



HAL
open science

The health and socioeconomic costs of exposure to soil pollution: evidence from three polluted mining and industrial sites in Europe

Pierre Levasseur, Katrin Erdlenbruch, Christelle Gramaglia

► To cite this version:

Pierre Levasseur, Katrin Erdlenbruch, Christelle Gramaglia. The health and socioeconomic costs of exposure to soil pollution: evidence from three polluted mining and industrial sites in Europe. *Journal of Public Health*, 2022, 30, pp.2533-2546. 10.1007/s10389-021-01533-x . hal-03205207

HAL Id: hal-03205207

<https://hal.science/hal-03205207v1>

Submitted on 23 Apr 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Title:

The health and socioeconomic costs of exposure to soil pollution: Evidence from three polluted mining and industrial sites in Europe

Authors:

Pierre Levasseur[‡], Katrin Erdlenbruch* and Christelle Gramaglia*

Affiliation:

[‡] SADAPT, INRAE, AgroParisTech, Univ. Paris-Saclay, Paris

*G-EAU, INRAE, Cirad, IRD, Institut Agro, Univ. Montpellier, Montpellier

Corresponding author:

Pierre Levasseur, pierre.levasseur@inrae.fr, SADAPT, 16 rue Claude Bernard 75231 Paris Cedex 05

DECLARATIONS

- **Funding:** This study was part of the SOILTAKECARE project (grant agreement ID: 677407) financed by INTERREG SUDOE funds 2017-2019 (funded by the European Commission through the FEDER). Please see the website for more information about the SOILTAKECARE project: <http://soiltakecare.eu/>.
- **Conflicts of interest/Competing interests (include appropriate disclosures):** The authors have no conflict of interest to declare.
- **Ethics approval (include appropriate approvals or waivers):** All ethical standards concerning data collection and analysis were respected. The database is anonymous and contains no personal information. The authors are responsible for all remaining errors.
- **Consent to participate (include appropriate consent statements):** Each participant filled a consent statement form before participation to the survey.
- **Consent for publication (consent statement regarding publishing an individual's data or image):** Not applicable. The publication does not contain individual's data. Data are anonymous.
- **Availability of data and material (data transparency):** Data under embargo until December 2022.
- **Code availability (software application or custom code):** Not applicable.
- **Authors' contributions:** Pierre Levasseur (50%): survey writing and design, data management, statistical analysis and article's writing; Katrin Erdlenbruch (30%): survey design and article's writing; Christelle Gramaglia (20%): fieldwork, survey writing and article's writing.

Acknowledgements: This study was part of the SoilTakeCare project (grant agreement ID: 677407) financed by INTERREG SUDOE funds 2017-2019. All ethical standards concerning data collection and analysis were respected. The database is anonymous and contains no personal information. The authors are responsible for all remaining errors.

Title: The health and socioeconomic costs of exposure to soil pollution: Evidence from three polluted mining and industrial sites in Europe

Abstract:

Aim. *This article aims at providing a better understanding of the health and socioeconomic costs induced by soil pollution exposure.*

Subject. *We conduct quantitative surveys in households living near mining and/or industrial sites in France, Spain, and Portugal, as well as those located in cleaner neighboring areas.*

Method. *We employ a complementary estimation approach based on ordinary least square, instrumental variables, and propensity score matching.*

Results. *Our results confirm significant life-long health risks for residents of polluted areas compared to those in control areas. We find lower birthweight and lower childhood health status, as well as a higher risk of chronic disease in adulthood and higher premature mortality. Regarding the socioeconomic costs, we find higher rates of school absenteeism and health service demand among residents from polluted areas compared to control areas. Furthermore, we observe heterogeneous effects according to sociodemographic characteristics. As expected, children and the elderly are the most sensitive age groups; in addition, materially deprived and uneducated households are particularly vulnerable to pollution. More surprisingly, there is some evidence of higher vulnerability of educated households with regard to birth outcomes.*

Conclusion. *Our results have important implications for public policy: they allow alerting about actually observed health risks in the exposed areas but they also call for designing awareness campaigns and remedial strategies that are targeted towards the most vulnerable.*

Key words: Mining and industrial sites; pollution; reported health outcomes; absenteeism; southwestern Europe.

JEL: Q53; I14; J24

1. Introduction

Pollution threatens the environment and human health worldwide. The epidemiological literature repeatedly draws attention to this fact. Pollution exposure was responsible for 16% of all deaths worldwide in 2015 (Forouzanfar et al. 2016; Landrigan et al. 2018). It is the largest environmental cause of disease and premature death around the world—three times higher than AIDS, tuberculosis, and malaria combined.

The majority of the existing literature focuses on air and water pollution, omitting health problems generated by exposure to other types of pollution. As underlined by the *Lancet* Commission on pollution and health (Landrigan et al. 2018, p.468), soil pollution, characterized by high concentrations of metals and toxic chemicals in the soil, is “a potentially important, but not yet fully characterized or quantified source of pollution-related disease”. Hence, in this article, we focus on soil pollution, which represents around 15% of the pollution-mortality burden in rich countries (Landrigan et al. 2018).¹ This research topic is particularly relevant in Europe since the European Commission failed to adopt concrete policies against soil degradation in 2007.

Specifically, we base our analyses on three contaminated sites located in France, Spain, and Portugal, where the soil has been highly polluted by the *long-term* presence of mining and other industrial activities. As in other mining and industrial sites around the world, the main hazardous substances persisting in soil are metals (e.g. mercury, cadmium, lead, zinc, and arsenic) and polyaromatic hydrocarbons (e.g. benzo(a)pyrene and naphthalene), producing a chemical cocktail that is potentially hazardous to human health (Vandermoere 2008). According to theoretical assessment methods using geoaccumulation models (Li et al. 2014),

¹ It is worth noting that the estimated number of deaths by risk factor varies according to national GDP. In high-income countries, air pollution is the primary cause of death (60%), followed by occupational pollutants (20%), soil pollution, including heavy-metal residues and toxic chemicals (15%), and finally water pollution (5%).

1 long-term exposure to metals and toxic chemicals is assumed to increase health risks through
2 several vectors of transmission: repeated and prolonged contact with residuals in the ground
3
4 (e.g. agricultural activity), air (e.g. breathing dust), water (e.g. swimming and drinking), and
5
6 local food (e.g. homegrown fruits and vegetables, wild food, and farm produce). However, few
7
8 studies analyze the cumulative effects of soil pollution exposure using empirical data and tools.
9
10

11
12 The health effects of soil pollution are of interest to social scientists in terms of both the overall
13
14 negative effect of illness on public welfare and the cost of pollution to society. First, soil
15
16 pollution exposure has a potential impact on medical and pharmaceutical expenditures by
17
18 increasing health service demand, including hospitalizations, physician visits, medication,
19
20 follow-up treatments, and other related services (see Deschênes et al. 2017; Moretti and Neidell
21
22 2011; Schlenker and Walker 2016 for results relying on air pollution exposure). Second,
23
24 pollution-related illness and mortality have potentially significant impacts on productivity,
25
26 lowering national aggregated income by reducing labor force participation and labor market
27
28 productivity. Soil pollution exposure may lead to individuals being too sick to work regularly
29
30 (absenteeism) or dying prematurely and no longer contributing to economic production
31
32 (Landrigan et al. 2018). Soil pollution exposure have also potential impacts on human capital
33
34 (Hunt et al. 2016). Indeed, since children are particularly sensitive to pollution-related disease,
35
36 even at extremely low doses of exposure, it is safe to assume that pollution exposure affects
37
38 productivity at school, not only in attendance rates, but through a loss of concentration during
39
40 the school day. There is convincing evidence that early-life exposure to neurotoxic pollutants
41
42 concentrated in soils, such as lead and mercury, has a significant effect on cognitive
43
44 development and the ability to focus (Landrigan et al. 2018). In short, any cause of productivity-
45
46 loss at school can affect academic outcomes and subsequent life opportunities. Hence, soil
47
48 pollution-related disease and morbidity that contribute to academic failure can perpetuate
49
50 intergenerational poverty (Hunt et al. 2016).
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

1 Our contribution to the literature is to provide empirical evidence about the health and
2 socioeconomic costs of soil pollution exposure, through the construction of an original
3 comparative database and the use of statistical models that limit bias induced by endogeneity
4 problems. Concretely, we conducted standardized quantitative surveys in households living
5 near mining and/or industrial sites known as pollution hot-spots in France, Spain, and Portugal,
6 as well as those located in ‘clean’ neighboring areas. The compilation of individuals from
7 polluted municipalities and clean municipalities nearby allows understanding health-related
8 outcomes of very specifically defined groups of individuals. In contrast, the use of
9 administrative health records would usually not allow matching individuals on such a small
10 scale to their place of residence. Also, the survey allows a large set of pollution-specific
11 variables to be collected such as original information about pollution perceptions. To identify
12 a non-biased effect of residential location (i.e. polluted vs. clean areas) and health outcomes,
13 we combine ordinary least square (OLS) estimations with two complementary endogeneity-
14 correcting estimators: instrumental variables (IV) and propensity score matching (PSM)
15 procedures. In addition, we investigate the potential presence of age-specific and
16 socioeconomic-specific heterogeneity in the relationship.

17 We find that living near mining or industrial sites has negative externalities on human health,
18 school absenteeism, and health service demand. While birthweight tends to be lower in polluted
19 areas, children growing up in control areas have a better health status and are less often absent
20 at school. We also find that adult residents of polluted areas have a higher risk of suffering from
21 chronic disease than control adults. Further, there is clear evidence that exposure to soil
22 pollution increases health service demand and premature mortality. Finally, we observe
23 heterogeneous effects according to demographic and socioeconomic factors. As expected,
24 health hazards tend to be greater for children and the elderly. Likewise, materially deprived and
25 uneducated households disproportionately report premature death in the immediate family when

1 they live in polluted areas. Conversely, we observe the impact on birthweight of living in a
2 polluted area to be greater among households where the respondent is educated. This
3
4 unexpected result might be explained by the higher consumption of local (but contaminated)
5
6 food products by relatively educated mothers than by uneducated mothers.
7
8
9

10 The structure of the article is as follows. We first describe the database and explain the empirical
11
12 strategy. We then present our results. Finally, we conclude and provide recommendations for
13
14 public policy.
15
16
17

18 **2. Materials and methods**

19 *2.1. An original comparative database*

20
21
22
23
24 Three cross-sectional standardized household surveys were conducted from October 2018 to
25
26 January 2019 among 1,194 families in France, Portugal, and Spain. These surveys resulted in
27
28 an original comparative database about the socioeconomic and health issues of soil pollution
29
30 exposure, called Comparative Survey on Pollution Exposure (CSPE). To collect representative
31
32 information for each study area, we employed a random selection of housings based on
33
34 cadastral geo-localized data (using a GIS software). The random selection of housings was
35
36 established at the town-level in France and Spain, and municipality-level in Portugal, using the
37
38 most recent census data provided by national institutes of statistics.² Consequently, the CSPE
39
40 database is representative of three polluted European areas (in rural and peri-urban settings):
41
42 the town of Viviez in France (156 households and 293 individuals), the municipality of
43
44 Estarreja in Portugal (300 households and 739 individuals), and three villages of the Spanish
45
46 Sierra Minera (Portman, Estrecho de San Ginès, and Alumbres) located to the east of Cartagena
47
48 (228 households and 557 individuals). Then, to construct a counterfactual group, we further
49
50 collected representative data in similar neighboring but cleaner areas as benchmarks for the
51
52
53
54
55
56
57
58
59

60 ² The response rate of selected and visited housings was between 20% and 25% depending on the survey area.
61
62
63
64
65

1 effects related to pollution exposure. The control areas are the town of Montbazens in France
2 (138 households and 309 individuals); the municipality of Vagos in Portugal (200 households
3 and 437 individuals); and a group of villages (Portus, Galifa, Perin, La Corona, Cantera, and
4 Molinos Marfagones) located to the west of Cartagena in Spain (172 households and 452
5 individuals).³ While the three polluted areas are hotspots of soil pollution in Europe, as
6 evidenced in geochemistry and mineralogy (Patinha et al. 2015; Pérez-Sirvent et al. 2016; Sivry
7 et al. 2010), the selection of control areas relied on the region-specific literature. In Portugal for
8 instance, Guihard-Costa et al. (2012) and Inácio et al (2014) explain that Estarreja and Vagos
9 had the same natural amenities before the installation of the chemical complex in Estarreja.
10 Similarly, in a health report, the French institute of public health considered Montbazens as
11 control area to infer the effects of pollution exposure in Viviez (Durand et al. 2011). As shown
12 in Figure 1, polluted and control areas are close enough to allow a certain comparability, but
13 far enough to not have the same concentrations of pollutants in soils.
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30

31 In short, our southwestern European sample makes it possible to observe three types of polluted
32 areas around the world: (i) ex-mining towns (Portman and Estrecho de San Ginès), (ii) former
33 metal industry undergoing technological reconversion (Viviez), and (iii) active (petro-)
34 chemical complexes (Estarreja and Alumbres). Not surprisingly, socioeconomic indicators of
35 the polluted locations tend to be worse in the two first groups and better in the third compared
36 to the controls. Further contextual information about our sample of study areas is available in
37 Appendix A.
38
39
40
41
42
43
44
45
46
47
48

49 [insert Figure 1]
50
51
52
53
54

55 ³ For example, Guihard-Costa et al. (2012) and Inácio, Neves, and Pereira (2014) explain that Estarreja and Vagos
56 had the same natural amenities before the installation of the chemical complex in Estarreja. Similarly, the French
57 Institute of Public Health uses Montbazens as a control area to establish the health effects of pollution exposure in
58 Viviez (Durand et al. 2011). We failed to find equivalent publications for the Spanish Sierra Minera. Hence, we
59 selected a group of control areas to the west of Cartagena, where the natural features would resemble the Sierra
60 Minera if the latter did not host large mining and petrochemical sites (Banos-Gonzales and Baños Paez 2013).
61
62
63
64
65

2.2. Outcome indicators

To measure the health and socioeconomic costs of exposure to soil pollution, several outcome indicators are computed for households and individuals. At the household level, we asked the respondent to report the birthweight in kg of the last child, and the number of family members (living the same community) to have died before the age of 65 years. Measuring early mortality by asking for familial premature death is an approach commonly used in epidemiological studies (Gami et al. 2007; Hendryx et al. 2010). Birthweight is a proxy for neonatal health, since maternal exposure to pollution may lead to lower birthweight, namely due to intrauterine growth retardation and/or prematurity (Bobak 2000). For each surveyed household, we also collected individual information on childhood health and healthcare demand (from 0 to 15 years old). More precisely, we asked the respondent to: (i) assess the health status of each child on a 5-point Likert scale (varying from very bad to very good health status), (ii) report whether a child needs frequent medical treatment (binary-response variable), (iii) report whether a child has been diagnosed with allergies by a doctor (binary-response variable). We also asked the respondent to report the occurrence of school absences during the last semester as a proxy for school productivity. Finally, we asked respondents to provide information on themselves, as well as other adult household members, in order to identify: (i) who has been diagnosed with a chronic disease (dummy), (ii) the number of work absences during the last semester (a proxy for labor productivity), (iii) the number of medical visits during the last semester (a proxy for health service demand and medical expenditure).

Epidemiologists are often concerned by reported health outcomes as recall bias can disproportionately affect respondent declarations between case and control areas (Coughlin 1990). This may lead to an overestimation of health externalities if respondents from polluted areas are better able to recall certain events (e.g. birthweight, death, absenteeism, medical visits) and illnesses (e.g. medical treatments, allergies, chronic diseases) compared to respondents

1 from control areas.⁴ However, Table 1 shows that there is no within-group correlation between
2 the respondents' perception of local pollution (on a 5-point Likert scale) and health-related
3
4 outcomes, except for reported child allergies, adulthood medical visits, and school absences in
5
6 control areas where positive correlations stand out. Hence, for most health indicators, we can
7
8 be confident that non-random recall errors are absent in our samples. If the reporting of health
9
10 outcomes was influenced by the respondents' perception of pollution, then an over-declaration
11
12 of health penalties and diseases would have been found among pollution-sensitive respondents.
13
14 We are also confident of the accuracy of declarations of birthweights, absences, and chronic
15
16 diseases. Moreover, we assume that having household members who die prematurely or suffer
17
18 from chronic disease is information that a respondent would be unlikely to forget—and
19
20 epidemiological studies support this.⁵
21
22
23
24
25
26

27 [insert here Table 1]
28
29

30 **2.3. A linear model**

31
32

33 Because mining and industry might also directly pollute air and water, or affect health through
34 exposure when working, we specify a reduced-form model of health effects from polluting
35 industries. In other words, we analyze the cocktail effect of soil pollution. Equation 1
36 mathematically poses the baseline multivariate regression model:
37
38
39
40
41
42
43

44 ▪ Eq.1: $Y_i = \beta_0 + \beta_1 T_i + \beta_2 X_i + \varepsilon_i$
45
46
47
48
49
50

51 ⁴ The risk of recall bias is especially high when the issue is commonly established in social media and popular
52 knowledge. For example, it is often the case with regard to the link between pollution and allergies.

53 ⁵ For instance, O'Sullivan et al. (2000) found that 75% of declared birthweights were within 50 grams of the
54 hospital record. Moreover, the birthweight gap (recalled birthweight vs. hospital record) correlates with neither
55 the sociodemographic indicators nor the time interval between childbirth and data collection. Similarly, a case-
56 control study assessing the effects of professional chemical exposure on cancer risk found that 82% of reported
57 work histories agreed with the records, with no statistical difference between cases and controls or according to
58 age, education level, or social class. Furthermore, comparing patients' self-reports to general practitioners'
59 information, Kriegsman et al. (1996) observed that the accuracy of reported chronic disease is generally
60 satisfactory.
61
62
63
64
65

1 Y_i refers to the outcome indicators that we previously described for statistical unit i . Depending
2 on the outcome variable, i refers to an individual or a household. For the sake of clarity, we use
3 an OLS estimator to run linear estimations when Y is continuous (e.g. birthweight, child health
4 status, occurrence of premature death, medical visits, and school and work absenteeism),⁶ but
5 run linear probability estimations when Y is binary (e.g. risk of child medical assistance, risk of
6 suffering from allergies for children and chronic diseases for adults). We also tested a Probit
7 binomial estimator for binary-outcome variables, but, as the results were similar, we chose only
8 to report OLS estimates.

9
10
11
12
13
14
15
16
17
18
19
20 X_i is a vector of control factors. First, we control for potentially important demographic and
21 socioeconomic confounders to the pollution-health relationship (Sheppard et al. 2012). More
22 specifically, we control for age, gender, marital status, household composition, ethnicity,
23 completed education level, a wealth index, and housing size. We also include a general
24 measurement of the respondent's risk attitude—another potential confounder between pollution
25 exposure and health outcomes. Indeed, risk lovers may take concomitant risks regarding
26 pollution exposure and health-related behaviors. In the survey, we measured risk aversion by a
27 5-point Likert scale, varying from having a very low to a very high willingness to take risks in
28 general. As suggested by Dohmen et al. (2011), measurements of general risk aversion generate
29 the best all-round predictor of risk behaviors. We then control for general health status by
30 integrating measurements of body mass index (BMI) and height. BMI is considered in the
31 literature as a *short-term* indicator of health status, since weight strongly depends on current
32 consumption and lifestyle. By contrast, height is considered a pre-existing *long-term* indicator
33 of health status, as it strongly reflects nutritional status and maternal risk behaviors occurring
34 from gestation to early childhood (Thomas et al. 1991). Finally, we control for the length of
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58

59 ⁶ We also ran regressions using logged values of birthweight rather than real birthweight values in order to limit
60 the influence of potential non-linearities in birthweight declarations. The results were similar.

1 residence in the area to fix the duration of pollution exposure across households, and we include
2 country fixed-effects.⁷ Obviously, the way in which we introduce control factors into the model
3
4 varies according to the targeted subject (i.e. the household-based, child-based, and adult-based
5
6 outcomes). Table B1 in Appendix B clarifies this matter by describing the way in which each
7
8 variable is included in the model.
9
10

11 T_i is a proxy for soil pollution exposure. To approximate this exposure we use a binary-
12
13 treatment variable that identifies whether the household lives in a polluted area or in a control
14
15 area. In addition to obvious measurement errors (due to possible individual heterogeneity in
16
17 soil pollution exposure), there is a potential selection bias due to an overrepresentation of
18
19 economically and socially deprived households in polluted areas (Banzhaf and Walsh 2008;
20
21 Levasseur et al. 2021). Yet the poorest tend to have lower health status, as well as riskier health
22
23 behaviors and lower interest in their children's education (Zuidema and Nentjes 1997).
24
25 Although we control for socioeconomic and demographic factors, as well as *long-term* and
26
27 *short-term* health status, unobserved factors ϵ_i may still bias OLS estimates (i.e. when ϵ_i is
28
29 correlated to T_i). For instance, those living in polluted areas could be less sensitive to
30
31 environmental health hazards, which may lead to risky health behaviors. In contrast, not
32
33 controlling for potential avoidance behaviors may understate the estimated effect of soil
34
35 pollution exposure on health outcomes (Moretti and Neidell 2011). To limit such selection bias
36
37 and subsequent endogeneity problems, we perform an IV strategy using the best available
38
39 instruments of our dataset. As the complete exogeneity of selected instruments is always
40
41 debatable in applied economics, we complete OLS and IV estimates with PSM estimates. Both
42
43 IV and PSM methodologies are described in Appendix C.
44
45
46
47
48
49
50
51
52
53

54
55
56
57 ⁷ We tested for various model specifications, by integrating covariates one by one, as well as analyzing more
58
59 careful specifications that exclude confounding factors for which the causality in the Eq.1 is less clear (e.g.
60
61 excluding height, BMI, housing size and length of residence). For each specification, the results were similar
62
63 (results available upon request).
64
65

1
2
3 **3. Results**

4 **3.1. Descriptive statistics**

5
6 Figures 2(a) and 2(b) show polynomial fractional estimates of the relationship between
7 residential location (polluted vs. clean areas) and health outcomes. More precisely, we model
8 soil pollution exposure by comparing lifetime in a polluted area (positive values on the right
9 side) versus a control area (negative values on the left side). For household-based outcome
10 indicators reported in Figure 2(a), we use the length of residence as lifetime indicator. Figure
11 2(a) clearly suggests that soil pollution exposure might reduce birthweight and increase the
12 occurrence of premature death. For instance, while the number of early deaths in the family
13 remains around 0.3-0.4 in control areas, this number significantly increases along with the
14 length of residence in a polluted area. After 30 years of residence in a polluted area, the number
15 of early deaths exceeds 0.6 local family members. Regarding individual-based health outcomes
16 (children and adults), in Figure 2(b) we use individual age as lifetime indicator. Figure 2(b)
17 shows increased health disadvantages for individuals aging in polluted areas. We find that the
18 perceived health status of children decreases with the child's age in polluted areas, whereas this
19 indicator increases with the child's age in control areas. Similarly, we observe that the
20 likelihood of requiring medical assistance increases with the child's age in polluted areas, but
21 decreases in clean areas. Interestingly, the likelihood of being diagnosed with allergies is low
22 (around 10%) and is uncorrelated with the child's age in control areas. By contrast, this
23 significantly increases with the child's age in polluted areas. By the age of 15, the risk reaches
24 almost 40%.

25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53 [Insert Figures 2(a) and 2(b)]

54
55
56 Mean-comparison tests are consistent with Figures 2(a) and 2(b). Table 2 exhibits significant
57 correlations between residential location (polluted vs. clean areas) and health problems. While
58
59
60
61
62
63
64
65

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

Figures 2(a-b) and Table 2 present a link between soil pollution exposure and health status, these results may be biased because of endogeneity issues. Hence, we implement multivariate regressions and endogeneity-corrector procedures to mitigate this.

[Insert Table 2]

3.2. Health costs of living in areas with soil pollution

Table 3 summarizes OLS, IV, and PSM estimates by only reporting the fitted coefficients of interest for all outcome indicators (i.e. the average treatment effect). Tables B2 to B6 in Appendix B exhibit full OLS and IV estimates for household-based outcome indicators (i.e. birthweight and premature deaths), for child-based outcome indicators (i.e. health status, medical assistance, allergies, and school absenteeism), and for adult-based outcome indicators (i.e. chronic disease, work absenteeism, and medical visits). In just-identified IV estimates, we use the respondent's local pollution perception (on a 5-point Likert scale) as a primary instrument. In over-identified IV estimates, we add the housing rent as a second instrument to test for the validity of the former.

Regarding the first IV requirement, first-stage IV estimates show that both instruments are strong predictors of the probability of living in polluted areas versus control areas (available in Table B7 of Appendix B).⁸ In addition, as illustrated at the bottom of Tables B2 to B6, F-statistics on the excluded instruments are systematically high. Another evidence of the strength of instruments is visible when we run reduced-form regressions, which regress outcome indicators on the instruments and covariates (available in Table B8 of Appendix B). As Angrist and Pischke (2008) argue, it is reassuring to note that the sign, the magnitude, and the

⁸ As expected, we observe a positive and significant correlation between the respondents' local pollution perception and the fact of living in a polluted area. Regarding housing cost, the sign of the correlation is also consistent: monthly rent is significantly and negatively correlated to the probability of living in polluted areas.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

significance of the fitted coefficients of instruments are consistent and proportional to the effect of treatment variables (polluted vs. clean areas) on health outcomes.

As evidenced in Tables B2 to B6, Hansen over-identification tests suggest that our instruments meet the exogeneity requirement. The null-hypothesis of exogeneity can never be rejected (i.e. no correlation between residuals and excluded instruments), except for the regression of birthweight on residential location (Column 3, Table B2). Furthermore, restricting the sample to control areas, there does not appear to be any correlation between the respondent’s pollution perception and family health outcomes that could skew our IV identification, except for the reported occurrence of medical visits among the elderly (Table B9 in Appendix B). In other words, the correlation between the respondent’s pollution perception and health outcomes is not due to omitted factors or reverse causality, at least for children and working-age adults.⁹

The third step of our empirical strategy is based on a PSM approach, comparing treated units (i.e. individuals or households living in polluted areas) to analogous control units weighted by observed characteristics. To verify the quality of the matching procedure, Figures B1(a) to B1(m) in Appendix B report the standardized percentage bias for each covariate before and after matching. Overall, the levels of standardized bias are relatively low after matching. In addition, for each covariate, there are no significant mean differences after matching between polluted and control areas (available upon request). Thus, we assume that, after matching, polluted and control areas are statistically comparable, conditional upon the set of factors included.

[Insert Table 3]

⁹ Among the elderly, one may suppose the presence of reverse causality. Indeed, older people tend to make more interactive attributions in negative situations than younger adults, for example, attributing their health problems to pollution exposure (Blanchard-Fields 1994).

1 Overall, we can observe the following facts. As evidenced in Table 3, OLS, IV, and PSM
2 estimates are consistent and reveal similar results.¹⁰ By fixing several important confounders,¹¹
3
4 OLS estimates show that living in a polluted area decreases the birthweight of the last child by
5
6 85 grams. Regarding PSM estimates, Table 3 exhibits a similar amplitude, with an average
7
8 treatment effect on the treated (ATT) of -82 grams. In contrast, just-identified IV estimates
9
10 show roughly twice the impact (-173 grams), suggesting potential estimation bias (recall that
11
12 we fail to accept the Hansen over-identification test in Table B2, Column 3). Table 3 also shows
13
14 a higher occurrence of premature family death (before the age of 65) in polluted areas compared
15
16 to cleaner areas. As suggested by IV and PSM procedures, this finding is robust to the potential
17
18 presence of unobserved heterogeneity and reverse causality, the effect varying from 0.16 to
19
20 0.26 extra deaths, depending on the estimator. Table 4 explores potential heterogeneous effects
21
22 by crossing residential location (i.e. polluted vs. clean areas) with household socioeconomic
23
24 status indicators (i.e. education and wealth).¹² As predicted, premature deaths due to exposure
25
26 to soil pollution are more frequent when the household is materially deprived (according to the
27
28 wealth index) and the respondent is not educated beyond primary level. More unexpectedly,
29
30 Table 4 presents a negative correlation between soil pollution exposure and birthweight that is
31
32 significantly higher for households with educated respondents compared to those with non-
33
34 educated respondents. We provide an interpretation for this result in the discussion.
35
36
37
38
39
40
41
42
43

44 In individual-based regressions, we observe that living in polluted areas also affects child health
45
46 during his/her growth and development. Table 3 shows that living in polluted areas decreases
47
48
49
50
51

52 ¹⁰ In line with Angrist and Pischke (2008), we prefer fitted coefficients from just-identified estimates as they are
53 median-unbiased and therefore less likely to be subject to a weak-instruments critique than over-identified
54 estimates.

55 ¹¹ Excluding residential location, the most important predictors of birthweight are respondents' BMI, height, and
56 origins, as well as accommodation size (Table B2 in Appendix B). These results highlight the importance of
57 parental anthropometric factors and sociodemographic factors regarding neonatal health.

58 ¹² Obviously, we also checked for potential heterogeneous effects nationally. However, we do not detect
59 heterogeneous effects across different cases, apart from in Alumbres, where respondents declare lower levels of
60 health status for their children compared to other polluted areas (results available upon request).
61
62
63
64
65

1 the 5-point perception score of child's health status by around 0.28-0.39 points, depending on
2 estimators. Moreover, there is convincing evidence that children living in polluted areas are
3 more likely to suffer from diagnosed allergies compared to those living in cleaner areas. For
4 instance, just-identified IV estimates indicate that living in a polluted area increases the risk of
5 a child suffering from allergies by 17.8 percentage points. Likewise, we find negative health
6 externalities for working-age and active adults, as well as for the elderly. For example, just-
7 identified IV estimates show that living in polluted areas increases the risk of suffering from
8 any chronic disease for working-age and active adults by 7.7 and 10.4 percentage points
9 respectively (Table 3). Note that, as expected, this risk is higher for the elderly (+16.9
10 percentage points, if we refer to just-identified IV). Using PSM, the estimates are only
11 significant for active workers and the elderly.
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26

27 ***3.3. Socioeconomic costs of soil pollution exposure***

28
29 To explore the socioeconomic consequences of health problems imputable to soil pollution
30 exposure, we estimate the effects of residential location (polluted vs. clean areas) on
31 school/work absenteeism, as well as potential impacts on healthcare demand. According to
32 Table 3, OLS, IV, and PSM estimates suggest a slight impact of soil pollution exposure on
33 school absenteeism (significant at the 15% level). Indeed, living in polluted areas increases by
34 1.5/1.6 days the occurrence of school absences over the last semester. However, we fail to find
35 an effect on work absences. Finally, Table 3 shows that working-age adults and active workers
36 who live in polluted areas visit medical practitioners significantly more frequently than their
37 counterparts in cleaner areas (between 0.3 and 0.8 extra visits, dependent on estimators and
38 samples). Surprisingly, this effect is more systematically significant for working-age adults and
39 active workers than for the elderly. One can for instance assume that this latter age group is
40 more likely to be hospitalized than to visit a physician. Similarly, OLS, IV, and PSM estimates
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

1 suggest that living in polluted areas significantly increases the risk of a child requiring regular
2 medical treatment (i.e. by around 8-15 extra percentage points, depending on estimators).
3
4

5 **4. Discussion and conclusion**
6

7
8 This article identifies the health and socioeconomic costs of exposure to soil pollution by
9 focusing on three polluted mining and industrial sites in France, Spain, and Portugal. We
10 employ an empirical approach by first collecting original household data in polluted and nearby
11 cleaner areas, and then analyzing this data using an empirical strategy that combines three
12 complementary estimators (i.e. OLS, IV, and PSM). In addition, we control for rarely observed
13 confounders influencing the link between residential location and health outcomes, such as
14 height and nutritional status. Even if the identification strategy performed does not rely on
15 perfect instruments, namely because the chronicity of soil pollution impedes the use of episodic
16 exogenous instruments, the convergence of findings across the three complementary
17 approaches is comforting.
18
19
20
21
22
23
24
25
26
27
28
29
30
31

32
33 Our results suggest that living near polluted sites not only may significantly affect human health
34 but also might impair the chance of socioeconomic success. With regard to health, we observe
35 negative effects of exposure to soil pollution throughout the whole lifecycle, from birthweight
36 and childhood health status to chronic diseases in adulthood and premature mortality. In
37 addition, we find some evidence that living in polluted areas may negatively influence school
38 attendance, probably due to illness. This result highlights the expected impact of soil pollution
39 exposure in terms of a loss of academic productivity, with the potential to impede human capital
40 accumulation during childhood and a subsequent professional career (Hunt et al. 2016). For
41 adults, we fail to identify an effect on work absenteeism, but we show that working-age
42 individuals and active workers living in polluted areas have higher occurrence of medical visits
43 than their counterparts who live in control areas. Similarly, children who live on polluted soil
44 are more likely to require regular medical treatment than control children. These latter findings
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

1 suggest that soil pollution exposure increases the demand for healthcare services and thus
2 contributes to increased public health expenditure.
3

4
5 Another contribution of this study is to explore potential heterogeneous effects across age and
6 socioeconomic status. In line with the existing literature, we find that the health effects of soil
7 pollution exposure are particularly salient for children (Landrigan et al. 2018; Makri and
8 Stilianakis 2008) and the elderly (Geller and Zenick 2005). In terms of socioeconomic
9 heterogeneity, the results are more ambiguous and call for further studies. Focusing on
10 premature mortality, we observed stronger effects for uneducated and materially deprived
11 households. This is in line with the findings of Morelli et al. (2016), who show a higher impact
12 of fine particulate matter (air pollution) on mortality in areas with a moderate to high social
13 deprivation index compared to areas with low deprivations. In contrast, the negative correlation
14 between soil pollution exposure and birthweight tends to be higher for households with a
15 minimum level of education (i.e. higher than primary) compared to uneducated households (i.e.
16 lower than primary education). A similarly surprising result is found in recent work by
17 Montazeri et al. (2019), who note that wealthier and more educated mothers and their children
18 are more exposed to certain toxic chemicals and metal residues (e.g. arsenic, mercury) than
19 their poorer counterparts in Europe. The authors speculate that there could be more regular
20 consumption of local food products containing high levels of heavy and chemicals (e.g. seafood,
21 organic food, homegrown fruits and vegetables, goods from short distribution channels, and
22 wild food) among higher social groups.¹³ Conversely, uneducated (and poorer) mothers are less
23 willing to pay for expensive (potentially contaminated) local products and more likely to buy
24 affordable processed and imported food products (Darmon and Drewnowski 2008; Dowler
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58

59 ¹³ See also Schupp and Sharp (2012) for more explanations about why the poor are less concerned by (risky) self-
60 provisioning activities.
61
62
63
64
65

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
2001).¹⁴ Differences in in-utero survival rates according to parental education could also
explain why educated parents reported lower birthweights. One could speculate that in-utero
survival is lower among less-educated parents than among more educated parents. If this
assumption is correct, the probability of preterm birth should be higher among educated parents.
However, there is no correlation between the risk of preterm birth and parental education (p-
value=0.732). Similarly, there is no bivariate correlation between birthweight and parental
education (p-value=0.148). Hence, this second possible explanation can be rejected.

17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
The identification of heterogeneous effects in the soil pollution-health relationship is highly
informative for public policy. First, it seems obvious for public policy providers to target the
most vulnerable demographic groups regarding anti-pollution interventions (i.e. children and
the elderly). Regarding social vulnerability to soil pollution exposure, targeting the poor only
could be a first strategy. However, more attention should be paid to higher social categories
too, who might disproportionally consume local (but contaminated) food products. It is
noticeable that the rapid succession of land uses and the reconversion of wasteland into
urbanized land can attract young households attracted by the wider environmental amenities of
the territory, exposing their children to risks they do not have to memory of (Levasseur et al.
2021). In villages and cities located near mining or industrial sites, these changes should be
accompanied by awareness campaigns on the avoidance of exposure to soil pollution (e.g.
washing hands, floors and toys frequently, avoiding that younger children play on the ground
outside, and not eating locally produced food or drink well water) and, more importantly,
remedial strategies (e.g. the storing/treatment of contaminant residues, ground covering in
schools, and the renovation of dwellings).

58
59
60
61
62
63
64
65
¹⁴ Obviously, we do not exclude the possibility that the regular consumption of cheaper imported processed food
(often rich in sugar, refined fat, and carbohydrates) generates further health issues, such as obesity and related
chronic diseases.

1 From a scientific perspective, this article helps to fill the literature gap by providing key insights
2 into the consequences of exposure to soil pollution, which still remains an underexplored issue
3
4 in the pollution-health literature. We provide clear evidence that metals and chemicals in
5
6 European soil may significantly affect human health when exposure is repeated over the long
7
8 term. Finally, this study makes a methodological contribution by demonstrating that
9
10 comparative study design (polluted vs. clean areas) through area-specific household surveys is
11
12 a powerful tool for environmental assessment analyses, especially when incorporating
13
14 endogeneity-corrector estimators.
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

REFERENCES

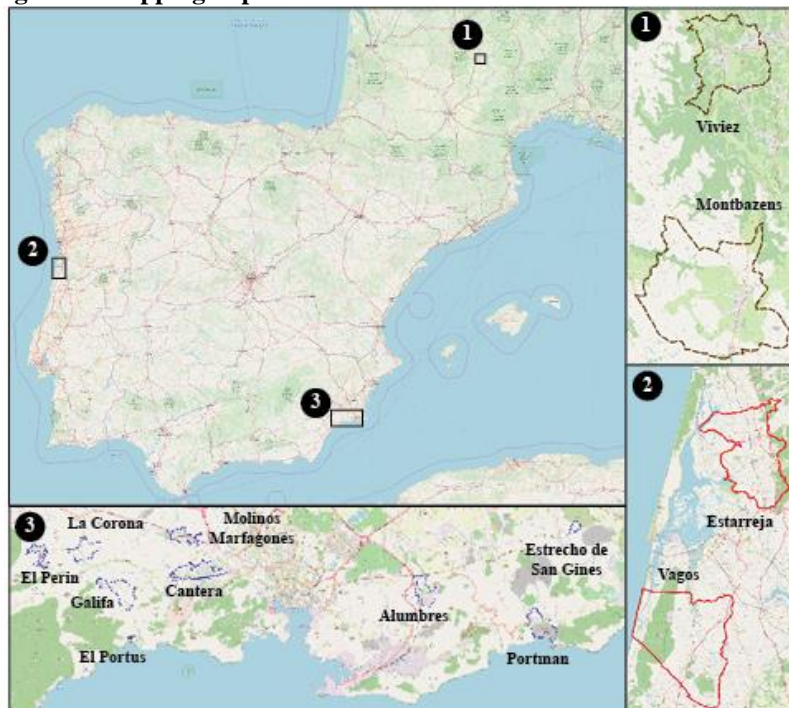
- 1 Angrist JD, Pischke J-S (2008) Mostly Harmless Econometrics: An Empiricist's Companion.
2 Princeton University Press
3
- 4 Banos-Gonzales I, Baños Paez P (2013) Portmán: De el Portus Magnus del Mediterráneo
5 Occidental a la Bahía Aterrada | Ediciones de la Universidad de Murcia (Editum),
6 Ediciones de la Universidad de Murcia. Baños Paez, Murcia, España
7
8
- 9 Banzhaf HS, Walsh RP (2008) Do People Vote with Their Feet? An Empirical Test of Tiebout.
10 Am Econ Rev 98:843–863. <https://doi.org/10.1257/aer.98.3.843>
11
- 12 Blanchard-Fields F (1994) Age Differences in Causal Attributions From an Adult
13 Developmental Perspective. J Gerontol 49:P43–P51.
14 <https://doi.org/10.1093/geronj/49.2.P43>
15
- 16 Bobak M (2000) Outdoor air pollution, low birth weight, and prematurity. Environ Health
17 Perspect 108:173–176. <https://doi.org/10.1289/ehp.00108173>
18
- 19 Coughlin SS (1990) Recall bias in epidemiologic studies. J Clin Epidemiol 43:87–91.
20 [https://doi.org/10.1016/0895-4356\(90\)90060-3](https://doi.org/10.1016/0895-4356(90)90060-3)
21
- 22 Darmon N, Drewnowski A (2008) Does social class predict diet quality? Am J Clin Nutr
23 87:1107–1117. <https://doi.org/10.1093/ajcn/87.5.1107>
24
- 25 Deschênes O, Greenstone M, Shapiro JS (2017) Defensive Investments and the Demand for Air
26 Quality: Evidence from the NOx Budget Program. Am Econ Rev 107:2958–2989.
27 <https://doi.org/10.1257/aer.20131002>
28
- 29 Dohmen T, Huffman D, Schupp J, et al (2011) Individual risk attitudes: measurement,
30 determinants, and behavioral consequences. J Eur Econ Assoc 9:522–550
31
- 32 Dowler E (2001) Inequalities in diet and physical activity in Europe. Public Health Nutr 4:701–
33 709. <https://doi.org/10.1079/PHN2001160>
34
- 35 Durand C, Sauthier N, Schwoebel V (2011) Évaluation de l'exposition à des sols pollués au
36 plomb, au cadmium et à l'arsenic en Aveyron. Étude Cassiopée (cadmium et arsenic
37 dans les sols : impact observé sur une population exposée). Santé Publique France,
38 France. <https://www.santepubliquefrance.fr/regions/occitanie/documents/rapport-synthese/2011/evaluation-de-l-exposition-a-des-sols-pollues-au-plomb-au-cadmium-et-a-l-arsenic-en-aveyron.-etude-cassiopee-cadmium-et-arsenic-dans-les-sols-i>
39
- 40 Forouzanfar MH, Afshin A, Alexander LT, et al (2016) Global, regional, and national
41 comparative risk assessment of 79 behavioural, environmental and occupational, and
42 metabolic risks or clusters of risks, 1990–2015: a systematic analysis for the Global
43 Burden of Disease Study 2015. The Lancet 388:1659–1724.
44 [https://doi.org/10.1016/S0140-6736\(16\)31679-8](https://doi.org/10.1016/S0140-6736(16)31679-8)
45
- 46 Gami AS, Rader S, Svatikova A, et al (2007) Familial Premature Coronary Artery Disease
47 Mortality and Obstructive Sleep Apnea. Chest 131:118–121.
48 <https://doi.org/10.1378/chest.06-1404>
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

- 1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
- Geller A, Zenick H (2005) Aging and the Environment: A Research Framework. *Environ Health Perspect* 113:1257–1262. <https://doi.org/10.1289/ehp.7569>
- Guihard-Costa A-M, Inácio M, Valente S, Ferreira da Silva E (2012) Méthode d'étude spatialisée des effets de la contamination industrielle sur la santé des populations locales, région d'Estarreja (Portugal). *Sud-Ouest Eur Rev Géographique Pyrén Sud-Ouest* 69–76. <https://doi.org/10.4000/soe.246>
- Hendryx M, Fedorko E, Halverson J (2010) Pollution Sources and Mortality Rates Across Rural-Urban Areas in the United States. *J Rural Health* 26:383–391. <https://doi.org/10.1111/j.1748-0361.2010.00305.x>
- Hunt A, Ferguson J, Hurley F, Searl A (2016) Social Costs of Morbidity Impacts of Air Pollution. OECD
- Inácio M, Neves O, Pereira V (2014) Acumulação de metais pesados em forragens e produtos agrícolas em Podzóis de um local industrial Português. *Comun Geológica* 1019–1022
- Kriegsman DMW, Penninx BWJH, Van Eijk JThM, et al (1996) Self-reports and general practitioner information on the presence of chronic diseases in community dwelling elderly: A study on the accuracy of patients' self-reports and on determinants of inaccuracy. *J Clin Epidemiol* 49:1407–1417. [https://doi.org/10.1016/S0895-4356\(96\)00274-0](https://doi.org/10.1016/S0895-4356(96)00274-0)
- Landrigan PJ, Fuller R, Acosta NJR, et al (2018) The Lancet Commission on pollution and health. *The Lancet* 391:462–512. [https://doi.org/10.1016/S0140-6736\(17\)32345-0](https://doi.org/10.1016/S0140-6736(17)32345-0)
- Levasseur P, Erdlenbruch K, Gramaglia C (2021) Why do people continue to live near polluted sites? Empirical evidence from Southwestern Europe. *Environ Model Assess*
- Li Z, Ma Z, van der Kuijp TJ, et al (2014) A review of soil heavy metal pollution from mines in China: pollution and health risk assessment. *Sci Total Environ* 468–469:843–853. <https://doi.org/10.1016/j.scitotenv.2013.08.090>
- Makri A, Stilianakis NI (2008) Vulnerability to air pollution health effects. *Int J Hyg Environ Health* 211:326–336. <https://doi.org/10.1016/j.ijheh.2007.06.005>
- Montazeri P, Thomsen C, Casas M, et al (2019) Socioeconomic position and exposure to multiple environmental chemical contaminants in six European mother-child cohorts. *Int J Hyg Environ Health* 222:864–872. <https://doi.org/10.1016/j.ijheh.2019.04.002>
- Morelli X, Rieux C, Cyrus J, et al (2016) Air pollution, health and social deprivation: A fine-scale risk assessment. *Environ Res* 147:59–70. <https://doi.org/10.1016/j.envres.2016.01.030>
- Moretti E, Neidell M (2011) Pollution, Health, and Avoidance Behavior: Evidence from the Ports of Los Angeles. *J Hum Resour* 46:154–175
- O'Sullivan JJ, Pearce MS, Parker L (2000) Parental recall of birth weight: how accurate is it? *Arch Dis Child* 82:202–203. <https://doi.org/10.1136/adc.82.3.202>

- 1 Patinha C, Reis AP, Dias AC, et al (2015) The mobility and human oral bioaccessibility of Zn
2 and Pb in urban dusts of Estarreja (N Portugal). *Environ Geochem Health* 37:115–131.
3 <https://doi.org/10.1007/s10653-014-9634-3>
- 4 Pérez-Sirvent C, Hernández-Pérez C, Martínez-Sánchez MJ, et al (2016) Geochemical
5 characterisation of surface waters, topsoils and efflorescences in a historic metal-mining
6 area in Spain. *J Soils Sediments* 16:1238–1252. <https://doi.org/10.1007/s11368-015-1141-3>
- 7
8
9
10 Schlenker W, Walker WR (2016) Airports, Air Pollution, and Contemporaneous Health. *Rev*
11 *Econ Stud* 83:768–809. <https://doi.org/10.1093/restud/rdv043>
- 12
13 Schupp JL, Sharp JS (2012) Exploring the social bases of home gardening. *Agric Hum Values*
14 29:93–105. <https://doi.org/10.1007/s10460-011-9321-2>
- 15
16
17 Sheppard L, Burnett RT, Szpiro AA, et al (2012) Confounding and exposure measurement error
18 in air pollution epidemiology. *Air Qual Atmosphere Health* 5:203–216.
19 <https://doi.org/10.1007/s11869-011-0140-9>
- 20
21
22 Sivry Y, Munoz M, Sappin-Didier V, et al (2010) Multimetallic contamination from Zn-ore
23 smelter: solid speciation and potential mobility in riverine floodbank soils of the upper
24 Lot River (SW France). *Eur J Mineral* 22:679–691
- 25
26
27 Thomas D, Strauss J, Henriques M-H (1991) How Does Mother’s Education Affect Child
28 Height? *J Hum Resour* 26:183–211. <https://doi.org/10.2307/145920>
- 29
30 Vandermoere F (2008) Psychosocial health of residents exposed to soil pollution in a Flemish
31 neighbourhood. *Soc Sci Med* 66:1646–1657.
32 <https://doi.org/10.1016/j.socscimed.2007.12.031>
- 33
34
35 Zuidema T, Nentjes A (1997) Health Damage of Air Pollution: An Estimate of a Dose–
36 Response Relationship for the Netherlands. *Environ Resour Econ* 9:291–308.
37 <https://doi.org/10.1023/A:1026411110952>
- 38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

FIGURES AND TABLES

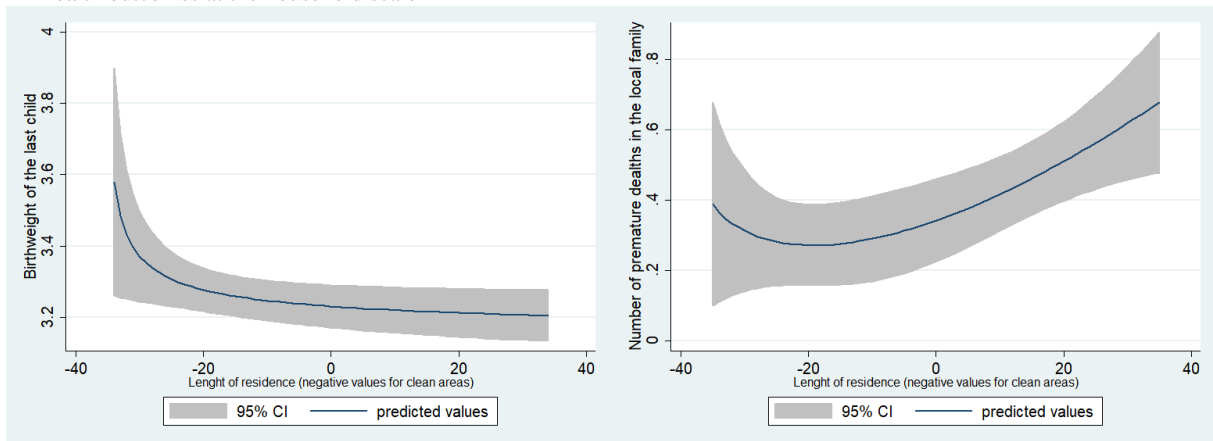
Figure 1: Mapping of polluted and control areas



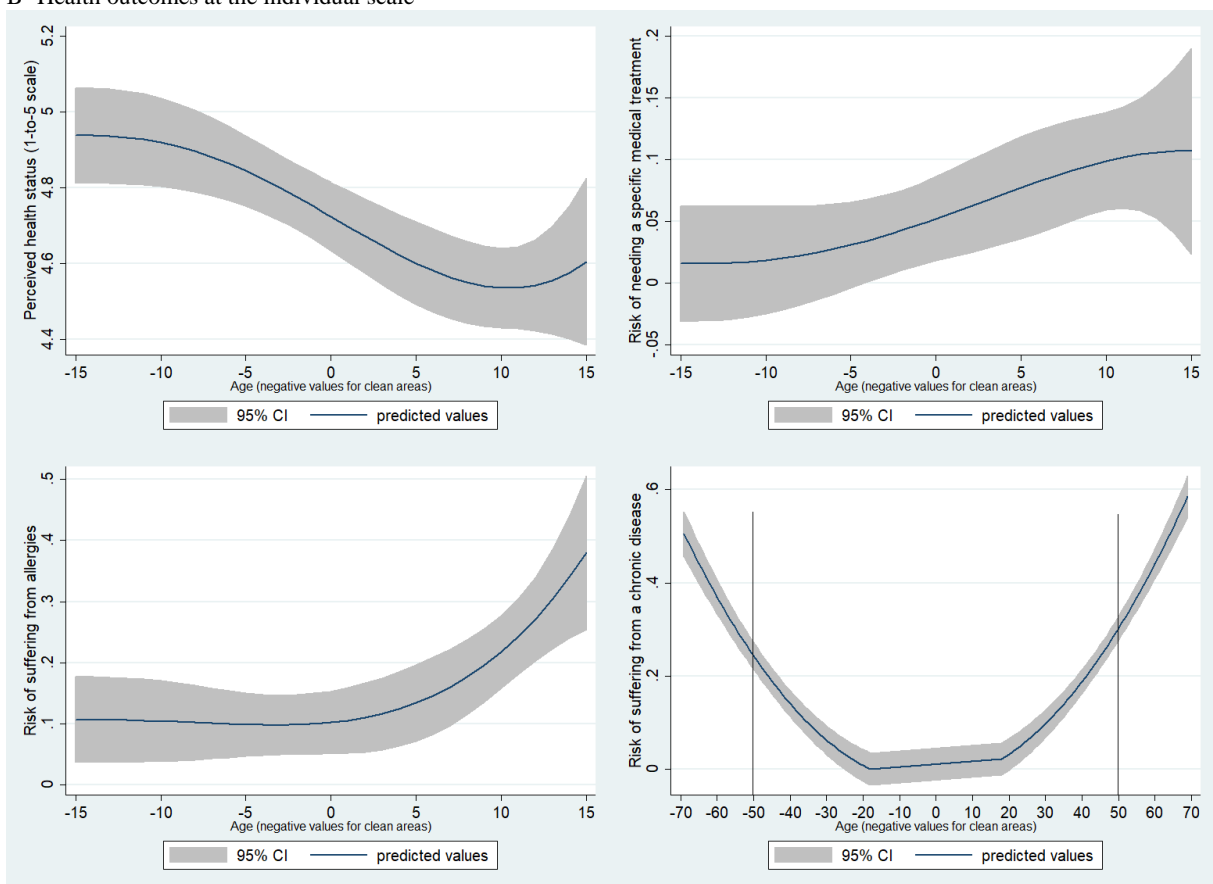
Source: OSM, authors' computations. This map has already been published in Levasseur et al. (2021).

Figure 2: Bivariate polynomial plots between soil pollution exposure and health outcomes

A- Health outcomes at the household scale



B- Health outcomes at the individual scale



Note: Fractional polynomial estimates are used. On the x-axis, negative values refer to control areas and positive values to polluted areas.

Source: Authors' calculations from the CSPE database.

Table 1: Within-group (polluted vs. clean areas) OLS regressions of the respondent's perception of area pollution on health outcomes

DEPENDENT VARIABLE: Level of area pollution perception by the respondent		
	Sample of polluted areas	Sample of control areas
Birthweight of the last child (in kg) during the last semester	-0.045 (-0.46) <i>N</i> =479	-0.133 (-1.30) <i>N</i> =342
Premature death	-0.001 (-0.02) <i>N</i> =595	0.050 (0.76) <i>N</i> =425
Health status perception (0-15yo)	-0.098 (-0.84) <i>N</i> =189	-0.489 (-1.63) <i>N</i> =137
Requiring medical treatment (0-15yo)	-0.056 (-0.14) <i>N</i> =189	0.129 (0.43) <i>N</i> =140
Suffering from allergies (0-15yo)	-0.182 (-0.77) <i>N</i> =189	0.563* (1.74) <i>N</i> =140
Suffering from a chronic disease (all adults)	-0.016 (-0.22) <i>N</i> =1,372	0.008 (0.11) <i>N</i> =1,035
School absences during the last semester	-0.007 (-0.24) <i>N</i> =102	0.055*** (4.90) <i>N</i> =80
Work absences during the last semester	0.001 (0.45) <i>N</i> =632	0.004 (1.36) <i>N</i> =485
Medical visits during the last semester (all adults)	-0.008 (-0.78) <i>N</i> =1,373	0.023** (2.33) <i>N</i> =1,034

Note: Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$ (t-statistics in parenthesis).
Source: CSPE database (2019).

Table 2: Mean-comparison tests between polluted and control areas across health outcomes

	Polluted areas		Control areas		Mean-difference test (p-value)
	Mean	S.E.	Mean	S.E.	
Birthweight of the last child (in kg) during the last semester	3.253	0.024	3.318	0.029	-0.065* (0.089)
Premature deaths in family living locally	0.579	0.042	0.333	0.0368	0.245*** (0.000)
Child health status assessed by the respondent (on a 5-point scale) (0-15yo)	4.587	0.055	4.883	0.033	-0.296*** (0.000)
Probability that a child requires medical treatment (0-15yo)	0.089	0.021	0.021	0.012	0.069*** (0.010)
Probability that a child suffers from allergies (0-15yo)	0.201	0.029	0.100	0.025	0.101** (0.013)
Probability that an adult suffers from a chronic disease (all adults)	0.355	0.013	0.288	0.014	0.067*** (0.001)

Note: Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$ (p-value in parenthesis).
Source: CSPE database (2019).

Table 3: Health effects of living in polluted areas versus control areas (summary results)

ESTIMATOR	OLS	Just-identified IV	Over-identified IV	PSM
<i>Household-based outcome indicators</i>				
Birthweight of the last child (in kg)	-0.085** (-2.08) N=729	-0.173** (-2.19) N=727	-0.225** (-2.43) N=555	-0.082* (-1.75) N=729
Premature death in the family living locally	0.177*** (2.96) N=914	0.255** (2.32) N=911	0.188 ^b (1.49) N=688	0.161** (2.51) N=914
<i>Child-based outcome indicators</i>				
Child's health status assessed on a 5-point Likert scale (0-15yo)	-0.369*** (-3.46) N=228	-0.386** (-2.23) N=228	-0.465** (-2.13) N=196	-0.283*** (-2.57) N=224
Probability that a child requires medical treatment (0-15yo)	0.077* (1.84) N=231	0.113 ^b (1.58) N=231	0.147* (1.82) N=197	0.085* (1.83) N=227
Probability that a child suffers from allergies (0-15yo)	0.154*** (2.75) N=231	0.178** (2.17) N=231	0.196** (2.19) N=197	0.118* (1.73) N=227
School absences during the last semester (kindergarten and primary)	1.036 ^b (1.55) N=135	1.561 ^b (1.55) N=135	2.005* (1.88) N=114	1.489 ^b (1.57) N=111
<i>Adult-based outcome indicators</i>				
Probability that a working-age adult (18-70yo) suffers from a chronic disease	0.044** (1.98) N=1,642	0.077** (1.99) N=1,637	0.061 (1.37) N=1,284	0.030 (1.21) N=1,635
Probability that an active worker suffers from a chronic disease	0.056** (2.30) N=955	0.104*** (2.59) N=953	0.110** (2.24) N=765	0.052* (1.84) N=951
Probability that an elderly person (>70yo) suffers from a chronic disease	0.098* (1.92) N=417	0.169* (1.66) N=417	0.370*** (3.18) N=236	0.118** (2.04) N=395
Medical visits during the last semester for working-age adults (18-70yo)	0.551*** (2.98) N=1,644	0.848** (2.51) N=1,639	0.755* (1.82) N=1,284	0.504*** (2.75) N=1,637
Medical visits during the last semester for active workers	0.389** (2.23) N=957	0.678** (2.08) N=955	0.747** (1.99) N=765	0.309* (1.74) N=953
Medical visits during the last semester for the elderly (>70yo)	1.087*** (2.67) N=415	1.176 (1.22) N=415	1.597 (1.36) N=235	1.310*** (2.88) N=392
Work absences during the last semester (active adults 18-70yo)	0.746 (0.88) N=954	1.294 (1.05) N=952	2.082 (1.18) N=762	1.947 (0.86) N=950

Note: Significance levels: *** p<0.01, ** p<0.05, * p<0.1, ^b p<0.15 (t-statistics are in parenthesis). T-statistics are cluster robust at the household level in OLS and IV estimates. All control variables are included. Country-fixed effects are included. Please refer to the Appendix for complete tables of OLS and IV estimates.

Source: CSPE database (2019).

Table 4: Heterogeneous birthweight and mortality effects across wealth and education (OLS estimates)

DEPENDENT OUTCOME	Birthweight of the last child (in kg)		Number of premature deaths in the local family	
	<i>Linear</i>	<i>Linear</i>	<i>Linear</i>	<i>Linear</i>
<i>Estimator</i>	(1)	(2)	(3)	(4)
Living in a polluted area	-0.057	-0.129***	0.524***	0.057
	(-0.44)	(-2.73)	(2.80)	(0.80)
LivingPollutedArea*WealthIndex	-0.007		-0.097**	
	(-0.24)		(-2.19)	
LivingPollutedArea*LowEducation		0.161*		0.340**
		(1.79)		(2.45)
Length of residence (in years)	0.001	0.001	0.003	0.002
	(0.61)	(0.62)	(1.57)	(1.50)
Age of the respondent (in years)	0.001	0.001	-0.000	-0.000
	(0.55)	(0.57)	(-0.09)	(-0.12)
Gender of the respondent (male)	-0.077	-0.076	-0.018	-0.019
	(-1.32)	(-1.31)	(-0.26)	(-0.26)
Body-mass index of the respondent (in kg/m ²)	0.025***	0.025***	0.008	0.007
	(5.10)	(5.06)	(0.95)	(0.85)
Height of the respondent (in meters)	0.872***	0.880***	-0.279	-0.257
	(2.97)	(3.00)	(-0.66)	(-0.61)
Respondent has foreign origins (dummy)	0.200**	0.202**	-0.064	-0.069
	(2.39)	(2.43)	(-0.69)	(-0.72)
Respondent is in a couple (dummy)	-0.023	-0.023	-0.089	-0.093
	(-0.49)	(-0.49)	(-1.26)	(-1.32)
Respondent's attitude to risk (a 5-point Likert scale)	-0.002	-0.001	0.085***	0.084***
	(-0.08)	(-0.05)	(2.68)	(2.67)
Respondent has a low level of education (dummy)	-0.025	-0.116	0.278***	0.077
	(-0.37)	(-1.39)	(2.90)	(0.80)
Respondent has an intermediate level of education (dummy)	-0.051	-0.043	0.188***	0.194***
	(-1.02)	(-0.86)	(2.75)	(2.83)
Parental education of the respondent (at least high-school level)	-0.032	-0.036	-0.021	-0.037
	(-0.59)	(-0.67)	(-0.28)	(-0.48)
Wealth index (0-to-7 score)	0.003	-0.002	0.018	-0.039
	(0.11)	(-0.09)	(0.62)	(-1.39)
Number of children in household	-0.034	-0.034	-0.013	-0.012
	(-1.13)	(-1.14)	(-0.30)	(-0.28)
Size of dwelling (number of rooms)	0.029*	0.029*	0.082*	0.082*
	(1.76)	(1.76)	(1.77)	(1.77)
Constant	1.065*	1.090**	-0.294	-0.024
	(1.95)	(2.01)	(-0.36)	(-0.03)
Country-fixed effect	Yes	Yes	Yes	Yes
Observations	729	729	914	914
R-squared	0.083	0.087	0.078	0.079

Note: Cluster robust t-statistics in parenthesis at the household level. Country-fixed effects are included. Significance levels: *** p<0.01,

** p<0.05, * p<0.1.

Source: CSPE database (2019).



Click here to access/download

Supplementary Material

APPENDIX_revisedversion_clean.docx

