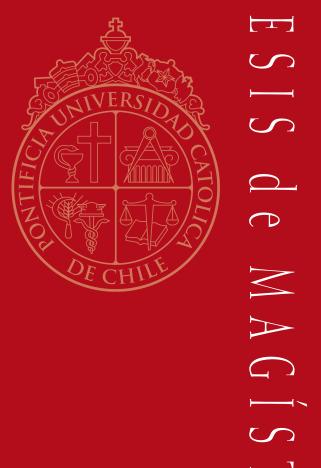
# INSTITUTO DE ECONOMÍA



2015

Toxic Disaster in Arica: Early Childhood Effects

Felipe Vial L.

# TESIS DE GRADO MAGISTER EN ECONOMIA

Vial, Lecaros, Felipe Andrés

Diciembre, 2015



**Toxic Disaster in Arica: Early Childhood Effects** 

Felipe Andrés Vial Lecaros

Comisión

Claudia Martínez Claudio Sapelli

Santiago, Diciembre de 2015

# Toxic Disaster in Arica: Early Childhood Effects

### Felipe Vial

Pontificia Universidad Católica de Chile

#### M.A. in Economics Thesis

December 2015

#### **Abstract**

In this paper I analyse the effect of lead contamination in Arica on the cognitive and non-cognitive skills of young children. Arica has received lead contamination from two main sources: toxic waste disposal and concentrated mineral storage related to the Arica-La Paz railroad. Using data from the ELPI 2010 and the ELPI 2012 and soil lead levels from a 2009 CONAMA study, I perform OLS regressions in both years to asses short and medium term effects. Results show negative effects on both cognitive and non-cognitive skills for children living closer to more contaminated areas.

<sup>\*</sup>Thesis written as a Masters student at Pontificia Universidad Católica de Chile, Vicuña Mackenna 4860, Macul. I would like to thank Tomás Rau for all the help and guidance during the semester, as well as Claudia Martínez and Claudio Sapelli for their comments and continuous feedback. I am also very grateful of all the discussions and conversations with Sebastián Montero, Matías Muñoz, Javier Martínez, family and friends; and the initial advances Sebastián Poblete made on this topic. Special thanks to Ximena Villarroel for proofreading with exceptional speed. Last but not least, I would like to thank the Ministerio del Medio Ambiente for the lead data, and the Comisión Nacional de Investigación Científica y Tecnológica for funding under the scolarship program of Beca Magíster Nacional 2015. Any comments or questions to this email address: fnvial@uc.cl.

## 1 Introduction

The scientific community has long been aware of the detrimental effects of lead on health, but more precise studies on the impact it has on early childhood are only recent. Identification threats such as endogeneity in exposure and the high costs of performing longitudinal studies are mainly responsible for this. Nonetheless, the Agency for Toxic Substances & Disease Registry (ATSDR) and a growing body of evidence<sup>1</sup> make it clear that there is a general consensus on the main effects lead has on children during early development stages (approximately ages 0-6). Effects on social behavior have been found (Reyes 2007), but most evidence focuses on the negative effects on cognitive skills (Currie et al., 2014).

Given that lead exposure can affect neurological functions, one next question logically comes to mind: what is the importance of cognitive and non-cognitive skills? The former have been more extensively studied; Murnane et al (1995) and Cawley et al (2001) find modest effects of cognitive skills in wage determination. Cawley et al find that one standard deviation increase in measured cognitive skills produced a \$0.74 hourly wage increase in a 1980 cohort. On the other hand, the relevance of non-cognitive skills is part of a more recent literature, but Heckman et al (2006), Heckman and Rubinstein (2001) and Borghans et al (2006) do find that non-cognitive abilities have an impact on wage determination, mainly through schooling decisions.

To analyse our results, a theoretical but more broad approach is useful. Cunha and Heckman (2007,2009) develop a model of the evolution of skills over the life cycle, where parental skills, investment and child skills in each period determine the child skills of the next period. This means that current stocks of skills help create future stocks of skills, making

<sup>&</sup>lt;sup>1</sup>The Yugoslavia Prospective Lead Study (Factor-Litvak et al. 1999) compared two nearby towns with different levels of lead exposure, and found prenatal and postnatal exposure to have significant negative effects on IQ levels. Schnaas et al (2000) report the Mexico City Prospective Lead study and find that prenatal blood lead levels have no impact in intellectual function, whereas postnatal blood lead levels were significantly correlated with it. Tong et al (2000), Emory et al (2003) and Canfield et al (2003) also perform longitudinal studies and find significant effects of lead on cognitive abilities in young children.

early childhood skills the foundation for building skills later in life. Their importance is accentuated by the following ideas established in recent literature:

1) Skills complement each other: skills in the next period don't only depend on the same kind of skills in the previous period, but of all skills in the previous period. This means that cognitive skills, non-cognitive skills and health complement each other to produce cognitive skills, non-cognitive skills and health in the next period<sup>2</sup>.

2) Skills complement investment: skills at early childhood facilitate the accumulation of skills later in life through dynamic complementarity, which is the higher productivity later investment has thanks to a higher skill stock<sup>3</sup>.

This framework helps understand the importance of early childhood skills, as well as the complementarity that cognitive and non-cognitive skills have.

In this paper I analyse the effects of lead exposure on cognitive and non-cognitive skills of children in Arica. For this I use soil lead levels, and cognitive and non-cognitive tests conducted by the Early Childhood Longitudinal Survey (Encuesta Longitudinal de Primera Infancia: ELPI). Lead contamination in Arica started in the 1980's and recent reports show that it still affects various sectors of the city. Rau et al. (2015) find that distance to one of the main lead exposure sources in Arica (Site F) has a significant impact in both SIMCE scores and PSU <sup>4</sup>.

My contribution is in three aspects. First of all, I contribute with quantitative evidence on the magnitude of the damage the people of Arica have suffered, therefore facilitating a more informed public policy response or monetary compensation. Secondly, the literature of lead poisoning in early childhood is not extensively developed, less so in the economic literature. In addition, if we were to combine the results found with an accumulative human

<sup>&</sup>lt;sup>2</sup>Heckman and Mosso (2014).

<sup>&</sup>lt;sup>3</sup>Cunha and Heckman (2008); Cunha et al. (2010)

<sup>&</sup>lt;sup>4</sup>The SIMCE tests are a set of standardized tests in math, language, history and science. They are taken in every school in Chile, every 2 years by students in 4th, 8th and 10th grade. The PSU is a college admission standardized test required by all higher education institutions in Chile. It consists of a math, language, science and history test, each one with it's own score from 0 to 850 points.

capital model like the one described, the consequences of lead poisoning could go far beyond a limited physiological damage. Finally, to the best of my knowledge, the effect of lead on non-cognitive skills has only been studied in the long term, through the proxy of social behavior such as crime, not in the short and medium term through a psychological test such as the CBCL. This contributes in identifying the channels through which lead affects educational and income outcomes.

The paper is organized as follows. In Section II I give details on the story of the different lead sources in Arica and the general scheme of lead poisoning. In Section III I describe the data used, in Section IV I present my empirical strategy, in Section V I present my results, in Section VI I perform robustness checks and Section VII presents the conclusion.

## 2 Lead in Arica

The lead contamination in Arica comes from two main sources: Site F and the Arica-La Paz railroad. In this section I present a chronological story of how these two sources came to be and the level of awareness and measures taken in each one. I also describe the process of lead intoxication and its biological consequences in section 2.3.

#### 2.1 Site F

The first known source of lead contamination in Arica is referred to as Site F. Between 1984 and 1985, Chilean company PROMEL imported approximately 20,000 tons of toxic waste from Sweden, depositing it in Site F, a government property on the northeast outskirts of Arica. The government authorized this importation and leased the property to PROMEL probably because of misinformation provided by PROMEL who qualified the waste as "nontoxic metallic mud unless eaten or handled without proper protection". The waste mainly contained lead, arsenic, zinc and copper. The initial intention was to extract gold from it,

but in 1993 PROMEL was given permission to abandon the waste since the business came to be considered non-viable. The city expanded rapidly in the 80's, and the Chilean government started building social housing around Site F, finishing the projects in 1989.

Due to residents' complaints, the first soil study was carried out in 1997 and found high levels of toxic waste in Site F. Relocation of the waste began in 1998, and blood tests were taken during 2000. This served as evidence for a collective legal action against PROMEL which ruled in 2007, compensating each claim with 8 million pesos. In 2009 the National Environmental Comission (Comisión Nacional de Medio Ambiente: CONAMA) performed a soil study (AGQ, 2009) and found that Site F had reduced its contamination radius, but still had high levels of lead in its soil. For a more detailed story of Site F, see Rau et al (2015).

#### 2.2 Ex-Maestranza and the harbor

The Arica-La Paz railroad, which exists since 1913, enters the city through the north. It finishes it's route at the harbor, ready to load and unload various shipments. For many decades it has been widely used as a means of transportation for lead exportation coming from Bolivia to the harbor. Despite this, it was only recently identified as a contamination source. Until 1993, an empty lot referred to as "Ex-Maestranza Chinchorros" served as storage place for lead concentrates. After this, the mineral was stored at the harbor itself. The storage of the lead had no wind protection whatsoever and, given the low rainfall in Arica, became a deadly combination for the spread of the lead particles. It was only around 1998 that these two sources of contamination (Ex-Maestranza and the harbor) were identified, mainly thanks to the work of Dr. Andrei Tchernitchin (Tchernitchin et al., 2006). Unfortunately, it has never gotten much public attention like Site F, and therefore few government measures have been taken.

#### 2.3 Lead Poisoning

Lead poisoning can occur through ingestation, inhalation, and through the skin. The two first are the most common and the ones taking place in Arica. Since mounds of concentrated minerals are left on the open, the wind carries lead particles and deposits them on the streets, roofs, playgrounds, etc. This can then be ingestated by children playing with dirt, or inhaled by any person who is on the outside and breathes in dust.

Lead then enters the bloodstream and fixes on different tissues, such as bones, kidneys and brain. Chronic exposure can cause cognitive and organ damage; acute exposure can cause convulsions, vomits, and even death (Hubbs-Tait et al. 2006). Lead is specially dangerous because since it fixes on bones, it can enter the bloodstream even after exposure to the exogenous source is over (ATSDR, 2015). Evidence suggests that the marginal effect of lead on cognitive skills is decreasing (Kordas et al. 2006) in the level of lead in the bloodstream.

Children are more likely to be poisoned due to the following reasons. They play with dirt, making ingestation of lead particles more likely. They are shorter, and therefore more likely to inhale dust with lead particles. Their stomachs absorb 5 to 10 times more lead than an adult's (ATSDR, 2015). In addition, because their brain is developing, the blood lead level (BLL) from which their neurological functions start being affected is much lower than for adults. Actually, there is no safe lower bound for children.

There is no specific minimum length of exposure that is defined to be hazardous; nevertheless, the 2 years of exposure we analyze in this paper is very likely to be enough exposure time to cause effects, when comparing with numerous lead pollution studies (Currie et al., 2014)<sup>5</sup>.

Given these precedents, I expect that children in Arica living closer to points of high lead concentration have a higher probability of suffering from lead intoxication. This should affect

<sup>&</sup>lt;sup>5</sup>More detailed information on the process of lead poisoning can be found in the Agency for Toxic Substances & Disease Registry's web page.

their cognitive and non-cognitive test scores, showing significant differences according to the level of lead in the soil close to their home. Since all the children I'll use in my sample were born before 2010, by 2012 the exposure length should definitely be enough for the lead to have an effect. This is not as clear with the 2010 test results, but it is possible that children already exhibit the effects of lead by then.

## 3 Data

### 3.1 Lead

CONAMA performed a soil study in Arica in 2009. In 440 points they collected 6 samples, 2 for each depth (surface, 15 cm and 30 cm) and calculated a mean lead concentration measured in mg/Kg. This provides us with 440 georeferenced points with their respective soil lead level (SLL)<sup>6</sup>.

Figure 1 shows the 440 georeferenced points, spread all over Arica. Arica was divided by a grid of 40 equal sections, and then specific points in each section were chosen randomly. This ensured the sampling of all of Arica. The mean lead level for the entire sample was 262 mg/Kg, ranging from 0.7 mg/Kg to 8169 mg/Kg. 38% of the sample presents a SLL over 140 mg/Kg, the maximum level considered safe by the Canadian Soil Quality Guidelines for the Protection of Environmental and Human Health. Figure 2 shows a density graph of the soil lead levels, with a line marking the 140 mg/Kg threshold.

In figure 3 we can see the map of Arica showing the spatial distribution of SLL using the 2009 CONAMA study. I performed an Inverse Distance Weighting interpolation<sup>7</sup> with the

<sup>&</sup>lt;sup>6</sup>For the same 440 points we have the measure of soil arsenic level, but to the best of my knowledge, there is no evidence of arsenic causing damage on cognitive or non-cognitive skills, in contrast to the extremely vast scientific literature on lead's effects. Therefore, the arsenic level will not be used.

<sup>&</sup>lt;sup>7</sup>Inverse Distance Weighting uses all the 440 CONAMA sample values to calculate an SLL value for each point of the map that doesn't have an SLL value already. To assign the SLL value for each point, it uses the following formula:  $SLL(x) = \{\frac{\sum_{i=1}^{440} w_i(x)SLL_i}{\sum_{i=1}^{440} w_i(x)}\}$  where  $w_i(x) = \frac{1}{d(x,x_i)}$  and  $d(x,x_i)$  is the distance between x (the point being interpolated) and  $x_i$  (each of the 440 points).

440 points of SLL to create a visual image of the level of contamination in the different areas of the city. We can clearly identify two sources of high SLL in the west, and one smaller zone on the east. The more northern source in the west is the Ex-Maestranza, and the more southern one is the harbor. The eastern source is Site F, clearly still present but quite small in size compared to the two western sources, mainly because of the government measures taken to remove the waste. This figure also shows the Arica-La Paz Railroad. We can notice higher lead levels around the rail, especially at the Ex-Maestranza and harbor area.

#### 3.2 ELPI

In 2010 the Chilean government performed the ELPI for the first time. It consists of a set of cognitive and non-cognitive tests conducted on one child per sampled household between the ages of 0 and 6. Approximately 15.000 infants were sampled in this first year. Analogous tests are conducted on the children's mother, and a wide array of household characteristics are obtained through a questionnaire. The sampling was done by the Microdata Center (Centro de Microdatos) of Universidad de Chile; using a two stage stratification procedure by comuna and age/sex. This ensures a representative sample of Chilean households. Naturally, I use only the households located in Arica. We can see a set of descriptive statistics of our sample in Table 1, such as income<sup>8</sup>, mother's education and the mother IQ tests; the WAIS number test and the WAIS vocabulary test<sup>9</sup>. Our sample presents an average of these two IQ tests of approximately 8, which is within the normal range. Household income and mother's

<sup>&</sup>lt;sup>8</sup>Exact income was not reported by all households, but those who didn't report, did report an income bracket (there are 10 income brackets, with approximately \$200.000 in each). This allows us to save some observations by imputing the mean of the bracket as the income for those households. When I only use the observations with exact income, the estimated lead coefficients barely change, but standard errors do, which suggests my estimates are not being biased by this imputation, just given more power. Nevertheless, I include an imputation dummy in all my estimations.

<sup>&</sup>lt;sup>9</sup>The mother of each child is evaluated with a cognitive test, the Weschler Adults Intelligence Scale (WAIS); which I use as a control variable. The WAIS is a test constructed to measure global intelligence, understood as the concept of IQ (intelligence quotient). It is subdivided into different scales to measure diverse skills. The ELPI only applied the vocabulary and digit retention subtests<sup>10</sup>. It is one of the most utilized intelligence tests, especially to clinically evaluate adults (Kaplan & Saccuzzo 2005).

education don't seem to correspond to either a particularly low or high income sample.

This survey was repeated in 2012, to 185 of the initial 233 households surveyed in 2010, and to 47 new households (this numbers only refer to Arica; the country-wide sample increased by approximately 3.000 children). I performed t-tests on income and mother's education to look for some common characteristic in the attrition, but the results showed no significant differences between the 2010 households that are followed in the 2012 sample and those that are not. These households are georeferenced as well, which provides us with a total of 280 georeferenced households, as we can see in figure 4. However, the evaluations applied on each child depend on their age and authorization of the mother, which is why the sample size for each estimation will be detailed for each test. The three tests I will use as outcomes will be the Batelle test, the Test de Aprendizaje de Desarrollo Infantil (TADI) and the Child Behavior Checklist (CBCL). The Batelle test and the CBCL are very popular international clinical measurement tools for youth (Martin & Volkmar 2007). The TADI is a Chilean instrument, created by the Centro de Investigación Avanzada en Educación (CIAE) and the Centro de Estudios para el Desarrollo y Estimulación Psicosocial (CEDEP). It takes into account the chilean cultural context and its development had the support of UNICEF, so despite being relatively new (2013), it should be useful and reliable.

The Batelle test is a measure of psychomotor development, and is divided into 5 categories: Social, Adaptive, Motor, Communication and Cognitive. Children from 6 months to 84 months are evaluated with this test. Each category receives a score, and a total score is also constructed; a higher Batelle score indicates a greater degree of development. All scores are standardized, providing t-scores. In 2010 the complete instrument was used, and therefore we have a t-score for each area as well as a total t-score. In 2012 only a screening version was used, so we don't have scores for each area, only a total t-score. These scores are then classified into "strong", "normal" and "weak". Table 2 show the values used to classify the t-scores and the percentage of the sample in each category; it is important to notice that

the complete test taken in 2010 differs in this classification from the screening test taken in 2012. The tables indicate that most children are classified as normal(71% in 2010 and 60% in 2012), followed far behind by weak (21% in 2010 and 27% in 2012) and last of all strong (8% in 2010 and 13% in 2012). There are 52 children that were evaluated with the Batelle test in both 2010 and 2012. We will use them as our sample to be able to compare the short and medium term effects of lead.

The TADI is an instrument designed to measure psychomotor skills in children from ages 6 months to 84 months. It is divided into cognitive, motor, language and socio-emotional skills. A t-score is obtained for each category as well as a total one. Higher t-scores mean a greater development; t-scores are then classified in "advanced", "normal", "at risk" and "low". Table 3 shows the values of classification and how the sample is distributed. The TADI was only used in 2012, and 189 children were evaluated (a few children were not evaluated in all 4 areas). Again we see that most children are classified as normal.

The CBCL measures non-cognitive skills, consisting of a socioemotional evaluation that is divided into two main areas: Internal (Emotional Reactivity, Anxiety, Somatic Complaints, Self Absortion) and External (Attention Problems, Agressive Conducts). A 7th category is Sleeping Problems, which is not included into any of the two main areas, but is included to calculate a total CBCL score. A higher CBCL score means a higher clinical risk. Children from 18 months to 72 months are evaluated with the CBCL. The scores are classified into "normal", "at risk" and "clinical range". Table 4 shows the classification and sample distribution. 85 children were evaluated with the CBCL both in 2010 and 2012; we will be using them as our sample. In this test we see a more even distribution of the sample between the normal classification and in clinical range (47% normal and 38% in clinical range for 2010; 45% normal and 42% for 2012).

# 4 Empirical Analysis

To measure the effect of lead exposure on the different tests, I first have to assign a level of SLL to each household. To do this I use Nearest Neighbor Matching. This means that each household will be assigned the SLL of the closest CONAMA soil sample<sup>11</sup>. As discussed before, lead intake in Arica mainly occurs through ingestation and inhalation. High levels of lead in the soil near the residence of the infant make it more likely for the child to play with dirt that is contaminated, or breath in contaminated dust. Thus, assigning each household the nearest SLL point should provide a good measure of exposure risk. This procedure will then provide us with the variable  $Pb_i$ , which is the SLL for household i.

Despite having panel data, I only have one measure of SLL in time, and thus can't perform panel regressions with fixed effects since the SLL variable would be linearly dependent with the fixed effects dummy<sup>12</sup>. And since contamination occurred before 2010, the test results of that year are not a baseline measure, and therefore I can't control for it. Because of these reasons, I will use OLS reggresions for 2010 and 2012 separately.

This way, our main equation will be:

$$y_i = \alpha + \gamma_2 P b_i + \gamma_3 P b_i^2 + \gamma_4 Z_i + \epsilon_i \tag{1}$$

Where  $y_i$  is the test result of the child in household i,  $Pb_i$  is the SLL assigned to household i, and  $Z_i$  is a series of controls containing the children's age (and age<sup>2</sup>), mother's education, mother IQ tests and household income.  $Pb_i^2$  is used since many studies have found the effect of lead poisoning to be non-linear in the level of lead (Rau et al. 2015, Kordas et al. 2006).

 $<sup>^{11}</sup>$ The average matching distance is 141 mts, and the median matching distance is 116 mts. Over 80% of the matches are under 200 mts, and over 97% of the matches are under 400 mts. Table 5 checks that there is no correlation between the matching distance and income.

<sup>&</sup>lt;sup>12</sup>Random effects could be used, but it is not convincing that the unobserved characteristics of the child are uncorrelated with the observable variables such as mother's education and income. Pooled OLS would be useful to gain sample size, but it is more interesting to see each year separately to compare the difference in short and medium term effects.

# 5 Addressing Endogeneity

Endogeneity is a problem present in all lead exposure studies, since it is plausible to believe that in knowledge of lead contamination, families with the necessary means will migrate away from contaminated areas. This would lead to confounding between the effect of lead and the effect of socio-economic related variables. There is no documentation of migration within Arica, and Soto and Torche (2004) document limited regional migration. Adding to this, the two largest sources of lead contamination in the 2009 CONAMA report are the harbor and the Ex-Maestranza, which did not receive much public attention like the F site, which now has a contamination radius very reduced in size in comparison to what it used to be and to these two other sources (figure 3 illustrates this well). This means that it is not likely that many people have moved away from these two sources of contamination. Besides this, I also control for many relevant variables such as income, mother's education, and mother IQ tests.

Analysing our sample more specifically, I can test if there is any significant correlation between the SLL matched with each home and socioeconomic variables. If the poorer households are left living near the contaminated areas, we should find a negative correlation between SLL and income, mother education, or some other socioeconomic variable. In table 6, I regress the lead level assigned to each household with various characteristics such as income, mother's education, and mother's IQ tests. We can see in Table 6 that none of the variables are significantly correlated with the lead level, and some signs are not even as expected. This suggests that at least in our sample, it is not the case that poorer or less educated households live in places with higher levels of lead concentration.

Figure 5 is a more visual representation of what is shown in Table 6. Households are divided in groups by Soil Lead Level, and the mean household income and mean mother's education (measured in years) of each group is calculated. We can see no downward trend in any of the variables.

I also perform a placebo test on the Mother IQ tests: the WAIS vocabulary test and the

WAIS number test. This tests were taken in the 2010 round. Table 7 shows the results of OLS regressions of Mother IQ tests on the same variables of our main model. Lead coefficients are not significant at all, showing no sign of endogeneity in the mother's cognitive development.

Figure 6 shows the spatial distribution of income in Arica, using the income data from the ELPI. Just as in figure 3, I used an IDW interpolation to create the figure. Figure 7 does the job of comparing figure 3 & 6. In this figure we can first of all notice that Arica does not appear to be highly stratified by income, and second, that there are households of all incomes in the different lead-intensity zones.

To look for evidence of migration movements or brain-drain from Arica as a city, I compare a set of variables with the average of 7 other northern cities: Iquique, Antofagasta, Calama, Copiapo, La Serena, Coquimbo and Ovalle. Figures 8 & 9 shows the evolution of population, and figures 10 & 11 show the share of university graduates and high school graduates. There is no clear trend of any kind of migration.

# 6 Results

I present the results for the Batelle test on tables 8 & 9. Table 8 has the results for the different Batelle tests for 2010. Since we have Pb and Pb<sup>2</sup>, we have to look at the joint significance of the parameters. The p-value for the F-Test can be seen in the last row of each table. The p-values reveal a significant negative effect of lead on every test except for the communications skills test. The coefficient of Pb<sup>2</sup> is positive in all cases, which coincides with the existing evidence of a non-linear effect in the level of lead. The magnitude of the effect is quite large. Using the results of the Batelle Cognitive test, a child living in an area with the average lead level for this sample (122 mg/Kg) presents almost 20 points less than a child living in a lead-free zone. This is 1.58 standard deviations<sup>13</sup> and is quite a large effect

<sup>&</sup>lt;sup>13</sup>All test standard deviations are calculated from same sample used in the estimations.

considering that the "normal" classification of this test has a range of 30 points. Figure 12 is a graph showing the marginal effect of lead on the 2010 Batelle Cognitive test. The diminishing marginal effect reveals the non-linear effect we expected to see.

Table 9 shows the results for the Batelle screening test in 2012. The screening test is an abbreviated version of the complete evaluation, used to asses if further evaluation is needed. The p-values on the last row also show a significant negative effect of lead on test scores. Unfortunately the screening test doesn't deliver an individual score for each area, which makes it hard to compare with the 2010 results<sup>14</sup>. In this case, a child living in an averagely polluted area scores almost 10 points less than a child living in a lead-free zone, which is 1.94 standard deviations. This is a bigger magnitude than in 2010, but it is not directly comparable since the 2010 complete test and the 2012 screening test use different classification scales.

Table 10 presents the results for the 2012 TADI test. Again in the last row of the table we find a significant negative effect of lead on all test scores, including the total t-score. In this test the effect is smaller than in the Batelle test. Using the SLL of 122 mg/Kg as a comparison point, the effect is of 2.7 points less than a child in a lead-free zone, which is only 0.35 standard deviations and clearly modest when looking at the scale.

Tables 11 & 12 show the results for the CBCL. Once again we must look at the p-values in the last row to evaluate the significance of lead on the test scores. The 2010 results in table 11 show only a significant effect in the internal area, raising the clinical risk. Using the same 122 mg/Kg, it increases the internal CBCL score by 5.6 points, which is 0.69 standard deviations. In table 12, I show results for 2012. Here we can observe significant lead coefficients in external, internal and total score. The coefficients are also quite higher. A child in a 122 mg/Kg SLL area obtains approximately 10 points more than one in a lead-

<sup>&</sup>lt;sup>14</sup>The screening version is used to diagnose if a child needs further evaluation, and thus is composed of fewer questions. Individual scores for each area can't be obtained, and to the best of my knowledge a screening score can't be obtained from the complete 2010 test.

free zone in both the internal and total score, which means 1.1 and 0.9 standard deviations respectively. In figure 13 we can see the marginal effect of lead on the 2012 CBCL Internal t-score, also showing a clear non-linear effect.

Because the CBCL is the exact same measure in 2010 and 2012, we can compare the results more easily, and draw some conclusions about the short and medium-term effects of lead. We see more significant and larger coefficients in 2012, suggesting that the effect of lead in non-cognitive skills could be larger in the medium-term than in the short-term; I test this in section VII. The psychomotor tests (Batelle and TADI) are harder to compare because it is not exactly the same instrument from sample to sample and even the classification scale is different.

The results found are very troubling. By 2012 we can see a deterioration in both cognitive and non-cognitive skills of at least 1 standard deviation (almost 2 in the case of the Batelle screening test) for children living in places with an average level of lead contamination. Using the model presented in Section I as our framework, this results imply serious consequences. Lower early childhood skills implies an initial setback that each period is harder to make up for. This is accentuated by the fact that cognitive and non-cognitive skills are complementary and both have been affected; and that less early childhood skilss means lower productivity of later investments, like school. Put in a practical example, a child with lower cognitive and non-cognitive skills is more likely to perform poorly in school, which will lead him to have less chance of finishing high school or pursuing higher education. Ultimately in the labor market, not only will the individual have lower skills, but will also be less qualified. This means the harm caused by lead is much more profound than the immediate physiological and neurological damage, and can have serious consequences later in life.

## 7 Robustness Checks

In table 13, I present the main results found in section 6 including an interaction of lead with age. The expected direction of the coefficient is not clear. Older children have been exposed longer, but younger children are in a more fundamental stage of brain development, so it is not clear which effect will predominate. We find only a significant coefficient of the interaction in the 2010 CBCL Internal test. The direction of the coefficient indicates that younger children are more affected by lead. The rest of the regressions don't exhibit any evidence of this, probably because of the neutralization of the two effects just described.

Another heterogeneity of interest has to do with site F. Site F is by far the contamination point with the most media attention. This could raise the concern that site F does present endogeneity in its nearby inhabitants; in other words, because of its media attention, people with enough economic resources have migrated away from site F. This endogeneity in Site F could potentially be responsible for the effects we are observing. To check for this, I constructed a variable with the distance from each home to site F, and Table 14 shows the main regressions of section 6, including the interaction between lead and distance to site F. The interaction term is not statistically significant in any of the regressions, and actually all except the coefficient on column (4) go in the same direction as the lead coefficient. The direction of this coefficient would suggest that children living in high SLL areas, but further away from site F, have worse scores. This goes against the endogeneity problem we discussed previously. In the last row we have the p-values for the F-Test on Pb, Pb<sup>2</sup> and the interaction term. Most are still significant, specially the 2012 test scores.

As another robustness check, I use the safety threshold of the Canadian Soil Quality Guidelines for the Protection of Environmental and Human Health (140 mg/Kg SLL) to create a "High Pb" dummy variable. Table 15 shows our main results found in Section 6 using this dummy. We find significant coefficients of the High Pb dummy only in the 2012 regressions, which goes in line with a larger medium-term effect.

Finally, I perform a pooled OLS regression for the CBCL test. This increases the number of observations, which are very limited for all of my estimations. The results are presented in table 16. We find in columns 1-3 significant lead effects on Internal and External t-scores; being Internal the most affected area by far (0.8 standard deviations if we use the 122 mg/Kg).

Column 4 of table 16 checks if the effect of lead on the 2010 CBCL internal t-score differs statistically from the effect in 2012. This is done by adding to the pooled regression a dummy assigned to the observations from the 2012 sample. I then see if the interactions of this dummy with the lead variables are jointly significant. We can see by the F-Test in the last row that they are not; meaning the lead coefficients from 2010 are not statistically different from the ones in 2012. This could be due to sample size.

## 8 Conclusions

In this paper I estimate the effects of early childhood exposure to lead in cognitive and non-cognitive skills. I use data from a soil sample conducted in 2009 in Arica and different cognitive and non-cognitive tests taken by the government through the ELPI (Encuesta Longitudinal de Primera Infancia) in 2010 and 2012. Arica has been contaminated with lead for many years, but drastic measures have not been taken and people continue to live in highly contaminated areas.

A recurring problem in lead studies is endogeneity of exposure. Families with more economic resources have the means to migrate away from contamination, leaving families of low socio-economic status living near the lead. This could lead the results to confound the effect of exposure with the attributes of the families. I find no correlation between soil lead level and any socio-economic variable. I also perform a placebo test with test scores from an IQ test taken by the mother of the child, finding no statistically significant coefficients. I support this with graphical analysis as well.

My results indicate that cognitive skills are affected quickly; the 2010 cognitive and psychomotor tests already presenting very large losses in children living in contaminated areas (1.58 standard deviations at the average SLL). The 2012 tests are not directly comparable to the ones taken in 2010 but a negative effect of lead is still found, and the magnitude appears to be equal or larger (1.94 standard deviations at the average SLL). Non-cognitive skills seem to take longer to be affected by lead, since 2010 socio-emotional tests are affected by soil lead level but in 2012 the effects are clear and larger. Nevertheless, there is no statistically significant difference between the short and medium-term effects. Despite the fact that the evidence is not conclusive, the main results combined with what was seen in the robustness checks lead to think that medium-term effects are larger than short-term effects.

These results are in line with the existing evidence of early childhood exposure to lead. They are very alarming, since early stages of development are fundamental in the accumulation of human capital. The Batelle test has been found to effectively predict academic performance in first grade (Guidubaldi and Perry, 1984). This means it is not only directly through less cognitive and non-cognitive skills that a lower income can be expected in the future, but also through a lower educational attainment. The evidence found in this paper can be helpful for more informed policy decisions not only in Arica but in other places suffering similar situations, including preventive measures as well as compensatory ones.

## References

AGQ (Agriquem America S.A.). 2009. Análisis Químico de suelos de la ciudad de Arica. Sistema Nacional de Información Ambiental, Santiago.

ATSRD. Agency for Toxic Substances & Disease Registry. http://www.atsdr.cdc.gov/

Borghans L, Duckworth A, Heckman J, Weel B. 2006. The Economics and Psychology of Personality Traits. *Journal of Human Resources* Vol. 43, 4.

Bustos C, Herrera M, Mathiesen M. 2001. Calidad del ambiente del hogar: inventario HOME como un instrumento de medición. *Estudios Pedagógicos*, 27:7-22.

Canfield RL, Henderson CRJ, Cory-Slechta DA, Cox C, Jusko TA, Lanphear BP. 2003. Intellectual impairment in children with blood lead concentrations below 10 mg per deciliter. N Engl J Med 348:1517–1526.

Cawley J, Heckman J, Vytlacil E. 2001. Three observations on wages and measured cognitive ability. *Labour Economics* 8: 419-442.

Cook D. 1979. Influential observations in linear regression. *Journal of the American Statistical Association* 74(365): 169-174.

Cunha F, and Heckman J. 2007. The technology of skill formation. *American Economic Review* 97 (2), 31–47.

Cunha F, and Heckman J. 2008. Formulating, identifying and estimating the technology of cognitive and noncognitive skill formation. *Journal of Human Resources* 43 (4), 738–782.

Cunha F, and Heckman J. 2009. The economics and psychology of inequality and human development. *Journal of the European Economic Association* 7 (2–3), 320–364.

Cunha F, Heckman J, and Schennach S. 2010. Estimating the technology of cognitive and noncognitive skill formation. *Econometrica* 78 (3), 883–931.

Currie J, Graff Zivin J, Mullins J, Neidell M. 2014. What Do We Know About Short and Long Term Effects of Early Life Exposure to Pollution? *Annual Review of Resource Economics*, 6: 217-247.

Emory E, Ansari Z, Pattillo R, Archibold E, Chevalier J. 2003. Maternal blood lead effects on infant intelligence at age 7 months. *Am J Obstet Gynecol* 188:S26–S32.

Factor-Litvak P, Wasserman G, Kline JK, Graziano J. 1999. The Yugoslavia Prospective Study of environmental lead exposure. *Environmental Health Perspectives* 107:9–15.

Guidubaldi J, Perry JD. 1984 Concurrent and predictive validity of the Battelle Development Inventory at the first grade level. *Educational and Psychological Measurement*. 44: 977–985.

Heckman J, Rubinstein Y. 2001. The Importance of Noncognitive Skills: Lessons from the GED Testing Program. *The American Economic Review* Vol. 91, No. 2 145-149.

Heckman J, and Mosso S. 2014. The economics of human development and social mobility. *Annual Review of Economics* 6 (1), 689–733.

Hubbs-Tait L, Nation J, Krebs N, Bellinger D. 2006. Neurotoxicants, Micronutrients, and Social Environments: Individual and Combined Effects on Children's Development. *Psychological Science in the Public Interest.* 6, 57-121.

Heckman J, Stixrud J, Urzua S. 2006. The Effects of Cognitive and Noncognitive Abilities on Labor Market Outcomes and Social Behavior. *Journal of Labor Economics* 2006, v24(3,Jul), 411-482.

Kaplan R, Saccuzzo D. 2005. Psychological Testing: Principles, Applications, and Issues.

Kordas K, Canfield R, Lopez P, Rosado J, García G. 2006. Deficits in cognitive function and achievement in Mexican first-graders with low blood lead level concentrations. *Environmental Research*. 100, 371-386.

Martin A, Volkmar F. 2007. Lewi's Child and Adolescent Psychiatry: A Comprehensive Textbook.

Murnane R, Willett J, Levy F. 1995. The growing importance of cognitive development in wage determination. *The Review of Economics and Statistics* Vol. 77, No 2: 251-266.

Rau T, Reyes L, Urzúa S. 2015. Early Exposure to Hazardous Waste and Academic Achievement: Evidence from a Case of Environmental Negligence. *Journal of the Association of Environmental and Resource Economists* Vol 4, No 2: 527-563.

Reyes JW. 2007. Environmental policy as social policy? The impact of childhood lead exposure on crime. B.E. J. of Econ. Anal. and Pol. 7(1) (Contributions), Article 51

Schnaas L, Rothenberg SJ, Perroni E, Martinez S, Hernández C, Hernández RM. 2000. Temporal pattern in the effect of postnatal blood lead level on intellectual development of young children. *Neurotoxicol Teratol* 22:805–810.

Tchernitchin A, Nina L, Molina L, Molina G, Tchernitchin N, Acevedo C, and Alonso P. 2006. Human Exposure to Lead in Chile. *Reviews of Environmental Contamination and Toxicology*, 185(6): 93-139.

Tong S, McMichael AJ, Baghurst PA. 2000. Interactions between environmental lead exposure and sociodemographic factors on cognitive development. *Arch Environ Health* 55:330–335.

Table 1: Descriptive Statistics of the ELPI 2010 & 2012 Arica sample

Variable	Mean	Standard Deviation	N
Income	\$498,963	$420,\!897$	276
Mother's Education	12.9	2.38	280
WAIS numbers test	7.94	2.65	211
WAIS vocabulary test	8.57	3.23	211

Source: ELPI 2010 & 2012.

Table 2: Batelle 2010 & 2012 t-scores Classification

Interpretation	2010 t-score	% of 2010 Sample N=52	2012 t-score	% of 2012 Sample N=52
Strong	>64	7.7%	>59	13.46%
Normal	36-64	71%	51-59	59.6%
Weak	<36	21.3%	<51	26.94%

Source: ELPI user manual, Batelle 2010 total t-scores and Batelle 2012 screening test t-scores.

Table 3: TADI 2012 t-scores Classification

Interpretation	t-score	Percentage of Sample N=189
Advanced	>59	11.65%
Normal	40-59	78.45%
At Risk	30-39	9.9%
Low	<30	0%

Source: ELPI user manual and TADI 2012 total t-scores.

Table 4: CBCL 2010 & 2012 t-scores Classification

Interpretation	t-score	% of sample 2010 N=85	% of sample 2012 N=85
Normal	<60	47%	44.7%
At Risk	60-63	15.29%	12.9%
Clinical Range	>63	37.71%	42.4%

Source: ELPI user manual, CBCL 2010 total t-scores and CBCL 2012 total t-scores.

Table 5: OLS Regression of Matching distance on Income

	(1)
	Matching Distance
Income	- 1.72e-08
	(1.64e-08)
$\overline{N}$	275
$R^2$	0.0004

Standard errors in parentheses

Notes: Matching Distance is the geographic distance between the household and the lead soil sample assigned to it through nearest neighbor matching. Income is household income. One outlier was removed from the sample as it presented a cook's distance of over 4 (Cook, 1979).

<sup>\*</sup> p < 0.1, \*\* p < 0.05, \*\*\* p < 0.01

Table 6: OLS Regressions Lead on Socioeconomic Characteristics

	(1)	(2)	(2)	(4)	<b>(F)</b>	(C)
	(1)	(2)	(3)	(4)	(5)	(6)
	Pb	Pb	Pb	Pb	Pb	Pb
Income	0.00000983			0.00000751		0.0000310
	(0.0000274)			(0.0000527)		(0.0000424)
Income Bracket		1.967		0.618		-8.162
		(5.913)		(11.67)		(9.798)
Mother's Education			-2.751	-0.0592		1.292
			(4.967)	(5.505)		(5.263)
Wais Numbers					0.0912	-2.250
					(3.660)	(4.342)
Wais Vocabulary					4.642	5.760
, rais rocas araij					(3.004)	(3.768)
N	232	232	280	232	211	177
$R^2$	0.001	0.000	0.001	0.001	0.013	0.041

Notes: Pb is the Soil Lead Level assigned to each home using a geographic Nearest Neighbor Matching. Mother's Education is measured in years, and the Wais tests are cognitive tests similar to IQ tests. Homes from both 2010 and 2012 are used.

<sup>\*</sup> p < 0.1, \*\* p < 0.05, \*\*\* p < 0.01

Table 7: Mother IQ Placebo Test

	(1)	(2)	(3)	(4)
Variables	Vocabulary	Vocabulary	Numbers	Numbers
Pb	0.00248	0.00662	0.00151	0.00114
	(0.0164)	(0.0142)	(0.0107)	(0.0104)
$Pb^2$	-6.72e-06	-1.79e-05	-1.09e-05	-9.60e-06
	(4.50e-05)	(3.71e-05)	(2.58e-05)	(2.55e-05)
SE Variables	NO	YES	NO	YES
Imputation Dummy	YES	YES	YES	YES
Observations	121	119	121	119
R-squared	0.000	0.165	0.029	0.047
Pb & Pb <sup>2</sup> F-Test p-value	0.9885	0.8902	0.6269	0.6955

Notes: Columns 1 & 2 correspond to different specifications of regressions on the WAIS vocabulary test, and columns 3 & 4 to the WAIS numbers test. In this table I regress the mother's IQ tests on the same variables used for the main estimations. Pb is the Soil Lead Level assigned to each home using a geographic Nearest Neighbor Matching. Socioeconomic Variables include age of the child, age<sup>2</sup>, household income and mother's education (measured in years). Imputation Dummy is a dummy assigned to those observations where income was estimated using a reported income bracket. The last row presents the F-Test p-value of the null hypothesis:  $\beta[Pb]=0$  &  $\beta[Pb^2]=0$ .

<sup>\*</sup> p < 0.1, \*\* p < 0.05, \*\*\* p < 0.01

Table 8: 2010 Batelle OLS Regressions

	(1)	(2)	(3)	(4)	(5)	(6)
Variables	Adaptative	Motor	Communication	Cognitive	Personal	Total
Pb	-0.155	-0.103	-0.162	-0.259**	-0.263***	-0.207**
	(0.124)	(0.106)	(0.101)	(0.115)	(0.0940)	(0.0986)
$\mathrm{Pb^2}$	0.000654*	0.000398	0.000563*	0.000789**	0.000775***	0.000687**
	(0.000344)	(0.000297)	(0.000300)	(0.000345)	(0.000272)	(0.000272)
SE Variables	YES	YES	YES	YES	YES	YES
Mother IQ tests	YES	YES	YES	YES	YES	YES
Imputation Dummy	YES	YES	YES	YES	YES	YES
Observations	52	52	52	52	52	52
R-squared	0.263	0.469	0.282	0.266	0.542	0.443
Pb & Pb <sup>2</sup> F-Test p-value	0.0005***	0.033**	0.1181	0.0850*	0.0243**	0.003***

Notes: A higher Batelle score means greater development. Columns 1-5 present the different subsections of the 2010 Batelle test. Column 6 shows the total t-score. Pb is the Soil Lead Level assigned to each home using a geographic Nearest Neighbor Matching. Socioeconomic Variables include age of the child, age<sup>2</sup>, household income and mother's education (measured in years). The mother IQ tests are the Wais Numbers test and the Wais Vocabulary test. Imputation Dummy is a dummy assigned to those observations where income was estimated using a reported income bracket. The last row presents the F-Test p-value of the null hypothesis:  $\beta[Pb]=0 \& \beta[Pb^2]=0$ . Robust standard errors are used.

<sup>\*</sup> p < 0.1, \*\* p < 0.05, \*\*\* p < 0.01

Table 9: 2012 Batelle Screening OLS Regressions

	(1)	(2)
Variables	Batelle Screening	Batelle Screening
Pb	-0.127***	-0.121**
	(0.0388)	(0.0534)
$\mathrm{Pb}^2$	0.000339***	0.000334*
	(0.000114)	(0.000172)
SE Variables	NO	YES
Mother IQ tests	NO	YES
Imputation Dummy	YES	YES
Observations	52	52
R-squared	0.189	0.423
Pb & Pb <sup>2</sup> F-Test p-value	0.0043***	0.0195**

Notes: A higher Batelle score means greater development. Columns 1 & 2 show regressions on the 2012 Batelle Screening test. Pb is the Soil Lead Level assigned to each home using a geographic Nearest Neighbor Matching. Socioeconomic Variables include age of the child, age², household income and mother's education (measured in years). The mother IQ tests are the Wais Numbers test and the Wais Vocabulary test. Imputation Dummy is a dummy assigned to those observations where income was estimated using a reported income bracket. We don't report the individual score of each category (adaptative, motor, etc) like in 2010 because in 2012 only a screening version was taken, which does not allow to obtain a score for each area, only a total score. The last row presents the F-Test p-value of the null hypothesis:  $\beta[Pb]=0$  &  $\beta[Pb^2]=0$ . Robust standard errors are used.

<sup>\*</sup> p < 0.1, \*\* p < 0.05, \*\*\* p < 0.01

Table 10: 2012 TADI OLS Regressions

	(1)	(2)	(3)	(4)	(5)
Variables	Socio-emotional	Motor	Language	Cognitive	Total
Pb	-0.0425*** (0.0152)	-0.00545 (0.0186)	-0.0277* (0.0158)	-0.0208 (0.0163)	-0.0244** (0.0123)
$\mathrm{Pb}^2$	3.27e-05*** (1.02e-05)	7.35e-06 (1.22e-05)	1.68e-05 (1.04e-05)	8.46e-06 (1.07e-05)	1.65e-05** (8.09e-06)
SE Variables	YES	YES	YES	YES	YES
Imputation Dummy	YES	YES	YES	YES	YES
Observations	179	186	186	187	179
R-squared	0.257	0.046	0.134	0.182	0.201
Pb & Pb <sup>2</sup> F-Test p-value	0.0021***	0.0638*	0.0733*	0.000***	0.0968*

Notes: A higher TADI score means greater development. Columns 1-4 present the different subsections of the 2012 TADI test. Column 5 shows the total t-score.Pb is the Soil Lead Level assigned to each home using a geographic Nearest Neighbor Matching. Socioeconomic Variables include age of the child, age<sup>2</sup>, household income and mother's education (measured in years). The mother IQ tests are the Wais Numbers test and the Wais Vocabulary test. Imputation Dummy is a dummy assigned to those observations where income was estimated using a reported income bracket. The last row presents the F-Test p-value of the null hypothesis:  $\beta[Pb]=0$  &  $\beta[Pb^2]=0$ . Robust standard errors are used.

<sup>\*</sup> p < 0.1, \*\* p < 0.05, \*\*\* p < 0.01

Table 11: 2010 CBCL OLS Regressions

	(1)	(2)	(3)	(4)	(5)
Variables	Total	Total	Total	Internal	External
Pb	0.0299	0.0355	0.0336	0.0700**	0.00196
	(0.0411)	(0.0359)	(0.0375)	(0.0316)	(0.0444)
$\mathrm{Pb}^2$	-6.30e-05	-6.98e-05	-6.39e-05	-0.000193**	3.90e-05
	(0.000108)	(8.33e-05)	(8.78e-05)	(7.61e-05)	(0.000101)
SE Variables	NO	YES	YES	YES	YES
Mother IQ tests	NO	NO	YES	YES	YES
Imputation Dummy	YES	YES	YES	YES	YES
Observations	85	83	83	83	83
R-squared	0.008	0.088	0.089	0.108	0.106
Pb & Pb <sup>2</sup> F-Test p-value	0.7196	0.6	0.6379	0.0423**	0.3511

Notes: A higher CBCL score means higher clinical risk. Columns 1-3 present regressions of the total 2010 CBCL total t-score, with different specifications. Columns 4 & 5 present the results for the internal and external subcategories respectively. Pb is the Soil Lead Level assigned to each home using a geographic Nearest Neighbor Matching. Socioeconomic Variables include age of the child, age<sup>2</sup>, household income and mother's education (measured in years). The mother IQ tests are the Wais Numbers test and the Wais Vocabulary test. Imputation Dummy is a dummy assigned to those observations where income was estimated using a reported income bracket. The last row presents the F-Test p-value of the null hypothesis:  $\beta[Pb]=0$  &  $\beta[Pb^2]=0$ . Robust standard errors are used.

<sup>\*</sup> p < 0.1, \*\* p < 0.05, \*\*\* p < 0.01

Table 12: 2012 CBCL OLS Regressions

	(1)	(2)	(3)	(4)	(5)
Variables	Total	Total	Total	Internal	External
Pb	0.0779* (0.0459)	0.0890* (0.0449)	0.0983** (0.0454)	0.117*** (0.0375)	0.0368 $(0.0582)$
$\mathrm{Pb}^2$	-0.000140 (0.000114)	-0.000155 (9.91e-05)	-0.000185* (0.000102)	-0.000256*** (8.57e-05)	-8.97e-06 (0.000132)
SE Variables	NO	YES	YES	YES	YES
Mother IQ tests	NO	NO	YES	YES	YES
Imputation Dummy	YES	YES	YES	YES	YES
Observations	85	83	83	83	83
R-squared	0.060	0.258	0.289	0.258	0.239
Pb & Pb <sup>2</sup> F-Test p-value	0.135	0.0558*	0.0435**	0.0099***	0.0056***

Notes: A higher CBCL score means higher clinical risk. Columns 1-3 present regressions of the total 2010 CBCL total t-score, with different specifications. Columns 4 & 5 present the results for the internal and external subcategories respectively. Pb is the Soil Lead Level assigned to each home using a geographic Nearest Neighbor Matching. Socioeconomic Variables include age of the child, age<sup>2</sup>, household income and mother's education (measured in years). The mother IQ tests are the Wais Numbers test and the Wais Vocabulary test. Imputation Dummy is a dummy assigned to those observations where income was estimated using a reported income bracket. The last row presents the F-Test p-value of the null hypothesis:  $\beta[Pb]=0$  &  $\beta[Pb^2]=0$ . Robust standard errors are used.

<sup>\*</sup> p < 0.1, \*\* p < 0.05, \*\*\* p < 0.01

Table 13: Main Results with Age & Lead Interaction

(1)	(2)	(3)	(4)	(5)
'10 Bat Tot	'10 Bat Cog	'10 CBCL Int	'12 CBCL Int	'12 CBCL Tot
-0.161	-0.316**	0.138***	0.245**	0.165
(0.122)	(0.143)	(0.0439)	(0.106)	(0.112)
( )	( )	,	,	-0.000190*
				(0.000102)
(0.000=11)	(0.000011)	(1.010 00)	(0.000 00)	(0.00010 <b>-</b> )
-0.00321	0.00392	-0.00205*	-0.00217	-0.00113
	(0.00567)			(0.00172)
(0.00 -0.)	(3.3333.)	(0.00_00)	(0.00-0-)	(0.00-1-)
6.524**	2.184	0.722	3.223	3.109
(2.830)	(3.100)	(1.090)	(2.059)	(2.004)
-0.235**	-0.110	-0.00729	-0.0259	-0.0275
(0.0950)	(0.103)	(0.0182)	(0.0180)	(0.0180)
YES	YES	YES	YES	YES
YES	YES	YES	YES	YES
YES	YES	YES	YES	YES
52	52	83	83	83
0.448	0.275	0.129	0.275	0.293
	(0.122) 0.000703** (0.000271) -0.00321 (0.00457) 6.524** (2.830) -0.235** (0.0950) YES YES YES YES 52	'10 Bat Tot '10 Bat Cog  -0.161	'10 Bat Tot       '10 Bat Cog       '10 CBCL Int         -0.161       -0.316**       0.138***         (0.122)       (0.143)       (0.0439)         0.000703**       0.000769**       -0.000209***         (0.000271)       (0.000344)       (7.04e-05)         -0.00321       0.00392       -0.00205*         (0.00457)       (0.00567)       (0.00106)         6.524**       2.184       0.722         (2.830)       (3.100)       (1.090)         -0.235**       -0.110       -0.00729         (0.0950)       (0.103)       (0.0182)         YES       YES       YES         \$52       \$83	'10 Bat Tot       '10 Bat Cog       '10 CBCL Int       '12 CBCL Int         -0.161       -0.316**       0.138***       0.245**         (0.122)       (0.143)       (0.0439)       (0.106)         0.000703**       0.000769**       -0.000209***       -0.000266***         (0.000271)       (0.000344)       (7.04e-05)       (8.63e-05)         -0.00321       0.00392       -0.00205*       -0.00217         (0.00457)       (0.00567)       (0.00106)       (0.00162)         6.524**       2.184       0.722       3.223         (2.830)       (3.100)       (1.090)       (2.059)         -0.235**       -0.110       -0.00729       -0.0259         (0.0950)       (0.103)       (0.0182)       (0.0180)         YES       YES       YES       YES         YES       283       83       83

Notes: A higher Batelle score means greater development. A higher CBCL score means higher clinical risk. Columns 1 & 2 show regressions on the 2010 Batelle total t-score and cognitive score respectively. Column 3 presents the 2010 CBCL internal t-score regression, and columns 4 & 5 present regressions on the 2012 CBCL internal and total t-score respectively. Pb is the Soil Lead Level assigned to each home using a geographic Nearest Neighbor Matching. Socioeconomic Variables include age of the child, age<sup>2</sup>, household income and mother's education (measured in years). The mother IQ tests are the Wais Numbers test and the Wais Vocabulary test. Imputation Dummy is a dummy assigned to those observations where income was estimated using a reported income bracket. AgexPb is an interaction of the children's age with the Soil Lead Level assigned to his or her household. The last row presents the F-Test p-value of the null hypothesis:  $\beta[Pb]=0$  &  $\beta[Pb^2]=0$ . Robust standard errors are used.

<sup>\*</sup> p < 0.1, \*\* p < 0.05, \*\*\* p < 0.01

Table 14: Main Results with Distance to Site F & Lead Interaction

	(1)	(2)	(3)	(4)	(5)	(6)
Variables	'10 Bat Tot	'10 Bat Cog	'12 Bat Tot	'10 CBCL Int	'12 CBCL Int	'12 CBCL Tot
Pb	-0.0952 (0.109)	-0.100 (0.111)	-0.105 (0.0625)	0.0551 $(0.0462)$	0.104** (0.0507)	0.0886 $(0.0570)$
$Pb^2$	0.000474* $(0.000252)$	0.000435 $(0.000268)$	0.000350* (0.000175)	-0.000119 (7.52e-05)	-0.000236*** (8.41e-05)	-0.000168 (0.000104)
DistSiteFxPb	-0.0138 (0.0148)	-0.0123 (0.0190)	-0.00848 (0.00616)	-0.00390 (0.0145)	0.00243 (0.0132)	0.00126 (0.0167)
DistSiteF	4.242 (2.905)	5.884* (3.138)	0.764 (1.002)	-1.712 (1.871)	-0.959 (1.974)	-0.710 (2.296)
SE Variables	(2.903) YES	(5.156) YES	YES	YES	(1.974) YES	(2.290) YES
Mother IQ tests	YES	YES	YES	YES	YES	YES
Imputation Dummy	YES	YES	YES	YES	YES	YES
Observations	52	52	52	83	83	83
R-squared	0.485	0.385	0.444	0.167	0.263	0.292
F-test p-value	0.0174**	0.2407	0.0009***	0.4071	0.038**	0.0624*

Notes: A higher Batelle score means greater development. A higher CBCL score means higher clinical risk. Columns 1 & 2 show regressions on the 2010 Batelle total t-score and cognitive score respectively. Column 3 shows the results for the regression on the 2012 Batelle screening test. Column 4 presents the 2010 CBCL internal t-score regression, and columns 5 & 6 present regressions on the 2012 CBCL internal and total t-score respectively. Pb is the Soil Lead Level assigned to each home using a geographic Nearest Neighbor Matching. Socioeconomic Variables include age of the child, age<sup>2</sup>, household income and mother's education (measured in years). The mother IQ tests are the Wais Numbers test and the Wais Vocabulary test. Imputation Dummy is a dummy assigned to those observations where income was estimated using a reported income bracket. DistSiteF is the distance from the household to site F (measured in kilometers). DistSiteFxPb is an interaction of this distance with the Soil Lead Level assigned to the household. The last row presents the F-Test p-value of the null hypothesis:  $\beta[Pb]=0$  &  $\beta[Pb^2]=0$  &  $\beta[DistSiteFxPb]=0$ . Robust standard errors are used.

<sup>\*</sup> p < 0.1, \*\* p < 0.05, \*\*\* p < 0.01

Table 15: Main Results with High Pb Dummy

-	(1)	(2)	(3)	(4)	(5)	(6)
Variables	10 Bat Tot	10 Bat Cog	12 Bat Tot	10 CBCL Int	12 CBCL Int	12 CBCL Tot
High Pb	-2.281 (3.854)	-4.160 (4.084)	-3.572** (1.487)	0.195 (2.008)	5.216** (2.071)	5.761** (2.275)
SE Variables	YES	YES	YES	YES	YES	YES
Mother IQ Tests	YES	YES	YES	YES	YES	YES
Imputation Dummy	YES	YES	YES	YES	YES	YES
Observations	52	52	52	83	83	83
R-squared	0.368	0.185	0.385	0.077	0.234	0.278

Notes: A higher Batelle score means greater development. A higher CBCL score means higher clinical risk. Columns 1 & 2 show regressions on the 2010 Batelle total t-score and cognitive score respectively. Column 3 shows the results for the regression on the 2012 Batelle screening test. Column 4 presents the 2010 CBCL internal t-score regression, and columns 5 & 6 present regressions on the 2012 CBCL internal and total t-score respectively. High Pb is a dummy that indicates if the Soil Lead Level assigned to each home using a geographic Nearest Neighbor Matching is above the safety threshold of the Canadian Soil Quality Guidelines for the Protection of Environmental and Human Health (140 mg/Kg SLL). Socioeconomic Variables include age of the child, age<sup>2</sup>, household income and mother's education (measured in years). The mother IQ tests are the Wais Numbers test and the Wais Vocabulary test. Imputation Dummy is a dummy assigned to those observations where income was estimated using a reported income bracket. Robust standard errors are used.

<sup>\*</sup> p < 0.1, \*\* p < 0.05, \*\*\* p < 0.01

Table 16: Pooled OLS Regressions for CBCL

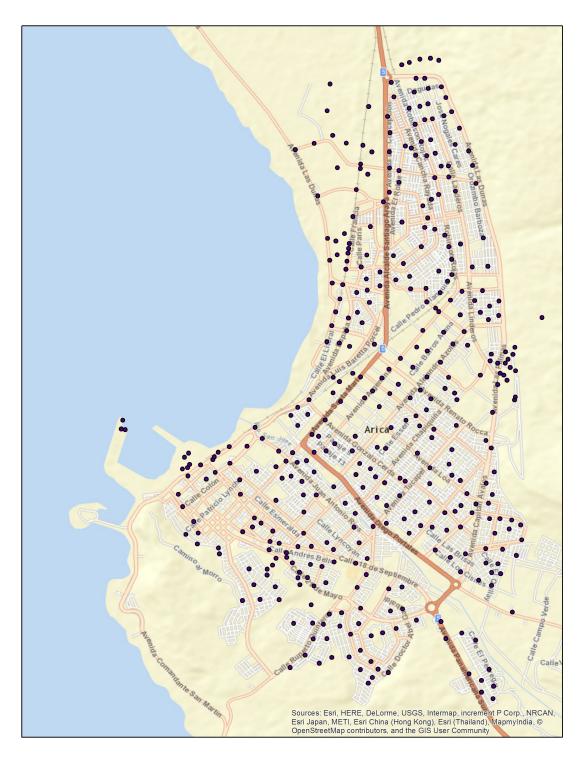
	(1)	(2)	(3)	(4)
Variables	Total	Internal	External	Internal
Pb	0.0517	0 .083***	0.00718	0.0627**
$\mathrm{Pb}^2$	(0.0319) $-9e-05$	(0.0237) -0.000198***	(0.04) $4.28e-05$	(0.0268) -0.00017***
Post Dummy	(0.00007)	(5.53e-05)	(9.27e-05)	(6e-05) -1.59
PbxPost				$(4.07) \\ 0.044$
$Pb^2xPost$				(0.039) $-6.9e-05$
				(8.6e-05)
SE Variables	YES	YES	YES	YES
Mother IQ Tests	YES	YES	YES	YES
Imputation Dummy	YES	YES	YES	YES
Observations	166	166	166	166
R-squared	0.166	0.166	0.144	0.186
Pb & Pb <sup>2</sup> F-Test p-value	0.1355	0.0024***	0.0091***	-
PbxPost & Pb²xPost F-Test p-value	-	-	-	0.36

Standard errors in parentheses

Notes: A higher CBCL score means higher clinical risk. Columns 1,2 & 3 show the pooled regressions for the 2010-2012 CBCL total, internal and external t-scores respectively. Column 4 shows the pooled regression of the 2010-2012 CBCL internal t-scores including a dummy and interactions. Pb is the Soil Lead Level assigned to each home using a geographic Nearest Neighbor Matching. Post Dummy takes the value of 1 for 2012 observations. PbxPost and Pb²xPost are interactions with this dummy. Socioeconomic Variables include age of the child, age², household income and mother's education (measured in years). The mother IQ tests are the Wais Numbers test and the Wais Vocabulary test. Imputation Dummy is a dummy assigned to those observations where income was estimated using a reported income bracket. The second to last row presents the F-Test p-value of the null hypothesis:  $\beta[Pb]=0$  &  $\beta[Pb^2]=0$ . The last row presents the F-Test p-value of the null hypothesis:  $\beta[PbxPost]=0$  &  $\beta[Pb^2xPost]=0$ . Standard error are clustered by individual.

<sup>\*</sup> p < 0.1, \*\* p < 0.05, \*\*\* p < 0.01

Figure 1: CONAMA 2009 Soil Lead Level Sample



Note: This figure presentes the 440 soil samples collected by CONAMA in 2009 throughout Arica.

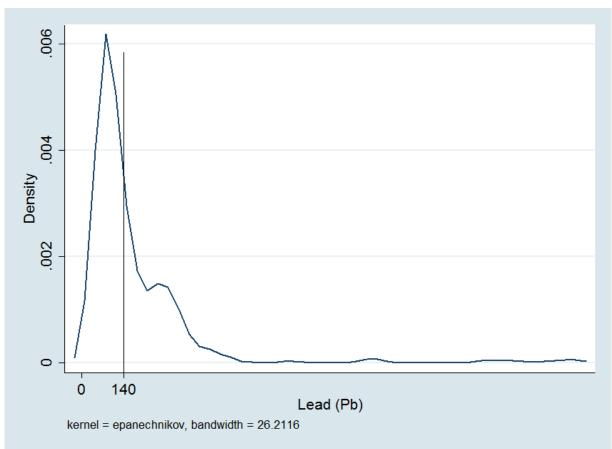
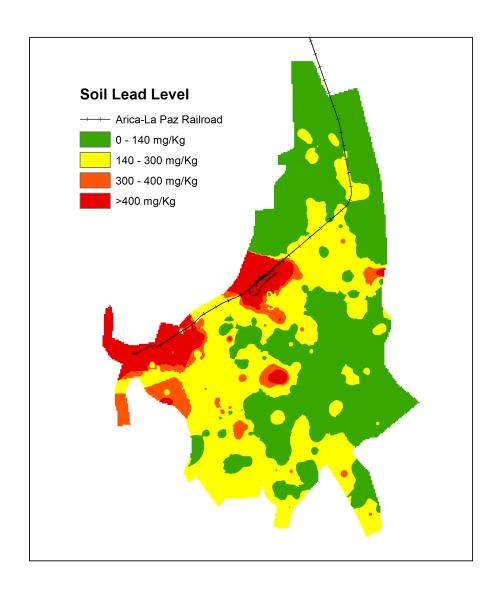


Figure 2: CONAMA 2009 Soil Lead Level Sample Density Plot

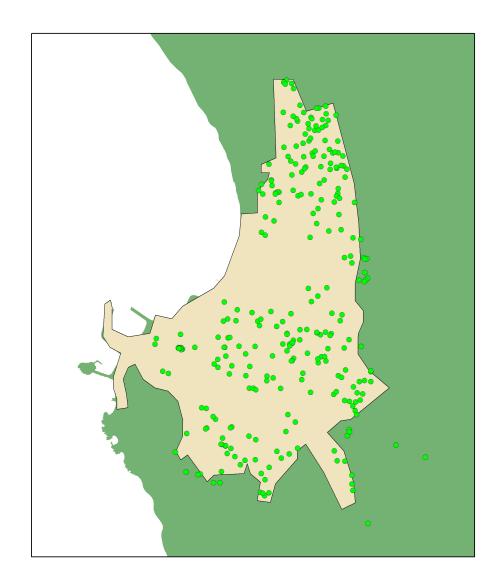
Note: This figure presents the density plot of the 440 soil samples collected by CONAMA in 2009 throughout Arica.

Figure 3: SLL distribution using IDW Interpolation



Note: This figure shows the spatial distribution of lead in Arica, using inverse distance weighting interpolation based on the 440 soil samples collected by CONAMA.

Figure 4: ELPI Households (Both 2010 & 2012)



Note: This figure shows the location of the 280 ELPI Arica households. This includes households from both 2010 and 2012.

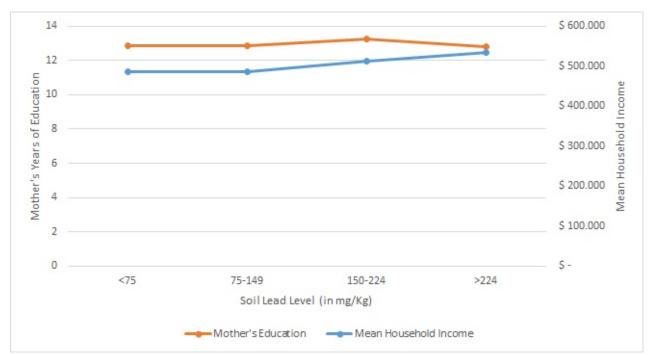
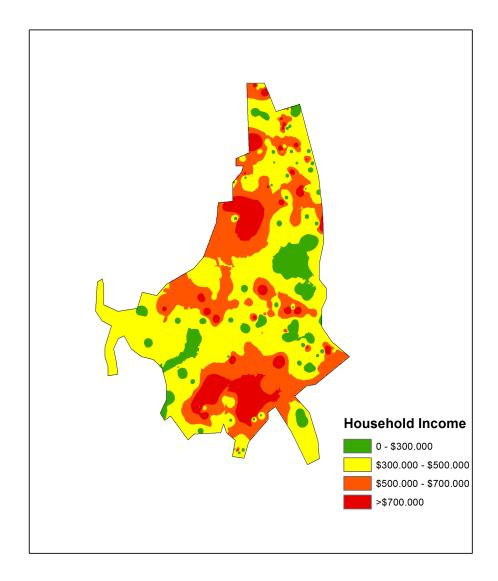


Figure 5: Mean Income and Mother's Education by SLL

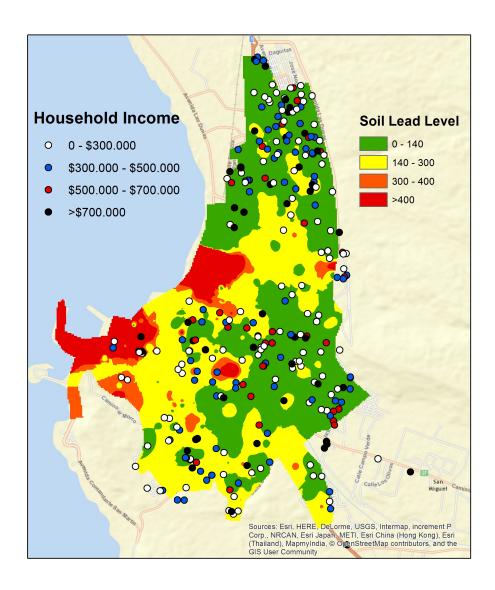
Note: This figure classifies households by the soil lead level assigned to them using geographic nearest neighbor matching. It then presents the average mother's education and household income for each group.

Figure 6: ELPI Households' Income Distribution using IDW Interpolation



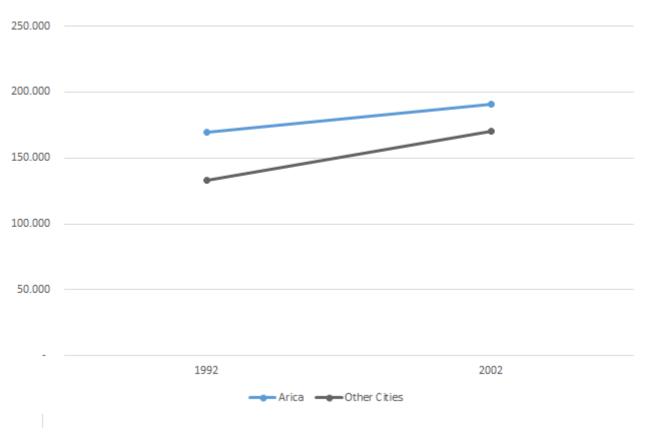
Note: This figure shows the spatial distribution of income in Arica, using inverse distance weighting interpolation based on the household income reported in the ELPI.

Figure 7: SLL distribution with IDW Interpolation and Household Income by Household



Note: This figure shows the spatial distribution of soil lead level in Arica using inverse distance weighting interpolation; and each household classified in different colors by the income they reported in the ELPI.

Figure 8: Total Population in Main Northern Cities and Arica: 1992-2002



Source: CENSO 1992 and CENSO 2002

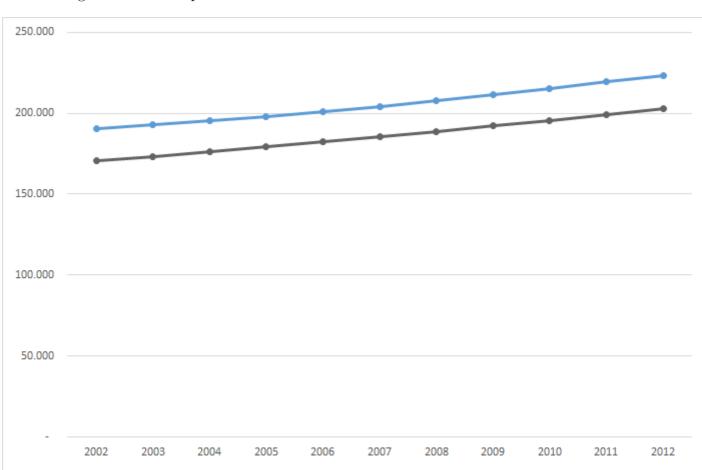
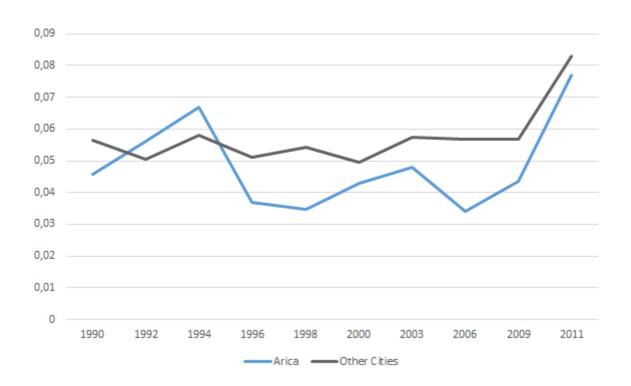


Figure 9: Total Population in Main Northern Cities and Arica: 2002-2012

Source: CENSO 2002 and estimations done by the INE for subsequent years based on the CENSO 2012.

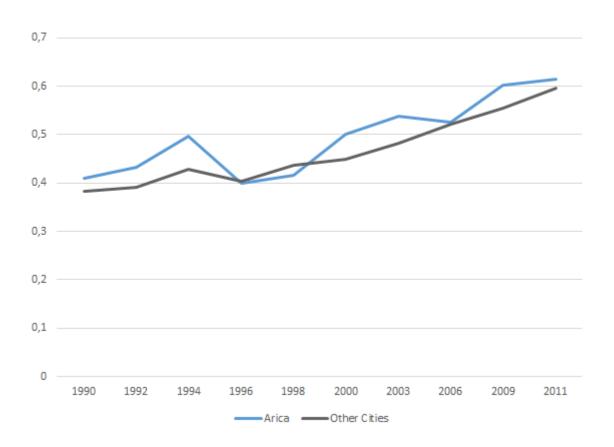
→Arica →Other Cities

Figure 10: University Graduate Share



Source: CASEN of each year shown.

Figure 11: High School Graduate Share



Source: CASEN of each year shown.

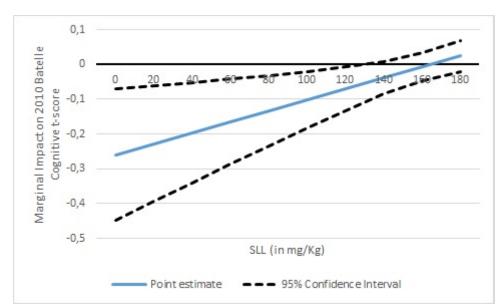


Figure 12: Marginal Effect of Lead on 2010 Batelle Cognitive t-score

Note: Estimated using results from table 7. A higher Batelle Score means greater development.

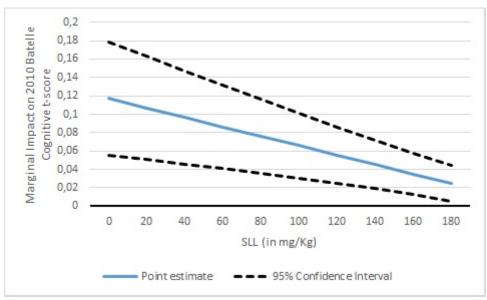


Figure 13: Marginal Effect of Lead on 2012 CBCL Internal t-score

Note: Estimated using results from table 11. A higher CBCL score means a higher clinical risk.