of Economics, University of Toronto; and BREAD. Address: 150 St. George St., Room 304, Toronto, Ontario, M5S 3G7, Canada. E-mail: [gustavo.bobonis@utoronto.ca](mailto:gustavo.bobonis@utoronto.ca).

<sup>‡</sup> INSEAD

<sup>§</sup> University of Toronto

## I. Introduction

Militaries around the world perform training exercises in preparation for war and, in general, they take place “at home”. The U.S. is no exception. As of 2001, approximately 950 formerly used defense sites (FUDS) located across the country associated with former military training ranges contained defense-related hazards such as unexploded ordnance and other explosive wastes. These sites represent health hazards to local populations exposed to contaminants and military debris (U.S. General Accounting Office 1996, 2001).<sup>1</sup>

Surprisingly, there is very limited research documenting the impacts of these military training exercises on the livelihoods of nearby populations. Given that these training exercises continue to take place around the world (Korea, Japan, Iraq to name but a few), we would benefit from a better understanding of the potential consequences of these practices on health and socioeconomic outcomes of nearby populations -- often vulnerable groups in society.

This study exploits a case study to provide evidence regarding the effects of these military exercises on the health outcomes of nearby populations. Over the span of six decades (1941-2001), the U.S. Navy utilized two-thirds of the territory of Vieques, Puerto Rico, to host a range of military exercises (including ship-to-shore gun fire, air-to-ground bombing by naval aircraft, and Marine amphibious landing) 12.5 kilometres away from residential population. Studying this setting allows us to explore whether exposure to military training in fact has consequences for the health and safety of nearby populations, including the health at birth for children exposed *in*

---

<sup>1</sup> FUDS are located throughout the United States, on remote islands such as Pacific Atolls or the Aleutian Islands in Alaska, while others include urban settings such as Hunter's Point Naval Shipyard in San Francisco.

*utero*.<sup>2,3</sup> We study whether the abrupt end of naval practices in July 2000 had short-term consequences for infants' health outcomes. To do so, we combine monthly data on tonnage of ordnance used in these naval exercises with the universe of births in Puerto Rico between 1990 and 2003. This allows us to study the relationship between *in utero* exposure to these military exercises and children's early-life health outcomes. Specifically, we examine infant health outcomes for cohorts conceived following the end of naval practices relative to those conceived in the preceding period within the municipality affected by bombings, and to those newborns conceived in municipalities not affected by the bombings. Our identification is therefore across treatment and control municipalities over time.

The military exercises have significant negative consequences for early life outcomes -- particularly for congenital anomalies. The sudden end of bombing practices is associated with a 5.6-7.9 per thousand (56-79 percent) decrease in the incidence of congenital anomalies as the main driver for improvements in neonatal health outcomes. We do not find robust evidence of improvements in other measures of neonatal health. We find direct support for the channel of environment pollution using data from U.S. Environmental Protection Agency (EPA) Discharge Monitoring Reports by the Atlantic Fleet Weapons Training Facility (U.S. Navy) of inorganic

---

<sup>2</sup> We also focus on neonatal health outcomes for two additional reasons: due to data availability as well as the fact that newborn health can respond rapidly to environmental conditions, which is important for an empirical strategy based on the high frequency timing of the practices. Prominent examples of the research on the effects of conflict for early-life and long-term outcomes are Alderman, Hoddinott and Kinsey (2006), Bundervoet, Verwimp, and Akresh (2009), Camacho (2008), León (2012), Mansour and Rees (2012), Galdo (2013), and Quintana-Domeque and Ródenas-Serrano (2017). A complementary literature analyzes how environmental factors that occur during the prenatal period have significant early life and long-term consequences; see Almond and Currie (2011) for a broad survey of the literature. Most epidemiological and medical research focuses on the role that nutrition plays for fetal development and early-life health outcomes (Kramer, 1987).

<sup>3</sup> This emerging body of research supports the "fetal origins" hypothesis first articulated by David Barker (1990): the idea that numerous external environmental factors during the mother's pregnancy can have important, long-lasting consequences on health outcomes. According to the fetal origins hypothesis, intrauterine exposure to environmental agents may program the fetus to have particular metabolic characteristics (Barker 1990). The specific biological mechanisms that manifest in different health outcomes of fetuses depend on the level of exposure to environmental factors as well as nutritional and genetic factors.

chemicals such as arsenic, cyanide, and lead, in waters surrounding the live impact area. In particular, bombing activity leads to short-term increases in arsenic levels above EPA limits in waters surrounding the live impact area: a one standard deviation increase in average ordnance levels leads to a 14 percentage point increase in this incidence risk, an effect that is 51.2 percent above the mean incidence risk. Given that arsenic exposure has been linked to increased frequency of spontaneous abortions and congenital malformations (Nordstrom et al. 1979; Hopenhayn-Rich et al. 1999) this evidence is suggestive of a link between water pollution resulting from the bombing activity and infant health outcomes. In contrast, we do not find evidence that possible disruptions in economic activity mediate these effects. The evidence thus supports the hypothesis that reductions in environmental pollution and other environmental factors lead to a substantial reduction in the risk of congenital anomalies among the infant population (Currie, Greenstone, and Moretti 2011; Currie, Ray, and Neidell 2011).

Our study makes several contributions to the health, conflict and human development literature. Prominent research on the effects of exposure to conflict for early-life and long-term outcomes document a link that is most plausibly explained by economic channels and changes in stress as well as mental and physical health of the adult population.<sup>4</sup> We demonstrate that an important plausible channel in the context of certain conflicts is one of an environmental nature, consistent with a complementary literature that analyzes how environmental factors that occur during the prenatal period have significant early life and long-term consequences (Almond and Currie 2011). The study also has important implications for both child policy and military policy. Infant health outcomes such as congenital anomalies and low birth weight are important predictors of child health (McCormick 1985; Pollack and Divon 1992; Almond, Chay, and Lee 2005) and

---

<sup>4</sup> For instance, Camacho (2008), León (2012), and Quintana-Domeque and Ródenas-Serrano (2017).

long-term outcomes such as educational attainment, labor market outcomes, and adult health (Currie and Hyson 1999; Behrman and Rosenzweig 2004; Black, Devereux, and Salvanes 2007). Given the well-documented relationship between neonatal health and later life outcomes, there is reason to believe that our substantial short-term effects may have longer-term consequences for this population. Moreover, previous research has documented the effects of both maternal stress and environmental pollution on infant health (c.f. Currie et al 2009) and our evidence strongly suggests that at least one of these channels is responsible for the detrimental effects on child health that we find here. Since the shutdown of the U.S. Navy in Vieques in 2001 formalized the start of a negotiation between the Commonwealth of Puerto Rico and the U.S. for an ecological and economic restoration strategy for the island, these findings have implications to expand the discussion to address child health and child development dimensions.

The paper is organized as follows. Section II provides background on U.S. Navy Activities in Vieques and the possible implications for health outcomes among the resident population. We follow with a description of the data and the empirical methodology in Section III. We present the central empirical results of the paper in Section IV. In Section V we explore potential mechanisms, followed in Section VI by a series of robustness tests and a discussion of potential impacts on other health outcomes. The paper concludes in Section VII with a discussion of findings and their broad implications.

## **II. Background**

### **II.A. U.S. Navy Activity in Vieques**

Vieques is an island off the eastern end of Puerto Rico with approximately 350 square kilometers (Figure 1). In 2010, the island hosted 9,301 habitants. Close to two-thirds of the Vieques

territory served as part of the U.S. Navy Atlantic Fleet Weapons Training Area from 1941 to 2001. Military training and operations were conducted in the eastern end of the island, while the western end was used to store munitions. The central part, the “civilian area” (approximately 45 square kilometers), was designated to accommodate local civilian residents.

The eastern naval area hosted a range of military exercises including ship-to-shore gun fire, air-to-ground bombing by naval aircraft, and Marine amphibious landing. Over the span of six decades, naval operations averaged between 180 to 250 days each year (approximately 6,300 shelling days) with an annual estimate of 3-14 million pounds of live ordnance detonated and dropped within the live impact area (189-662 million pounds in total) (Porter, Barton, and Torres 2011). The live impact area encompasses an area of about 900 acres and is 12.5 kilometers away from residential population.

In addition to conventional weaponry, the composition of munitions used during bombardment exercises posed risks to the health of the population and ecology of the island. The U.S. Agency for Toxic Substances and Disease Registry (ATSDR) notes that naval training involved handling of Napalm and Agent Orange at various stations within the eastern naval area (ATSDR 2001). Despite serious concerns of the radiological and toxicological effects of depleted uranium once it vaporises in the air, over 250 rounds (88 lb) of ammunition tipped with depleted uranium were fired in 1999 (Wargo 2009). According to a U.S. Congress report by the U.S. Department of Defense, biological weapons were tested in Vieques but no further details of the operation are publicly disclosed (Porter, Barton, and Torres 2011).

The U.S. Navy reduced its operations in April 1999 following a widely publicized campaign when David Sanes, a civilian employee, was killed during a bombing accident. No military training exercises took place on Vieques for approximately thirteen months. In May 2000,

the Navy resumed military training exercises but only with “practice” bombs and other non-explosive ordnance for a brief period of less than fifty days (ATSDR 2003, p. 13). All military training exercises at Vieques officially ceased on May 1, 2003, when the Navy turned its lands over to the U.S. Fish and Wildlife Service.

Since various areas of the island remained contaminated by solid and hazardous waste resulting from decades of military activity, in 2005 the U.S. Environment Protection Agency (EPA) declared these lands a superfund site. This required the U.S. Navy to partner with the EPA, Fish and Wildlife Service and the Puerto Rico Environmental Quality Board to determine and implement cleanup actions. The effects of decontamination practices may pose further risks to Vieques residents and stress on pregnant women as these involve, among other things, detonating defective ordnance in the air.<sup>5</sup> Current projections indicate that work at the site will be completed in 2022 for the land areas and in 2029 for the underwater effort (EPA 2013).

## **II.B. Implications for Health Outcomes among the Resident Population**

Most research efforts on the impacts of military activity on the health profile of the Vieques’ resident population (“Viequenses”) have focused on documenting the unusually high cancer incidence rates in the municipality. Reports produced by the Puerto Rico Department of Health have identified an upward trend since 1960 in cancer incidence rates in Vieques relative to the rest of Puerto Rico.<sup>6</sup> The U.S. ATSDR has produced public health assessment studies on drinking groundwater (released in 2001), ingesting or touching soil (2003), breathing air (2003),

---

<sup>5</sup> McCaffrey (2009) highlights this and several environmentally degrading practices by the U.S. Navy during the cleanup process, including burning excess materials and waste and dumping toxic chemicals and substances in sensitive wetland areas.

<sup>6</sup> See Zavala Zegarra (2000) for the period covering 1960-2004; and Figueroa, Suarez, De La Torre, Torres, and Perez (2009) for the period covering 1995-2004. The latter documents that Vieques residents were 26 per cent more likely to develop cancer in 1995-1999, and 19 per cent in 2000-2004 compared to residents in the rest of Puerto Rico.

and eating fish and shellfish (2003), and they all conclude that the resident population has not been exposed to harmful levels of chemicals resulting from U.S. Navy training activities.<sup>7</sup> A small number of independent research studies have documented exposure of the population to higher levels of mercury, lead, copper, and nickel than those clinically recommended by the World Health Organization (Ortiz-Roque and López-Rivera 2004; Massol-Deyá et al 2005). Although based on small samples, this literature is suggestive of an environmental link that can help explain the Vieques population's poor health profiles compared to those of residents of other municipalities in Puerto Rico.

The effects of naval aircraft and bombing exercises on the psychological and psychosocial profile of Vieques is an area that remains unexplored despite documented qualitative evidence that these activities disrupted regular life activities. According to the 1999 Special Commission on Vieques, officials from the P.R. Department of Education reported that “the vibrations caused by bombing practices shudder educational facilities, affecting the physical structure of buildings and interrupting classes”. The Department concluded that it is evident that “this sort of activity and the noise generated cause anxiety and concern among students and school staff in general” (Puerto Rico 1999, p. 10). Even without definitive empirical evidence on the levels of physical and mental health among the population of Vieques during the period of interest, the limited documentation available suggests that pregnant women resident in the municipality may have been exposed to disruptive environmental factors that would have affected fetal development.

### **III. Data and Methodology**

---

<sup>7</sup> The impartiality of ATSDR studies have been questioned by the academic community and journalists alike. For an analysis on the narrative and language used by different U.S. departments and agencies on handling the Vieques file see Davis, Hayes-Conroy, and Jones (2007).

### **III.A. Data Description and Sample Selection Criteria**

Data on tonnage of ordnance used in these naval exercises is available from Discharge Monitoring Reports submitted by the U.S. Department of the Navy's Atlantic Fleet Weapons Training Facility (Roosevelt Roads Base, Puerto Rico) to the U.S. Environmental Protection Agency (EPA). Information on total weight of ordnance used by the U.S. Navy and other parties for all military training exercises (including air-to-ground, ship-to-ground, and land-based activities) on a monthly basis is available for the period May 1985 - July 1999.<sup>8</sup> From 1988 to 1999, between 1,359 and 2,667 tons of ordnance were used in training exercises, of which between 124 and 469 tons were considered high explosives.

We combine these data with the universe of birth records in Puerto Rico between 1990 and 2003, available from the Puerto Rico Statistics Institute. Specifically, we have data on approximately 855,429 births in the territory of P.R. with information on birth outcomes such as sex, month of conception, exact date of birth, gestation period, and detection of congenital anomalies, among other characteristics. In addition, we have data on the mother's municipality of residence at the time of birth, as well as a number of characteristics of the mother (i.e., age and educational attainment).

We use the sample of births of mothers resident in P.R. at the time of birth for which the date of birth and the period of gestation, as well as the municipality of residence, are known.<sup>9</sup> After

---

<sup>8</sup> These reports were made available in August 2001 to Dr. Arturo Massol Deyá via a Freedom of Information Act (FOIA) request to the U.S. Environmental Protection Agency. We thank Dr. Massol Deyá for sharing these reports with us. Additional data on ordnance used in the naval exercises is based on a study prepared for the Secretary of the U.S. Navy in 1999, which was later reproduced in an ATSDR public health assessment of pollution via air pathways (2003, pp. 96-97). The dataset contains two measures of live-fire range utilization: (1) total weight of ordnance that the U.S. Navy and other parties used for all military training exercises, including air-to-ground, ship-to-ground, and land-based activities; and (2) total weight of high explosives used at the live impact area. The measures are available for the fiscal years 1983 to 1998 (October 1<sup>st</sup> - September 30<sup>th</sup>).

<sup>9</sup> Among the sample of births with known date of birth and gestation period, there is only one (1) observation for which we do not have information on the mother's municipality of residence.

combining these data with the monthly tonnage of ordnance, we lose 50,081 observations (5.8 percent) and an additional 59 observations with missing information on maternal age or educational attainment. The remaining sample is composed of 805,348 births (94.1 percent of the population), with item non-response rates for our various outcomes of interest in the 0.2-0.3 percent range. We first collapse the data to the municipality-by-month level and weight the observations by the number of live births in the month, a process that yields identical point estimates to the micro data. Our dataset is thus composed of 12,228 municipality-by-month cells between January 1990 and March 2003, covering 159 months per jurisdiction.

Summary statistics on the ordnance measures are reported in Table 1. Monthly ordnance used in the exercises (in the complete 1985-1999 period) ranged from none (zero) to 658 tons; on average 120.1 tons of ordnance were used. The average tonnage of ordnance during each child's potential gestation period is similar, as it is a 9-month moving average of the monthly measure. It ranges from 0 to 277 tons of ordnance, and averages 122.0 tons per month.<sup>10</sup> We use these measures of ordnance to both document the large drop in ordnance following the end of naval exercises and to inform our evidence of potential mechanisms, discussed in Section V below.

Table 2 reports summary statistics from the birth records data. Regarding child health outcomes at birth, one percent of live births are born with a congenital anomaly. Also, 14.4 percent of infants are born prematurely, most of them being moderately or late preterm, but a substantial proportion (0.6 percent) are extremely preterm. Almost eleven (10.6) percent of live births are considered low weight births, and 1.1 percent of children have a low 5 minute Apgar score (defined as less than 7 out of 10). These health indicators are similar for infants in Vieques than those in

---

<sup>10</sup> Density plots of the distributions for both measures show that the measure of monthly ordnance is skewed to the right, with most of the observations in the 0-2 tons per month range. In contrast, the measure of ordnance exposure during each child's potential gestation period is more symmetrically shaped and in the 0-3 tons per month range.

the rest of Puerto Rico: although the incidence of births with congenital anomalies on average is lower in Vieques than in the rest of Puerto Rico (0.5 percent vs. 1.0 percent), the remaining health measures are similar across these groups. These indicators are generally worse than those for the overall U.S. population during this time period.<sup>11</sup> The proportion of female live births is slightly higher in Vieques than in the rest of Puerto Rico (50.0 percent vs. 48.6 percent).

The average characteristics of mothers in Vieques are reasonably different from those in the rest of the sample: 37.1 percent of mothers in Vieques have only a high school degree versus 29.7 percent in the rest of Puerto Rico. Mothers are slightly younger (ages 23.7 versus 24.7, on average) and more of the mothers are giving birth as teenagers (14.4 percent versus 9.2 percent).

Finally, we include municipality-level monthly employment and unemployment data estimates from the U.S. Bureau of Labor Statistics (BLS) Local Area Unemployment Statistics (LAUS) as further local controls.

### **III.B. Empirical Methodology**

Our empirical strategy consists of a difference-in-differences design. We use the end of practices in the year 2000 to compare outcomes for cohorts conceived following the end of naval practices to those conceived in the preceding period within the municipality using the health outcomes of newborns in other municipalities to control for period-specific effects. Specifically, we estimate linear models of the form:

$$y_{jmt} = \theta(\text{End}_{mt} \times \text{Vieques}_j) + X_{jmt}\beta + \alpha_j + \gamma_{mt} + \varepsilon_{jmt}, \quad (1)$$

---

<sup>11</sup> See National Center for Health Statistics, National Vital Statistics System.  
<http://www.childstats.gov/americaschildren/health1.asp>

where  $y_{jmt}$  is the proportion of children with a congenital anomaly, born to mothers residing in municipality  $j$ , conceived in the month  $m$  of year  $t$ ;  $\text{End}_{mt}$  is a pre-post dummy for the end of the exercises;  $\text{Vieques}_j$  is an indicator variable for mother's residence in Vieques;  $\alpha_j$  are municipality fixed effects;  $\gamma_{mt}$  are month-by-year fixed effects; and  $\varepsilon_{jmt}$ , is the error term. Each specification includes controls for maternal characteristics at that municipality-month-year level.

Our identification strategy is focused on comparing outcomes pre and post the ceasing of naval exercises in Vieques relative to other municipalities in Puerto Rico. In order to provide a first examination of the research design, in Figure 2 we examine trends in the incidence of congenital anomalies for births conceived by residents of Vieques (solid blue line) and those in the rest of Puerto Rico (dotted green line), from early 1989 until early 2003 (births in 1990-2003), together with the volume of ordnance during this time period (dotted red line) (see Figure 2). Due to the small cell sizes in the case of Vieques, we smooth the series by presenting 24-month moving averages of the outcomes for both Vieques and control municipalities. We observe a clear relationship between the volume of ordnance—and its abrupt end—and incidences of congenital anomalies among birth cohorts to Viequense mothers. In contrast, we see no relationship between ordnance levels and these across cohorts in the rest of Puerto Rico.

We also report a series of estimates from an event study to (i) document the year-specific effects as well as (ii) to test our identifying assumption: that the outcome in Vieques would not have evolved differently to other municipalities in the territory in the absence of the ceasing of naval exercises. We estimate equation (2) to explore this:

$$y_{jmt} = \theta_t \cdot \mathbf{I}\{m = \text{Vieques}_j\} \cdot \mathbf{I}\{t = 1990, 1991, 1992, \dots, 2003\} + X_{jmt}\beta + \alpha_j + \gamma_{mt} + \varepsilon_{jmt}, \quad (2)$$

where  $\mathbf{I}\{m = \text{Vieques}_j\} \cdot \mathbf{I}\{t = 1990, 1991, 1992, \dots, 2003\}$  is a vector capturing the interaction of the Vieques indicator with an indicator variable for each year from 1990 to 2003, with 1999 – the year before the end of bombing exercises – as the base year. All other variables are as defined above in equation (1). The vector  $\theta_t$  captures the year-specific effects for each year from 1990 to 2003 (with 1999 as the reference year).

Inference based on this differences-in-differences research design relies on asymptomatic approximations associated with the assumption that the number of jurisdictions – municipalities in this case – grows large. However, this assumption does not apply in our setting since treatment occurred in only one jurisdiction.<sup>12</sup> We conduct inference based on two methodologies that appropriately take this design into account. First, we report p-values from a cluster residual bootstrap test of significance of the treatment effect (in a differences-in-differences setting) when there are few treated groups of heterogeneous groups sizes even with one treatment cluster, recently developed by Ferman and Pinto (2019).<sup>13</sup> Their method – a cluster residual bootstrap with a heteroskedasticity correction – provides valid hypothesis testing asymptotically when the number of control groups goes to infinity, even when there is only one treated group.

Second, following the literature on randomization inference (Fisher 1925; Neyman 1990), we implement a variant of Fisher’s permutation or randomization test against the exact null

---

<sup>12</sup> Inference in difference-in-differences models is complicated by the fact that errors might exhibit intra-group and serial correlations; the primary concern when using data aggregated at the group level, such as ours, is accounting for possible serial correlation (Bertrand, Duflo, and Mullainathan 2004). Although we use data from 77 municipalities, we cannot use either (i) a standard cluster robust variance estimator (CRVE) or (ii) inference based on the wild cluster bootstrap (Cameron, Gelbach, and Miller 2008) because the relevant degrees of freedom are the number of treatment units – which in this case is a single municipality (Imbens and Kolesar 2012; MacKinnon and Webb 2017). An additional concern is that ignoring spatial correlations that do not vanish between municipalities may lead to bias in standard errors (Barrios et al 2012).

<sup>13</sup> With a small number of groups, it might still be possible to obtain tests with correct size using alternate methods. However, no other existing methods perform well when there is only one treated group (i.e., Cameron Gelbach, and Miller (2008), Brewer et al. (2013), Canay et al. (2017), Ibragimov and Müller (2010, 2016), and MacKinnon and Webb (2017)).

hypothesis of no effect of the end of bombing on health outcomes ( $\theta = 0$ ). To implement the procedure, we estimate equation (1) an additional 76 times replacing Vieques with an indicator for each one of the other municipalities. Then we compare the Vieques estimate to the other placebo estimates obtained. With seventy-six placebo estimates, achieving 10% significance from a one-sided test requires that Vieques be ranked seventh from the bottom of the placebo distribution, while 5% significance requires that it be ranked third. We report the percentile rank of the coefficient from the permutation test as well as the approximate p-value. In addition, we provide graphical illustrations (histograms) from the placebo-based inference results for the main outcomes of interest. This is a very demanding statistical test to achieve statistical significance at conventional levels (Buchmueller et al. 2011). Finally, we construct a summary index of all health outcomes to pool multiple outcomes in a single test to address the issue of multiple inference (e.g., O'Brien 1984; Kling, Liebman, and Katz 2007).

An alternative identification strategy that could be implemented would be to use a continuous measure of ordnance during the bombing period to identify the effects of the bombing to identify intensive marginal effects. The estimates generated by this approach are quantitatively similar in magnitude to the ones from our preferred difference-in-differences design; this is perhaps not surprising given the strong relationship between ordnance levels over time and the incidence of congenital anomalies in Vieques (see Figure 2). However, the point estimates from this intensive-margin design are not as precisely estimated using the permutation inference procedure described above (results for this alternative strategy are available upon request).

#### **IV. Results**

Following our graphical exploration of the results described above, we present regression-based evidence of these effects; these are reported in Table 3. The first column reports the effects of the end of bombing during the pregnancy period on the risk of birth with congenital anomalies. We find a large and significant effect: the sudden change in bombing practices is associated with a 7.7 per thousand decrease in the incidence of congenital anomalies, a 77 percent reduction relative to the baseline mean. The magnitude of this impact is consistent with existing evidence on the consequences of reductions in environmental pollution from Superfund site clean-ups on neighbouring populations in the United States (Currie, Greenstone, and Moretti 2011).

Inference based on the Ferman-Pinto test of significance of the treatment effect indicates that we have sufficient precision to reject the null hypothesis of no relationship ( $p\text{-value} = 0.070$ ). Inference based on the stringent permutation test leads to rejection of the null hypothesis, with a permutation percentile rank of six out of 77 (approximate  $p\text{-value} = 0.078$ ). In Figure 3, we provide a graphical illustration (histogram) of the distribution of placebo-municipal effects and of the permutation test-based inference results for the incidence of congenital anomalies (reported in Table 3, columns 1). The (red and green) dashed lines are respectively the 5th and 10th percentile values (other than Vieques) of the distribution of estimated placebo effects; the solid vertical line is the Vieques value. This very demanding statistical test achieves statistical significance at conventional levels.

Figure 4 plots the coefficients ( $\theta_t$ ) of the Vieques-specific year effects on the incidence of congenital anomalies, generated from the even study regression model (equation (2)). The solid vertical (light blue) line in the figure denotes the end of naval practices. The dashed vertical lines are the sampling distributions for the placebo estimates from the 5th–95th percentile for each year. The point estimates for the post-period imply a reduction in the incidence of congenital anomalies

of 1.8-2.1 percentage points (relative to year 1999), with permutation test-based approximate p-values in the 0.026-0.078 range. Also, consistent with the reduction in ordnance volume during the two years preceding the end of naval practices in Vieques, the point estimates imply a reduction in congenital anomalies of 1.8-2.1 percentage points during the years 1997-1998 (permutation test-based approximate p-values = 0.013, 0.039).<sup>14</sup> In contrast, the point estimates of the 1990-1996 pre-period effects are smaller in magnitude and not statistically significant at conventional confidence levels.

Coupled with the evidence cited above on the link between environmental factors and congenital anomalies in particular, we believe our inference is most robust to the effects of the bombings on congenital anomalies.<sup>15</sup> For completeness, we also study potential impacts on other neonatal health outcomes. We examine a variety of other possible outcomes including premature births, birthweight, and APGAR scores (Table 3, columns 2-5). Overall, we do not find robust statistically significant evidence of improvements in child health in these areas following the ending of the bombing: a 2.8 percentage point *increase* in the incidence of premature births (19.5 percent), but a 0.18 percentage point (30 percent) reduction in that of extremely preterm births; and a 0.2 percentage point (1.9 percent) reduction in the incidence of low birthweight births. Given that we test for impacts on multiple health outcomes, we construct a summary index of neonatal

---

<sup>14</sup> This reduction is also consistent with (and captured by) the models that measure intensive-margin effects of the naval practices on the incidence of congenital anomalies, mentioned in Section III.B.

<sup>15</sup> We also explore whether the relationship between bombing ordnance and newborn health differs for different trimesters of the pregnancy, as highlighted by previous literature (e.g., Almond and Mazumder 2011). In our context, mothers can be exposed to bombing during the entire pregnancy or, among those pregnant around the time of the end of naval exercises, in the first or first and second trimester of pregnancy. In order to explore this potential heterogeneity, we estimate models that take into account the different exposure for the mother in each of the trimesters: (a) first and second trimesters; (b) first trimester only; and (c) not exposed. The relationship between end of exposure to ordnance and congenital anomalies appears to be strong and stable: exposure in both the first and second trimester or in only the first trimester reduces the incidence of congenital anomalies by 0.9 percentage points, with permutation percentile rank approximate p-values of 0.09 and 0.10, respectively. Given the small number of cells for these cohorts, we have insufficient precision to detect differences based on the Ferman-Pinto test. The estimates are available upon request.

health outcomes and conduct a single test to address the issue of multiple inference (Kling, Liebman, and Katz 2007). Estimates using this normalized score as dependent variable indicate that the end of naval exercises led to no overall improvement in the neonatal health outcomes index; the point estimate implies a reduction in the health outcomes z-score of  $0.018\sigma$ . The Ferman-Pinto test of significance of the treatment effect indicates that we have sufficient precision to reject the null hypothesis of no relationship (p-value = 0.063). However, inference based on the stringent permutation test leads to no rejection of the null hypothesis, with a permutation percentile rank of 52 out of 77 (approximate p-value = 0.675). Analogous event study estimates of the year-specific effects on the neonatal health outcomes index show suggestive, but weaker, evidence of an improvement in health outcomes as a result of the reduction/end of naval exercises (reported in Figure A2 in the Appendix).<sup>16</sup>

## **V. Potential Mechanisms**

The measured impacts are consistent with existing evidence on the consequences of comparable changes in environmental pollution on neighbouring populations in the United States (Currie, Greenstone, and Moretti 2011; Currie and Schwandt 2014) as well as with effects of terrorist attacks (Quintana-Domeque and Ródenas-Serrano 2017). There are several possible mechanisms by which these bombings could have affected mothers during pregnancy: the bombings could disrupt economic activity affecting households' access to or capacity to invest in human capital. They could cause maternal stress, fear, and anxiety, which can have negative effects on physical and mental health, and in turn have indirect negative effects on infant health outcomes

---

<sup>16</sup> We find no evidence of impacts on other neonatal health outcomes based on models that allow for heterogeneous effects to exposure. This is not surprising given the lack of precision in the estimates of average effects and the smaller number of cohorts in each group of pregnancy with distinct exposure times. Estimates are available upon request.

at birth. The bombings could have a negative effect on the environment (air, soil, and water) thereby affecting maternal and child health. In this section, we present and discuss the available evidence regarding these potential channels.

### *Environmental Pollution*

Increases in solid and hazardous waste, resulting in potential contamination of soil, water, and/or air may also have affected infant health. Previous literature has shown a direct link between air pollution and infant health (Chay & Greenstone, 2003; Currie, Neidell, & Schmieder, 2009). To investigate the potential role of environmental contamination, we examine water pollution levels. The Discharge Monitoring Reports contain information regarding tests of water quality in the training area's Inner Range, the waters surrounding the live impact area. These reports contain information on tests to measure the levels of inorganic chemicals in the water (arsenic, cyanide, lead, among others) in addition to other characteristics of the waters in the range. We use the information available on a quarterly basis during the same period (May 1985 - July 1999). For measures of dissolved oxygen, acidity, alkalinity and nitrogen, while average levels are within the range deemed safe by the Environmental Protection Agency, the maximum level reported exceeds the range (data available upon request). For many of the inorganic chemicals tested, average levels reported exceed the maximum level for safety.

Given that arsenic exposure has been linked to increased frequency of spontaneous abortions and congenital malformations (Nordstrom et al. 1979; Hopenhayn-Rich et al. 1999), we document the extent to which there is a relationship between the level of bombing activity and arsenic in the waters surrounding the live impact area. Specifically, we estimate time series models to measure the relationship between naval bombing activity and the incidence that maximum

arsenic levels surrounding the live impact area are above permitted EPA limits, of the following linear form:

$$1(y_t > y^{EPA}) = \theta_B(\text{Ordnance}_t) + \varepsilon_t, \quad (2)$$

where  $y_t$  captures the maximum arsenic level measured in quarter  $t$ , and  $\text{Ordnance}_{mt}$  is the mean level of ordnance in the quarter. We allow for leads and lags of the ordnance volume measure to evaluate the degree of correlation of arsenic pollution based on bombing activity in previous time periods. Standard errors are corrected for heteroskedasticity.

The estimates of the contemporaneous (same quarter) correlation are reported in Table 4. The estimates are remarkably stable. For instance, the model with the simple correlation implies that the effect of a one standard deviation change in average ordnance levels leads to a 14 percentage point increase in this incidence risk, an estimated effect that is 51.2 percent above the mean incidence risk. Estimates that allow for lags and leads of ordnance show a quantitatively similar pattern. This evidence is suggestive of a link between water pollution resulting from the bombing activity and infant health outcomes. However, a number of other chemicals were also present (although not correlated with bombing activity), and thus it is difficult to assess the role of arsenic in the etiology of these effects.

### *Maternal Stress*

It is feasible that the bombings increased stress levels or sleep deprivation among pregnant women. The medical literature indicates that prenatal stress increases levels of corticotrophin releasing hormone, which regulates the duration of pregnancy and fetal maturation. Increases in prenatal stress levels have been associated with a decrease in infant birth weight, an increased likelihood of LBW, and a decrease in gestational age at birth (Wadhwa et al 1993). Studies have

also suggested that stress induced during the first trimester tend to have more significant effects on birth weight and preterm birth (Zhu et al 2010). Therefore, it is certainly feasible that increased stress due to the bombings had negative effects on infant health in Vieques.

Mothers in Vieques are exposed to the bombings before they conceive – most of them for the better part of their adult lives. Recent research suggests that, maternal health long before conception can influence child health. Mother characteristics such as chronic disease or obesity can have negative consequences for child outcomes (Stephenson et al 2018). If this exposure has an effect on their own health which in turn affects their child’s health at birth, then our estimates of the effects of in-utero exposure on child health can be interpreted as the combined effect of both this in-utero exposure and the longer maternal exposure.

## **VI. Robustness Tests and Selection**

### *Economic Downturns*

Although, given our research context, we can rule out the destruction of physical capital as a determinant of these adverse infant health consequences (e.g., Abadie and Gardeazabal 2003), another plausible pathway is through disruption of economic activity affecting households’ access to or capacity to invest in human capital (León 2012). While our results are robust to including controls for unemployment levels, here we examine the relationship between naval bombing exercises and the municipal unemployment rate, using municipality-level monthly unemployment data estimates from the U.S. Bureau of Labor Statistics (BLS) Local Area Unemployment Statistics (LAUS). We estimate models analogous to equation (1), excluding maternal controls, to predict unemployment rates as a function of the end of naval practices in Vieques.

The estimates are reported in the first two columns of Table 5. If anything, the evidence is consistent with the end of naval activities leading to a short-term deterioration in local economic conditions. We find a positive and significant relationship between the end of naval practices and the unemployment rate – a 3.2 percentage point / 0.22 log point increase in the unemployment rate. Given that these short-term fluctuations could lead to improved neonatal health outcomes attributable both to selection (changes in the type of mothers who conceive based on local economic conditions) and to improvements in health behaviors during recessions (Dehejia and Lleras-Muney 2004), we include average unemployment rates as controls in our main specification above. As will be shown below, our estimates are robust to specifications that control for maternal characteristics and these local economic conditions. This evidence is quite suggestive that in our context the link between bombing activity and infant health is not driven by a pernicious effect in economic activity.<sup>17</sup>

*Mother characteristics, and other possible selection effects*

Since the characteristics of mothers in Vieques are reasonably different from those in the rest of the sample, it is plausible that the estimated effects could be partially driven by sample selection. For instance, mothers with children with a worse underlying health status may be more likely to reside in Vieques during periods of more military activity, causing a downward bias in our estimates of interest. Also, if bombing activity were to cause miscarriages or stillbirths, and pregnancies that terminated in these were more likely to be of children with worse underlying health conditions, this would cause our estimates to be biased downwards.

---

<sup>17</sup> As a further test of the effects of the bombing on the local environment (and potentially broader than health at birth) we investigate whether there is a relationship between yearly average ordnance and yearly crime statistics by municipality. We find no significant relationship between ordnance levels and reported crime.

In order to evaluate these concerns, we directly estimate the relationship between our main bombing variables in both specifications and mother and cohort characteristics. Specifically, we estimate linear models of the form:

$$x_{jmt} = \theta(\text{End}_{mt} \times \text{Vieques}_j) + \alpha_j + \gamma_{mt} + \varepsilon_{jmt}, \quad (3)$$

where  $x_{jmt}$  is the average (or aggregate) characteristic for mothers residing in municipality  $j$ , who conceived a child in month  $m$  of year  $t$ ; and the other variables are defined as above. We also report analogous estimates of a model using the child's sex as the dependent variable.

Columns 3 and 4 of Table 5 reports results for estimates of cohort size effects. We find a 0.079 log point decrease in the number of live births following the ending of naval practices; this is consistent with the short-term worsening of economic conditions in the municipality causing lower fertility rates. However, this relationship is not robust to estimating effects in levels. In this case, we estimate an increase of 6.9 births in the municipality (approximately 10 percent). We also examine trends in the number of live births conceived by residents of Vieques (solid blue line) and those in the rest of Puerto Rico (dotted green line), from early 1989 until early 2003 (births in 1990-2003), together with the volume of ordnance during this time period (dotted red line) (see Appendix Figure A1). We observe no relationship between the volume of ordnance—and its abrupt end—and cohort sizes among birth cohorts to Viequense mothers or those in the rest of Puerto Rico. The evidence is thus not indicative of changes in cohort sizes due to fertility or mortality explaining our results.

Columns 5 through 13 of Table 5 report results for estimates of selection on observable child or maternal characteristics. Our results show evidence of negative gender-based selection. We find a 5.5 percentage point drop in the probability of female births following the end of naval

practices, statistically significant based on the permutation and Ferman-Pinto tests (p-values = 0.013 and <0.001, respectively) (column 5). The estimate is consistent with prior evidence that suggests that male foetuses are more vulnerable to detrimental influences *in utero* and therefore more likely to terminate prior to term (Kraemer 2000; Eriksson et al. 2010; Almond and Mazumder 2011; Dinkelman 2013).

Again, given that we test for impacts on multiple observable selection outcomes we construct a summary index of cohort/child characteristics and conduct a single test to address the issue of multiple inference. Estimates using this normalized score as dependent variable indicate that the end of naval exercises led to an overall deterioration in selection of  $0.255\sigma$  (column 6). The Ferman-Pinto and permutation tests of significance of the treatment effect indicate that we have enough precision to reject the null hypothesis of no relationship (p-values = 0.012 and 0.026, respectively). These negative post-bombing selection patterns would tend to lead to a downward bias in our estimated impacts on congenital anomalies and other neonatal health outcomes.

Columns 7 through 9 report results for the educational attainment of mothers giving birth, broken down by high school dropouts, high school graduates, and those with some or completed higher education. There is a *negative* and significant relationship between end of naval exercises and mothers' educational attainment. The end of naval practices predicts a 6.9 percentage point increase in births by high school graduates, with a corresponding 8.6 decrease in the probability by mothers with tertiary education. Since the relationship between maternal educational attainment and their children's health status at birth is positive, this is indicative of negative selection based on educational attainment.

Columns 10 through 12 of Table 5 report results on the differential correlation of end of naval practices in Vieques with mothers' age. There is, in general, a negative relationship between

the end of bombing practices and mothers' age (column 10). There is also an increase in the likelihood of teenage births and a reduction in births to women ages 40 and above (columns 11-12). Again, since the relationship between teenage pregnancies and children's health status at birth is negative, these correlations are indicative of negative selection based on maternal age.

Finally, we construct a summary index of maternal characteristics to conduct a single test. Estimates using this normalized score as dependent variable indicate that the end of naval exercises led to an overall deterioration in selection based on maternal characteristics of  $0.31\sigma$  (column 13). The permutation and Ferman-Pinto tests of significance of the selection effect indicate that we have enough precision to reject the null hypothesis of no relationship (p-values = 0.039 and  $<0.001$ , respectively). Again, these negative post-bombing selection patterns tend to lead to a downward bias in our estimated impacts.

In addition to looking directly at the relationship between end of naval practices and these characteristics, we also re-estimate the linear models on child health outcomes (equation 1) excluding these maternal controls to assess the sensitivity of the estimates to these control variables as well as the inclusion of the unemployment rate as control. The estimates are reported in Table 6. We report estimates of the effects when we (i) exclude maternal age controls (columns 2 and 7), (ii) exclude maternal schooling attainment controls (columns 3 and 8), and (iii) exclude all maternal controls (columns 4 and 9). For purposes of comparability, we report our baseline estimates with maternal controls – with and without economic activity proxy controls – in columns 1 and 6, respectively. The estimates are remarkably stable. In general, this sensitivity analysis suggests that there is a small degree of correlation between these observable characteristics and the incidence of congenital anomalies, and as a result the degree of selection on observables

explains a minimal share of the relationship between the volume of ordnance and the risk of congenital anomalies.<sup>18</sup>

For completeness, we also report coefficients ( $\theta_t$ ) of the Vieques year-specific selection effects on the two summary indices of cohort/child and maternal characteristics, generated from analogous event study regression models (equation (2)) (see Figure 5). We find mixed evidence of negative selection based on cohort/child characteristics: there is a statistically significant reduction in the cohort characteristics index z-score in years 2000 and 2002 (permutation test-based p-values are respectively  $<0.001$ ,  $0.039$ ), as well as significant reductions decreases in years 1993 and 1998, in which there were substantial drops in the volume of ordnance exposure (p-values =  $0.013$ ,  $0.013$ ). Relatedly, we find slight, but statistically insignificant, evidence of a change in the degree of positive selection of births based on maternal characteristics. This characterization of year-specific correlates of births again suggests that we our treatment effect estimates represent an underestimate of the true effects, given this documented pattern of negative selection post-end of naval exercises.

### *Alternative Inference Methods*

Inference based on this differences-in-differences research design relies on two robust methodologies that appropriately take the research design into account. For completeness, we also report as a robustness check OLS estimates of the impacts on the incidence of congenital anomalies from models that include a lagged dependent variable (with heteroskedasticity-robust standard errors) to account for autocorrelation (Table 6, columns 5 and 10). The estimates are of similar

---

<sup>18</sup> Our data includes information on whether mothers smoked during pregnancy (1.7% of mothers in Vieques smoked versus 1.3% for the rest of PR). We re-estimate our main specifications including this measure as an additional control and our results remain unchanged. However, we do not include this control in the specifications reported in our main results given concerns about the endogeneity of maternal smoking.

magnitude and are estimated with similar precision. Estimates looking at the effect of the end of the bombing suggest a reduction of between 5.2 and 5.6 per thousand in the incidence of congenital anomalies. Finally, we rely on the large time dimension of the data to conduct inference based on asymptotics as the time dimension becomes large. Specifically, we estimate standard errors using a nonparametric covariance matrix estimator that produces standard errors that are robust to general forms of cross-sectional (spatial) and temporal dependence when the time dimension becomes large (Driscoll and Kraay 1998) (henceforth DK). The point estimates and standard errors from these models are reported in Appendix Table A1. According to this alternate approach, the reduction in the incidence of congenital anomalies is significant at the 1 percent level. These various robustness checks imply that we find large and robust effects of bombing ordnance on congenital anomalies.

## **VII. Conclusion**

We identify the relationship between frequent explosions and high-ordnance military exercises and child health at birth. Our results indicate that there is a negative effect of these exercises on health outcomes at birth with the most robust effects on congenital anomalies. Our analysis helps inform the literature regarding why stress and environmental factors may be important channels via which conflict affects human development and also speaks to the direct effect of military practice sessions on the local population – a practice that continues today by militaries around the world.

While it is challenging to make exact comparisons with other bombing sites (active war zones), it is perhaps interesting to compare the magnitude of ordnance with those in war zones studied in the literature. Vieques experienced between 100,000 and 300,000 tons of ordnance

versus approximately 454,000 in the Korean war and over 6 million in Vietnam (Miguel and Roland, 2011). There are, to our knowledge, few comparable estimates of the effects of bombing on health outcomes at birth. However, our estimates on pre-term birth are consistent to the effects of terrorism on health at birth (i.e., 1/1000 in Quintana-Domeque and Ródenas-Serrano, 2017) and not inconsistent with estimates of conflict on child height (Bundervoet, Verwierp and Akresh, 2009). To our knowledge there are no other studies of the impacts of military preparation exercises on the health of the local population.

While previous literature looking at the longer-term effects of war on economic outcomes has found small economic impacts (c.f. Miguel and Roland, 2011), there is reason to suspect that in this case there may be longer lasting effects given the areas in which we find effects. Previous health economics research has documented a number of long term consequences related to being born in poor health including long term effects on education, welfare receipt, earnings and adult health (Black, Devereux, and Salvanes 2007; Oreopoulos, Stabile, Roos, and Walid 2008) and specifically, the effects of prenatal exposure to environmental pollution from radioactive fallout on adult outcomes (IQ, educational attainment, labor earnings) (Black et al 2017). One study has even documented long-term consequences of fetal health on adults in Puerto Rico (Sotomayor 2013). Therefore, there is reason to believe that our findings of short-term effects on infant health in this context may have longer-term effects on educational attainment, labor force attachment, and adult health. Further study is required to better understand the mechanisms through which the bombings affected infant health and to inform public policy.

## References

- Al-Hadithi, T., J. K. Al-Diwan, A. M. Saleh, and N. P. Shabila (2012). Birth defects in Iraq and the plausibility of environmental exposure: A review. *Conflict and Health*, 6(3), 1-7.
- Al-Sabbak, M., Sadik Ali, S., Savabi, O., Savabi, G., Dastgiri, S., & Savabieasfahani, M. (2012). Metal Contamination and the Epidemic of Congenital Birth Defects in Iraqi Cities. *Bulletin of Environmental Contamination and Toxicology*, 89(5), 937–944.  
<http://doi.org/10.1007/s00128-012-0817-2>
- Alderman, H., Hoddinott, J., and Kinsey, B. (2006). Long Term Consequences of Early Childhood Malnutrition. *Oxford Economic Papers*, 58(3), 450–74.
- Almond, D., & Currie, J. (2011). Killing Me Softly: The Fetal Origins Hypothesis. *The Journal of Economic Perspectives*, 25(3), 153-172.
- Almond, D., Chay, K. Y., & Lee, D. S. (2005). The Costs of Low Birth Weight. *Quarterly Journal of Economics*, 120(3), 1031-1083.
- ATSDR. (2001). *Public Health Assessment: Drinking Water Supplies and Groundwater Pathway Evaluation, Isla de Vieques Bombing Range, Vieques, Puerto Rico*. Atlanta: Agency for Toxic Substances and Disease Registry. Retrieved Aug. 2014, from [http://www.atsdr.cdc.gov/HAC/PHA/reports/isladevieques\\_10162001pr/printview.html](http://www.atsdr.cdc.gov/HAC/PHA/reports/isladevieques_10162001pr/printview.html)
- ATSDR. (2003). *Public Health Assessment: Air Pathway Evaluation, Isla de Vieques Bombing Range, Vieques, Puerto Rico*. Atlanta: Agency for Toxic Substances and Disease Registry.
- Barker, D. J. (1990). The Fetal and Infant Origins of Adult Disease. *BMJ*, 301(6761), 1111.

- Barrios, T. Diamond, R., Imbens, G. W., & Kolesár, M. (2012). Clustering, Spatial Correlations, and Randomization Inference. *Journal of the American Statistical Association*, 107(498), 578-91.
- Behrman, J., & Rosenzweig, M. (2004). Returns to Birthweight. *Review of Economics and Statistics*, 86(2), 586-601.
- Bertrand, M., Duflo, E., & Mullainathan, S. (2004). How Much Should We Trust Differences-in-Differences Estimates? *Quarterly Journal of Economics*, 119(1), 249-75.
- Beydoun, H., and Saftlas, A. (2008). Physical and Mental Health Outcomes of Prenatal Maternal Stress in Human and Animal Studies: A Review of Recent Evidence. *Paediatric and Perinatal Epidemiology* 22(5), 438–466.
- Black, S., Devereux, P., & Salvanes, K. (2007). From the Cradle to the Labor Market? The Effect of Birth Weight on Adult Outcomes. *Quarterly Journal of Economics*, 122(1), 409-439.
- Black, S., Bütikofer, A., Devereux, P.J., & Salvanes, K.J. (2017). This is Only a Test? Long-Run Impacts of Prenatal Exposure to Radioactive Fallout. NBER Working Paper No. 18987.
- Brewer, M., Crossley, T.F., and Joyce, R. (2013). Inference with Difference-in-Differences Revisited, IZA Discussion Papers 7742, Institute for the Study of Labor (IZA), November.
- Buchmueller, T.C., DiNardo, J., and Valletta, R.G. (2011). The Effect of an Employer Health Insurance Mandate on Health Insurance Coverage and the Demand for Labor: Evidence from Hawaii, *American Economic Journal: Economic Policy*, 3, 25–51.

- Bundervoet, T., Verwimp, P., & Akresh, R. (2009). Health and Civil War in Rural Burundi. *Journal of Human Resources*, 44(2), 536-63.
- Camacho, A. (2008, May). Stress and Birth Weight: Evidence from Terrorist Attacks. *The American Economic Review*, 98(2), 511-515.
- Cameron, C. Gelbach, J., & Miller, D. (2008). Bootstrap-Based Improvements for Inference with Clustered Errors. *Review of Economics and Statistics*, 90(3), 414-27.
- Canay, I.A., J.P. Romano, and A.M. Shaikh (2014). Randomization Tests under an Approximate Symmetry Assumption, *Econometrica*, 85(3), 1031-1030.
- Carmichael, S. L., Shaw, G. M., Yang, W., Abrams, B., & Lammer, E. J. (2007). Maternal Stressful Life Events and Risks of Birth Defects. *Epidemiology* (Cambridge, Mass.), 18(3), 356–361. <http://doi.org/10.1097/01.ede.0000259986.85239.87>
- Carmichael, S. L., Ma, C., Tinker, S., Rasmussen, S. A., Shaw, G. M., & the National Birth Defects Prevention Study. (2014). Maternal stressors and social support as risks for delivering babies with structural birth defects. *Paediatric and Perinatal Epidemiology*, 28(4), 338–344. <http://doi.org/10.1111/ppe.12123>
- Chay, K., & Greenstone, M. (2003). The Impact of Air Pollution on Infant Mortality: Evidence from Geographic Variation in Pollution Shocks Induced by a Recession. *Quarterly Journal of Economics*, 118(3), 1121-1167.
- Currie, J., Greenstone, M., & Moretti, E. (2011). Superfund Cleanups and Infant Health. *American Economic Review Papers and Proceedings*, 101(3), 435-41.

- Currie, J., & Hyson, R. (1999). Is the Impact of Health Shocks Cushioned by Socioeconomic Status? The Case of Low Birth Weight. *American Economic Review Papers and Proceedings*, 89(2), 245-250.
- Currie, J., Neidell, M., & Schmieder, J. F. (2009). Air Pollution and Infant Health: Lessons from New Jersey. *Journal of Health Economics*, 28, 688-703.
- Currie, J., Ray, S. H., & Neidell, M. (2011). Quasi-Experimental Studies Suggest That Lowering Air Pollution Levels Benefit Infants' And Children's Health. *Health Affairs*, 30(12), 2391-99.
- Currie, J., & Schwandt, H. (2014). The 9/11 Dust Cloud and Pregnancy Outcomes: A Reconsideration. NBER Working Paper #20368.
- Davis, J. S., Hayes-Conroy, J.S., & Jones, V.M. (2007). Military Pollution and Natural Purity: Seeing Nature and Knowing Contamination in Vieques, Puerto Rico. *GeoJournal*, 69(3).
- de Weerrth, C., & Buitelaar, J. (2005). Physiological Stress Recreatiity in Human Pregnancy. A Review." *Neuroscience and Behavioral Review*, 29:295-312.
- Dehejia, R., & Lleras-Muney, A. (2004). Booms, Busts, and Babies' Health. *Quarterly Journal of Economics*, 119, 1091-1130.
- Driscoll, J. C., & Kraay, A. C. (1998). Consistent Covariance Matrix Estimation with Spatially Dependent Panel Data. *Review of Economics and Statistics*, 80(4), 549-60.
- Dustmann, C., and Fasani, F. (2016). The Effect of Local Area Crime on Mental Health. *Economic Journal*, 126(593), 978-1017.

- EPA. (2013). *May 2013 Vieques Activities Update*. Environmental Protection Agency. Retrieved Aug. 2014, from [http://www.epa.gov/region2/vieques/may2013viequesupdate\\_.pdf](http://www.epa.gov/region2/vieques/may2013viequesupdate_.pdf)
- Ferman, B., and Pinto, C. (2019). "Inference in Differences-in-Differences with Few Treated Groups and Heteroskedasticity." *Review of Economics and Statistics*, 101(3), 452-467.
- Figueroa, N., Suarez, E., De La Torre, T., Torres, M., & Perez, J. (2009). *Incidencia y Mortalidad de Cancer en Vieques, 1990-2004*. San Juan: Registro Central de Cancer de Puerto Rico.
- Fisher, R. A. (1925). *The Design of Experiments* (1st ed.) London: Oliver & Boyd.
- Galdo, J. (2013). Long-Run Labor-Market Impacts of Early-Life Exposure to Civil War: Evidence from the Shining Path in Perú. *Economic Development and Cultural Change*, 61(4), 789-823.
- Hadi, A. S. (1992). Identifying Multiple Outliers in Multivariate Data. *Journal of the Royal Statistical Society, Series B* 54, 761-771.
- Hadi, A. S. (1994). A Modification of a Method for the Detection of Outliers in Multivariate Samples. *Journal of the Royal Statistical Society, Series B* 56, 393–396.
- Hansen, D., Lou, H.C., and Olsen, J. (2000). Serious life events and congenital malformations: a national study with complete follow-up. *The Lancet*, 356(9233), 875-880.
- Hopenhayn-Rich, C., Hertz-Picciotto, I., Browning, S., Ferreccio, C., Peralta, C. (1999). Reproductive and developmental effects associated with chronic arsenic exposure. In Chappel, W.R., Abernathy, C. O., & Calderon, R. L. (eds.), *Arsenic Exposure and Health*

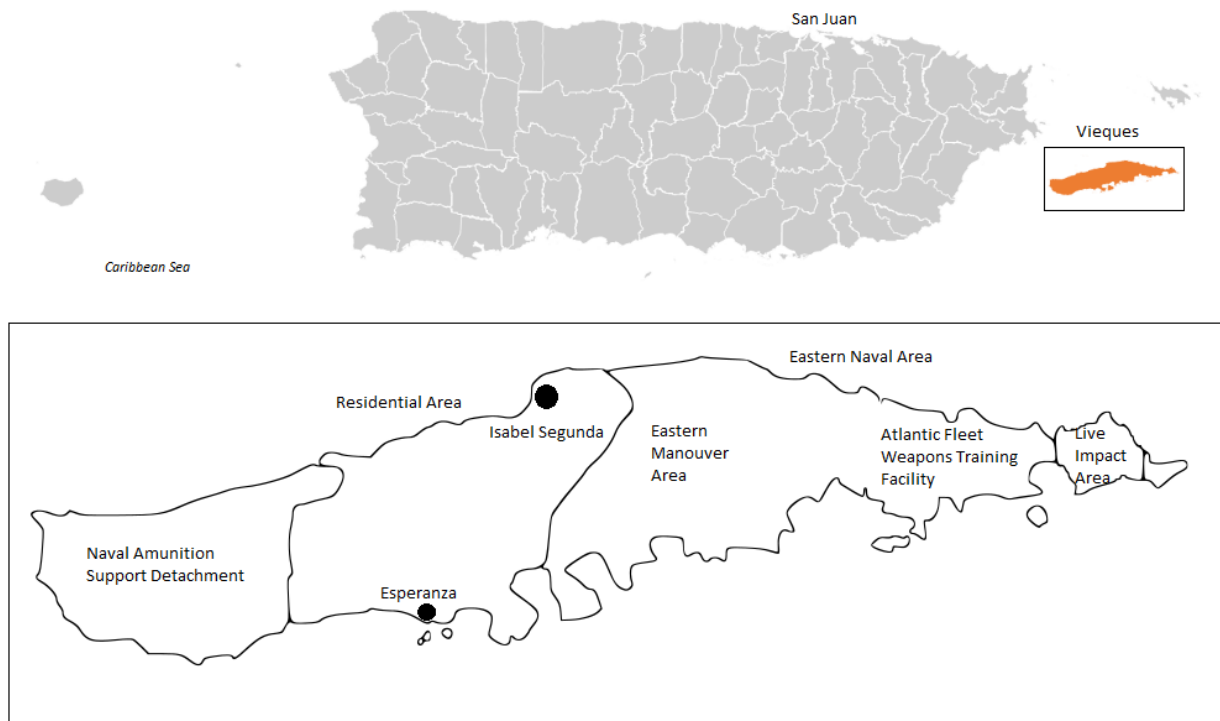
- Effects: Proceedings of the Third International Conference on Arsenic Exposure and Health Effects*, San Diego, California, 12-15 July, 1998, 115-64.
- Ibragimov, R., and U.K. Müller (2010). t-Statistic Based Correlation and Heterogeneity Robust Inference, *Journal of Business & Economic Statistics*, 28(4), 453–468.
- Ibragimov, R. and U.K. Müller (2016). Inference with Few Heterogeneous Clusters, *Review of Economics and Statistics*, 98(1), 83–96.
- Imbens, G., & Kolesár, M. (2012). Robust Standard Errors in Small Samples: Some Practical Advice. NBER Working Paper No. 18478.
- Kling, J., Liebman, J., and Katz, L. (2007), “Experimental Analysis of Neighborhood Effects,” *Econometrica*, 75, 83–119.
- Kramer, M. S. (1987). Determinants of Low Birth Weight: Methodological Assessment and Meta-Analysis. *Bulletin of the World Health Organization*, 65(5), 663-737.
- León, G. (2012). Civil conflict and Human Capital Accumulation: Long Term Consequences of Political Violence in Perú. *Journal of Human Resources*, 47(4), 991-1023.
- Mansour, H., & Rees, D. I. (2012, Sept.). Armed Conflict and Birth Weight: Evidence from the al-Aqsa Infatada. *Journal of Development Economics*, 99(1), 190-199.
- Massol-Deyá, A., Pérez, D., Pérez, E., Berríos, M., & Díaz, E. (2005). Trace Elements Analysis in Forage Samples from a U.S. Navy Bombing Range (Vieques, Puerto Rico). *International Journal of Environmental Research and Public Health*, 2(2), 263-266.
- MacKinnon, J. G., & Webb, M. D. (2017). Wild Bootstrap Inference for Wildly Different Cluster Sizes, *Journal of Applied Econometrics*, 32(2), 233-254.

- McCaffrey, K. (2009, Sept-Oct). Fish, Wildlife, and Bombs: The Struggle to Clean Up Vieques. *NACLA Report*, 35-41.
- McCormick, M. C. (1985). The Contribution of Low Birth Weight to Infant Mortality and Childhood Morbidity. *New England Journal of Medicine*, 2(2), 263-266.
- Miguel, E., Roland, G. (2011). The Long-Run Impact of Bombing Vietnam. *Journal of Development Economics* 96, 1-15.
- Neyman, J. (1990). On the Application of Probability Theory to Agricultural Experiments: Essay on Principles--Section 9. *Statistical Science*, 5(4), 465-80.
- Nordstrom, S., Beckman, L., & Nordenson, I. (1979). Occupational and environmental risks in and around a smelter in northern Sweden. *Hereditas*, 90, 291-6.
- O'Brien, P. (1984), "Procedures for Comparing Samples with Multiple Endpoints," *Biometrics*, 40, 1079-1087.
- Oreopoulos, P., Stabile, M., Roos, L., & Walld, R. (2008). The Short, Medium, and Long Term Effects of Poor Infant Health. *Journal of Human Resources*, 43(1), 88-138.
- Ortiz-Roque, C., & Lopez-Rivera, Y. (2004). Mercury Contamination in Reproductive Age Women in a Caribbean Island: Vieques. *Journal of Epidemiology and Community Health* (1979-), 58(9), 756-757.
- Pollack, R., & Divon, M. (1992). Intrauterine Growth Retardation: Definition, Classification, and Etiology. *Clinical Obstetrics and Gynecology*, 35(1), 99-107.
- Porter, J. W., Barton, J. V., & Torres, C. (2011). Ecological, Radiological, and Toxicological Effects of Naval Bombardment on the Coral Reefs of Isla de Vieques, Puerto Rico. In G.

- E. Machlis, T. Hanson, Z. Špirić, & J. E. McKendry, *Warfare Ecology: A New Synthesis for Peace and Security: Proceedings of the NATO Advanced Study Institute / Advanced Research Workshop* (pp. 65-122). Dordrecht: Springer Netherlands.
- Puerto Rico. (1999). *Informe de la Comisión Especial de Vieques al Gobernador de Puerto Rico, Hon. Pedro Rosselló*. San Juan.
- Quintana-Domeque, C., & Ródenas-Serrano, P. (2017). The Hidden Costs of Terrorism: The Effects on Health at Birth. *Journal of Health Economics*, 56, 47-60.
- Sotomayor, O. (2013). Fetal and Infant Origins of Diabetes and Ill Health: Evidence from Puerto Rico's 1928 and 1932 Hurricanes. *Economics and Human Biology*, 11(3), 281-293.
- Stephenson, J. et al (2018). Before the Beginning: nutrition and lifestyle in the preconception period and its importance for future health. *Lancet*; 391, 1830-41.
- U.S. Navy. (1999). *The National Security Need for Vieques: A study prepared for The Secretary of the Navy*.
- Wadhwa, P. D., Sandman, C. A., Porto, M., Dunkel-Schetter, C., & Garite, T. (1993, October). The Association between Prenatal Stress and Infant Birth Weight and Gestational Age at Birth: A Prospective Investigation. *American Journal of Obstetrics & Gynecology*, 169(4), 858-65.
- Wargo, J. (2009). *Green Intelligence: Creating Environments that Protect Human Health*. New Haven: Yale University Press.
- Zavala Zegarra, D. (2000). *Incidencia de Cancer en Vieques*. San Juan: Registro Central de Cancer de Puerto Rico.

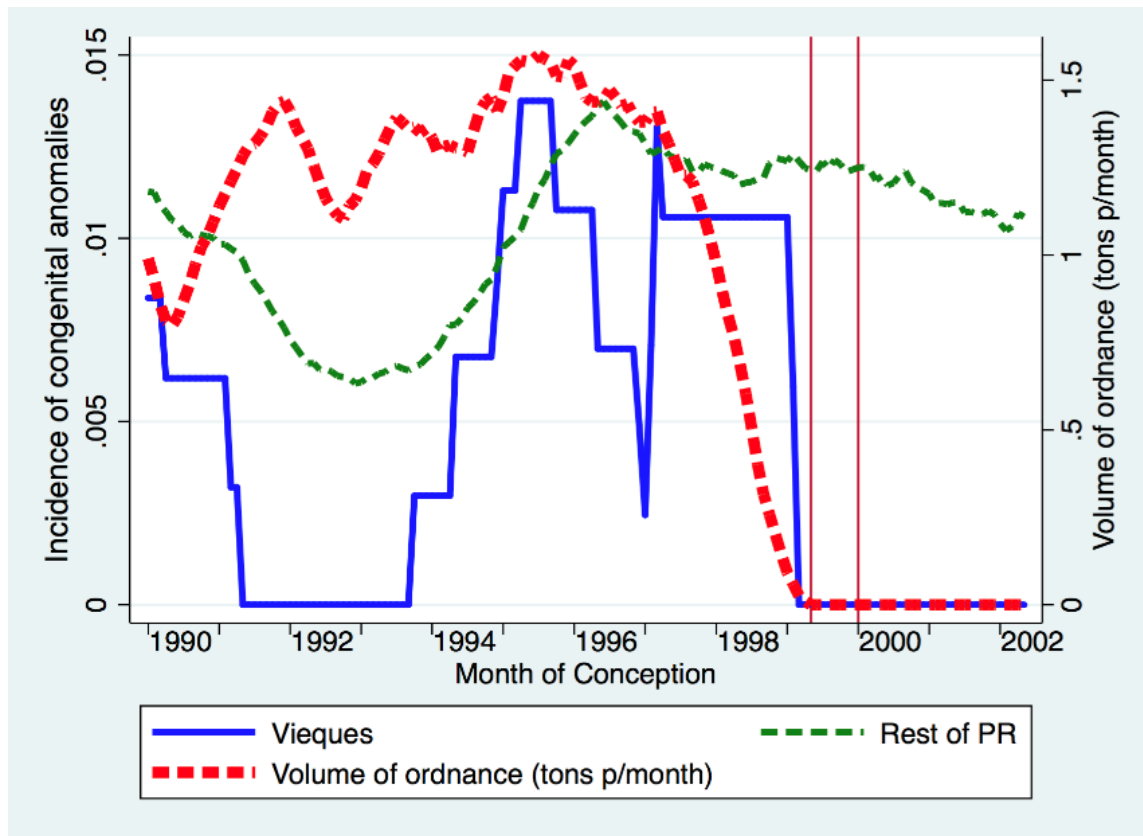
Zhu, P., Tao, F., Hao, J., Sun, Y., & Jiang, X. (2010, July). Prenatal Life Events Stress: Implications for Preterm Birth and Infant Birthweight. *American Journal of Obstetrics & Gynecology*, 203(1), 1-8.

**Figure 1:** Map of Puerto Rico and Vieques with Former Division of Residential and Military Zones



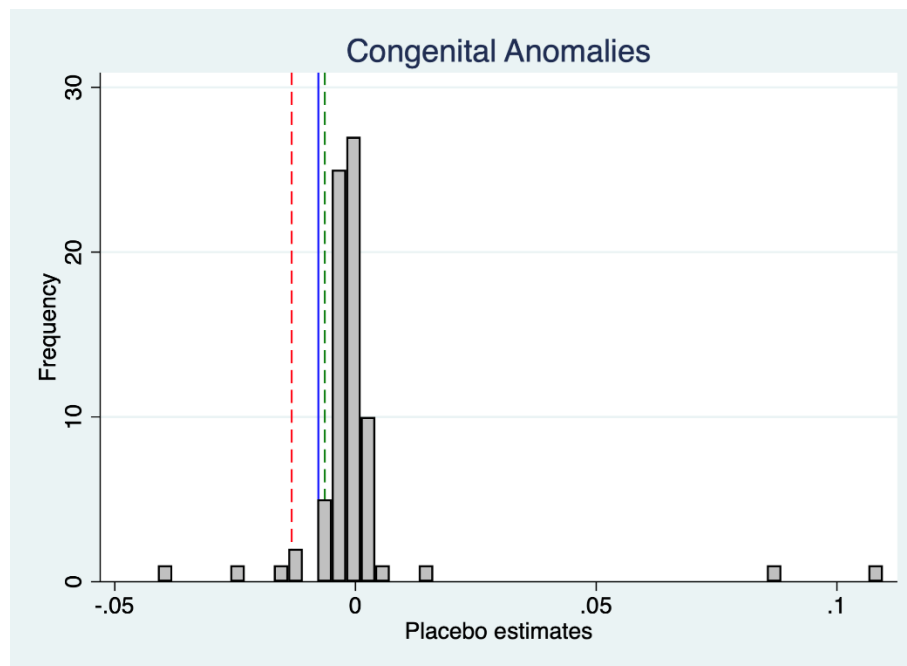
Notes: Map is for illustrative purposes only. Map may not be drawn to scale.

**Figure 2:** Ordnance Volume During Naval Exercises and Incidence of Congenital Anomalies, 1990-2003



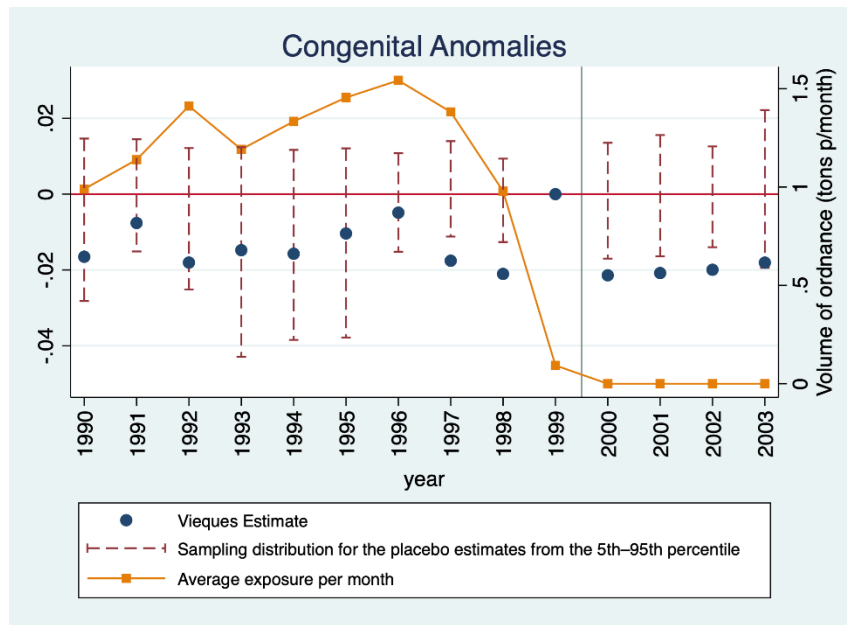
Notes: Ordnance volume during naval exercises for potential (9-month) gestation period for births conceived in each month (bold dashed line) (12-month moving average); incidence of congenital anomalies by month of conception (24-month moving average), for live births of mothers resident in Vieques (solid) and in the rest of PR (dashed line). The first vertical line marks April 1999, when U.S. Navy reduced its operations; the second vertical line marks the end of the year 1999. The correlation between Vieques congenital anomalies and the Volume of ordnance is 0.43. For the rest of PR it is -0.34.

**Figure 3:** Distribution of Placebo Estimates – Permutation Tests for Inference on Effects of End of Bombing Activity and the Incidence of Congenital Anomalies, 1990-2003



Notes: Distribution of placebo municipality effects on the incidence of congenital anomalies, estimated from a permutation test, reported in Table 3 (columns 1). The (red and green) dashed lines are respectively the 5th and 10th percentile values (other than Vieques) of the distribution of estimated placebo effects; the solid vertical line is the Vieques value.

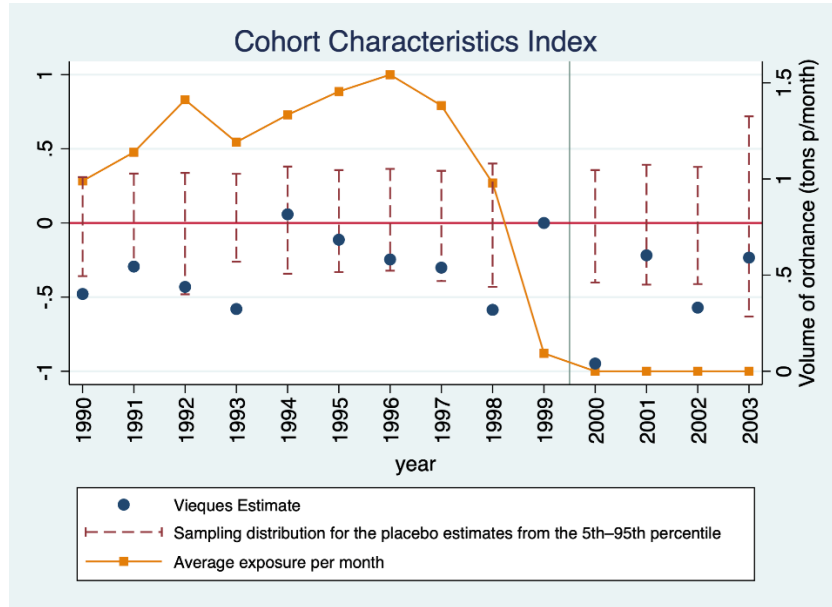
**Figure 4: Year-Specific Effects – End of Bombing Activity and Incidence of Congenital Anomalies, 1990-2003**



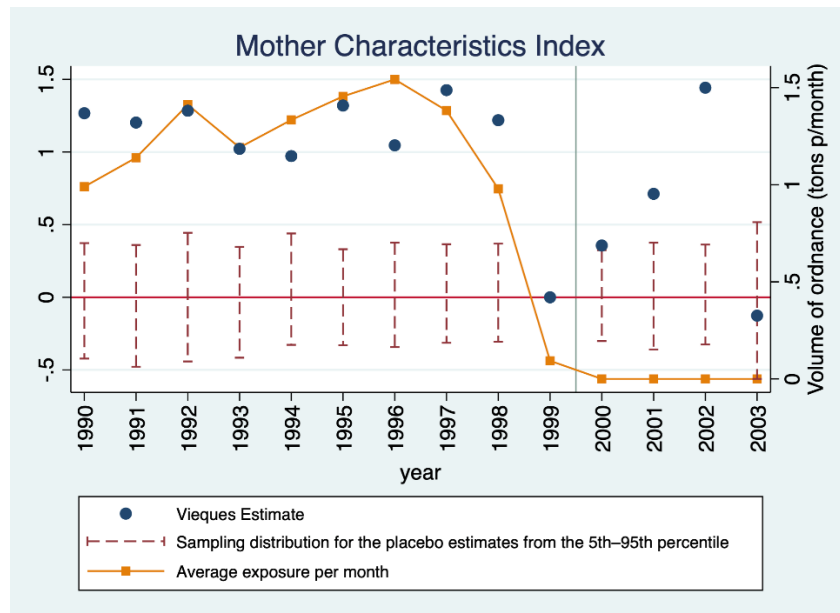
**Notes:** The figure plots the coefficients on Vieques-specific year effects ( $\theta_t$  from equation 2) for the incidence of congenital anomalies. The solid vertical line denotes the end of naval practices. The dashed vertical lines are the sampling distribution for the placebo estimates from the 5<sup>th</sup> to the 95<sup>th</sup> percentile for each year. The solid orange line represents (the average of the) average tonnage of ordnance used in practices during each child's 9-month potential gestation period, for newborns conceived in each calendar year.

**Figure 5: Year-Specific Correlates of End of Bombing Activity (Selection on Observables), 1990-2003**

*Panel A: Cohort Characteristics Index z-Score*



*Panel B: Mother Characteristics Index z-Score*



**Notes:** Each figure plots the coefficients on Vieques-specific year effects ( $\theta_t$  from equation 2) for each “selection on observables” measure of interest: (a) the cohort characteristics index z-score, and (b) the mother characteristics index z-score (reported in Panels A and B, respectively). The solid vertical line denotes the end of naval practices. The dashed vertical lines are the sampling distribution for the placebo estimates from the 5<sup>th</sup> to the 95<sup>th</sup> percentile for each year. The solid orange line represents (the average of the) average tonnage of ordnance used in practices during each child’s 9-month potential gestation period, for newborns conceived in each calendar year.

**Table 1:** Volume of Ordnance – Summary Statistics, 1985-1999

	<b>Mean</b>	<b>Std. Dev.</b>	<b>Min.</b>	<b>Max.</b>	<b>N</b>
	(1)	(2)	(3)	(4)	(5)
Volume of ordnance (monthly) [Hundreds of tons per month]	1.201	1.405	0.000	6.585	184
Volume of ordnance (9-month potential gestation per.) [Hundreds of tons per month]	1.220	0.549	0.000	2.767	176

Notes: Reported are the sample mean, standard deviation, minimum and maximum of each variable; based on aggregated data at the month level for the period 1985-1999.

**Table 2:** Descriptive Statistics, Puerto Rico Natality Files, 1990-2003

Sample	All Residents		Vieques Residents		Other Residents	
	Mean / (SD)	N	Mean / (SD)	N	Mean / (SD)	N
	(1)	(2)	(3)	(4)	(5)	(6)
<b>Panel A: Neonatal Outcomes</b>						
Congenital anomaly	0.010 (0.027)	12,228	0.005 (0.020)	159	0.010 (0.027)	12,069
Premature birth (< 37 weeks)	0.144 (0.076)	12,228	0.147 (0.111)	159	0.144 (0.076)	12,069
Extremely premature birth (< 28 weeks)	0.006 (0.014)	12,228	0.007 (0.026)	159	0.006 (0.013)	12,069
Low birth weight	0.106 (0.061)	12,228	0.115 (0.109)	159	0.106 (0.061)	12,069
Low APGAR score (5-min)	0.011 (0.020)	12,228	0.010 (0.029)	159	0.011 (0.020)	12,069
<b>Panel B: Child and Mother Characteristics</b>						
Female birth	0.486 (0.091)	12,228	0.500 (0.154)	159	0.486 (0.090)	12,069
High school graduate	0.298 (0.095)	12,228	0.371 (0.156)	159	0.297 (0.093)	12,069
Higher education (some/completed)	0.394 (0.120)	12,228	0.216 (0.134)	159	0.396 (0.118)	12,069
Mother's age	24.7 (1.3)	12,228	23.7 (2.1)	159	24.7 (1.2)	12,069
Teenage mother	0.093 (0.059)	12,228	0.144 (0.109)	159	0.092 (0.057)	12,069
Mother's age 40+	0.012 (0.020)	12,228	0.015 (0.044)	159	0.012 (0.019)	12,069

**Notes:** Reported are the sample mean and standard deviation of each variable; based on aggregated data at the municipality-by-month level. The dataset is composed of 12,228 municipality-by-month cells.

**Table 3:** End of Bombing Activity and Health Outcomes at Birth, 1990-2003

	Dependent variables:					
	Congenital anomaly	Premature birth	Extremely preterm birth	Low birth weight	Low APGAR score	Health Outcomes Index z-Score
	(1)	(2)	(3)	(4)	(5)	(6)
<i>End of Naval Practices</i> × <i>Vieques</i>	-0.0077*	0.028*	-0.0018*	-0.0020*	0.0062*	-0.018*
Ferman-Pinto test p-value	[0.070]	[0.070]	[0.064]	[0.073]	[0.060]	[0.063]
Permutation test (one-sided)						
Percentile rank	6/77	71/77	20/77	38/77	75/77	52/77
Approximate p-value	[0.078]	[0.922]	[0.260]	[0.493]	[0.974]	[0.675]
Mother Characteristics	Yes	Yes	Yes	Yes	Yes	Yes
Municipality & Month-Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Outcome Mean	0.010	0.144	0.006	0.106	0.011	0.000
N	12,228	12,228	12,228	12,228	12,228	12,228

Notes: Coefficient estimates from municipality fixed effects regressions with individual level birth data collapsed into 159 month-municipality cells; each set of estimates (by column) is from a different regression. To conduct inference, we report the (a) p-value for the Ferman and Pinto (2019) test for inference in differences-in-differences regression models with few treated groups and heteroskedasticity; statistically significant at (\*) 10 percent, (\*\*)5 percent, (\*\*\*) 1 percent levels, respectively; and (b) p-values from a one-sided test based on the distribution of placebo effects from a permutation exercise, as well as the percentile rank of the coefficients based on the distribution of these placebo estimates. Mother characteristics' controls include (month-municipality cell averages of) indicators for high school graduate, higher education (some/completed), teenage pregnancy, mother's age at childbirth of 40 years or greater, and maternal age.

**Table 4:** Bombing Activity and Arsenic Pollution Levels in Water – Live Impact Area Inner Range Water, 1985-1999

Dependent variables:	Maximum Arsenic Levels Above EPA Limit				
	(1)	(2)	(3)	(4)	(5)
Tons per month, Quarter $t$	0.171** (0.072)	0.172** (0.071)	0.183** (0.072)	0.177** (0.075)	0.176** (0.078)
$\Delta$ Outcome from One SD $\Delta$ in TPM	0.14	0.14	0.15	0.15	0.14
<i>Other Controls</i>					
Tons per month, Quarter $t-1$		0.086 (0.072)	0.087 (0.072)	0.090 (0.076)	0.091 (0.078)
Tons per month, Quarter $t-2$			0.003 (0.071)	0.004 (0.073)	0.005 (0.076)
Tons per month, Quarter $t-3$				0.010 (0.085)	0.010 (0.087)
Tons per month, Quarter $t+1$					-0.011 (0.086)
Municipality Fixed Effects	Yes	Yes	Yes	Yes	Yes
Month-Year Fixed Effects	Yes	Yes	Yes	Yes	Yes
Outcome mean	0.275	0.280	0.286	0.292	0.292
N	51	50	49	48	48

Notes: Coefficient estimates from time series regressions; each set of estimates (by column) is from a different regression. Heteroskedasticity-robust standard errors in parentheses; significant at (\*) 90 percent, (\*\*) 95 percent, (\*\*\*) 99 percent confidence levels. The change in each of the variables from a one (1) standard deviation increase in the tons of ordnance per month is calculated as the product of the relevant coefficient estimate and the standard deviation of the ordnance volume measure (= coefficient estimate  $\times$  0.822).

**Table 5:** Correlates of End of Bombing Activity (Selection on Observables), 1990-2003

Dependent variables:						
	Unempl. Rate	ln(Unempl. rate)	Cohort Size	ln(Cohort size)	Female birth	Cohort Characteristics Index
	(1)	(2)	(3)	(4)	(5)	(6)
<i>End of Naval Practices</i> × <i>Vieques</i>	3.23***	0.215**	6.917*	-0.079*	-0.055***	-0.255**
Ferman-Pinto test p-value	[0.009]	[0.015]	[0.081]	[0.074]	[<0.001]	[0.012]
Permutation tests (one-sided)						
Approximate p-value (< 0)	[0.909]	[0.948]	[0.818]	[0.169]	[0.013]	[0.026]
Approximate p-value (> 0)	[0.104]	[0.065]	[0.195]	[0.844]	[1.000]	[0.987]
Mother Characteristics	No	No	No	No	No	No
Municipality Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Month-Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
N	12,228	12,228	12,228	12,228	12,228	12,228

Notes: Coefficient estimates from municipality fixed effects regressions with individual level birth data collapsed into 159 month-municipality cells; each set of estimates (by column) is from a different regression. To conduct inference, we report the (a) p-value for the Ferman and Pinto (2019) test for inference in differences-in-differences regression models with few treated groups and heteroskedasticity; statistically significant at (\*) 10 percent, (\*\*)5 percent, (\*\*\*) 1 percent levels, respectively; and (b) p-values from one-sided tests based on the distribution of placebo effects from a permutation exercise.

**Table 5 (continued):** Correlates of End of Bombing Activity (Selection on Observables), 1990-2003

	Dependent variables:						
	High school dropout (7)	High school graduate (8)	Higher education (9)	Mother's age (10)	Teenage mother (11)	Mother's age 40+ (12)	Maternal Characteristics Index (13)
<i>End of Naval Practices</i> × <i>Vieques</i>	0.018	0.069***	-0.086***	-0.951***	0.061***	-0.005**	-0.313***
Ferman-Pinto test p-value	[0.268]	[<0.001]	[<0.001]	[<0.001]	[<0.001]	[0.033]	[<0.001]
Permutation tests (one-sided)							
Approximate p-value (< 0)	[0.740]	[0.987]	[0.013]	[0.039]	[1.000]	[0.130]	[0.039]
Approximate p-value (> 0)	[0.273]	[0.026]	[1.000]	[0.974]	[0.013]	[0.883]	[0.974]
Mother Characteristics	No	No	No	No	No	No	No
Municipality Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Month-Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	12,228	12,228	12,228	12,228	12,228	12,228	12,228

**Notes:** Coefficient estimates from municipality fixed effects regressions with individual level birth data collapsed into 159 month-municipality cells; each set of estimates (by column) is from a different regression. To conduct inference, we report the (a) p-value for the Ferman and Pinto (2019) test for inference in differences-in-differences regression models with few treated groups and heteroskedasticity; statistically significant at (\*) 10 percent, (\*\*)5 percent, (\*\*\*) 1 percent levels, respectively; and (b) p-values from a one-sided test based on the distribution of placebo effects from a permutation exercise, as well as the percentile rank of the coefficients based on the distribution of these placebo estimates.

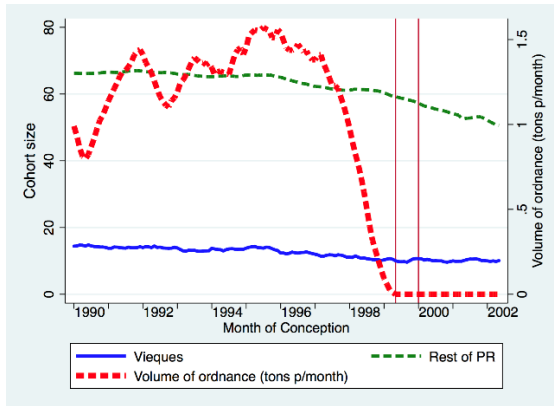
**Table 6: Robustness Tests**

	Dependent variable: Congenital anomaly (1/0)									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
End of Naval Practices × Vieques	-.0077*	-.0074*	-.0078*	-.0079*	-.0056*	-.0072*	-.0069*	-.0072*	-.0073*	-.0052*
Ferman-Pinto test p-value	[0.071]	[0.068]	[0.065]	[0.061]	[0.084]	[0.077]	[0.096]	[0.076]	[0.083]	[0.067]
Permutation test (< 0)										
Approximate p-value	[0.078]	[0.091]	[0.078]	[0.078]	[0.078]	[0.078]	[0.091]	[0.078]	[0.078]	[0.078]
Mother Characteristics										
Education Controls	Yes	Yes			Yes	Yes	Yes			Yes
Age Controls	Yes		Yes		Yes	Yes		Yes		Yes
Municipality FE's	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Month-Year FE's	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Lag of Dep. Variable					Yes					Yes
Local Economic Activity Controls						Yes	Yes	Yes	Yes	Yes
N	12,228	12,228	12,228	12,228	12,137	12,228	12,228	12,228	12,228	12,137

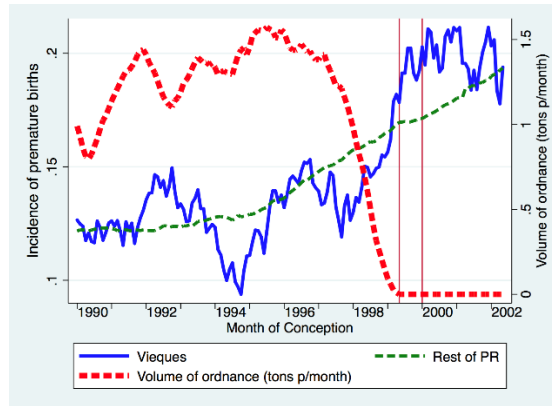
**Notes:** Coefficient estimates from municipality fixed effects regressions with individual level birth data collapsed into 159 month-municipality cells; each set of estimates (by column) is from a different regression. To conduct inference, we report the (a) p-value for the Ferman and Pinto (2019) test for inference in differences-in-differences regression models with few treated groups and heteroskedasticity; statistically significant at (\*) 10 percent, (\*\*)5 percent, (\*\*\*) 1 percent levels, respectively; and (b) p-values from a one-sided test based on the distribution of placebo effects from a permutation exercise, as well as the percentile rank of the coefficients based on the distribution of these placebo estimates. Mother characteristics' controls include (month-municipality cell averages of) indicators for high school graduate, higher education (some/completed), teenage pregnancy, mother's age at childbirth of 40 years or greater, and maternal age.

**Appendix Figure A.1: Ordnance Volume During Naval Exercises and Other Neonatal Health Outcomes, 1990-2003**

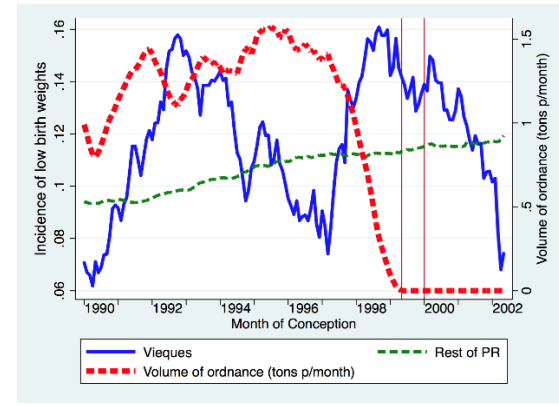
**Panel A: Live Births**



**Panel B: Premature Births**

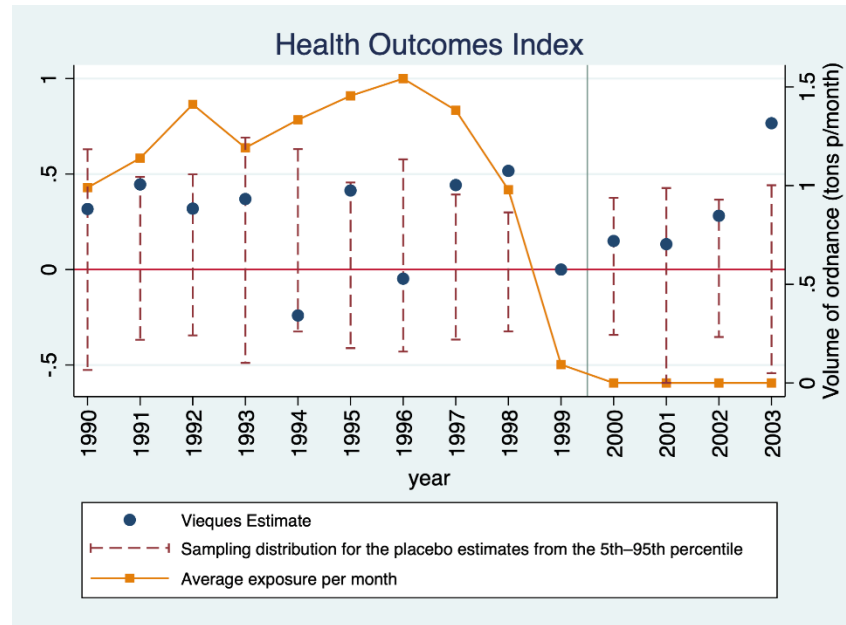


**Panel C: Low Birth Weight**



**Notes:** Ordnance volume during naval exercises for potential (9-month) gestation period for births conceived in each month (12-month moving average) (bold dashed line); number of live births (Panel A), incidence of premature births (Panel B), and incidence of low birthweight births (Panel C) by month of conception (24-month moving average), for live births of mothers resident in Vieques (solid) and in the rest of PR (dashed line).

**Appendix Figure A.2: Year-Specific Effects – End of Bombing Activity and Health Outcomes Index z-Score, 1990-2003**



**Notes:** The figure plots the coefficients on Vieques-specific year effects ( $\theta_t$  from equation 2) for the health outcomes index z-score. The solid vertical line denotes the end of naval practices. The dashed vertical lines are the sampling distribution for the placebo estimates from the 5<sup>th</sup> to the 95<sup>th</sup> percentile for each year. The solid orange line represents (the average of the) average tonnage of ordnance used in practices during each child's 9-month potential gestation period, for newborns conceived in each calendar year.

**Appendix Table A1:** Alternative Inference Method - End of Bombing Activity and Health Outcomes at Birth, 1990-2003

	Dependent variables:					
	Congenital anomaly	Premature birth	Extremely preterm birth	Low birth weight	Low APGAR Score	Health Outcomes Index z-Score
	(1)	(2)	(3)	(4)	(5)	(6)
End of Naval Practices × Vieques	-0.0077*** (0.0024)	0.028 (0.020)	-0.0018 (0.0048)	-0.0020 (0.0183)	0.0062 (0.0054)	-0.018 (0.130)
Mother Characteristics	Yes	Yes	Yes	Yes	Yes	Yes
Municipality & Month-Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Outcome Mean	0.010	0.144	0.006	0.106	0.011	0.000
N	12,228	12,228	12,228	12,228	12,228	12,228

Notes: Coefficient estimates from municipality fixed effects regressions with individual level birth data collapsed into 159 month-municipality cells; each set of estimates (by column) is from a different regression. To conduct inference, we report robust standard errors to general forms of cross-sectional and temporal dependence (Driscoll and Kraay 1998) in parentheses; statistically significant at (\*) 10 percent, (\*\*) 5 percent, (\*\*\*) 1 percent levels, respectively. Mother characteristics' controls include (month-municipality cell averages of) indicators for high school graduate, higher education (some/completed), teenage pregnancy, mother's age at childbirth of 40 years or greater, and maternal age.