Journal of Agricultural and Resource Economics: Preprint ISSN: 1068-5502 (Print); 2327-8285 (Online)

doi: 10.22004/ag.econ.338994

Giant Oil Discoveries and Long-Term Health Effects: Evidence from China

Ohad Raveh and Yan Zhang *

Does the discovery and operation of a nearby giant oil field carry long-term health consequences? Capitalizing on the 2011-2018 waves of the China Health and Retirement Longitudinal Study, we find that a giant oil discovery, occurring within a 60km radius, significantly decreases the relative average long-run health conditions of individuals born after it. Specifically, their average share of individuals diagnosed with a chronic disease increases, in relative terms, by 22%. This effect is observed most notably in diseases related to the respiratory, digestive, and urinary systems, and may be driven by changes in the consumption habits of alcohol and tobacco.

Key words: body mass index (BMI), chronic diseases, long-term analysis, resource discoveries

Introduction

What are the long-run health effects of an oil field operating in close proximity? On one hand, it induces local economic development, which improves the health conditions; on the other hand, it introduces a multitude of environmental concerns that deteriorates health.[1] The examination of this health-wealth trade-off, within the context of natural resources, has been of perennial interest to economists and policy makers.[2] The literature, however, largely focused on its relatively short-term dynamics, while giving little attention to its long run implications, although both the resource-triggered health and wealth effects are potentially persistent.[3] Moreover, related, purely cross-sectional analyses, which may be interpreted as examining long-term implications, disregarded potential mobility-driven external influences that may be crucial for identification.[4] Indeed, examining long-term health effects represents a challenging endeavor because it requires accounting for the location of individuals over a prolonged period.

Ohad Raveh is a senior lecturer in the Department of Environmental Economics and Management, Hebrew University of Jerusalem, Israel. Yan Zhang (z0005352@zuel.edu.cn) is a lecturer in School of Economics, Institute of Ecological Civilization, Zhongnan University of Economics and Law, Wuhan, China.
We acknowledge the financial support from the National Natural Science Foundation of China (No. 72303243), and the Fundamental Research Funds for the Central Universities in China (No. 31513110407).

Review coordinated by Christian A. Langpap.

[1] Cust and Poelhekke (2015) summarize the evidence on the generally positive local impact of resource abundance. Okeke and Abubakar (2020) illustrate the potential positive impact of wealth effects on health services and outcomes. On the other hand, Johnston, Lim, and Roh (2019) outline the evidence on the potential negative effects of nearby oil facilities on health, arising through a number of possible channels, including air pollution, water and soil contamination, and increased work-related stress.

[2] See Baum and Tolonen (2021) and references therein for a summary of the related discussion.

[3] The persistence of both effects has been the subject of growing literatures. See Almond Currie, and Duque (2018) for the case of long-term health outcomes, and van der Ploeg (2011) for that of long-term resource outcomes.

[4] See O'Callaghan-Gordo, Orta-Martínez, and Kogevinas (2016), and references therein.

This work fills this gap by capitalizing on the unique features of the China Health and Retirement Longitudinal Study (henceforth, CHARLS), which tracks the location of surveyed individuals over their lifetime. We draw on this survey, and spatially match individuals' place of residence to the location of giant oil fields, which via their discovery and operation represent major, plausibly exogenous, economic shocks that affect both local development and health in the long term. We find that, on average, living in relatively close proximity to such an oil extraction site significantly deteriorates individuals' long-run health conditions, suggesting that the negative health effects of oil discoveries and operation are prominent in the long term, despite inducing economic development.

The CHARLS survey is a comprehensive longitudinal health survey on a representative sample of Chinese residents in continental China, aged 45 and above (Zhao et al., 2014). The survey, taken in five main waves in 2011, 2013, 2014, 2015, and 2018, covers more than 25,000 individuals located in 125 cities, and 28 provinces. Importantly, it offers a wealth of economic and demographic measures at the individual and city levels, including a host of measures on the physical, mental, and cognitive wellbeing of individuals during their adulthood as well as childhood periods. We exploit these rich data in our analysis.

Employing the CHARLS survey is appealing for our analysis for several reasons. First, a key requisite for effectively estimating the long-run health impacts of nearby oil sites is to consider individuals who lived next to them (or otherwise away from them) throughout the examined period. The survey uniquely identifies individuals who lived in more than one location for more than 6 months throughout their lives, thus enabling to focus the analysis on those who did not move for long periods since birth. Notably, given that the survey addresses the elderly, born at times of low population mobility within China, close to 80% of individuals in the sample have never changed locations for long periods. The latter mitigates concerns related to sample selection, given that the survey employs a representative sample of the Chinese elder population, as well as concerns related to excessive short-term travel. Last, the richness of the survey, in terms of both the various health measures and the socio-economic indicators available, enables considering various dimensions of health, and controlling for a host of confounding factors at the individual level, ranging from the health conditions early in life, to subjective happiness, and physical disabilities. Notably, the CHARLS survey has been used previously to examine patterns of long-term health (see, e.g., Cui, Smith, and Zhao, 2020); however, to our best knowledge none exploited the locational features of CHARLS, as well as examined the role of oil fields in this context.

While CHARLS provides both objective and subjective measures of long-term health, which are examined in the analysis, our primary focus is on the former as they are less susceptible to measurement errors. Specifically, we consider the diagnoses of 14 major chronic diseases, and individuals' weight status at the time of the survey, measured via the World Health Organization's Body Mass Index (BMI) measure. Both measures rely on objective and consistent diagnoses, and reflect health conditions of long-term nature. To examine the former, we compose an individual-level *Health Index* which measures the extent to which an individual is diagnosed with chronic diseases. In effect, this index is an average of the (binary) existence of each of the considered chronic diseases. Importantly, the timing of the earliest diagnoses of each of these chronic diseases may be earlier than the survey year; however, the survey reports the timing of the earliest diagnoses for about half of the individuals in the sample. We exploit these data to show that the earliest diagnoses exhibit long-term characteristics, despite occurring pre-survey in various cases, and also control for them directly in the analysis.

We consider giant oil discoveries as natural experiments of major positive local economic shocks. In particular, we employ data from Horn (2014) on the location and timing of giant oil field discoveries.[5] Horn (2014) defines a giant oil field to be one for which the estimate of ultimately recoverable oil is at least 500 million bbl of oil or gas equivalent. As such, these oil fields provide extraordinarily large potential profits. Based on that, we follow Arezki, Ramey, and Sheng (2017), Lei and Michaels (2014), and Perez-Sebastian and Raveh (2016b), among others, and regard their discovery, development, and exploitation to be plausibly exogenous. In addition, their size and scope are expected to induce a sizeable impact on individuals' physical and mental wellbeing in the long-term, either through their effect on the environment, or via their impact on local economic development. We focus on the discoveries made in China, within our sample period. Up to 2018 there were 37 giant oil discoveries across China, with the first being in 1938, and the last in 2006.[6]

We thus seek to identify and estimate the impact of giant oil fields on health outcomes measured many years, approximately decades, after the occurrence of discoveries; i.e., long-term health. To do so, we spatially match the location of individuals in the survey to the giant oil discovery made closest to them,[7] and adopt a within-city difference-in-differences setup in which the treated group includes the individuals who live in relatively close proximity to their closest giant oil field,[8] and the treatment is the discovery itself.

Focusing our analysis on those individuals who, to a large extent, never moved (approximately 21,000 individuals), and considering their year of birth, the treatment (discovery) induces variation with respect to the extent to which individuals were exposed to the effects of the nearby giant oil field. Our long-term approach requires considering exposure time, relative to non-exposure time, thus adopting a relative perspective. Individuals born after the discovery were exposed to its effects their whole life, whereas individuals born prior to it were exposed to its effects only a fraction of their life, in a magnitude that decreases with age. Our main conjecture in the analysis is that the long-term health effects of giant oil fields intensify with the *relative* extent to which individuals are exposed to it. The within-city approach, which controls for fixed cross-city differences (including in income, technology, and pollution, among others), enables focusing the analysis on the variation in birth-distance from discovery which we exploit to identify and estimate the impact of oil discoveries on long-term health conditions.

Our identification strategy rests on three key features. First, the plausible exogeneity of the discovery and exploitation of giant oil fields. Second, the unique feature of the CHARLS survey which crucially enables focusing the analysis on individuals who did not move locations for considerable amounts of time throughout their lifetime. Last, common pre-trends in our identifying variation. As noted, our identification is derived from examining the difference in the health outcome differences between those born before and after a discovery, across proximities. As we illustrate in the analysis, the differences across proximities are not observed when examining the individuals born prior to their closest discovery; conversely, within the post-discovery cases the differences become evident, as we

---

[5] Horn (2014) provides data on the geolocation and timing of all giant oil field discoveries from the 19th century and throughout our sample period.

[6] The earliest discovery, made in 1938, was not in close proximity to any of the cities in the sample. Conversely, the earliest such discovery was made in 1964. Assuming that giant oil fields remain active for 50 years on average, as reported by Horn (2014), this suggests that in all the cases in our sample in which a giant oil field was discovered in relatively close proximity to a city, it remained active until the time the CHARLS survey was conducted.

[7] We do so irrespective of the timing of the discovery. Notably, due to the distribution of discoveries across China, each individual in the sample has one clearly identifiable closest discovery, with the second-closest one being no less than a 100km away from the closest one, in all cases.

[8] We define close proximity to be within a 60km radius. This definition is based on our spatial estimations presented initially in the empirical analysis section, and it is generally in line with estimates provided previously in the literature (e.g., Aragon and Rud, 2013; De Haas and Poelhekke, 2019).

observe a deterioration in the long-term health conditions of individuals living in close proximity to a giant oil field.

Our main finding is that living in close proximity to a giant oil field significantly deteriorates the long-term health conditions, despite the positive, general (i.e., not individual-specific) wealth effect. Specifically, we find that individuals who have a giant oil field located nearby, and have been born after its discovery, experience worse long-term health conditions related to the respiratory, digestive, and urinary systems, although they are younger, and on average (i.e., across all proximities) are happier, smoke less, and consume fewer alcoholic beverages. The magnitude is large. A giant oil field discovery induces an increase of approximately 22% in the average share of individuals born after the discovery who are diagnosed with chronic diseases, compared to those born before the discovery. We find that this result is robust to a range of tests, and may be driven by changes in the consumption behavior of alcohol and tobacco goods, calling for corresponding mitigating policy measures in regions located near operating giant oil fields.

The paper is structured as follows. The next section reviews the related literature and places the current contribution within it. The data and empirical design are outlined in Section 3. Section 4 presents the main results, and Section 5 covers the additional robustness tests. Section 6 offers concluding remarks, and the appendices present data and technical details.

Related literature

The paper is related to number of literature strands. First is the literature on the long-term effects of resource booms on development and economic growth. Economists have long noticed that natural resource abundance can turn out to be a blessing as well as a curse. This literature is surveyed by van der Ploeg and Poelhekke (2017), van der Ploeg (2011), Venables (2016), and Cust and Poelhekke (2015). The latter focus on the local effects, which as the evidence they summarize indicate, have been shown to be generally positive and economically significant (e.g., Allcott and Keniston, 2018; Perez-Sebastian and Raveh, 2016a; Raveh, 2013). The majority of empirical evidence in this literature focuses on national, regional, and more recently firm-level data. Little attention, however, was given to individual-level analysis, due to data limitations.[9] We contribute to this literature by first, highlighting a new potential transmission channel, namely long-term health conditions, and second, via an adoption of a micro-level perspective that examines individual-level outcomes.

Second is the literature on the health outcomes of resource booms. Studies that examine resource-triggered health outcomes, including empirical evidence for their sign and magnitude, are surveyed by Baum and Tolonen (2021), and Johnston, Lim, and Roh (2019). Notably, empirical evidence so far focused either on short-term effects, under which contemporaneous correlates between health outcomes and resource booms are examined (Cockx and Francken, 2014), or on cross-sectional evidence that compares between outcomes observed in resource rich and poor areas while disregarding mobility (O'Callaghan-Gordo, Orta-Martínez, and Kogevinas, 2016). Under either perspective, the majority of evidence point at a negative impact on various health factors that range from physical to mental conditions. The present effort contributes to this literature by first, estimating the impact on long-term health, and illustrating that it is also overall negative; second, doing so while crucially accounting for mobility; and third, by examining a central health outcome that have been largely overlooked in these previous studies, namely the Body Mass Index.

Third is the literature on the impact of giant oil discoveries and exploitation. There have been a number of recent empirical studies that examined the impact of giant oil fields on

[9] An exception is Jacobsen, Parker, and Winikoff (2023), who examined the Resource Curse Hypothesis via the lens of individual-level data.

Table 1. Descriptive Statistics

Variable	Mean	Sd	Min	Max	Variable	Mean	Sd	Min	Max
Overweight	0.312	0.463	0	1	Stroke	1.93	0.259	1	2
Health	1.78	0.36	1.07	2	Kidney Disease	1.9	0.303	1	2
Disability	1.82	0.295	1.2	2	Stomach Disease	1.69	0.461	1	2
Mental health	1.91	0.33	1	2	Memory-related Disease	1.96	0.199	1	2
Childhood health	0.352	0.477	0	1	Arthritis	1.6	0.49	1	2
Smoke	0.415	0.493	0	1	Asthma	1.94	0.239	1	2
Alcohol	0.142	0.349	0	1	Liver disease	1.93	0.255	1	2
Exercise	0.373	0.484	0	1	Before	0.797	0.402	0	1
Insurance	0.96	0.196	0	1	Close	0.06	0.245	0	1
Vaccinations	0.857	0.35	0	1	Earliest diagnosis	2,008	12.9	1,900	2,018
Emotional condition	1.97	0.178	1	2	Size of discovery	839.9	530.75	500	4823
Hypertension	1.61	0.488	1	2	Gender	0.471	0.499	0	1
Dyslipidemia	1.78	0.414	1	2	Education	3.45	1.97	1	10
Diabetes	1.87	0.335	1	2	Marriage status	1.65	1.41	1	6
Cancer	1.98	0.149	1	2	Siblings	0.118	0.323	0	1
Chronic lung disease	1.85	0.362	1	2	Age	61.7	10.6	18	108
Heart disease	1.8	0.4	1	2	Wage	25,604	43,276	0	50,000

Notes: See Variable Definitions for detailed description of variables.

various outcomes, including the current account (Arezki, Ramey, and Sheng, 2017), conflicts (Lei and Michaels, 2014), exchange rates (Harding, Stefanski, and Toews, 2020), congressional voting (Perez-Sebastian and Raveh, 2019), tariff patterns (Perez-Sebastian, Raveh, and van der Ploeg, 2021), and fiscal decentralization (Perez-Sebastian and Raveh, 2016b). Our analysis examines their impact on a new outcome, namely long-term health, illustrating that giant oil fields may yield long-term impacts.

Last is the vast literature on the long-term health effects of early life shocks. Related studies examine the impact of various shocks during childhood, including disease outbreaks (Almond, 2006), famine (Almond et al., 2010), rare natural disasters (Almond, Edlund, and Palme, 2009), and economic policies (Hoynes, Schanzenbach, and Almond, 2016), on health outcomes later in life. Almond, Currie, and Duque (2018) provided a synthesis of this literature. A central theme within this literature relates to the Fetal Origins Hypothesis which examines the long-term effects, health related in particular, of shocks that occur while in-utero (see Almond and Currie, 2011) for a review of the related evidence). In this study we examine the long-term health effects of a new factor: giant oil fields, and we go beyond the Fetal Origins Hypothesis by considering the long-term impacts of consistent exposure from birth and over the life time.

Data and Empirical Design

We seek to estimate the long-term health impacts of giant oil discoveries. Hence, our analysis utilizes two primary measures, namely giant oil discoveries and health measures at the individual level, in addition to further individual, and regional covariates. Next, we discuss each of these components in more detail, and describe the empirical setup we adopt. Further details are provided in the data Appendix. Descriptive statistics for the key variables are

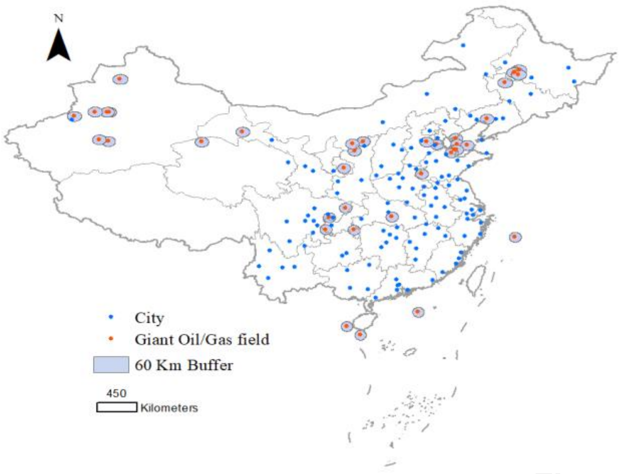

Figure 1. The Distribution of the Giant Oil or Gas Fields and Cities, within 60 KM Radius
Notes: The figure shows the distribution of our sample's 37 giant oil or gas fields (red dots) in China and 125 Chinese cities (blue dots) covered in the CHARLS survey. The grey shaded areas present the zones with a 60 Km radius around the giant oil or gas fields.

provided in Table 1 for the full sample, and in Appendix Table A1 for proximity and birth related sub-samples respectively, as we further describe below.

Giant Oil Discoveries

Starting with some related background on the nature of oil and gas development and policy in China, throughout most of our sample period China has been importing oil and gas to meet the high local demand, despite being considered relatively abundant in oil and gas reserves.[10] This led to large investments in local oil exploration and production, as well as adoption of related technologies (Wu, 2014). Consequently, the oil and gas sector steadily grew over recent decades. More specifically, data from the National Bureau of Statistics of China (NBSC) indicate that over the post-2000 period oil and gas production increased by almost 8-fold, and the labor share of the oil and gas apparatus, within the larger mining sector, averaged at 11%. These patterns yield major implications for local economies near exploited reserves. Large discoveries have been shown to be a catalyst of local development due to new infrastructure, incoming capital, and job creation (e.g., Aragon and Rud, 2013), especially as the latter is directed at local populous given that its demand for specialist, skilled workforce is relatively low (Raveh and Reshef, 2016). In the context of the Chinese economy, local development is also manifested via increases in the local fiscal budgets, as local governments accrue resource taxation and utilization fees. Indeed, as the NBSC reports, indirect resource income takes an average share of 3% in the fiscal budgets of local governments in recent years. Hence, we examine giant oil discoveries in China, regarded as natural experiments of major positive local economic shocks.

We employ data from Horn (2014) on discoveries of giant oil and gas fields, which are defined to be fields for which the estimate of ultimately recoverable oil is at least 500 million barrels of oil or gas equivalent. Horn (2014) reports that oil production commences, on average, five years following the discovery. Production can then spread over a period of approximately

[10] Reserves are generally spread across the country, but are most notably located in western China, in the Tarim, Ordos, and Sichuan basins. Regionally, the provinces of Xinjiang, Gansu, Shaanxi, Heilongjiang and Hebei have relatively larger oil and gas sectors.

50 years.[11] Our focus is on the discoveries that occurred within China. The earliest discovery in the sample occurred in 1938, and the last one in 2006; however, the majority of discoveries located next to cities in the sample occurred during the 1960s and 1970s. Altogether, 37 discoveries were made in China during our sample period, 1938–2018, four of which being offshore.[12] The distribution of these discoveries across China is plotted in Figure 1.

Employing data on the discoveries of giant oil and gas fields is appealing for our purpose for various reasons. First, it is plausible that their timing is exogenous due to the uncertain nature of oil exploration success. The return on oil exploration is uncertain because of the relatively limited (ex-ante) knowledge of geological features of exploration locations. This feature represents one of the key aspects of our identification strategy.[13] Second, giant oil fields provide a significant trigger for economic development, especially at the local level, due to being a major potential source of oil revenues and profits.[14] The latter also incentivizes their continuous operation irrespective of the local environment, thus suggesting that once discovered their further development and exploitation are likely to be exogenous to the local institutional and economic setting.[15] Third, giant oil field discoveries are expected to have a sizeable impact on the physical and mental health conditions of residents living in close proximity for a multitude of potential reasons that could induce either a positive or negative impact, ranging from air pollution and increased work-related stress to improved economic conditions and infrastructure (see, e.g., von der Goltz and Barnwal, 2018). Importantly, these effects may have long term implications, or rather may sink in, and hence be observed, only in the long term, namely our focus in the analysis to follow.

The CHARLS Survey

We rely on data from the China Health and Retirement Longitudinal Study (CHARLS), a rich biennial longitudinal health survey on a representative sample of Chinese residents in continental China aged 45 and above as well as their spouses (of any age).[16] The first wave of CHARLS was conducted in 2011, following by further waves in 2013, 2015, and finally 2018; an additional wave of life history was undertaken in 2014.

In each wave respondents answered a detailed questionnaire related to their health, personal, and demographic circumstances, ranging from the existence of chronic diseases, weight, and mental health, to fitness, physical disabilities, wages, and income. In addition, the special life history wave, conducted in 2014, includes additional questions related to early life conditions, health during childhood, income and employment, as well as housing characteristics.[17] Our analysis focuses on objective health outcomes, such as BMI and the existence of chronic

[11] While our focus in the analysis is on the discoveries, due to their plausibly exogenous nature (as we further outline in the related section), we follow previous related studies (e.g., Arezki, Ramey, and Sheng, 2017) and also examine the distinction between discovery and production within our context.

[12] The giant oil discoveries data is available up to 2012; however, data from the statistical yearbooks of natural resources in China reveals that up to 2018 no further giant oil discoveries were made in China.

[13] In effect, we follow the approach adopted in previous related studies, including Arezki, Ramey, and Sheng (2017), Lei and Michaels (2014), Harding, Stefanski, and Toews (2020), and Perez-Sebastian, Raveh, and van der Ploeg (2021) among others.

[14] For instance, Arezki, Ramey, and Sheng (2017) report that the mean value of the GDP share of the net present value of a giant oil field discovery is about 67% and can get higher than 6000% (the case of Qatar). A more concrete example for the scope of local development induced by giant oil fields is provided by Toews and Vezina (2022).

[15] As argued also in various previous related studies, including Perez-Sebastian and Raveh (2019), among others.

[16] The CHARLS data is maintained at the National School of Development at Peking University, and is available at http://charls.pku.edu.cn.

[17] See Zhao et al. (2014) for further details.

diseases, as they enable making clearer inference; however, we exploit the full extent of this data, by examining additional more subjective health indicators, childhood conditions, physical disabilities, and economic factors.

Albeit being a longitudinal study, these waves do not represent identical samples; while some of the surveyed individuals were excluded in later waves, others were added.[18] Therefore, we exploit the full extent of this survey by merging the data over all waves to create a complete sample of individuals. Since each individual appears only once in our panel data set, and because we seek to consider the longest period available, individuals appear in our data in the most recent wave in which they were involved.[19] Altogether, our sample covers 20,722 individuals across 125 cities.[20]

A unique feature of the CHARLS survey is that it enables tracking the (general) location of respondents throughout their lifetime; specifically, the survey reports whether individuals were born in their current location (i.e., during the survey year), and whether they lived in a different location for more than six months throughout their lifetime. This feature, therefore, enables identifying individuals who, to a large extent, never moved since birth. Being able to do so is central to our analysis, which seeks to estimate the long-term health impacts of a local shock (oil discovery), and hence requires considering individuals who generally remained in the same location during the examined period. Notably, the vast majority of respondents in the survey did not move locations throughout their life[21]; we restrict the analysis to these individuals.[22]

The CHARLS survey has been used widely to study patterns of long-term health (see, e.g., Kim, Fleisher, and Sun, 2017; Cui, Smith, and Zhao, 2020). It is appealing for this purpose due to its wide coverage of current and early health related as well as other personal measures, in conjunction with tracking respondents' (general) location over time. In addition, its focus on the middle-age and older population provides an adequate sample to study long-term health impacts (Kim, Fleisher, and Sun, 2017). An additional advantage related to the examination of these population segments in China is that their internal and international mobility, in the periods considered, is low in general (see Shi et al., 2020, and references therein). Hence, it is not only the case that about 80% of the individuals sampled in CHARLS have never lived in a different location for more than 6 months, but also that within these 80% short-term travelling is not a common feature, thus minimizing outside effects. We exploit these features in the analysis undertaken next.

Methodology

Our main objective is to estimate the long-term health effects of giant oil discoveries and their operation thereafter. To do so, we employ a standard difference-in-differences approach in which the treated group is represented by the individuals who live in close proximity to a giant oil discovery (that then becomes an active giant oil field), which represents the treatment.[23]

---

[18] Panel attrition in the CHARLS sample is relatively low, about 13% on average. New individuals were then added in each wave to balance this attrition.

[19] As an example, individuals included in the 2011 and 2013 waves, and excluded from the later ones, appear in our panel data within the 2013 wave.

[20] A city may also be referred to in the survey as a county or district; however, throughout the paper we adopt the term city for convenience.

[21] Specifically, 20,722 individuals out of 25,949 respondents.

[22] However, we do show in a later part of the analysis that the main results hold also for the unrestricted sample, and we consider further immigration-related implications.

[23] We consider all oil fields that fall under the 'giant' definition outlined earlier. In that sense, we do not account for additional differences in size (measured via estimates of ultimately recoverable oil barrels), to minimize potential measurement errors. Nonetheless, in a later section we also consider variations in size, and show that the main results are robust to this addition.

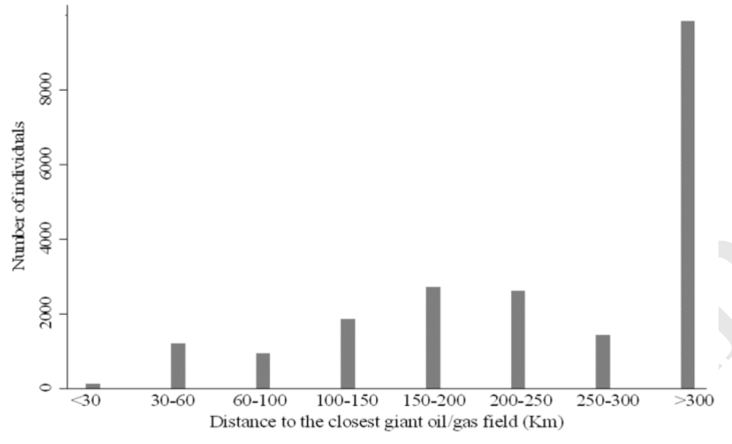

Figure 2. The Distribution of Individuals Across the Proximity Dimension
Notes: The figure presents the distribution of individuals in the sample across their distance from their closest active giant oil field. This distance ranges from about 20km to close to 800km.

We compare between the differences in the effects of the treatment on individuals born before and after a discovery across the groups of individuals living relatively close-to and far-from a discovery, conditional on remaining in the same residential location since birth.[24]

We measure proximity as follows. Using the geo-location of discoveries, we compute the distance of the closest discovery to each of the individuals in our sample, irrespective of its timing. We consider an individual to live in close proximity to a discovery if the computed distance is less than 60km.[25] We consider a threshold of 60km for a number of reasons. First, it is consistent with our baseline spatial findings, discussed in the next section, which indicate that the spatial impacts of giant oil discoveries are observed over a range of approximately 60km. Second, it is generally consistent with findings of previous studies on the spatial effects of natural resources (e.g., Aragon and Rud, 2013). Third, it represents a reasonable match to the spatial distribution of discoveries and cities in our sample.[26] Out of the 20,722 individuals in our sample, 1,330 live within the 60km threshold. The distribution of individuals across the proximity dimension is plotted in Figure 2, ranging from about 20km to close to 800km. Table A1 presents descriptive statistics of the main variables across the two proximity groups; notably, this division does not yield major observable differences in the basic statistical properties of the key variables.

Our treatment, namely a giant oil field discovery, divides the sample of individuals within each city to those born before and after it. Fixing mobility, this division provides variation in the extent to which individuals are exposed to the effects of a nearby giant oil field. While those born after a discovery are exposed to the effects of a giant oil field their whole life (up

[24] As noted, this is defined by the CHARLS survey as not living in a different location for more than 6 months.
[25] We adopt a binary difference based on the assertion that the variation in distance in close and large proximities is not informative; i.e., the difference in health outcomes of individuals living 200km from a discovery to those living 300km from one is not expected to differentiate due to the differences in distance, as the distance is already sufficiently large to yield any effect.
[26] Within the 60km range the differences in distance to the closest discovery across cities is relatively small; however, above this threshold there is a noticeable jump in the distance of the next group of cities. In addition, within this range it is possible to match a single discovery to each individual who lives in close proximity, mitigating concerns related to multiple close discoveries at different periods.

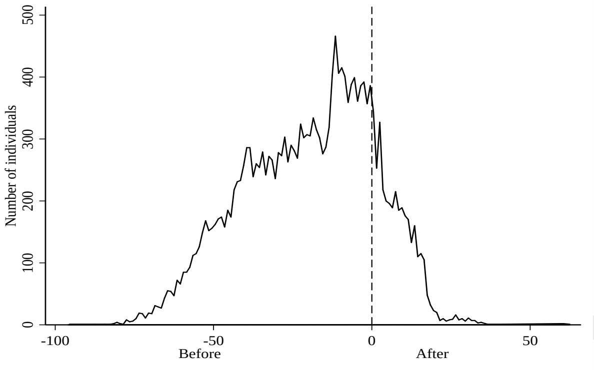

Figure 3. Distribution of Individuals Born Before and After Discovery
Notes: The figure plots the distribution of individuals across the discovery threshold (marked by the dashed line, at 0), ranging from 96 years prior to a discovery up to 63 years following one.

to the survey year),[27] those born before it are exposed to these impacts only a fraction of their life. Importantly, this fraction decreases with age. We conjecture that the long-term health impacts of a giant oil field, located in close proximity, increases with the *relative* extent to which individuals are exposed to it.[28] Adopting a relative approach, we focus on this fraction, rather than on the absolute exposure years, because we aim to estimate long-term impacts, over the span of a lifetime, and hence in effect compare between the extent of exposure years to that of non-exposure years, with the shift between the two states being marked by the discovery.[29]

The distribution of individuals across this fraction, observed via the difference between the birth and discovery years, is plotted in Figure 3. The range is quite wide. Our sample includes one individual who was born 96 years prior to its closest discovery, and another who was born 63 years after one; however, as observed in Figure 3, the majority of individuals in the sample were born within the range of 50 years prior to a discovery, to about 15 years following one. Altogether, 4,200 individuals, out of our 20,722, were born after the discovery made closest to them. Table A1 provides descriptive statistics along this dimension, noting minor differences in the averages of the key variables across the groups.

To that end, we estimate different versions of the following baseline specification, for individual i, living at city c, and at survey wave w:[30]

(1) $$O_{i,c,w} = \alpha + \beta\left(after\right)_i + \gamma\left(after * close\right)_{i,c} + \boldsymbol{\delta}\left(\mathbf{X}\right)_i + \upsilon_c + \eta_w + \varepsilon_{i,c,w}$$

where O is the health-related outcome variable (outlined below), *after* is a dummy variable that captures the individuals born before the discovery made closest to them, *close* is a dummy that captures the cities that are regarded as being close to a giant oil discovery (i.e., within

[27] With the exception of the earliest discovery in the sample (1938), which is not within close proximity to any of the cities in the sample, this applies to all the discoveries in the sample, based on the reported observation that once production starts (about 5 years post-discovery) extraction is spread over a period of approximately 50 years.
[28] Nonetheless, in the analysis we primarily consider the binary distinction between fractional exposure and full exposure, via a discovery-based division, rather than a continuous one, while controlling for age. The latter enables focusing the estimation on the key identifying variation induced specifically by the discovery.
[29] Notice that with age exposure years increase, yet non-exposure years increase as well. Hence, to examine the long-term impacts of giant oil discoveries we need to consider the relative standing of each state (exposure vs. non-exposure), and thus follow a relative perspective.
[30] Where $i \in [1, 20722]$, $c \in [1, 125]$, and $w \in \{2011, 2013, 2015, 2018\}$.

the 60km range),[31] **X** is a vector of controls (outlined below), and υ_c and η_w are city and wave fixed effects, respectively. The latter control for wave-specific differences such as within-year price or fiscal shocks that may affect wave quality, as well as city-specific differences such as pollution levels or discovery and extraction technologies employed (recall each city is assigned one giant oil field, which is closest to it). Standard errors are clustered by city. Our coefficient of interest is γ. The latter provides the discovery-triggered difference in the differences in the outcome variable between the treated and control groups. Importantly, the health-related outcome variable is reported at the survey year, hence γ represents an estimate of the long-term health impact of the treatment, namely a nearby giant oil discovery that turns to an oil-producing giant oil field.

Baseline outcome and control variables

We examine five main outcome variables, each representing a major health category, which together cover the wide range of health measures provided in CHARLS. Specifically, we examine the existence of chronic diseases, the BMI, and the extent of physical disabilities, mental health, and health during childhood. While the analysis examines each, our focus is on the first two. We maintain this focus primarily because, as we further explain below, unlike the remaining variables the first two represent objective outcomes of long-term nature, which in turn provide clearer inferences concerning the impact on long-term health. Next, we outline the details of each variable. Further details concerning the components and descriptions of each variable, including the additional co-variates outlined below, are provided in the Data Appendix.

Our first, and main outcome variable is what we term the *Health Index*, namely an index that measures the extent to which an individual is diagnosed with chronic diseases. The CHARLS survey covers diagnoses over 14 major chronic diseases, namely hypertension, dyslipidemia, diabetes, cancer, stroke, arthritis, asthma, as well as chronic diseases related to the lungs, liver, heart, kidney, stomach, psychiatric condition, and finally memory-related condition. The diagnoses of each type of chronic disease are done externally using objective, standard medical procedures, and are reported in CHARLS as a discrete variable that indicates the positive or negative diagnoses.[32] The *Health Index* we construct is then a simple average of these reports across the range of chronic diseases, in which a higher value indicates greater positive diagnoses, i.e., a worse health condition.

A key aspect of the diagnoses that compose this index is the timing of the earliest diagnosis, which is the relevant outcome for our purposes. Although reported in the survey we examine, the earliest diagnoses may have been undertaken prior to the survey year, and hence may not necessarily be indicative of long-term impacts. We address this concern by observing the year of the earliest diagnosis directly, which is available for about half of the individuals in the survey. Figure 4 plots the distribution of the earliest diagnoses across the age dimension (left figure), and the discovery-threshold dimension (right figure). The patterns observed in the latter distribution demonstrate that the majority of the earliest diagnoses occurred post-discovery, most notably within the range of 10 to 60 years past the discovery,[33] while those observed in the former distribution indicate that most of the earliest diagnoses occur between the ages of 40-75. Put together, these patterns suggest that the treatment's impact on the

[31] Notably, *close* is absorbed by the city fixed effects, and hence not added separately in equation (1).

[32] In all cases a value of 1 (0) indicates positive diagnosis (no disease).

[33] It is also quite evident, more generally, that the discovery represents a discontinuity with respect to the number of positive diagnoses as they start increasing shortly after the discovery, supporting its role as a viable treatment.

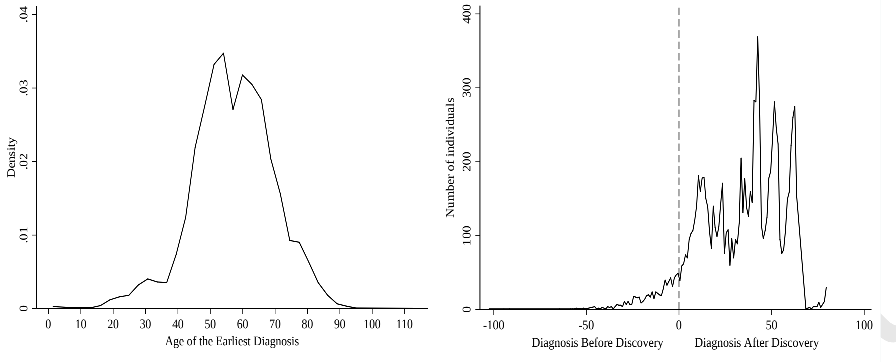

Figure 4. The Age Distribution of the Earliest Diagnosis
Notes: The left figure plots the distribution of the earliest diagnoses across the age dimension (i.e., the age at which the earliest diagnosis occurred); the right figure plots the distribution of the earliest diagnosis across the discovery-threshold dimension (i.e., the earliest diagnosis relative to the time of discovery, marked by the dashed line, at 0).

affected population (individuals born post-discovery) is indeed of long-term nature, despite having diagnoses prior to the survey year.[34]

Second, the BMI, computed as the body mass divided by the square of the body height, and measured at the survey year. The BMI provides a convenient rule of thumb used to broadly categorize individuals as being overweight, namely having a BMI of 25 and above. Being overweight, in turn, is associated with worse health conditions and higher mortality (e.g., Whitlock et al., 2009). To consider overweight status more clearly, we adopt a binary variable that indicates whether the individual is categorized under the said overweight BMI-based definition. We examine this measure separately because it provides a broad indication for individuals' health condition based on factors that are measured consistently across individuals, mass and height, and are distinct from the chronic diseases captured in the *Health Index*.

Third, physical and mental disabilities, including intellectual, vision, and hearing disabilities, as well as speech impediments, and more general physical disabilities. Individuals report having one or more of these disabilities, and we examine this discrete variable, which takes the value of 0 in case of no disabilities, and 1 otherwise. Notably, the roots of these disabilities often date back to birth, making their interpretation as a long-term outcome less clear; hence, the analysis does not go beyond an initial consideration of it.

Fourth, mental health. We construct a mental health index, which is an average of the self-reported mental conditions related to depression, mental intactness, and episodic memory. Similarly, we also examine a self-reported measure on the health status during childhood. Nonetheless, being based on self-reported factors, which may be prone to measurement errors, these measures do not take the focus throughout the analysis.

Finally, we also consider various co-variates, included in each specification within the vector $\mathbf{X}$ in equation (1). The baseline group of controls, included in all specifications, includes various standard individual-level demographic and socio-economic measures that have been shown to be associated with health status, including gender, education level, marriage status, number of siblings, wage level, and finally age (see, e.g., Folland, Goodman, and Stano, 2016; Phelps, 2017).[35] Given the latter's relative importance to the analysis, and

[34] In addition, in later parts of the analysis we also control for year of the earliest diagnosis and demonstrate that the main patterns are robust to this examination.
[35] Notably, each may also operate as a potential transmission channel for the effects of oil discoveries on long-term health. We note, however, that the main results are robust to their exclusion. In addition, we

its potential for inducing non-linear effects, we also include age-squared, as well as age-cubed measures. Additional, case-specific co-variates are also considered; we discuss their details within the analysis.[36]

Identification strategy

The identification of the impact of giant oil discoveries and production on long-term health is based on various factors. First, the appealing characteristics of the treatment; specifically, as noted, its plausible exogenous nature with respect to both giant oil discoveries, as well as their exploitation and production over time (e.g., Arezki, Ramey, and Sheng, 2017; Perez-Sebastian and Raveh, 2019). Second, the unique feature of the CHARLS survey, that enables tracking the location of individuals over their lifetime. This is a particularly central feature for identification since it ensures that we examine individuals who consistently stayed in close proximity to an active giant oil field, and hence were not exposed to other potential long-term health-affecting factors elsewhere.

Third, the identifying variation we examine. We adopt a within-city difference-in-differences setup. The within-city approach, provided via the city fixed effects, ensures that location-based differences in pollution, income, development, technology (in particular that related to oil search and extraction), genetics, and other related factors are, importantly, controlled for. Consequently, it focuses our identifying variation on the difference in the health outcome differences between individuals born before and after a discovery within the same city, across proximities to the closest giant oil field.

The validity of this identifying variation is not only based on the exogenous nature of the treatment, and the immobility of the individuals considered, but also on two additional factors. First, on the assertion that lengthier relative exposure, measured as a fraction of time since birth, to a nearby giant oil field is associated with a stronger impact of it over the long term. Second, on having common pre-trends; i.e., trends of similar nature among those born prior to discoveries across different proximity levels.

Figures 5 and 6 presents these trends. In Figure 5, the left (right) graph examines the BMI (*Health Index*). In each, the dots marked by a *x* (*o*) represent the average value of those living far from (close to) an active giant oil field, conditional on the baseline set of controls noted above. In effect, we take the average of the residuals from regressing each of the outcome variables on the baseline set of controls. The X-axis presents age-bins, separated by the threshold at 0, at which a discovery was made; i.e., the X-axis groups individuals by birth-distance from the closest discovery, dividing them to those born before and after it. The patterns in both cases show common trends amongst those born before a discovery, which then diverge across proximities in the cases of individuals born after a discovery, thus supporting the validity of the variation we examine.

Figure 6 addresses the concern that the outcome variables examined in Figure 5 are, in effect, post-treatment by construction. It does so by considering our measure of childhood health (outlined in the Data Appendix), which measures health conditions up to age 15, while restricting the sample to individuals born up to 15 years prior to their closest discovery. This restriction ensures that the outcome examined in this case is pre-treatment for all individuals (i.e., we consider health conditions prior to an occurrence of a discovery). Similar to Figure 5, we plot the average values, across distance proximities, conditional on the baseline set of controls. The pattern illustrates similarities across all age-bins, pointing at a common trend.

examine their role, together with that of additional co-variates, as potential intermediate channels in a separate sub-section in the analysis.

[36] Note that the number of observations differ across estimations throughout the analysis, depending on the availability of data across the outcome variables, and co-variates, examined.

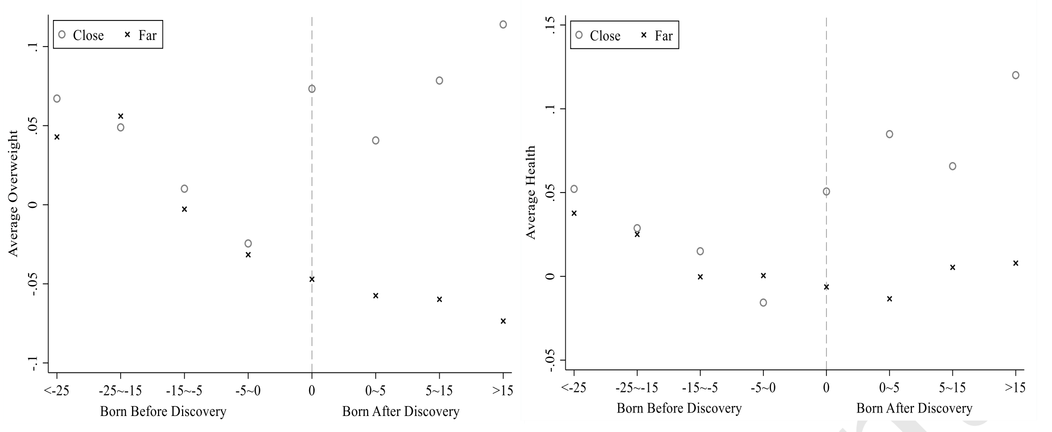

Figure 5. Average Health -- Close VS. Far (Common Pre-trends)
Notes: The figure plots the average levels of health outcomes ('Overweight' in the left panel and 'Health' in the right panel), conditional on the baseline set of controls (namely, gender, education, marriage status, siblings, age, age squared, age cubed, and wage) of two groups: 'Close' and 'Far'; the former (latter) includes individuals living no more (farther) than 60km from their closest giant oil field. The discovery is marked by the dashed line, at 0. Marked on the X-axis, across the discovery threshold, are time-bins that denote individuals' time of birth relative to the discovery (i.e., 0-5 includes individuals born between 0 to 5 years of the discovery made closest to them).

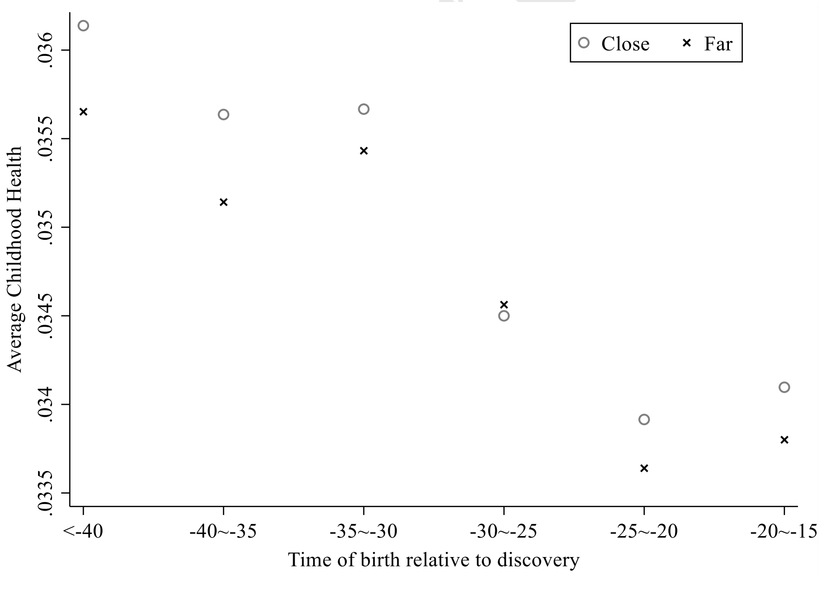

Figure 6. Average Childhood Health – Close VS. Far (Common Pre-trends)
Notes: The figure plots the average levels of childhood health (measuring health conditions up to age 15), of individuals born up to 15 years prior to their closest discovery, conditional on the baseline set of controls (namely, gender, education, marriage status, siblings, age, age squared, age cubed, and wage) of two groups: 'Close' and 'Far'; the former (latter) includes individuals living no more (farther) than 60km from their closest giant oil field. Marked on the X-axis, are time-bins that denote individuals' time of birth relative to the discovery (i.e., -15~-20 includes individuals born between 15 to 20 years prior to the discovery made closest to them).

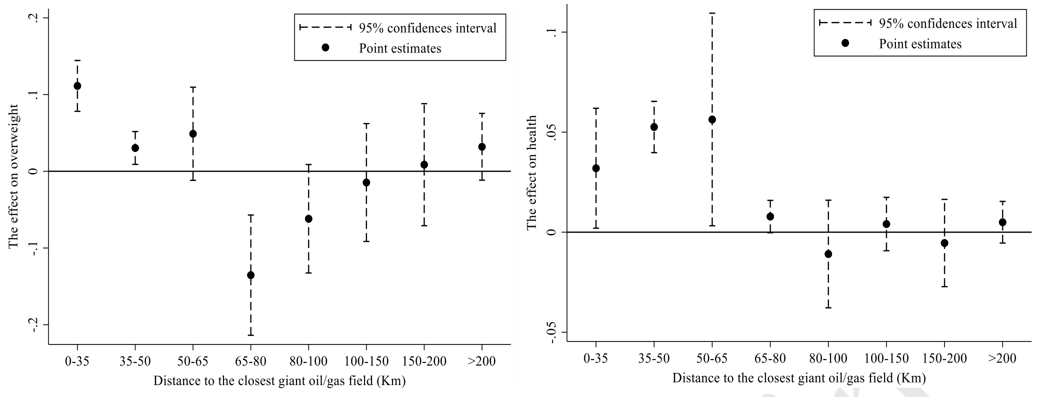

Figure 7. The Effect on Physical Health over Distance
Notes: The figure plots the coefficients of the interactions between 'After' and the dummy variables that capture the distance-bins appearing in the X-axis (0-35km, 35-50km, etc.), following the specification noted in equation (2), for the two health outcomes ('Overweight' ('Health') in the left (right) figure). The dashed lines represent 95% confidence intervals. All regressions include an intercept, city and wave fixed effects, and the following controls: gender, education, marriage status, siblings, age, age squared, age cubed, and wage. Standard errors are clustered by city. See Data Appendix for detailed description of variables.

Estimation

In this section we present the estimation results. We start with an initial analysis that examines the spatial distribution of the treatment's impact, based on which we then undertake the baseline analysis. Thereafter, we dig deeper into the main results by examining some disaggregated categories, and heterogeneous effects.

Initial analysis

As an initial step, we examine the spatial characteristics of the impact of giant oil fields on long term health, focusing on our two main outcome variables, the *Health Index* and the BMI. To do so, we estimate the following version of equation (1):

$$(2) \quad O_{i,c,w} = \alpha + \beta\left(after\right)_i + \gamma\left(\mathbf{after} * \mathbf{dist\_bin}\right)_{i,c} + \boldsymbol{\delta}\left(\mathbf{X}\right)_i + \upsilon_c + \eta_w + \varepsilon_{i,c,w}$$

where γ is a vector of coefficients, and **dist_bin** is a vector of 8 distance bins, outlined on the X-axis of Figure 7.[37] The latter's left (right) graph examines the case of the BMI (*Health Index*). In each case, we plot the γs, together with their 95% confidence intervals.

In both cases, the results point at similar patterns. We observe that distance from the giant oil field matters in general. However, we can also observe the extent to which it matters more specifically. Both cases indicate that the impact of a giant oil field on the two health indicators is apparent within the range of 60km, after which it generally wears off.

In the case of the BMI, we notice a positive and significant coefficient, with a decreasing magnitude, up to the 60km range. This means that within this area, individuals born after a discovery are relatively more overweight in the long term, in a magnitude that decreases with distance. Similarly, in the case of the *Health Index*, we observe as well a positive and

[37] Each bin represents distance from the closest giant oil field. E.g., the first distance-bin, namely 0-35, covers individuals who live in cities that are 0-35km from the closest giant oil field.

Table 2. Baseline Analysis

	Overweight (1)	Health (2)	Overweight (3)	Health (4)	Disability (5)	Mental health (6)	Childhood health (7)
After*Close	0.07***	0.04***			0.0042	-0.031	0.042
	(0.025)	(0.01)			(0.015)	(0.054)	(0.053)
After *Far			0.007	0.002			
			(0.02)	(0.004)			
After *Close			0.08***	0.043***			
			(0.02)	(0.015)			
Observations	10,921	18,874	10,921	18,874	19,324	11,484	16,189
R-squared	0.095	0.071	0.095	0.075	0.159	0.035	0.038

Notes: Standard errors appear in the parentheses for the independent variables, clustered by city. Superscripts *, **, *** correspond to 10, 5 and 1% levels of significance. 'After*Close' is the interaction term indicating that individuals are born after and live close to the giant oil/gas discovery (within 60km). In columns (3)-(4), we break down the variable 'Close' into two variables, 'Far' and 'Close'. 'Far': a dummy variable that identifies people who live far from a giant oil/gas field (beyond 60km). 'Close': a dummy variable that identifies people who live close to a giant oil/gas field (within 60km). All regressions include an intercept, city and wave fixed effects, and the following controls: gender, education, marriage status, siblings, age, age squared, age cubed, and wage. See Data Appendix for detailed description of variables.

significant coefficient, with a decreasing magnitude, up to the 60km range, indicating that within this area individuals born after a discovery have relatively more positive diagnoses of chronic diseases in the long term, despite being younger.[38]

Based on these results we consider individuals living within the 60km range as being in close proximity to a giant oil field, and those living farther than 60km as being far from one. Adopting this division, we next turn to the baseline analysis.

Baseline results

We estimate various versions of equation (1). The results are presented in Table 2. Columns 1 and 2, the baseline specifications, examine the cases of the BMI, and the *Health Index*, respectively. The reported negative (positive) and precise γ in column 1 (2) provides an average estimate that points at the same patterns observed in Figure 7. proximity, either through the BMI or the up-rise of chronic diseases. In addition, these coefficients provide an estimate for the magnitude of the effect, demonstrating that it is relatively significant.

Specifically, in the case of the (BMI) *Health Index*, the estimated γ coefficient indicates that a giant oil field discovery induces an increase of approximately (7%) 4% in the share of individuals born after a discovery who are (overweight) diagnosed with a chronic disease in the long term. This, in turn, amounts to an increase of (23%) 22% in the average share of (overweight) diagnosed individuals born after the discovery, compared to those born before the discovery.

In columns 3 and 4 we examine the source of the relative effect, for the cases of the BMI and the *Health Index* respectively, by estimating in each case the following variation of equation (1):

$$(3) \quad O_{i,c,w} = \alpha + \beta\left(far * after\right)_{i,c} + \gamma\left(close * after\right)_{i,c} + \boldsymbol{\delta}\left(\mathbf{X}\right)_i + \upsilon_c + \eta_w + \varepsilon_{i,c,w}$$

[38] In a later subsection we also examine additional characteristics of those born after a discovery (in general, as well as more specifically with respect to those living in close proximity to a giant oil field).

Table 3. Disaggregated Health

	Hyper-tension	**Dyslipidemia**	**Diabetes**	**Cancer**	**Lung**	**Liver**	**Heart**
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**
After*Close	0.018	-0.018	-0.013	0.004	0.068***	0.004	0.106*
	(0.05)	(0.04)	(0.016)	(0.016)	(0.024)	(0.019)	(0.06)
Observations	18,456	18,317	18,274	18,229	18,322	18,261	18,340
R-squared	0.096	0.08	0.036	0.015	0.066	0.035	0.129

	Stroke	**Kidney**	**Stomach**	**Psychiatry**	**Memory**	**Arthritis**	**Asthma**
	(8)	**(9)**	**(10)**	**(11)**	**(12)**	**(13)**	**(14)**
After*Close	0.014	0.082***	0.153***	-0.017	-0.059***	0.042	0.028
	(0.024)	(0.016)	(0.023)	(0.01)	(0.015)	(0.026)	(0.027)
Observations	18,239	18,286	18,461	18,230	18,233	18,542	18,249
R-squared	0.043	0.042	0.06	0.027	0.051	0.125	0.038

Notes: Standard errors appear in the parentheses for the independent variables, clustered by city. Superscripts *, **, *** correspond to 10, 5 and 1% levels of significance. 'After*Close' is the interaction term indicating that individuals are born after and live close to the giant oil/gas discovery (within 60 km). All regressions include an intercept, city and wave fixed effects, and the following controls: gender, education, marriage status, siblings, age, age squared, age cubed, and wage. See Data Appendix for detailed description of variables.

In effect, we exclude *after*, and instead include its disaggregation to *close* and *far* dummies, each capturing individuals living within and outside the 60km range, respectively. This disaggregation enables observing the direct impact of each. The results in columns 3 and 4, reporting the estimated β and γ, indicate that the observed relative decrease in the health status of individuals in the long term is entirely driven by effects undertaken within the 60km range, further clarifying the spatial impact of giant oil fields, and consistent with the patterns observed in Figure 7.

In columns 5-7 we examine the remaining main outcome variables related to physical and mental disabilities, mental health, and health during childhood, respectively, under the baseline specification. The γ estimates in each of the cases indicate that giant oil fields do not induce an observable long-term impact on these outcomes. Of particular interest, however, is the result in column 7, related to childhood health, which provides further support to the assertion that the impacts of giant oil fields appear at later stages in life, namely in the long term.

Last, to comment on the estimated βs, under the baseline specification. In all cases, namely columns 1-7, the coefficient on *after*, not reported for clearer exposition, is not revealing of patterns with clear statistical inference, with the exceptions of the *Health Index* and mental health measure. We observe, not surprisingly, that younger individuals, born past their closest discovery, tend to have less chronic diseases and better mental health, on average, irrespective of the proximity to a giant oil field.

Health Categories

Next, we disaggregate the *Health Index* to its different components, in an attempt to pinpoint the type of chronic diseases that drive the main patterns. In effect, we estimate the baseline specifications for each of the 14 chronic diseases that compose the index, separately. The results appear in Table 3. Each column examines a case in which the outcome variable is an indicator that captures a positive diagnosis of the disease noted in the column's title, reporting the estimated γ.

The results indicate that the main patterns are observed most notably in health outcomes related to chronic lung disease, heart disease, and kidney and stomach diseases. Conversely, memory-related chronic diseases exhibit opposite patterns. These outcomes are consistent with the reported types of environmental pollution induced by oil fields, including air pollution, and environmental contamination via industrial waste and oil spills (Baum and Tolonen, 2021; Brisbois et al., 2019).

Heterogenous effects

We examine heterogenous effects, driven by each of the baseline controls, with the exception of age.[39] To do so, we estimate the following variation of equation (1):

$$\begin{aligned} O_{i,c,w} = {} & \alpha + \beta\left(after\right)_{i,c} + \gamma\left(z\right)_{i,c} + \delta\left(close * after\right)_{i,c} + \theta\left(z * after\right)_{i,c} + \\ & \lambda\left(close * z\right)_{i,c} + \mu\left(close * after * z\right)_{i,c} + \boldsymbol{\pi}\left(\mathbf{X}\right)_{i} + \upsilon_{c} + \eta_{w} + \varepsilon_{i,c,w} \end{aligned} \tag{4}$$

in which z denotes one of the (non-age) baseline controls. We examine, and report in Table 4, the characteristics of μ, for each of the cases, and separately for the two main outcome variables.

The results indicate that the main patterns observed do not differentiate across the distribution of the majority of baseline controls. The exceptions are education, in the case of the BMI, and gender in the case of the *Health Index*. The results demonstrate that the effects are weaker amongst educated and male individuals. The difference in the share of overweight individuals is mitigated if the comparison between those born before and after a discovery considers relatively more educated individuals; similarly, this is also the case when examining males, compared to females, under the outcome related to chronic diseases.

Robustness

In this section we undertake various robustness tests. We start with examining the role of numerous potential underlying intermediate channels for the baseline patterns observed. Thereafter, we undertake a number of additional tests which look into further aspects in the empirical design.

Potential Intermediate Channels

Living close to a giant oil field may affect long term health across exposure levels via various channels. While our within-city approach accounts for oil triggered effects on pollution, development, technology, and income across cities, within them their change may affect the population differentially across the exposure distribution. For example, individuals born after a close-by discovery are born into a relatively wealthier and more developed environment compared to those born prior to one. Given the average discovery birth distance in our sample, under which the average individual was born about 20 years prior to the closest discovery, the development difference continues during the critical childhood phases and also afterwards. Put together, individuals with lengthier exposure to the nearby oil field, also experience greater relative exposure to an improved environment.

This, in turn, may have both positive and negative implications for the long-term health outcomes, via various intermediate channels. For instance, a greater exposure to an improved environment may be associated, on one hand, with greater exposure to improved health

[39] Age is not examined separately because the DiD setup already exploits its heterogeneity.

Table 4. Heterogeneous Effects

	Overweight (1)	Overweight (2)	Overweight (3)	Overweight (4)	Overweight (5)	Health (6)	Health (7)	Health (8)	Health (9)	Health (10)
After*Close*Education	-0.05***					-0.002				
	(0.02)					(0.08)				
After*Close*Marriage Status		0.007					0.016			
		(0.07)					(0.02)			
After*Close*Gender			-0.192					-0.02**		
			(0.124)					(0.01)		
After*Close*Wage				-0.004					0.002	
				(0.006)					(0.003)	
After*Close*Siblings					-0.07					0.12
					(0.13)					(0.09)
Observations	10,921	10,921	10,921	10,921	10,921	18,874	18,874	18,874	18,874	18,874
R-squared	0.096	0.096	0.096	0.096	0.096	0.075	0.075	0.075	0.075	0.075

Notes: Standard errors appear in the parentheses for the independent variables, clustered by city. Superscripts *, **, *** correspond to 10, 5 and 1% levels of significance. ‘After*Close’ is the interaction term indicating that individuals are born after and live close to the giant oil/gas discovery (within 60km). All regressions include an intercept, city and wave fixed effects, double interaction terms between ‘After’, ‘Close’ and the heterogenous effect examined, as well as the following controls: gender, education, marriage status, siblings, age, age squared, age cubed, and wage. See Data Appendix for detailed description of variables.

services, which may yield better relative long-term health conditions; however, on the other hand, greater exposure to improved health services may also induce a moral hazard effect in which individuals undertake less healthy activities, such as smoking, drinking alcoholic beverages in excessive amounts, and foregoing physical exercises, which may deteriorate the relative long-term health position.

We examine a number of potential intermediate channels, which may manifest both the positive and negative effects of an improved environment. Specifically, we consider the following options, which exploit the relevant available data options in CHARLS: smoking habits, consumption of alcoholic beverages, physical exercise, the extent of medical insurance, the extent of vaccinations, and the wage level. The description of each is provided in the Data Appendix. The estimation results appear in Table 5. Each column follows the baseline specification, in which the examined channel enters as the outcome, dependent variable, rather than as a regressor; each channel is tested separately, in each of the columns.

The results point at some interesting patterns. First, via the coefficients on $after$ we observe that individuals born after their closest discovery tend on average to smoke less, and consume fewer alcoholic beverages; when accounting also for the results reported in the baseline cases, we can in addition conclude that they are on average happier, have less chronic diseases, and are younger (by construction). It is, therefore, surprising to observe that (some of) these trends change when a giant oil field is in close proximity, to the extent that the health conditions of these individuals deteriorate, when comparing the difference with those born prior to a discovery to the pre-post difference observed in farther locations.

The coefficient on the interaction term, γ, points at possible directions. It is evident that under close proximity to a giant oil field, individuals born after their closest discovery tend to rather smoke more, and consume more alcoholic beverages. These results can be the outcome of various possible effects, out of which we consider three potential options, which are consistent with the observations and the literature. First, while a giant oil field brings new employment opportunities to the younger, working population, it also produces a more stressful environment as a result, which in turn encourages less healthy habits.[1] Second, the more developed environment may induce a stronger moral hazard effect on the younger, post-discovery born individuals, which also may encourage following a less healthy lifestyle, as described in Einav and Finkelstein (2018). Third, the general increase in wealth may induce differences in the tobacco and liquor related consumption behavior of the pre- and post-discovery born individuals, which is not manifested in farther locations that did not undergo a development boom; this alternative is consistent with evidence provided, for example, by Apouey and Clark (2015), who find that lottery winnings in the United Kingdom are associated with higher consumption of alcohol and tobacco goods.

The results in column 6 deserve further comment. Notably, we do not observe a wealth effect for the average individual born prior to a discovery (i.e., irrespective of the distance from a giant oil field). However, importantly we also do not observe such an effect across the distance dimension, suggesting that the income differences between pre- and post-discovery born individuals who live in close proximity to a giant oil field are not statistically different than those differences elsewhere. This observation strengthens the notion that the wealth effect is primarily a cross-sectional phenomenon; nevertheless, it does not preclude the option that a uniform increase in wealth may induce heterogenous effects across the age distribution, as suggested above.

[1] In their review, Baum and Tolonen (2021) discuss this channel, its implications, and related evidence.

Table 5. Potential Intermediate Channels

	Smoke (1)	Alcohol (2)	Exercise (3)	Insurance (4)	Vaccinations (5)	Wage (6)
After	-0.023**	-0.03**	0.011	0.006	0.004	-0.02
	(0.04)	(0.01)	(0.024)	(0.007)	(0.01)	(0.05)
After*Close	0.025***	0.078***	-0.002	-0.02	-0.007	0.004
	(0.02)	(0.023)	(0.03)	(0.02)	(0.03)	(0.051)
Observations	19,317	15,348	16,038	19,404	16,087	18,874
R-squared	0.56	0.17	0.11	0.06	0.11	0.1

Notes: Standard errors appear in the parentheses for the independent variables, clustered by city. Superscripts *, **, *** correspond to 10, 5 and 1% levels of significance. 'After*Close' is the interaction term indicating that individuals are born after and live close to the giant oil/gas discovery (within 60km). All regressions include an intercept, city and wave fixed effects, and the following controls: gender, education, marriage status, siblings, age, age squared, age cubed, and wage (excluded in column 6). See Data Appendix for detailed description of variables.

Additional Tests

We undertake various additional robustness tests for our main results on the *Health Index*.[2] Results of this sub-section appear in Table 6, focusing on the estimated γs. Each estimation follows the baseline specification, under the modifications outlined in each case. First, we address the earliest diagnosis concern more directly. As noted, the majority of diagnoses that compose the *Health Index* have been taken prior to the survey year. About half of the individuals in the sample report the year of the earliest diagnosis, with which we have shown that the majority of earliest diagnoses occur post-discovery, and during adulthood, thus supporting the long-term interpretation adopted for the complete sample.

In columns 1 and 2 we undertake two direct tests with the earliest diagnosis measure. First, we include it as a regressor (column 1); second, we restrict the sample to individuals with a post-discovery earliest diagnosis (column 2).[3] The first examination accounts for the earliest diagnosis directly. The second examination restricts the sample to individuals that either had negative diagnoses, or post-discovery positive diagnoses, thus excluding cases with positive diagnoses that occurred prior to a discovery (and hence were not triggered by it). Both cases yield an outcome similar to that estimated under the baseline case in terms of sign and preciseness, albeit with a relatively larger magnitude, reflecting the mitigating impact of the pre-discovery diagnosed individuals included in the complete sample.

Second, we account for the individuals who moved locations (i.e., lived in a different location for more than 6 months) throughout their lifetime. A central aspect of the main analysis is restricting it to individuals who lived to a large extent in the same location throughout their life. Nonetheless, immigrants may affect the health conditions of the city's permanent residents via inflows of external influences (Giuntella and Mazzonna, 2015). We address that in two ways. First, we compare the averages of our outcome variables, including an additional measure for parents' health conditions (outlined in the Data Appendix), between the group of movers and non-movers in the total sample. This comparison is presented in Table A2. The comparison shows no

[2] We focus on the *Health Index* as it is the relevant outcome measure for some of the tests outlined in this section (specifically, those related to earliest diagnoses), and since it represents the main, comprehensive measure. For the relevant cases, not related directly to the *Health Index*, the results are qualitatively similar under the BMI measure.
[3] Notably, both cases include also the individuals with no positive diagnoses.

Table 6. Additional Tests

	Earliest diagnosis (1)	Post-discovery diagnosis (2)	Immigration (3)	Coal cities excluded (4)	Size of discovery (5)	Production (6)	Controls (7)	Minimizing difference (8)
After*Close	0.052***	0.054***	0.03**	0.04***	0.058***	0.021***	0.032**	0.049**
	(0.018)	(0.019)	(0.014)	(0.015)	(0.014)	(0.055)	(0.015)	(0.02)
Earliest diagnosis	-0.0013***							
	(0.0001)							
Observations	9,310	7,360	21,620	17,431	18,874	18,874	13,904	2,873
R-squared	0.08	0.07	0.08	0.075	0.075	0.075	0.082	0.11

Notes: Standard errors appear in the parentheses for the independent variables, clustered by city. Superscripts *, **, *** correspond to 10, 5 and 1% levels of significance. 'After*Close' is the interaction term indicating that individuals are born before and live close to the giant oil/gas discovery (within 60km). All regressions include an intercept, city and wave fixed effects, and the following controls: gender, education, marriage status, siblings, age, age squared, age cubed, and wage (in addition to earliest diagnosis in column 1). The sample is restricted to post-discovery diagnosis in column 2. Column 3 examines the complete sample, including movers. Column 4 excludes the following major coal-producing cities from the sample: Hulunbuir, Loudi, Suzhou, Pingdingshan, Xuzhou, Zaozhuang, Yulin, Huainan, Jiaozuo, Chifeng, Yangquan, and Jixi. Column 5 accounts for the size of discoveries. Column 6 examines the case of production (rather than discovery). Column 7 includes the following additional controls: mining share, and parents' health. Column 8 restricts the sample to individuals born 5 years pre and post discoveries. See Data Appendix for detailed description of variables.

notable difference in the mean values across the two groups, thus mitigating the concern of an immigration-driven bias. Moreover, similarities in parental health points at lack of genetically-driven differences (further addressed in a later examination). Last, to the extent that the majority of movers represent intra-regional cases, provided that during large parts of the twentieth century international mobility in China was relatively marginal, these similarities also mitigate concerns related to potential biases of out-migration (since out-migrants in one region are captured in the data as incomers in other regions).[1]

Next, to more formally address this, we estimate in column 3 the complete sample provided by CHARLS, including the individuals excluded from the main analysis, namely those who were living in other locations for more than 6 months throughout their lifetime. The results indicate that the main outcome is robust to this inclusion. The estimated magnitude, however, is relatively smaller than that estimated in the baseline case, suggesting that incoming migrants have an overall mitigating effect on the health status of younger individuals in cities located close to a giant oil field.

Third, we account for the existence of nearby large coal mines. Coal production takes the largest share in China's natural resources sector. Several of the cities we examine, regardless of the proximity to a giant oil field, may be located next to major coal mines; in addition, discovery and production of coal mines may be correlated with those of giant oil field. This, in turn, may bias the estimated results to the extent that coal mines may also have an impact on long-tern health. To address that, we identify 12 cities, from those included in our sample, as being major coal producers.[2] In column 4 we estimate the baseline specification under their exclusion. The estimated outcome is almost identical to that estimated in the baseline case, in sign and significance, but notably also in magnitude. This outcome further supports the assertion that the long-term health impacts observed are an outcome of the treatment examined, namely the discovery and operation of giant oil fields.

Fourth, we consider the size of discoveries. The main analysis adopted a uniform category for the treatment, based on the defined threshold of estimated recoverable oil barrels. This was done to mitigate concerns related to possible measurement errors. However, these estimates vary across giant oil fields. Some giant oil fields are larger than others, and hence may induce a stronger impact, compared to other giant oil fields. In column 5 we exploit this additional variation. We do so by accounting for the estimated recoverable oil barrels directly; in effect, we interact the *close*after* term with these estimates, thus inducing further variation across cases.[3] In this case it does not only matter whether individuals were born prior to their closest discovery, or the proximity of their location to a giant oil field, but also whether the relevant discovery was of a relatively large or small giant oil field. The results indicate that the main outcome is robust to this modification. The estimated magnitude, however, is larger, suggesting that the extent of the observed impact is indeed positively associated with the size of the discovery.

Fifth, we examine the case of production. Doing so enables gaining deeper insights on the underlying channels. As noted, once discovered a giant oil field becomes operational after five years, on average. The commencement of production induces a steep increase in pollution,

[1] Nonetheless, we do note that to the extent that out-migrants represent a select group with respect to their vulnerability to a nearby giant oil field, then this would potentially bias the estimated effect. Specifically, if out-migrants have been more vulnerable to the impacts of a nearby giant oil field (compared to the general population) then our main estimate represents a lower-bound. Otherwise, in a case in which, for instance, wealthier families with better health conditions tend to out-migrate, then our estimates over-estimates the negative health impact reported.

[2] In effect, we follow the definition adopted in the 2006 Statistical Yearbook of Chinese Mining Industries for categorizing a city as being coal-rich, namely having at least a 10% share of coal output to total industrial output (in 2006 data). The 12 cities in our sample that follow this definition are: Hulunbuir, Loudi, Suzhou, Pingdingshan, Xuzhou, Zaozhuang, Yulin, Huainan, Jiaozuo, Chifeng, Yangquan, and Jixi.

[3] This specification also includes all the additional interactions between *close*, *after*, and the estimates of recoverable oil (notably, the latter is absorbed by the city fixed effects, and hence excluded).

compared to the pre-production, post-discovery stage in which extraction has not yet begun. Conversely, wealth is already affected upon discovery, albeit being merely a news shock (Arezki, Ramey, and Sheng, 2017). Hence, examining the case in which the treatment is the movement to the production phase may be revealing of the difference between the impact of wealth and pollution on long-term health. In column 6 we estimate a specification in which the treatment is the shift to production of the closest giant oil field, measured as occurring five years from its discovery (the baseline treatment). The estimated outcome points at a qualitatively similar result to that estimated in the baseline case, yet quantitively smaller. Specifically, the magnitude of the impact drops by approximately 50%. This suggests that differences in the long-term health conditions materialize more strongly upon discovery, than upon production, thus further implying that the triggering factor for these differences is the sudden increase in wealth and development.

Sixth, we examine two further controls. First, to better capture potential boom and bust or policy changes in the oil industry over time, which in turn may impact the main effect captured, we include a city-by-year measure of the mining sector's labor share. Second, we include a proxy for parents' health condition (outlined in the Data Appendix). The latter mitigates concerns related to potential genetically-transmitted diseases. In addition, it also addresses concerns related to systematic genetic differences of incoming migrant-parents, residing nearby a giant oil field, whose child (or children) is included in the sample as a non-mover. The outcome in column 7 indicates that the main result is robust to the inclusion of these measures.

Last, to better identify the discovery as the triggering factor for affecting long-term health, and to further address concerns related to the impact of age on the estimated outcomes, we restrict the pre- and post-discovery time span examined to 5 years. In effect, we estimate a restricted sample that only includes individuals born up to 5 years before and after their closest discovery. The outcome, presented in column 8, indicates that the main result is robust to this examination; furthermore, the estimated magnitude is relatively higher, further strengthening the role of the discovery in triggering the main effects observed.

Conclusion

This work examined empirically the long-term health effects of giant oil discoveries and operation, via the case of China. To do so, we spatially matched giant oil discoveries across China to individual-level data from the CHARLS surveys of 2011–2018. The analysis capitalized on a unique feature of CHARLS, namely tracking the location of survey participants over their lifetime, which enabled examining a sample of individuals who did not leave their place of birth for prolonged periods over the course of their life, thus continuously living in close proximity to, or far from, a giant oil field.

To identify the impact of giant oil fields on long-term health we adopted a difference-in-differences setup in which the treated group covered the individuals living in close proximity to a giant oil field (<60km), and the treatment was the discovery of one. In effect, we undertook a within-city approach which compared between the health differences of individuals born before and after the discovery made closest to their city of residence, across proximities, exploiting the variation in the time exposed to the giant oil field driven by the birth date. Identification was based on the plausible exogeneity of the discoveries, the restriction of the analysis to individuals who, to a large extent, never moved, and the existence of common pre-trends.

Our main finding was that giant oil fields deteriorate the long-term health conditions, in a fairly large magnitude, despite the positive (general, rather than individual) wealth impact, by primarily affecting the respiratory, digestive, and urinary systems. These patterns, plausibly driven by changes in the consumption habits of alcohol and tobacco goods, were observed via the relatively stronger impact on the individuals located near a giant oil field and born after the discovery, although in general, on average and across all locations, these individuals are younger, happier, smoke less, consume fewer alcoholic beverages, and have less chronic diseases.

These results shed new light on the health impacts of nearby oil fields, which to this point focused primarily on the short term, or disregarded mobility across prolonged periods. In addition, they provide a new, long-term, perspective on the wealth-health trade-off of oil extraction, which may carry policy implications concerning the operation of large natural resource reserves and their impact on the economy. Policy makers should anticipate the reported effects and apply mitigating policy measures in regions nearby operating giant oil fields, most notably with respect to the consumption of alcohol and tobacco goods, and the management of a regional development boom.

[First submitted May 2023; accepted for publication October 2023.]

References

Allcott H., and D. Keniston. 2018. "Dutch Disease or Agglomeration? The Local Economic Effects of Natural Resource Booms in Modern America." *Review of Economic Studies* 85(2): 596–731. doi: 10.1093/restud/rdx042.

Almond D. 2006. "Is the 1918 Influenza Pandemic Over? Long-Term Effects of In Utero Influenza Exposure in the Post-1940 U.S. Population." *Journal of Political Economy* 114: 672–712. doi: 10.1086/507154.

Almond D., and J. Currie. 2011. "Killing Me Softly: The Fetal Origins Hypothesis." *Journal of Economic Perspectives* 25(3): 153–172. doi: 10.1257/jep.25.3.153.

Almond D., J. Currie, and V. Duque. 2018. "Childhood Circumstances and Adult Outcomes: Act II." *Journal of Economic Literature* 56(4): 1360–1446. doi: 10.1257/jel.20171164.

Almond D., L. Edlund, and M. Palme. 2009. "Chernobyl's Subclinical Legacy: Prenatal Exposure to Radioactive Fallout and School Outcomes in Sweden." *The Quarterly Journal of Economics* 124(4): 1729–1772. doi: 10.1162/qjec.2009.124.4.1729.

Almond D., L. Edlund, H. Li, and J. Zhang. 2010. "Long-Term Effects of the 1959-1961 China Famine: Mainland China and Hong Kong." In I. Takatoshi and R. Andrew, eds., *The Economic Consequences of Demographic Change in East Asia*. National Bureau of Economic Research: University of Chicago Press, pp. 321–350. doi: 10.7208/chicago/9780226386881.001.0001.

Apouey B., and A. E. Clark. 2015. "Winning Big but Feeling no Better? The Effect of Lottery Prizes on Physical and Mental Health." *Health Economics* 24(5): 516–538. doi: 10.1002/hec.3035.

Aragon M. F., and P. J. Rud. 2013. "Natural Resources and Local Communities: Evidence from a Peruvian Gold Mine." *American Economic Journal: Economic Policy* 5: 1–25. doi: 10.1257/pol.5.2.1.

Arezki R., A. V. Ramey, and L. Sheng.2017. "News Shocks in Open Economies: Evidence from Giant Oil Discoveries." *Quarterly Journal of Economics* 132(1): 103–155. doi: 10.1093/qje/qjw030.

Baum S., and A. Tolonen. 2021. "Extractive Industries and Gender Equality." *Review of Environmental Economics and Policy* 15(2): 195–215. doi: https://doi.org/10.1086/715525

Brisbois B. W., J. Reschny, T. M. Fyfe, H. G. Harder, M. W. Parkes, S. Allison, C. G. Buse, R. Fumerton, and B. Oke. 2019. "Mapping Research on Resource Extraction and Health: A Scoping Review." *The Extractive Industries and Society* 6(1): 250–259. doi: 10.1016/j.exis.2018.10.017.

Cockx L., and N. Francken. 2014. " Extending the Concept of the Resource Curse: Natural Resources and Public Spending on Health." *Ecological Economics* 108: 136–149. doi: 10.1016/j.ecolecon.2014.10.013.

Cui H., J. P. Smith, and Y. Zhao. 2020. "Early-Life Deprivation and Health Outcomes in Adulthood: Evidence from Childhood Hunger Episodes of Middle-Aged and Elderly Chinese." *Journal of Development Economics* 143: 102417. doi: 10.1016/j.jdeveco.2019.102417.

Cust J., and S. Poelhekke. 2015. " The Local Impact of Mining." *Annual Review of Resource Economics* 7: 251–268. doi: 10.1146/annurev-resource-100814-125106.

De Haas R., and S. Poelhekke. 2019. " Mining Matters: Natural Resource Extraction and Firm-Level Constraints." *Journal of International Economics* 117: 109–124. doi: 10.1016/j.jinteco.2019.01.006.

Einav L., and A. Finkelstein. 2018. "Moral Hazard in Health Insurance: What We Know and How We Know It." *Journal of the European Economic Association* 16(4): 957–982. doi: 10.1093/jeea/jvy017.

Folland S., A. C. Goodman, and M. Stano. 2016. *The Economics of Health and Health Care: Pearson New International Edition*. 7th ed. Routledge.

Giuntella O., and F. Mazzonna. 2015. " Do Immigrants Improve the Health of Natives?" *Journal of Health Economics* 43: 140–153. doi: 10.1016/j.jhealeco.2015.06.006.

Harding T., R. Stefanski, and G. Toews. 2020. "Boom Goes the Price: Giant Resource Discoveries and Real Exchange Rate Appreciation." *Economic Journal* 130: 1715–1728. doi: 10.1093/ej/ueaa016.

Horn M. K. 2014. Giant Fields Data Set, Revision 16: AAPG-GIS Open File. Available online at http://archives.datapages.com/.

Hoynes H., D. W. Schanzenbach, and D. Almond. 2016. "Long-Run Impacts of Childhood Access to the Safety Net." *American Economic Review* 106(4): 903–934. doi: 10.1257/aer.20130375.

Jacobsen G. D., D. N. Parker, and J. B. Winikoff. 2023. "Are Resource Booms a Blessing or a Curse? Evidence from People (Not Places)". *Journal of Human Resources* 58(2): 393–420. doi: https://doi.org/10.3368/jhr.58.4.0320-10761R1.

Johnston J. E., E. Lim, and H. Roh. 2019. "Impact of Upstream Oil Extraction And Environmental Public Health: A Review of the Evidence." *Science of The Total Environment* 657: 187–199. doi: https://doi.org/10.1016/j.scitotenv.2018.11.483.

Kim S., B. Fleisher, and J. Y. Sun. 2017. "The Long-Term Health Effects of Fetal Malnutrition: Evidence from the 1959-1961 China Great Leap Forward Famine." *Health Economics* 26(10): 1264–1277. doi: 10.1002/hec.3397.

Lei Y., and G. Michaels. 2014. "Do Giant Oilfield Discoveries Fuel Internal Armed Conflicts?" *Journal of Development Economics* 110: 139–157. doi: https://doi.org/10.1016/j.jdeveco.2014.06.003.

O'Callaghan-Gordo C., M. Orta-Martínez, and M. Kogevinas. 2016. " Health Effects of Non-Occupational Exposure to Oil Extraction." *Environmental Health* 15: 56. doi: 10.1186/s12940-016-0140-1.

Okeke E. N., and S. Abubakar. 2020. " Healthcare at the Beginning of Life and Child Survival: Evidence from a Cash Transfer Experiment in Nigeria." *Journal of Development Economics* 143: 102426. doi: https://doi.org/10.1016/j.jdeveco.2019.102426.

Perez-Sebastian F., and O. Raveh. 2016a. "Natural Resources, Decentralization, and Risk Sharing: Can Resource Booms Unify Nations?" *Journal of Development Economics* 121: 38–55. doi: https://doi.org/10.1016/j.jdeveco.2016.02.003.

———. 2016b. "The Natural Resource Curse and Fiscal Decentralization." *American Journal of Agricultural Economics* 98(1): 212–230. doi: https://doi.org/10.1093/ajae/aav051.

Perez-Sebastian F., and O. Raveh. 2019. “Federal Tax Policies, Congressional Voting, and Natural Resources.” *Canadian Journal of Economics* 52(3): 1112–1164. doi: https://doi.org/10.1111/caje.12404.

Perez-Sebastian F., O. Raveh, and F. van der Ploeg. 2021. “Oil Discoveries and Protectionism: Role of News Effects.” *Journal of Environmental Economics and Management* 107: 102425. doi: https://doi.org/10.1016/j.jeem.2021.102425.

Phelps E. C. 2017. *Health Economics*. 6th ed. Routledge.

Raveh O. 2013. “Dutch Disease, Factor Mobility, and the Alberta Effect: The Case of Federations.” *Canadian Journal of Economics* 46(4): 1317–1350. doi: https://doi.org/10.1111/caje.12050.

Raveh O., and A. Reshef. 2016. “Capital Imports Composition, Complementarities, and the Skill Premium in Developing Countries.” *Journal of Development Economics* 118, 183–206. doi: https://doi.org/10.1016/j.jdeveco.2015.07.011.

Shi L., W. Chen, J. Xu, and L. Ling. 2020. “Trends and Characteristics of Inter-Provincial Migrants in Mainland China and its Relation with Economic Factors: A Panel Data Analysis from 2011 to 2016.” *Sustainability* 12: 610. doi: https://doi.org/10.3390/su12020610.

Toews G., and P.-L. Vezina. 2022. “Resource Discoveries, FDI Bonanzas, and Local Multipliers: Evidence from Mozambique.” *The Review of Economics and Statistics* 104(5), 1046–1058. doi: https://doi.org/10.1162/rest_a_00999.

van der Ploeg F. 2011. “Natural Resources: Curse or Blessing?” *Journal of Economic Literature* 49(2): 366–420. doi: 10.1257/jel.49.2.366.

van der Ploeg F., and S. Poelhekke. 2017. “The Impact of Natural Resources: Survey of Recent Quantitative Evidence.” *The Journal of Development Studies* 53(2): 205–216. doi: 10.1080/00220388.2016.1160069.

Venables A. 2016. “Using Natural Resources for Development: Why Has It Proven So Difficult?” *Journal of Economic Perspectives* 30(1): 161–184. doi: 10.1257/jep.30.1.161.

von der Goltz J., and P. Barnwal. 2018. “Mines: The Local Wealth and Health Effects of Mineral Mining in Developing Countries.” *Journal of Development Economics* 139: 1–16. doi: https://doi.org/10.1016/j.jdeveco.2018.05.005.

Whitlock G., S. Lewington, P. Sherliker, R. Clarke, J. Emberson, J. Halsey, N. Qizilbash, R. Collins, and R. Peto. 2009. “Body-Mass Index and Cause-Specific Mortality in 900000 Adults: Collaborative Analyses of 57 Prospective Studies.” *The Lancet* 373(9669): 1083–1096. doi: https://doi.org/10.1016/S0140-6736(09)60318-4.

Wu K. 2014. “China’s Energy Security: Oil and Gas.” *Energy Policy* 73: 4–11. doi: https://doi.org/10.1016/j.enpol.2014.05.040.

Zhao Y. H., Y. S. Hu, J. P. Smith, J. Strauss, and G. Yang. 2014. “Cohort Profile: The China Health and Retirement Longitudinal Study (CHARLS).” *Journal of Epidemiology* 43(1): 61–68. doi: 10.1093/ije/dys203. Epub 2012 Dec 12.

Appendix

We employ cross-sectional data on 25,949 individuals, distributed over 125 cities and 28 provinces in China. Unless otherwise specified, variables are based on data from the 2011–2018 CHARLS survey waves (Zhao et al., 2014). The CHARLS data is maintained at the National School of Development at Peking University, and is available at: http://charls.pku.edu.cn, including more detailed descriptions of the survey variables employed in our analysis. Descriptive statistics for all variables appear in Table 1, as well as in Appendix Table A1 for proximity and birth related sub-samples respectively.

Variable Definitions

Overweight: A binary indicator that captures whether an individual is categorized as being overweight, based on the World Health Organization's BMI-based definition (being a BMI measure of 25 and above). BMI (Body Mass Index) is measured as the body mass divided by the square of the body height, at survey year.

Health Index: The *Health Index* is an average of 14 indicators, each capturing the existence of a chronic disease: hypertension, dyslipidemia, diabetes, cancer, lung disease, heart disease, stroke, kidney disease, stomach disease, memory-related disease, arthritis, asthma, liver disease, and emotional condition. Each indicator takes either the value of 0 (no disease) or 1 (positive diagnosis). Hence, the index takes a value between 0 and 1, with higher values representing worse health conditions.

Disability: An indicator that captures whether an individual has one or more of the following disabilities – intellectual, vision, and hearing disabilities, as well as speech impediments, and more general physical disabilities. The indicator takes the value of 0 in case of no disabilities, and 1 otherwise.

Mental health: An indicator that captures whether an individual has one or more of the following mental health conditions – depression, self-reported mental intactness, self-reported low ability of episodic memory. The indicator takes the value of 0 in case of no mental conditions, and 1 otherwise.

Childhood health: A self-reported indicator about the health status prior to age 15. Specifically, respondents replied to the following question: "Before you were 15 years old (including 15 years old), would you say that compared to other children of the same age, you were 1 Much healthier; 2 Somewhat healthier; 3 About average; 4 Somewhat less healthy; 5 Much less healthy." The indicator takes the value of 0 in case of an answer of 1 or 2, otherwise the indicator takes the value of 1.

Smoke: An indicator that takes the value of 1 if a person answered 'yes' to the following question: "Have you ever chewed tobacco, smoked a pipe, smoked self-rolled cigarettes, or smoked?", and 0 otherwise.

Alcohol: An indicator that takes the value of 1 if an individual reported consuming alcoholic beverages more than once a month, and 0 otherwise.

Exercise: An indicator that takes the value of 1 if an individual reported practicing physical exercise activity for at least 10 minutes each week, and 0 otherwise.

Insurance: An indicator that takes the value of 1 if an individual reported being the primary beneficiary of any type of health insurance, and 0 otherwise.

Vaccinations: An indicator that takes the value of 1 if an individual reported receiving any vaccinations prior to (and including) being 15, and 0 otherwise.

Age: An individual's age, based on its reported year of birth in the survey.

After: An indicator that takes the value of 1 if an individual was born after the discovery of its closest giant oil field, and 0 otherwise. The matching undertaken between individuals and their closest discoveries is outlined in the main text. Data on giant oil discoveries is retrieved from Horn (2014).

Close: An indicator that takes the value of 1 if an individual's place of residence is within close proximity to its closest giant oil discovery. Close proximity is defined as being within the 60km range. The matching undertaken between individuals and their closest discoveries is outlines in the main text. Data on giant oil discoveries is retrieved from Horn (2014).

Earliest diagnosis: The earliest year in which an individual received positive diagnosis on one of the 14 chronic diseased covered by the *Health Index*.

Size of discovery: The present value of the ultimately recoverable oil barrels (computed at the time of discovery), retrieved from Horn (2014).

Gender: An indicator that takes the value of 1 if an individual is male, and 0 if female.

Education: An indicator that captures an individual's education level. The indicator can take the value of 1 to 10, with a higher value indicating higher level of education, according to the following categories (numbered 1 to 10, respectively): 'Illiterate', 'Did not finish primary school but capable of reading or writing', 'Sishu/home school', 'Elementary school', 'Middle school', 'High school', 'Vocational school', 'Two-/Three-Year College/Associate degree', 'Master's degree', and ' Doctoral degree'.

Marriage status: An indicator that captures an individual's marriage status. The indicator can take the value of 1 to 6, according to the following categories (numbered 1 to 6, respectively): 'married with spouse present', 'married but not living with spouse temporarily for reasons such as work', 'separated', 'divorced', 'widowed', 'never married'.

Siblings: An indicator that takes the value of 1 if an individual is an only child, and 0 otherwise.

Wage: The total sum of wages, social income transfers, agricultural income, and self-employed income, an individual earned in the year prior to the survey

Table A1. Descriptive statistics, by proximity and by birth relative to discovery

	Close				**Far**				**Before**				**After**			
Variables	**Mean**	**Std. Dev.**	**Min.**	**Max.**	**Mean**	**Std. Dev.**	**Min.**	**Max.**	**Mean**	**Std. Dev.**	**Min.**	**Max.**	**Mean**	**Std. Dev.**	**Min.**	**Max.**
Overweight	0.351	0.477	0	1	0.308	0.462	0	1	0.312	0.463	0	1	0.308	0.462	0	1
Health	0.24	0.367	0	0.93	0.22	0.359	0	0.91	0.23	0.364	0	0.93	0.19	0.342	0	0.92
Disability	0.22	0.333	0	0.6	0.17	0.292	0	0.8	0.19	0.301	0	0.8	0.14	0.268	0	0.6
Mental health	0.09	0.33	0	1	0.08	0.32	0	1	0.09	0.33	0	1	0.07	0.32	0	1
Childhood health	0.344	0.475	0	1	0.352	0.478	0	1	0.355	0.479	0	1	0.336	0.472	0	1
Smoke	0.417	0.493	0	1	0.415	0.493	0	1	0.426	0.495	0	1	0.372	0.483	0	1
Drink	0.143	0.35	0	1	0.142	0.349	0	1	0.146	0.353	0	1	0.124	0.33	0	1
Exercise	0.366	0.482	0	1	0.374	0.484	0	1	0.399	0.49	0	1	0.274	0.446	0	1
Insurance	0.959	0.199	0	1	0.96	0.196	0	1	0.962	0.19	0	1	0.95	0.218	0	1
Vaccinations	0.866	0.341	0	1	0.857	0.35	0	1	0.858	0.35	0	1	0.857	0.35	0	1
Hypertension	0.43	0.496	0	1	0.39	0.487	0	1	0.42	0.493	0	1	0.28	0.45	0	1
Dyslipidemia	0.28	0.449	0	1	0.22	0.411	0	1	0.23	0.424	0	1	0.16	0.363	0	1
Diabetes	0.16	0.37	0	1	0.13	0.332	0	1	0.14	0.346	0	1	0.09	0.281	0	1
Cancer	0.03	0.183	0	1	0.02	0.146	0	1	0.02	0.151	0	1	0.02	0.141	0	1
Chronic lung disease	0.2	0.398	0	1	0.15	0.359	0	1	0.16	0.366	0	1	0.14	0.342	0	1
Liver disease	0.06	0.241	0	1	0.07	0.256	0	1	0.07	0.255	0	1	0.07	0.258	0	1
Heart disease	0.27	0.444	0	1	0.2	0.397	0	1	0.22	0.414	0	1	0.12	0.327	0	1
Stroke	0.08	0.265	0	1	0.07	0.259	0	1	0.08	0.274	0	1	0.03	0.18	0	1
Kidney disease	0.1	0.303	0	1	0.1	0.303	0	1	0.1	0.305	0	1	0.09	0.293	0	1
Stomach Disease	0.35	0.478	0	1	0.3	0.459	0	1	0.31	0.461	0	1	0.3	0.46	0	1
Emotional condition	0.05	0.219	0	1	0.03	0.175	0	1	0.03	0.178	0	1	0.03	0.18	0	1
Memory-related Disease	0.07	0.253	0	1	0.04	0.194	0	1	0.05	0.21	0	1	0.02	0.147	0	1
Arthritis	0.43	0.496	0	1	0.4	0.489	0	1	0.4	0.489	0	1	0.41	0.492	0	1
Asthma	0.08	0.287	0	1	0.05	0.235	0	1	0.06	0.246	0	1	0.05	0.208	0	1
Parents' health	0.073	0.132	0	1	0.075	0.116	0	1	0.074	0.114	0	1	0.072	0.128	0	1
Mining share	0.07	0.07	0	0.58	0.055	0.08	0	0.58	0.058	0.065	0	0.58	0.055	0.08	0	0.58
After	0.155	0.362	0	1	0.206	0.404	0	1	0	0	0	0	1	0	1	1
Close	1	0	1	1	0	0	0	0	0.068	0.252	0	1	0.049	0.216	0	1
Earliest diagnosis	2,008	12.3	1,953	2,018	2,008	13	1,900	2,018	2,007	13.5	1,900	2,018	2,010	9.92	1,960	2,018
Size of discovery	1,010	602	500	2,000	829	518	500	4,823	828	49.7	550	4,823	888	62.3	500	4,810
Gender	0.479	0.5	0	1	0.47	0.499	0	1	0.479	0.5	0	1	0.436	0.496	0	1
Education	3.44	1.86	1	9	3.45	1.98	1	10	3.44	1.99	1	10	3.51	1.89	1	10
Marriage status	1.68	1.46	1	6	1.64	1.41	1	6	1.69	1.46	1	6	1.46	1.15	1	6
Siblings	0.11	0.313	0	1	0.119	0.323	0	1	0.119	0.324	0	1	0.114	0.318	0	1
Wage	32,250	38,533	0	50,000	24,990	47,023	0	43,950	25150	35,550	0	50,000	27,020	46,290	0	47,550
Age	62.5	10.1	18	94	61.6	10.6	18	108	63.7	10.3	19	108	62.4	5.99	18	79

Notes: See Data Appendix for detailed description of variables.

Table A2. Non-movers VS. Movers – Key Measures

Variables	Non-movers Mean	Movers Mean
Overweight	0.31	0.314
Health	0.23	0.22
Disability	0.18	0.19
Mental health	0.09	0.09
Childhood health	0.36	0.346
Parents' health	0.077	0.074

Notes: See Data Appendix for detailed description of variables.