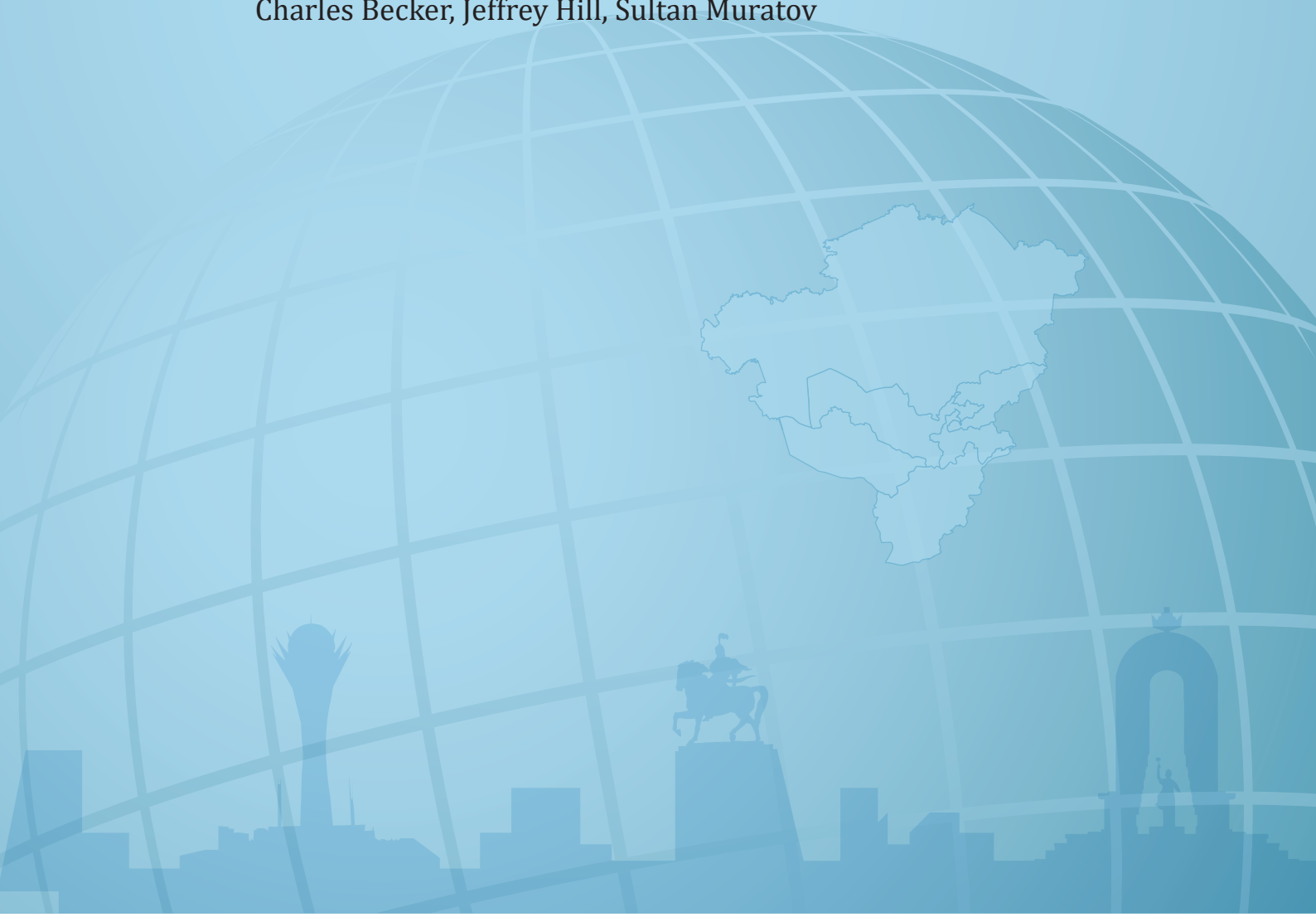
UNIVERSITY OF CENTRAL ASIA
GRADUATE SCHOOL OF DEVELOPMENT
Institute of Public Policy and Administration

Brighter than a Million Suns: Contemporary Health Consequences of Atomic Testing in the Semipalatinsk Nuclear Polygon

Charles Becker, Jeffrey Hill, Sultan Muratov



Working Paper #70, 2022

UNIVERSITY OF CENTRAL ASIA
GRADUATE SCHOOL OF DEVELOPMENT
Institute of Public Policy and Administration

WORKING PAPER #70, 2022

Brighter than a Million Suns: Contemporary Health Consequences of Atomic Testing in the Semipalatinsk Nuclear Polygon

Charles Becker, Jeffrey Hill, Sultan Muratov

Abstract: Between 1949 and 1962 there were over 450 surface, air, and underground nuclear tests in the Semipalatinsk nuclear polygon in north-eastern Kazakhstan, with underground testing continuing through 1989. As the region was heavily populated, there were major health consequences for the "treated" population, and this has been studied extensively. However, the full, long-run scope of the effects remains unexplored and this paper addresses the gap. Using a remarkable population health database from the Kazakhstan Ministry of Health from 2000 onward, and matching treated regions with similar but untreated areas, we find exceptionally large consequences for a wide range of health conditions many decades after the explosions ceased. Our propensity matching technique links demographic and economic conditions at the district – raion – level between treated and untreated areas, using a carefully constructed series of satellite night-light data from 1992 onward to measure economic characteristics.

Keywords: Kazakhstan, health, nuclear polygon

JEL classification: I10, P3

ISSN 2617-9245

University of Central Asia
138 Toktogul Street, Bishkek 720001, Kyrgyz Republic
Tel.: +996 (312) 910 822, E-mail: ippa@ucentralasia.org
www.ucentralasia.org

The Institute of Public Policy and Administration was established in 2011 to promote systematic and in-depth research on issues related to the socio-economic development of Central Asia, and to explore policy alternatives.

The Institute of Public Policy and Administration is part of the Graduate School of Development, University of Central Asia. The University of Central Asia (UCA) was founded in 2000. The Presidents of Kazakhstan, the Kyrgyz Republic, and Tajikistan, and His Highness the Aga Khan signed the International Treaty and Charter establishing this secular and private university, ratified by the respective parliaments, and registered with the United Nations.

The Institute of Public Policy and Administration's Working Papers is a peer-reviewed series that publishes original contributions on a broad range of topics dealing with social and economic issues, public administration and public policy as they relate to Central Asia. It is intended as a dynamic avenue for authors to share ideas with audiences concerned with current and emerging issues facing the wider Central Asian region. Comments on papers or questions about their content should be sent to ippa@ucentralasia.org. 'Working Papers' may be cited without seeking prior permission.

IPPA Working Paper Series Editors:
Dr Bohdan Krawchenko and Dr Roman Mogilevskii.

About the authors:
Professor **Charles M. Becker** joined the Duke faculty in 2003, where he directed the American Economic Association's Summer Program and Minority Scholarship Program (2003-2007). He previously directed this program at the University of Colorado at Denver, where he taught from 1999-2003. Becker earned his PhD in Economics at Princeton University (1981) and was an undergraduate at Grinnell College (BA, Economics and Russian language, 1976). He previously taught at Vanderbilt University and at the University of Colorado at Boulder, where he directed the Economics Institute from 1990-96. In 1998-99, Becker served as Team Leader for the Asian Development Bank's pension reform project in Kyrgyzstan; from 1999-2006 he served as an advisor to the Kazakhstan government on its pension reform program. International Academic Council of the Kyiv School of Economics from 2012 to 2015. (member from 2005 to 2017)

Jeffrey Hill is currently obtaining his Ph.D in Economics from UCLA. His interests include the economics of education, networks, and existential risk. He earned his Masters in Economics at Duke University and was an undergraduate at Wheaton College (BA, Economics and Chemistry).

Sultan Muratov is now working as an Associate in Research at Duke University. He received his Bachelor of Science degree in Mathematics with minor in Economics from Nazarbayev University and obtained Master of Arts in Economics degree from Duke.
His research interests lie in the fileds of health, environmental, urban and development economics.

The authors are thankful to Rob Garlick, Alex van Nievelt, and Erik Zitser for valuable comments and assistance.

Table of Contents

Figures

Tables

List of Acronyms

ATT	Average treatment effect on the treated
CC CPK	Central Committee of the Communist Party of Kazakhstan
CDR	Crude Death Rate
CRS	Chronic Radiation Syndrome
DMSP	Defense Meteorological Satellite Program
DN	Digital Number
EKO	East Kazakhstan Oblast
EOG	Earth Observation Group
GoK	Kazakhstan Government
ICRP	International Commission on Radiological Protection
IRSE	Institute of Radiation Safety and Ecology
MBs	Monthly Benefits Pension Fund of the Russian Federation (PFR)
PSM	Propensity Score Matching
SNP	Semipalatinsk Nuclear Polygon
STS	Semipalatinsk Test Site
ZEMR	Zone of Extreme and Maximal Radiation Risk
ZHR	Zone of High Radiation Risk
ZMR	Zone of Minimal Radiation Risk
ZRR	Zones of radiation risk

The Republican Commission on Investigation into the State of Ecological Conditions in Areas Testing Nuclear Weapons that I created in March 1989 conducted a series of research and analytical tasks. It concluded that there was enormous harm as a result of the tests, affecting not only currently living people, but extending many generations forward.

N. A. Nazarbaev[1]

1 Созданная мной в марте 1989 года Республиканская комиссия по наблюдению за состоянием экологической обстановки в местах испытания ядерного оружия, проведя серию исследований и анализов, сделала заключение: огромный вред от последствий испытаний оказывается не только на ныне живущих людей, но будет распространяться ещё на многие поколения вперёд… (Назарбаев, 2001, стр. 64)

1. Introduction

The effect of exposure to ionizing radiation on increased cancer prevalence of various types is well established.[2] Numerous studies have investigated the health outcomes due to exposure to radiation following the first detonations of nuclear bombs over Hiroshima and Nagasaki, Japan. Since then, other areas have been exposed to similar radiation from nuclear detonations due to testing as opposed to warfare. Populations affected by such nuclear testing include people living in the vicinity of the Nevada Test Site in the USA, the Marshall Islands in the Pacific Ocean, and the Semipalatinsk Nuclear Polygon (SNP) in northeastern Kazakhstan.

Of these regions, SNP had the largest exposed population and exposure levels were most profound. Bauer (2005), Markabayeva et al. (2018) and many others have examined the cancer and overall mortality patterns of the affected population in the vicinity of the Semipalatinsk Nuclear Polygon. Most of these papers track individuals directly exposed to the radiation or those born in the area during the period of testing. Their findings – and the myriad of photographs of children with horrendous physical deformations – are shocking. Indeed, the SNP has been the topic of both a detailed investigative study and an accompanying film Poligon by the thorough and intrepid Vlast.kz investigative team headed by Svetlana Romashkina, as well as a documentary by British Director Antony Butts, After the Apocalypse.[3] And, as the quotation above from Nursultan Nazarbayev, Kazakhstan Communist Party First Secretary and later first President of independent Kazakhstan, shows, the central theme of this paper – that there are severe long term consequences to massive atomic testing – was well known even in the Soviet era. What was not fully known, and what we seek to document, is the catastrophic extent of the "nuclear testing's long shadow."

To document these effects it is essential to address several limitations of prior studies. First, causal inference is largely imperfect. Above-ground atomic tests did not involve subjects in randomized control trials, and "treated" populations may well be unobservably different from control groups. Conversely, control groups also may have been substantially treated as well, though generally less so. In addition, both treatment and control groups may have received other treatments that are correlated with exposure to atomic test fallout. Finally, most prior studies focus on a limited set of outcome variables, and often restrict attention to a fairly short period following exposure.

This study addresses – albeit imperfectly – all of these issues thanks to a remarkable coincidence of rules and data sets. The SNP, which experienced nuclear tests from 1949 to 1989, appears to have been chosen carelessly or with utter disregard for potentially exposed populations – indeed, ignorance by decision-makers of the test site's proximity to Semipalatinsk (now Semey) city vastly increased the size of the exposed population, while the extent of risks were not well understood in 1949. Nor was the population free to leave: by 1949 the USSR had adopted the Tsarist-era internal passport (propiska) system that greatly restricted population movements. Consequently, as we document below, the affected population was substantial and remained stable or growing in treated areas until movement of ethnic Europeans back to Russia and other European republics began in the 1980s and became widespread in the 1990s. We are able to address the effect of nuclear testing in a comprehensive way thanks to the Kazakhstan Ministry of Health's exceptional public health database, available at the Medinfo.kz site.[4] This database provides detailed health and health services data at the raion (county, or district) level from 2000 through 2018, and can be used to track long-lag health

2 Gilbert 2009
3 https://polygon.vlast.kz/
4 MedInform, n.d.

effects ranging from birth anomalies to distinct major illnesses (both incidence and prevalence) and resulting mortality for the nation's 197 raions – small entities with a mean population, depending on the year, in the range of 75,000 to 90,000, and administratively separate cities. Few if any other middle or upper-middle income countries report these data regularly.

In addition, we are able to identify treated regions without difficulty. Not only is wind direction and hence fallout dispersion of the largest tests well-know; better still, the Kazakhstan Government (GoK) has instituted its own definition of extreme, maximal, high, low, and zero-exposure raions. Moreover, GoK also instituted a system of supplemental pension payments for those who were living in high + exposure areas during the period of testing, and the number of supplemental (l'gotnyi) pension recipients is published annually at the oblast (provincial) level. Combining this information with Census data, we are able to get a reasonable sense of population movements. Census and other government data are also helpful in allowing us to determine changes in raion-level ethnic composition and age structure over time.

This information is critical in enabling us to use kernel propensity score matching to compare different gradations of treated areas with matched counterparts (and also an "untreated" group that has very low propensities to be matched with any of the treated raions at any level). We start with information on population age and gender structure, as well as indicators of health care infrastructure (any endogeneity bias will work against us). However, industrial and mining cities and towns in Soviet Kazakhstan experienced a variety of ghastly environmental treatments in addition to atomic fallout, and it is important to control for these, which we do. Unfortunately, Kazakhstan does not have raion-level economic data that we can use. Therefore, we construct a consistent satellite night-light series – by far the hardest part of the paper – at the raion level from 1992 through 2018 and use the moments of these distributions to match districts by economic structure.

Our results are staggering: for many adverse health conditions, risks in the polygon remain greatly elevated some 40 to 50+ years after the most damaging tests had been completed. These elevated risks range from pregnancy complications to a range of immune disorders to circulatory illnesses to muscular-skeletal conditions and deformities. However, as we show below, risks are not universally elevated in the risk zones.

2. Background and History of the SNP

The Semipalatinsk Nuclear Polygon is named after the city of Semipalatinsk and sits roughly 150 kilometers southwest of the city. The SNP's origins go back to the nuclear arms race between the USA and USSR. In an attempt to end the US's monopolization of nuclear arms, the Soviet government urgently demanded its scientists to create an atomic bomb and search for the lands "suitable" for construction of a nuclear test-site.[5] Back then, most of the USSR military test sites were located in steppe areas, and the Kazakh SSR was especially preferred for its favorable geophysical conditions.[6]

In principle, the selection of the test site was based on the following criteria: minimally populated area, no agricultural lands, and proximity to some transportation arteries to allow future construction.[7] After a long selection process, the choice fell on steppes of Semipalatinsk Region, about

5 Voloshin, 2002
6 Logachev, 2002
7 Shkolnik 2002; Logachev, 2002

140 kilometers away from the city of Semipalatinsk.[8] This location on the Kazakh steppe was selected for being large, and, allegedly, relatively unpopulated.[9]

Even acknowledging that nuclear testing had a short history in 1940s and USSR government was under pressure of nuclear arms race, selection of the test site was highly questionable for several reasons. First, the areas in the vicinity of the proposed SNP test site were not unpopulated. USSR census data of the Semipalatinsk Region (and East Kazakhstan Region) reveals that population of Semipalatinsk city alone was around 150,000 in 1939 and close to 200,000 in 1959.[10] Apart from Semipalatinsk city, the area neighboring the proposed test site contained villages and seasonal settlements used for stockbreeding. Secondly, the harmful effects of ionizing radiation already were well-known, especially after the explosions in Hiroshima and Nagasaki.[11] The obvious conclusion is that in fact the Soviet government's choice was guided instead by utilitarian motives aiming to minimize construction and transportation costs.

The Nuclear Polygon was established in 1947 under the resolution of the USSR Cabinet of Ministers and the CPSU Central Committee.[12] The first test took place in August 1949 involving an above-ground plutonium bomb test. In August 1953, there was the world's first test of a thermonuclear device and the first hydrogen bomb was tested in August 1955.[13] In the literature on SNP, testing is commonly split into two periods: 1949-1962 and 1963-1989. Such division is conditioned on the location of the explosions – atmospheric (air and above-ground) in the former and underground in the latter periods.

The period of atmospheric testing includes 117 nuclear tests, of which 86 were in the air and 30 were above ground.[14] Following the first test in 1949, the intensity and power of explosions steadily increased. The intensity peaked in 1961-62 when 68 explosions were carried out, including 15 tests in September 1961 alone.[15] Atmospheric nuclear tests, and especially ground tests, pose the greatest threat to humans and environment as radioactive particles are uncontained and spread with the wind and explosion wave way beyond SNP boundaries. During some of the atmospheric explosions, radioactive plumes reached the city of Ust-Kamenogorsk (about 300 km from SNP) and Altai Territory of the Russian Federation (up to 570 km from SNP), far beyond the city of Semipalatinsk.[16] It is the consensus that the aboveground tests between 1949 and 1962 are responsible for the majority (up to 95%) of radiation exposure and environmental contamination.[17]

Given the established dangers of above-ground testing at SNP, it is reasonable to question whether the Soviet government grasped the effects of testing on the nearby population and environment. The answer is unequivocal. Based on reports published in From the History of Semipalatinsk Polygon, 1951-1992, which contains declassified CPSU documents that were "top secret," such information was unambiguously available.[18] Medical professionals from Dispensary Number 4 in Semipalatinsk

8 Shkolnik, 2002
9 The head of the Soviet atomic project (and also NKVD head) was one of Stalin's key associates, Lavrentiy Beria. Beria did not con duct a thorough survey and simply presumed that the area was relatively unpopulated – a disaster for the treated population and their descendants, but important for causal inference. The story and a description of the area is available (in English) at the Caravanistan.com website – in normal times, and officials willing, they also conduct tours of the SNP. For a detailed discussion of Beria's role (in Russian) see Sudarikov (2017).
10 Demoscope, n.d.
11 Nazarbayev, 2001, p61
12 Shkolnik, 2002
13 Shepel, 2007
14 Logachev, 1997
15 Shepel, 2007
16 Shkolnik, 2002
17 Vakulchuk et al., 2014
18 Shepel, 2007

and scientists from Institute of Biophysics of the Academy of Medical Sciences of the USSR in 1958 were reporting that about 22% of observed individuals had symptoms consistent with chronic radiation sickness. There is also evidence that radiation has spread into the environment and food products. It was noted that "specific activity" of food products consumed on a daily basis was exceeding norms 2-9 times, and for butter up to 58 times.[19]

It is possible that the effects of testing would have been smaller if public safety provisions were undertaken properly. Logachev and Logacheva (2004), who studied public safety provisions at SNP during atmospheric testing split the 1949-1962 period into three time periods: 1949-1951, 1953-1957, and 1958-1962. They mention that public safety measures were practically absent during 1949-1951 but improved and included temporary evacuation of people from the potential trajectory of radioactive clouds during 1953-1957. Despite mentioned improvements in public safety provisions, it also appears that fallout problems were not considered until the last moment and that the military directors of the SNP preferred "evacuating people in a rush" to postponing a test of the first thermonuclear device. Moreover, there is an evidence that some residents of Karaul (about 95 km from SNP) and Chagan (about 110 km from SNP) villages were ordered to stay during the testing and later were used for an assessment of the effects of radiation exposure.[20]

During the 1958-1962 period, regulations and restrictions on nuclear tests are claimed to have become stricter still. At the time, it was believed that following those measures imposed sufficed to ensure public safety; therefore, based on the decision of the USSR government, the population in the vicinity of the SNP was not notified about the upcoming tests.[21] Admittedly, due to those restrictions, most of the ground tests in that period did not exceed the TNT equivalent of 0.5 kT with the exception of two explosions in August (10kT) and September (7kT) 1962.[22] These explosions occurred at the ground, and not in the air as originally planned, and are claimed to have "insignificantly" polluted the Semipalatinsk Region and Altai Krai of the Russian Federation.[23] However, contrary to public reports at the time, evidence emerged that these ground explosions had a significant effect on environment and human health. Declassified documents 10-12, which mention the faulty explosion in August 1962, report that grain in districts near the SNP had "specific activity" levels exceeding norms by 10-60 times.[24] Moreover, military personnel and representatives of the USSR Ministry of Defense recommended exporting grain out of the Semipalatinsk Region to reduce the levels of radioactive pollution.[25]

The second period of testing started in 1963, when the Partial Test Ban Treaty banned surface and air tests, restricting detonations to underground shafts and tunnels. In that period, 340 nuclear tests and "peaceful" nuclear explosions took place. Most of them occurred in the Degelen Mountains (Zone G), Balapan Zone (Zone B), and Sary-Ozen and Murzhik (Zone C). Among those tests, only 5% are acknowledged to have resulted in the release of gaseous radioactive products that exposed districts near SNP to ionizing radiation.[26] One of such tests was the 140-kiloton February 1965 "Chagan" explosion – an experiment aimed to create a man-made reservoir for water storage, which later became known as an "Atomic Lake". After the test, medical professionals of Semipalatinsk Region have reported that some localities near SNP including Semipalatinsk city were exposed to gamma-contamination.[27] Women of

19 Shepel, 2007, documents 3 and 4
20 Kassenova, 2022
21 Logachev & Logacheva, 2004)
22 Shkolnik, 2002
23 Logachev & Logacheva, 2004
24 Shepel, 2007
25 Shepel, 2007, document 12
26 Logachev, 1997
27 Shepel, 2007, document 13

exposed Semipalatinsk-4 (Chagan village) even reached out to Brezhnev and conducted demonstrations to stop the tests, at which point a commission from Moscow assured them that released dose of radiation was harmless and notified that the testing will continue.[28] Similarly, SNP authorities admitted minor release of radioactive materials, but assured the First Secretary of the Central Committee of the Communist Party of Kazakhstan (CC CPK) that the test posed no threat to the public health.[29] Moreover, the area around the explosion funnel was not properly restricted for public access, and starting from the spring of 1965 these territories were used for pasturing livestock.[30] It is likely that the health and environmental effects of the Atomic Lake explosion were understated or undercounted, and serious contamination of the areas nearby occurred. Even today, ground near the lake and the water itself still contain traces of nuclear pollution.[31]

Despite the redirection of the explosions into shafts and tunnels, tension between the SNP authorities, representatives of military-industrial complex, and the local population grew steadily. During 1964-66 alone, the Semipalatinsk regional KGB collected more than 340 letters containing information about negative reactions of local population on the tests and their environmental/public health consequences.[32] Relations between SNP authorities and local Party representatives were also tense. Local authorities were not notified about radioactive gas leaks during underground tests, and medical data acquired by the SNP authorities was secretly collected only for statistical purposes and not shared with local health representatives.[33]

One such underground explosion with the release of radioactive gases occurred on February 12, 1989. A couple of days after the explosion, radioactive gases spread through densely populated areas and even reached the city of Pavlodar (about 200-250 km from the SNP).[34] Not surprisingly, the Soviet news agency TASS simply reported that "radiological situation on the test site and outside it is normal".[35] Only because of an unexpected change in the wind direction, radioactive gases exceeding background levels by 100 times were noticed by chemical protection forces located in Chagan village and information about the release of radioactive gases became publicly available.[36] As later become known, such radioactive gas leaks were frequent and happened in about every third underground explosion: in 1987-88 period alone, radioactive gases have reached the city of Semipalatinsk eight times.[37]

In total, over 450 nuclear detonations were carried out between 1949 and 1989.[38] The last nuclear test took place in October of 1989. Still, radiation originating from the site impacted large portions of the region's population, affecting between 500,000 to one million civilians up to 400 kilometers away.[39] The site was officially closed in 1991 by President Nursultan Nazarbayev following the independence of Kazakhstan.

Despite the closure of the SNP, issues of radiological safety of the test site are still present. Today, signs warning about the radioactive pollution surround the SNP boundaries as well as the boundaries of specific test sites. Access to Degelen zone is physically restricted and protected by military forces.[40]

28 Shepel, 2007, document 18
29 Shepel, 2007, document 14
30 Logachev, 2001
31 Lukashenko et al., 2017
32 Shepel, 2007, document 18
33 Shepel, 2007, documents 19 and 21
34 Shepel, 2007, document 21
35 Shepel, 2007, document 25
36 Shepel, 2007, documents 19 and 21
37 Nazarbayev, 2001, pp 55-56
38 Mikhailov 1996; Grosche et al. 2015
39 Brunn, 2011
40 Lukashenko et al., 2017

However, territories of SNP are still illegally used for livestock pasturing, farming, hay collection, and extraction and re-use of construction materials.[41] Moreover, whether or not territory in the vicinity of Semipalatinsk nuclear test site can be used for industrial needs remains a matter of scientific dispute. Scholars from the Institute of Radiation Safety and Ecology (IRSE), in particular Lukashenko et al. (2017), argue that licensed operation and following the rules of radiation safeness guarantees that radiation levels do not exceed the bound of Hygienic Standards on Radiation Safety. This result may imply that background radiation levels are tolerable and pose no threat to human health. On the contrary, Stawkowski (2017) questions the results from IRSE. She concludes that authorities may be unwilling to restrict access to radioactive areas as territory in the SNP is valuable for industrial purposes, especially coal mining.

3. Prior Literature and Mortality Estimates

The academic literature surrounding the Semipalatinsk Nuclear Polygon has evolved over the years. One of the first studies exploring the effects of SNP were carried out by Professors S. Balmukhanov and B. Atchabarov covering the 1954-59 period. Their study was undertaken using the sample of 6000 adults from three "treated" districts – Abai, Beskaragai, Eguindy Bulak, and three "control" – Chubar Tau, Bayan Aul, Ulu Tau districts.[42] They found out that blood disorders, pathologies of skin, nails and hair, vegetovascular dystonia, asthenic syndrome, arterial hypotension, liver disorders, pathologies of the gastrointestinal tract, and irregular menstruation are more frequently observed in exposed districts compared to the controls.[43] This set of symptoms has later became known in academic literature as "Kainar Syndrome," named for the village where those symptoms were first observed.[44]

Health effects on the exposed population were also studied by a classified medical institution named 'Dispensary No. 4', which was established in Semipalatinsk in 1957. This institution was disguised as an antibrucellosis facility, but mainly served to collect health data on effects of radiation and conducted long-term studies on the health effects in the exposed population.[45] In particular, they analyzed the carcinogenic effects of radiation over 1954-1989 period, and found that the incidence of malignant tumors was 33.4% higher in contaminated areas.

Following the complaints and concerns of people near the SNP, an Interdepartmental Commission Study was carried in 1989 by Professor A.F. Tsyb.[46] The Commission revealed the following: 50% of the examined residents of the exposed villages of Kainar and Sarzhal had immune systems lower than normal, Semipalatinsk city had an incidence of pediatric diseases three times that of the rest of Kazakhstan, mental retardation rates among children in the vicinity of SNP were three times higher than the national average, and gynecological pathologies resulting in pregnancy or puerperium complications occurred 39% more than national level.[47]

Early studies after the closure of SNP were descriptive in nature. Bul'bulian and Tokareva (1991) examine the age specific morbidity rates for multiple types of malignant tumors across Kazakhstan and the former Soviet Union. They find that tumor prevalence was higher in the Semipalatinsk region

41 Lukashenko et al., 2017; Butts, 2010
42 Balmukhanov et al., 2006
43 Логинова, (н.д.)
44 Balmukhanov et al., 2006
45 Kassenova, 2022
46 Balmukhanov et al., 2006
47 ibid

than in the rest of Kazakhstan or the U.S.S.R. Gusev et al. (1998) compare cancer rates in 9 highly-exposed villages relative to a less-exposed control sample, tracking the peaks of differing types of cancer. Both the treatment and control groups had approximately 10,000 individuals each. The control area was located in central and south East Kazakhstan Province, east of the SNP. Katayama et al. (2006) document progress in constructing a database of affected individuals and health conditions, with potential use for examining trans-generational effects. They highlight problems the database faces, such as large sources of bias.

As data and methods improved, more thorough analysis of the issue was performed. Implementing the first analysis of the Semipalatinsk historical cohort, Bauer et al. (2005) find significantly higher cancer mortality rates for the most exposed population across a wide range of cancers. More recently, Grosche et al. (2011) and Markabayeva et al. (2018) examine the cardiovascular health of the affected population. Both find high rates of hypertension and cardiovascular disease present in the exposed populace relative to reference groups. These studies examine the "historical cohort," finding that the exposed population showed higher mortality due to cardiovascular disease. However, the difference in outcomes between the exposed and unexposed groups also could reflect differences in baseline rates, so no relationship between radiation dose and cardiovascular-related mortality could be established with the historical cohort. Bauer et al. (2013) look at the differences between birth cohorts of those in East Kazakhstan finds that those born during the dates closest to the nuclear tests were at higher risk of health complications, such as cardiovascular disease and cancer. The control group were persons who immigrated to the examined areas after 1990.

Teleuov (2007) examines the breast cancer morbidity among patients from zones of extreme radiation risk in East Kazakhstan and maximal radiation risk. Thyroid glands are highly radiosensitive organ, and are generally one of the first bodily locations to develop abnormalities due to high radiation exposure. He finds that thyroid nodules were significantly more likely to develop for individuals in exposed groups, and that breast cancer incidence was 1.5 times higher in East Kazakhstan than Kazakhstan overall, and 2.5 times higher than in South Kazakhstan. Land et al. (2008) use a cross-sectional study of 2994 residents in eight villages to explore thyroid disease prevalence via ultrasound screening. By reconstructing fallout deposition patterns, they determine the extent to which degree radiation exposure was external or internal (through consuming exposed food). They find increased prevalence of thyroid nodules comparable to acute radiation from the Hiroshima and Nagasaki atomic bombings.

There have also been studies looking into possible radiation effects on the children of the exposed population. Mudie et al. (2007) focus on whether radiation exposure led to change in the sex ratio of newborns. Using a sample of 3,992 exposed mothers, they find the sex ratio of the newborns of exposed mothers was comparable to the overall sex ratio of Kazakhstan at that time. However, Mudie et al. (2010) explore 191 twin deliveries from the same sample of mothers and find that there was a statistically significant increase in the odds of having different sex twins for births occurring within five years after exposure compared to twenty years after exposure in all villages in the sample.

Given the vast amount of research in the area by specialists in many fields, we close with reference to two surveys. Grosche et al. (2015) serves to provide an overview of previous studies evaluating the health effects of nuclear testing and the impact dosimetry may have on these results. For most recent historical epidemiological studies, the key data: consists of an "historical cohort" comprising 10,000 individuals from exposed villages and 10,000 individuals from an unexposed comparison area. They also note a cross-sectional study of 2,994 residents from eight exposed villages. This study involved ultrasound screening for thyroid diseases. Other highlights include discussion of Katayama (2006), which documents a team working on assembling a database of affected individuals, poten-

tially useful for transgenerational work. Discussion of limitations highlights that no studies have focused on mental disability or congenital malformation (with congenital malformation being one of the characteristics in the Medinform raion-level data we use below).

Vakulchuk et al. 2014 (NUPI Humanitarian report) provide a detailed historical account of the Semipalatinsk Test Site (STS) and the progression of the current environmental situation. They also include testimonies and perspectives on the nuclear tests and their impacts. Of note is the consensus that the aboveground tests between 1949 and 1962 are responsible for the majority (up to 95%) of radiation exposure and environmental contamination.

As we focus on long-term effects of radiation exposure, we close this section with a brief reference to prior studies on this topic. Not surprisingly, the key long-run outcome of radiation exposure is its association with risk of solid cancer. A cohort study of Atomic bomb survivors of Hiroshima and Nagasaki reveals that the risk of developing cancer throughout the life is elevated after exposure to ionizing radiation. This risk was especially elevated for bladder, female breast, lung, brain, thyroid, colon, esophagus, ovary, stomach, liver, and skin cancers.[48]

Non-cancer effects of early and late exposure to radiation are thoroughly explored by the International Commission on Radiological Protection (ICRP). ICRP's researchers conclude that long-term effects of radiation exposure include immunosuppression (under high doses), negative influence on bone marrow, inhibiting effects on gonads, skin disorders, elevated risk of mortality from circulatory and respiratory diseases, risk of developing cataracts (even at low exposure doses), lung and renal damage, disorders in musculoskeletal system (especially in children), and cognitive and behavioral defects (especially if exposure was during childhood).[49]

These ICRP findings are also supported by other studies. An analysis of Chronic Radiation Syndrome (CRS) in the Southern Ural region of Russia reveals that those diagnosed with CRS had higher risk of mortality from cardiovascular diseases, malignant neoplasms and diseases of respiratory system.[50] In addition, CRS affected eye, nervous and musculoskeletal systems in the long-run.[51] These findings are consistent with those of Stubblefield (2011), whose assessment of radiation-induced toxicity reveals that later consequences of radiation treatment may result in neuromuscular and musculoskeletal disorders. In sum, radiation exposure has been found to have various severe long-run effects.

The long-run effects of exposure of populations near the Semipalatinsk Nuclear Test Site have not been extensively studied. The novelty of our approach lies in the use of district-level population data, propensity score matching for improved comparative purposes, and a large array of health outcomes that gives us some sense of relative risks. To our knowledge, this sort of analysis has not been undertaken before in the academic literature. Using dataset from MedInfo, we can trace the consequences of atomic testing 11-29 years after the last test at SNP (38-56 years after the last atmospheric explosion). The fairly lengthy series also makes it possible to get a sense as to which elevated health risks are diminishing with time – and which are not. As Section V details, our results are consistent with the existing literature on the long-lasting effects of radiation exposure.

48 Kamiya et al., 2015
49 ICRP, 2012
50 Akleyev, 2012
51 Akleyev, 2012

4. Data: Health Indicators, Population Structure, and Night Lights

Our detailed health data come from Medinfo, which collects and maintains medical and demographic indicators on the Kazakhstani population. The data are at the raion (tertiary level: district or county) level covering 2000-2018. Included are all 170 districts, as well as certain cities. The cities included in the sample are all designated as official Kazakhstan localities and are treated as separate entities. However, not every officially designated locality is in the sample. As this sample was established in 1999, the sample includes cities that had locality status in 1999. Since then certain cities have gained and lost locality status. Any city that lost locality status was dropped from the sample, while cities that gained locality status were not added to the sample. This was verified using information from the Kazakhstan Bureau of National Statistics, often referred to as Kazstat or Goskomstat.[52] For brevity, hereafter we use "district" and "district and city" interchangeably. These data include mortality, disease incidence and healthcare variables.

Figures 1 and 2 contain maps generated by Medinfo's DPS data program. Figure 1 displays the mid-year estimate of 2018 population by oblast in 2018, and, at the district/city level, East Kazakhstan oblast. Figure 2 displays new cases of all cancers for the same areas registered in 2018. The blue polygon on each map indicates the location of the Semipalatinsk Nuclear Polygon. Evidently, cancer incidence comoves with population, but not perfectly. These maps in addition to the significant body of existing literature motivate the examination of negative health outcomes as a function of distance from the Semipalatinsk Nuclear Polygon.

Figure 1. Kazakhstan Provincial and East Kazakhstan District-Level Populations, 2018

(blue diamond demarks Semipalatinsk Nuclear Polygon)

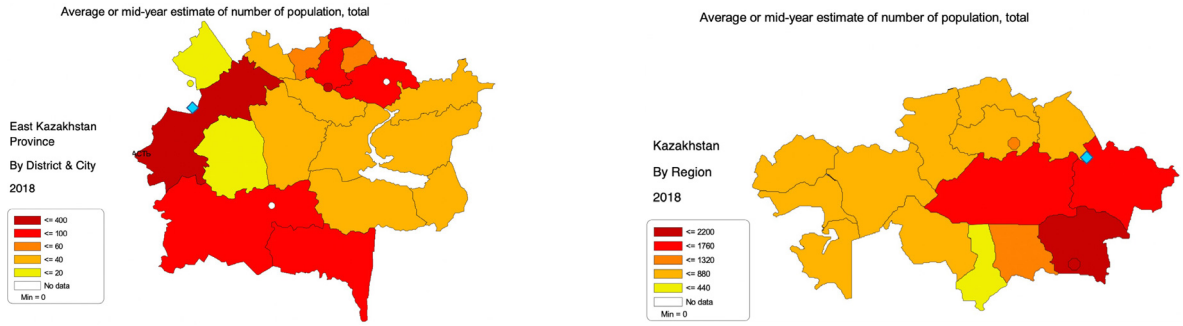
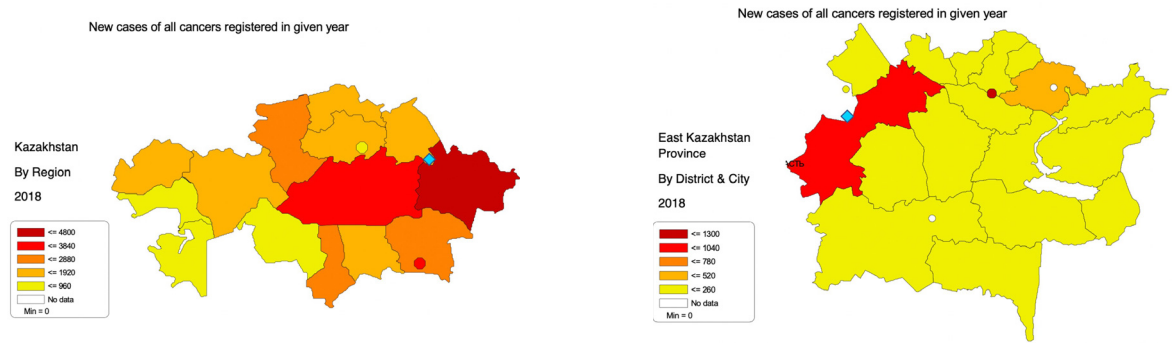


Figure 2. Kazakhstan Provincial and East Kazakhstan District-Level Cancer Incidence, 2018

(blue diamond demarks Semipalatinsk Nuclear Polygon)



52 stat.gov.kz/

4.1. Selective outmigration

A critical issue to be addressed concerns possible the outmigration of population from the East Kazakhstan Oblast where the Semipalatinsk Nuclear Polygon is located. It is plausible that people could have left those districts located adjacent to the SNP for safety purposes. If that was indeed the case, then the population that moved out of the treated areas either would have amalgamated with control or non-treatment groups, or simply resulted in a decline in the observed treated population, if those treated emigrated from Kazakhstan. In both of these scenarios, our results would be biased downward and the effects of radiation exposure would be underestimated.

However, available evidence is that outmigration of population from East Kazakhstan Oblast (EKO) was not primarily driven by the fear of exposure to radiation, and was mainly caused by urbanization and abolishment of the propiska system after the dissolution of the USSR. As Appendix Table A3 documents, all of the districts in East Kazakhstan Oblast experienced a sharp population decline between 1989 and 1999 but did not have large population declines prior to 1989 – that is, during the period of atomic testing.

Most of the districts continued to experience decline in population throughout 1999-2018 with the exception of the cities of Semey (Semipalatinsk), Oskemen (Ust-Kamenogorsk) and Kurchatov. Such trends can be attributed to movement of population from rural to urban areas. This pattern occurred throughout Kazakhstan, and was not limited to EKO.

However, while urbanization can explain movement of population from countryside to cities, it does not explain EKO's overall population decline from 1989-2018 to. Our explanation for this phenomenon is revealed by Appendix Table A4, which provides oblast population dynamics by nationality (unfortunately, such data by nationality are not available for earlier years). Table A4 reveals that huge share of Kazakhstan's European population emigrated (mainly to Russia, Ukraine, and Germany) during 1989-1999, with more continuing to leave thereafter. Such outmigration was not just limited to EKO: it was common across Kazakhstan and can be attributed to the dissolution of USSR. Upon closer inspection of Table A4, it is evident that the ethnic Kazakh population retains stable growth over 1970-2018 period for most of the oblasts.

Apart from urbanization and collapse of the USSR, there is further reason to believe that most of the radiation-exposed population did not leave the region or mix with control or non-treatment groups. Appendix Table A5 contains information on the number of beneficiaries of supplemental "Semipalatinsk Polygon" pensions from the Government of Kazakhstan. In order to be eligible for such pensions, people had to reach the retirement age by January 1, 1998 and reside in the treated areas at the moment they entered retirement.[53] After such pensions were assigned, these people could potentially move anywhere within Kazakhstan and still receive the pensions. However, as the Table A5 shows, most of the "Semipalatinsk pensioners" resided in East Kazakhstan and adjacent Pavlodar Oblasts even after 2002.

In short, the Soviet system prevents large-scale population movements prior to the 1980s, and from 180 onward, migration was dominated by ethnic rather than radiation exposure concerns, and by rural-urban migration within treated areas (since the largest cities in EKO are in high exposure zones). Those raions in EKO with the largest population losses tended to be in minimally exposed areas near the Chinese border and far from the SNP.

53 https://egov.kz/cms/en/articles/ecological_disaster_zones

4.2. Determining exposure and "treated" regions

There are two broad options to determine exposure. First, we can use official Kazakhstan Government assessments of exposure in determining eligibility for supplemental pensions and health care support. There are five official designations for those living in (1) maximal, (2) extreme, (3) high, (4) low, and (5) no exposure raions during the era when tests were conducted. We focus on this approach, since it reflects more careful assessment than alternatives, and also because we have oblast-level information on the number of recipients of these supplemental payments who have moved to other areas. While we cannot use this information directly in the propensity score matching approach discussed below, we can use it to exclude certain regions (namely, the nation's largest cities).

Alternatively, we could use estimates of wind direction and contamination levels of the largest tests. Gusev et al. (1997) provide a helpful map, and nearly all studies until recently (since the supplemental payments are recent) used this approach. Results are not sensitive to the type of treatment measure used, and we provide supplemental regressions using both approaches.

Our measure of exposure merits further detail, and, fortunately, government websites provide this information in detail. As a consequence of the continuous nuclear tests in the SNP, the Kazakhstani government offers social benefits and monetary compensations for people who live, have lived, worked in, or completed their military service in territories recognized as "zones of radiation risk".[54] More precisely, "zones of radiation risk" are defined as those territories where the dose of the population's exposure exceeded the amount of 0.1 rem over the 1949-1990 period. These zones are divided into 4 categories depending on the level of radiation exposure: zone of emergency radiation risk (>100 rem), zone of maximal radiation risk (35-100 rem), zone of high radiation risk (7-35 rem), and zone of minimal radiation risk (0.1-7 rem).[55] Additionally, the government defines a territory with "beneficial social economic status." This territory is adjacent to the zone of minimal radiation risk, and though the dose of the population's exposure was less than 0.1 rem over the entire testing period, people experienced a significant stress of living near radiation and seismic activities.[56] Altogether, these territories are specified below:[57]

- Zone of emergency [extreme] radiation risk: Sarzhal rural area of Abay district, Dolon rural area of Beskaragay district. Settlements Sarapan and Isa of Zhanasemey district. All of these territories are part of East Kazakhstan oblast.
- Zone of maximal radiation risk: Abay, Beskaragay and abolished Zhanasemey region, Akbulak, Abralin, Algabas, Ainabulak, Karaolen, and Tanat rural areas of Semey city - East Kazakhstan oblast; Akzhar and Maldar rural areas of Mayskiy district - Pavlodar oblast.
- Zone of high radiation risk: Borodulikha, Zharma, Ayaguz, Glubokovskiy, Shemonaikha, Ulan districts of Semey city - East Kazakhstan oblast. Kurchatov city, Ust-Kamenogorsk city, Ridder city - East Kazakhstan oblast. Karkaraly district of Karaganda oblast within the territory of the now-abolished Kazybek district. Maiskiy district of Pavlodar oblast.
- Zone of minimal radiation risk: Urdzharskiy, abolished Taskesken district, Kokpekty, Aksuat, Altai, Zaisan and Tarbagatay districts of East Kazakhstan oblast. Karkaraly region not including the territory of the abolished Kazybek region. Akkuly district of Pavlodar oblast.
- Territory with beneficial social economic status: Bayanaul district Pavlodar oblast.

54 https://egov.kz/cms/en/articles/ecological_disaster_zones
55 https://adilet.zan.kz/eng/docs/Z920003600_
56 ibid
57 https://egov.kz/cms/en/articles/ecological_disaster_zones

Apart from zones of radiation risk, some parts of Kazakhstan are classified as ecological disaster zones. Depending on the degree of severity of ecological conditions, these territories are split into three categories (More on that on website):[58] Zone of ecological catastrophe, Ecological crisis zone, and Zone of ecological pre-crisis condition. These zones are described in detail below:[59]

- Zone of ecological catastrophe: Aral and Kazaly districts of Kyzylorda oblast, Chelkar district of Aktobe oblast.
- Ecological crisis zone: Kyzylorda oblast (except for the Aral and Kazaly Districts), the city of Kyzylorda and the city of Baikonur, including settlements that are part of its administrative and territorial subordination.
- Zone of ecological pre-crisis condition: Baigany, Irgiz, Mugalzhar (within the boundaries of the settlements of the former Mugalzhar district), and Temir districts of Aktobe oblast. Arys (including the city of Arys), Otrar, Suzak, Chardara districts and the cities of Turkestan. Ulytau district (within the boundaries of settlements of the former Zhezdy district of the Zhezkazgan oblast) of the Karaganda oblast.

A complicating factor is that the SNP was not the only part of Kazakhstan to experience nuclear explosions. At various times, the USSR conducted 124 non-military ("peaceful") underground nuclear explosions for industrial purposes in areas outside of military polygons; 39 of these were in Kazakhstan (32 outside of SNP).[60] These "peaceful nuclear explosions" occurred in Mangistau, West-Kazakhstan, Aktobe, Kostanay, Akmola, South-Kazakhstan and East Kazakhstan oblasts.[61] The main reasons were opening up oil or gas fields, capping oil or gas plumes, and accessing other mineral deposits. Since extraction was intended, radiation was more likely to escape.

Among these, the Azgirskii Polygon lies on the border of West Kazakhstan and Atyrau oblasts (provinces). As Urgushbaeva et al. (2015) detail, medical experts claim that health in this area was 2.0-2.5 times worse (a very unclear statement) in this region than in the rest of Atyrau, especially for children.[62] Apart from peaceful nuclear explosions and nuclear tests for military purposes, Kazakhstan has uranium mine tailings located in North Kazakhstan, Almaty oblasts. Unfortunately, these other irradiated and polluted areas have not been graded for level of exposure. Therefore, we create a single, separate dummy variable eco_fnd containing these polluted districts, which are represented as "zone of ecological findings". Altogether, zones of radiation risk and zones of ecological disasters are presented in Figures 3a and 3b below:

58 https://adilet.zan.kz/rus/docs/Z920002600_
59 https://egov.kz/cms/en/articles/ecological_disaster_zones
60 Abishev et al., 2016
61 ibid
62 Urgushbaeva et al. (2015) focus on this last case, where there was an industrial nuclear explosion in 1987. Specifically, they explore outcomes in the village of Kaldaibek in Baiganinskii raion in Aktobe oblast. Naturally, there was a river running through it as well to assist in downstream contamination. Estimated life expectancy at birth in Baiganinskii raion after the explosion: 47 years, vs. 55 in Aktobe oblast overall, and 65 in the USSR (all horrendous numbers). Even though observed radiation by 2009-13 was no longer elevated, cancer incidence in the raion remained elevated. Hence, control areas may not have been all that controlled (though the last 1987 explosion was only 8.5 kilotons, far from the record, which was approximately 480, in 1950s SNP). Readers interested in details are directed to an excellent Wikipedia piece on the Nuclear_Explosions_for_the_National_Economy program, which in turn links to the American counterpart, Operation Plowshare. The Russian Wikipedia entry (also excellent) on the topic lists the location of each of these explosions: Мирные_ядерные_взрывы_в_СССР (peaceful atomic explosions in the USSR).

Figure 3a. Zone of Radiation Risk

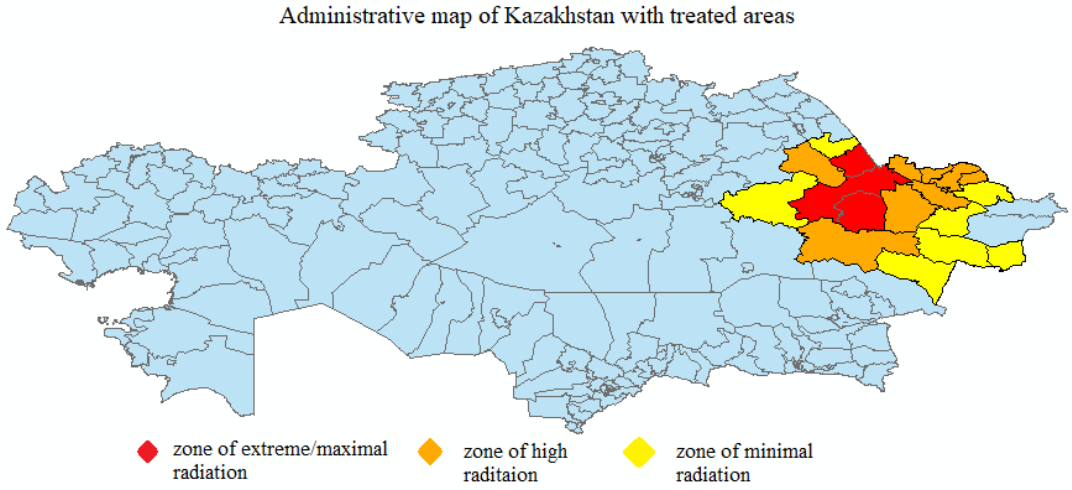
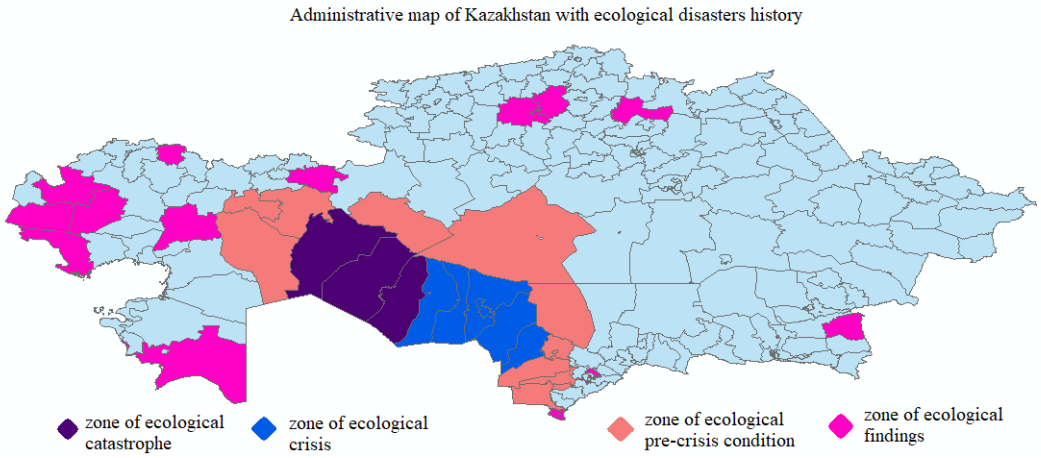


Figure 3b. Zone of Ecological disasters



Knowing (and having official designation for) these other disaster zones is particularly useful for our matching efforts, since it is important that the control districts are not areas that have simply received different, disastrous treatments. The sheer number of radiation-exposed and other ecological disaster regions is staggering, and a grave indictment of Soviet (since virtually all disaster zones reflect Soviet rather than post-Soviet treatments) government indifference to the populace at large, and especially to those outside of the Russian SFSR.

Semipalatinsk pensions in Russia: Today the story is slightly different, as the radiological effects of nuclear tests at the SNP are recognized outside Kazakhstan as well. In Russia, in accordance with the Federal law № 2 dated 10.01.2002, those who were affected by the nuclear tests are eligible for Monthly Benefits (MBs) payable by the Pension Fund of the Russian Federation (PFR).[63] Individuals who resided in the period 1949-1963 in one of the localities recognized by Russian Government as "exposed to radiation due to nuclear tests at SNP", and were exposed to the amount of radiation exceeding 5 rem, are eligible for MBs.[64] There are two types of MBs: one for those whose radiation exposure exceeded 25 rem, the other for those whose radiation exposure is less than 25 rem, but

63 https://pfr.gov.ru/grazhdanam/federal_beneficiaries/
https://pfr.gov.ru/en/pens_system/types_soc/
64 https://docs.cntd.ru/document/901808295#7DK0K9

greater than 5 rem.[65] Appendix Table A1 details pension policies toward SNP radiation exposed individuals in both Kazakhstan and Russia.

4.3. Comparing demographic structures: age-gender and nationality distributions

In order to match treated and untreated districts, it is essential to capture characteristics of the health care system, demographic structure, and economic structure.[66] Few if any health measures, whether relating to mortality, morbidity incidence, or morbidity prevalence, are age- and gender invariant. In principle, we could address this either by calculating age/sex-standardized health measures, or by using population distributions as characteristics on which we match. We choose the former approach in the belief that differences in population structure may be correlated with unobserved factors that matter as well, and that matching will pick up some of this effect.

Kazstat's demographic statistics page provides raion-level age-gender distribution data.[67] These data are available for 2005 through 2021; we use the 2005-2018 subset. Data for 2000-2004 are absent at the district level. Consequently, we take age-gender distributions from the 1999 Census of Kazakhstan and linearly interpolate the missing years.[68]

These data are not perfectly aligned. For 2005-18, data for each gender are split into three bins. For males, these bins are years 0-15, 16-62, and 63+. For females, these bins are years 0-15, 16-57, and 58+. These distributions represent children, working-age population, and retirees (women are eligible to retire five years earlier than men). However, the 1999 census data contains 16 bins. Each gender is split as follows: less than 1, 1-4, 5-9, 10-14, 15-19, 20-24, 25-29, 30-34, 35-39, 40-44, 45-49, 50-54, 55-59, 60-64, 65-69, and 70+. In order to make the 1999 census data comparable with the 2005-2018 dataset, we assume that each specific year is equally represented within the 5-year bins and then allocate them to the three bins per gender available from 2005 onward.

A second problem we face is that different raions will have different nationality compositions. Nationality matters for two reasons. First, Slavic Europeans in Kazakhstan have lower life expectancies than

65 https://docs.cntd.ru/document/901808295#7DK0K9
The list of localities of former KazSSR recognized by Russian Government as "exposed to radiation due to nuclear tests at SNP" (https://docs.cntd.ru/document/9010855#64U0IK) (where some Russians lived includes:
Exceeding 25 rem:
Semipalatinsk Oblast (now East Kazakhstan Oblast) Beskaragay district: Алгобас (Алгабас), Бодене (Буденя), Долонь, Жана-Куш (Жанакуш), Канонерка, Карабаш, Комсомольский, Кордон Беркаин, Кордон Вышка, Кордон Встречный, Кордон Новый, Подорел, Сольпром, Ферма Планке, Чаган, Станция Чаган, Черемушка. Borodulikha district: Алексеевка, Верхняя Жайма, 2-е отделение (бригада) совхоза Коростелевский, Жайма, Казбек, Коростели, Разъезд N 39, Семеновский, Степановка, Тарск (Тарский), Толумгожа. Zhanasemey district: Байтанат, Кордон Бугорок, Кордон Ерусалик, Кордон Литовченковский, Кордон Теплый Угол, Лесничество Тюмень, Мещанка (Мещанский), Николаевка, Оторвановка.
From 5 rem to 25 rem:
Semipalatinsk Oblast (East Kazakhstan Oblast) Beskaragay district: Бестерек, Известковый, Карамурза (Карамырза), Кара-Тагай, Кордон Гилик, Кордон Джемур, Мостик, Кордон Тополька.Borodulikha district: Ивановский, Киякпай (Сарбас). Zhanasemey district: Зыряновский, Кордон Дальний, Кордон Пограничный, Курмангужа, Молдары (Курчатов), Станция Конечная.
Pavlodar Oblast Mayskiy district: Кызылкудук.
There is also a list of localities in Russia affected by SNP radiation (https://docs.cntd.ru/document/9007771#64U0IK). For this reason, data on MBs for "Semipalatinsk polygon" (https://pfr.gov.ru/opendata/~7706016118-edinovremennayavyplata) may represent local population in addition to those who possibly moved to Russia from Kazakhstan. It is also noteworthy that Russian Social Benefits are only cover victims of exposure to radiation, but not victims of ecological, chemical, biological disasters (https://pfr.gov.ru/branches/moscow/news~2019/03/28/178877, https://pfr.gov.ru/grazhdanam/federal_beneficiaries/ and https://pfr.gov.ru/en/pens_system/types_soc/.

66 available from MedInform and also from KazStat

67 https://stat.gov.kz/official/industry/61/statistic/5

68 https://stat.gov.kz/for_users/national/1999

ethnic Kazakhs, at least for part of this period, and especially for men.[69] Second, ethnic Russians, Ukrainians, and Belarusians are far more likely to emigrate than Kazakhs or other Asian populations, and emigration is strongly age- and health-dependent. Since young, healthy adults are more likely to emigrate than elderly, frail counterparts, observed death and morbidity rates are likely to rise more rapidly in heavily European areas, especially as we do not have age-specific health data.

The best we can do is to observe overall population numbers in treated and matched untreated districts, and also to compare nationality compositions from Census data. As it turns out, and as we discuss below, matched regions tend to be more European than treated regions, thereby biasing our findings toward zero rather than inflating them.

5. Matching and Regression Models for Districts and Satellites

We focus on the potential relationship between radiation exposure as measured by zone status and the list of health variables. To explore this relationship, we explore the following regression specifications:

1. Random Effects models

$$PopHealthChar_{it} = \alpha + \beta_1 radiation\ exposure\ zone_i + \beta_2 environmental\ disaster\ zone_i + \beta_3 age\ structure_{it} + \beta_4 economic\ development\ level_{it} + \beta_5 nationality\ structure_{it} + \varepsilon_{it}$$

2. OLS regressions with time dummies

$$PopHealthChar = \alpha + \beta_1\ radiation\ exposure\ zone + \beta_2\ environmental\ disaster\ zone + \beta_3\ age\ structure + \beta_4\ economic\ development\ level + \beta_5\ nationality\ structure + \beta_6\ time + \beta_7\ time^2 + \varepsilon$$

3. OLS regressions with time dummies and zone interactions

$$PopHealthChar = \alpha + \beta_1\ radiation\ exposure\ zone + \beta_2\ environmental\ disaster\ zone + \beta_3\ age\ structure + \beta_4\ economic\ development\ level + \beta_5\ nationality\ structure + \beta_6\ time + \beta_7\ time^2 + \beta_8\ radiation\ exposure\ zone * time + \beta_9\ radiation\ exposure\ zone * time^2 + \beta_{10}\ environmental\ disaster\ zone * time + \beta_{11}\ environmental\ disaster\ zone * time^2 + \varepsilon$$

4. Year fixed effect models

$$PopHealthChar_{it} = \alpha + \beta_1\ radiation\ exposure\ zone_i + \beta_2\ environmental\ disaster\ zone_i + \beta_3\ age\ structure_{it} + \beta_4\ economic\ development\ level_{it} + \beta_5\ nationality\ structure_{it} + \beta_6 year + \varepsilon_{it}$$

69 Becker & Urzhumova, 2005

5. Year fixed effect models with zone interactions

$$PopHealthChar_{it} = \alpha + \beta_1\, radiation\ exposure\ zone_i + \beta_2\, environmental\ disaster\ zone_i + \beta_3\, age\ structure_{it} + \beta_4 economic\ development\ level_{it} + \beta_5 nationality\ structure_{it} + \beta_6\, year + \beta_7 radiation\ exposure\ zone_i * year + \beta_8\, environmental\ disaster\ zone_i * year + \varepsilon_{it}$$

where *i* is district or city, *t* is year, *PopHealthChar* is a district-level measure of one of the available population health characteristics. *Radiation exposure zone* is a vector of binary variables indicating radiation exposure level ("no exposure" is the omitted term). Correspondingly, *environmental disaster zone* is a vector of binary variables indicating the presence and severity of other environmental disasters ("no exposure" is the omitted term).

We also seek to control for population structure and economic development levels. These terms are intended to sop up unobserved heterogeneity that could be correlated with variables of interest and outcomes, but there is no causal implication. The control variables are as follows: *age structure* is a vector of terms reflecting the district's population age composition (children, working age, retirees; we also control for gender composition), *nationality structure* is a vector of terms reflecting the district's population ethnic composition. Finally, since we do not have direct measures of economic structure or per capita incomes, we use luminosity (night light) characteristics as our measures of *economic development level*.

Given the fact that we have a long list of explanatory variables, it is possible that some of them may be unnecessary for our regression analysis. In order to narrow down the list of the right-hand side variables, we analyzed the correlation table presented below.

Table A1. Correlation coefficients between potential explanatory variables

	STD	MEAN	MAX	VARIETY	MEDIAN	SUM	pct_kaz	pct_rus	pct_ukr	pct_ger	pct_tat	pct_ urban	city	pop_ln
STD	1.0000													
MEAN	0.7757	1.0000												
MAX	0.6753	0.4630	1.0000											
VARIETY	0.6581	0.3282	0.9530	1.0000										
MEDIAN	0.6752	0.9849	0.3937	0.2497	1.0000									
SUM	0.4312	0.1647	0.6354	0.6832	0.1029	1.0000								
pct_kaz	-0.3852	-0.3033	-0.3070	-0.2998	-0.2707	-0.2879	1.0000							
pct_rus	0.4176	0.3541	0.3461	0.3362	0.3254	0.3025	-0.9096	1.0000						
pct_ukr	0.0493	0.0337	0.0261	0.0236	0.0295	0.0966	-0.6261	0.4492	1.0000					
pct_ger	0.0442	-0.0053	0.0744	0.0884	-0.0149	0.1818	-0.7320	0.6145	0.7544	1.0000				
pct_tat	0.3783	0.3511	0.2883	0.2655	0.3228	0.2860	-0.6402	0.6209	0.4293	0.4521	1.0000			
pct_ urban	0.6997	0.6303	0.5827	0.5367	0.5694	0.2930	-0.2348	0.3278	-0.0559	-0.0793	0.3404	1.0000		
city	0.8153	0.7578	0.5119	0.4543	0.6896	0.2797	-0.2771	0.3420	-0.0182	-0.0634	0.3830	0.8009	1.0000	
pop_ln	0.6299	0.4657	0.5445	0.5471	0.4092	0.4660	-0.0981	0.0642	-0.2368	-0.2121	0.0875	0.4785	0.4846	1.0000

Age-gender population structure variables are not listed in the table above because we include them in all regression models. The primary reason for doing so is that virtually all illnesses and other health characteristics vary substantially over the life cycle and across genders. An additional reason for including age-gender population structure variables is selective outmigration. For the similar reasons, we also include ethnic composition variables. Given the fact that the Kazakh and Russian populations are dominant across a large majority of districts in Kazakhstan, we decided to retain both *pct_kaz* and *pct_rus* variables that reflect percentage of Kazakh or Russian population of each district, despite their high correlation. We also added percentage of Ukrainian population, because apart from Kazakhs and Russians, Ukrainians are one of the most important ethnic groups in Kazakhstan, and especially in some of the exposed areas.[70] Many of the exposed regions will have higher Slavic population shares, and they would be likely to have been the ones to experience more selective outmigration. Since we do not know the nationality composition of outmigrants (mostly emigrants), we control for population change and also nationality composition. Fortunately for our purposes, the *propiska* internal passport system would have greatly constrained these outflows in the Soviet era, and the passing of time means that selective out-migration likely would have been less in the Independence post-Soviet era.

Healthcare quality measurements available to us include number of *physicians per 100000*, number of *nurses per 100000* and *total number of hospital beds per 100000*. These terms are excluded from regressions because of the endogeneity concerns, though reading through the released classified documents suggests that the entire region suffered from shocking neglect, regardless of productivity.

Endogeneity concerns also apply to economic development measures. Higher income and more educated populations are likely to receive more medical treatment and to be self-aware of health conditions, and thus to have them diagnosed. These effects will be offset for some measures by earlier and superior diagnosis and treatment. By implication, both economic development and health care quality measures may have markedly different effects on measures involving diagnosis than on those involving survival. Yet another complication arises in the (post)-Soviet context: highly developed districts are likely to experience a higher level of other toxins that affect health outcomes. This makes controlling for development all the more important, though it also is critical that this relationship has changed over time as Kazakhstan's economy has become noticeably "cleaner." For these reasons, we add economic development variables, which are substituted by satellite night light data (NLD). However, since there are possible biases in both directions, we are cautious in interpreting coefficients and include the terms largely as controls.

In an effort to capture both overall level of economic development and income/wealth distribution, we include terms related to the first two moments of the distribution of the night-lights for each district, incorporating both the mean *(MEAN)* and standard deviation *(STD)*. Among other commonly reported NLD measures, it can be noticed that maximum *(MAX)* and variety *(VARIETY)*, and mean and median *(MEDIAN)* values have correlation coefficients very close to one. Hence, we exclude variety and median out from the list of regressors. Conversely, the sum of the district night lights (SUM) is not as correlated with *MEAN*, *MAX* and *STD*, and is added to the list of right-hand side variables. Other measures of economic-demographic activity such as percent urban population *(pct_urban)*, indicator for city *(city)* and natural logarithm of population of the district *(pop_ln)* are excluded because of the relatively high correlation coefficient values with the variables selected already.

70 Perhaps the most prominent Ukrainian-Kazakhstani is Wladimir Klitschko. Born in Semipalatinsk, he was twice world heavyweight boxing champion and is the brother of Kyiv Mayor Vitali Klitschko.

As the data are time-series in nature but the key variables of interest *(Radiation exposure zones)* are time-invariant, a random effects model is first examined. Table 1 displays coefficients for dummies of radiation exposure zones and environmental disaster zones under the random effects model. All other regression specifications are provided in the Online Appendix. Regressions for every health characteristic of interest are organized as follows: Columns 1-3 display random effects models for restricted time domains – 2000-2005, 2006-2012 and 2013-2018, respectively. Column 4 represents an OLS regression model with time and time squared coefficients. The variable *time* is defined as *time = year - 2000.* Column 5 provides an extension of the model in column 4 with additional radiation and environmental disaster zone interactions with time. Column 6 displays a year fixed effects model as a method for analyzing the time trends of the variables of interest. Column 7 provides an extension of the year fixed effects model. We create two joint treatment dummies *zrr (zones of radiation risk)* and *eco (ecological zones)* for all districts in radiation and environmental zones and analyze their interactions with year fixed effects. Such a model allows us to analyze the differences in time trends between "treated" and "untreated" areas. Additionally, we run the year fixed effects models with interactions for each radiation or environmental zone, but did not add it to the already vast Online Appendix due to space concerns. These sets of regressions are available upon request.

As noted above, we employ two approaches in estimation. The first is regression analysis, using the models outlined above. The second is to engage in a careful matching exercise, using an Epanechnikov kernel approach by first estimating a series of Probit treatment propensity models over our vector of controls. We estimate this sequentially considering maximal and extreme exposure districts (combined) first, followed by high treatment, and then low or "minimal" treatment regions. Health outcomes between treated and matched neighbors are then compared, giving an estimate of the effect of exposure to different levels of radiation for the years 2000-18.

5.1. Matching satellites to create a consistent luminosity database by *raion* using ArcGIS

The satellite night-lights data originate from Defense Meteorological Satellite Program (DMSP) of the US Air Force. This program has focused on collecting visible and near infrared lights from nighttime human economic activity. The main dataset spans 1992-2013 period and combines observations from six satellites. In some years, data is collected by one satellite, while for the majority of the dataset years overlap between two satellites. There are 34 satellite-year observations in total. For the purposes of this research, we use the "stable_lights.avg_vis" product that contain images of nighttime world with persistent source of lighting.[71] These images rely on a Digital Number (DN) to represent luminosity on 0-63 scale, where 63 represents most luminous parts of the Earth, while 0 represents no lights/background noise. Despite preprocessing, these datasets require removal of gas flares (which are common in parts of Kazakhstan).

Recently, the DMSP dataset has been extended through 2019 by the Earth Observation Group (EOG). This extended dataset contains 10 additional satellite-years and spans from 2014 until 2019. In the original DMSP dataset, the satellites that collected the data were following a day/night orbit. Over time, these satellites started to shift to a dusk/dawn orbit, which resulted in DMSP extension series images becoming dimmer than the original 1992-2013 series.[72]

71 Elvidge et al., 1997
72 https://eogdata.mines.edu/products/dmsp/#v4_dmsp_download

Intercalibration: DMSP satellites lack on-board calibration and the dataset presents DNs rather than radiance. Consequently, it is difficult to perform any sensible comparison between images from different satellites or different years. For this reason, an intercalibration process needs to be performed.[73] This paper adopts the intercalibration method used in Elvidge et al. (2009) to make yearly satellite night lights comparable to one another.

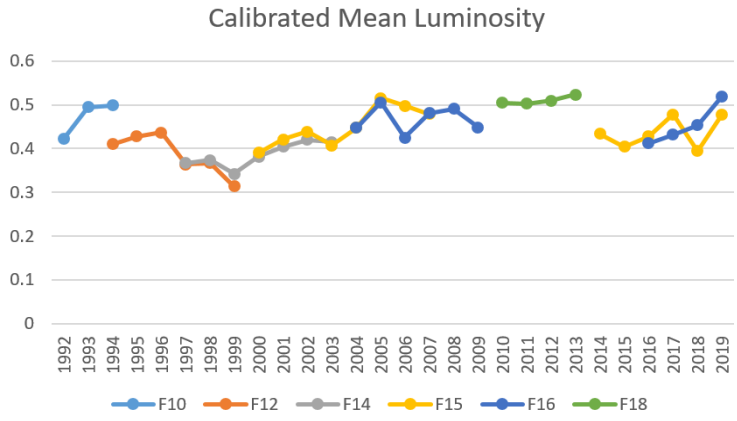
The intercalibration procedure described in Elvidge et al. (2009) relies on the selection of the reference satellite-year and reference region. After reviewing the data for multiple candidate satellite-years, we observed that satellite F18 from the year 2010 had the highest average DN and had the highest digital values. For these reasons, data from F182010 are used as the reference and all the other images are calibrated to adjust the range of the reference satellite-year.

In order to do this adjustment, the intercalibration procedure requires selection of a reference region. The data from this region are then used to fit a second order polynomial that adjusts the DN for all other satellite years. Important characteristics of a reference region include a full dynamic range of DN values and a reasonably even distribution of DN values. We ultimately select Bukhar-Zhyrauskiy and Pavlodarskiy districts, including the cities of Pavlodar, Karagandy, Temirtau and Saran, for their favorable characteristics.

These districts are heavily industrialized and remain "luminous" throughout the time span of the dataset with maximum DNs not falling below 61.

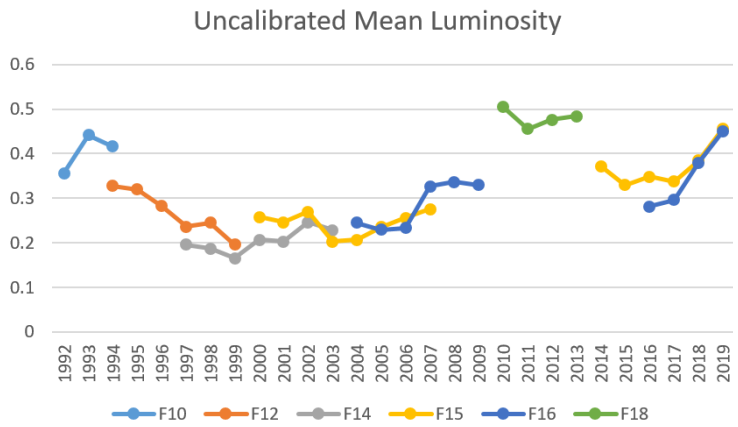
Elvidge et al. (2009) argue in favor of comparison of satellite measures from overlapping years as an effective way to assess the successfulness of an intercalibration process. Most obviously, measures from the calibrated overlapping satellites should be closer than the raw measures from the same satellites. Figures 3a and 3b represent calibrated and uncalibrated mean luminosity of Kazakhstan throughout 1992-2019. From the figures, it is evident that intercalibration improves the measures from the satellites significantly. However, for 4 years out of 28 full convergence was not achieved. One example of incomplete convergence are satellites F15 and F16 for the years 2018 and 2019. This absence of full convergence can result from the difference in the overpass times of the satellites.[74]

Figure 4a. Calibrated mean luminosity, Kazakhstan 1992-2019



73 ibid
74 Elvidge et al., 2009

Figure 4b. Uncalibrated mean luminosity, Kazakhstan 1992-2019

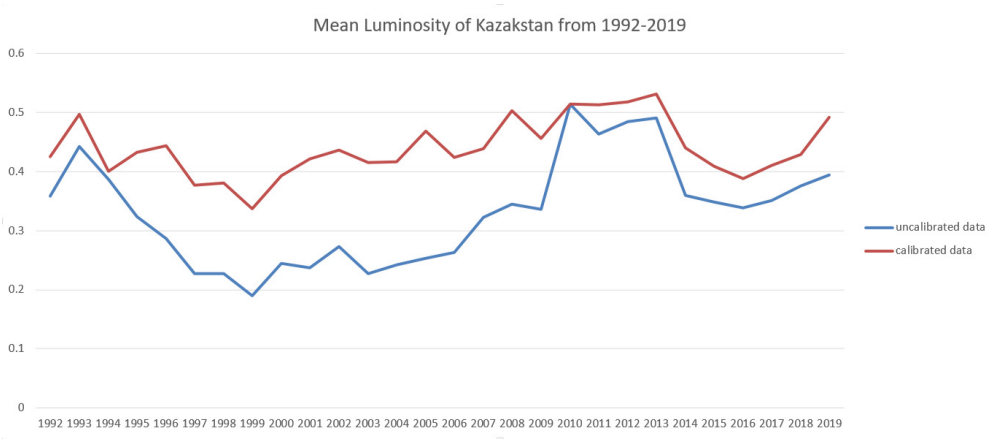


Specifically, the intercalibration process creates adjusted DN values from application of the formula as expressed in equation:[75]

$$DN_{-adj} = C_0 + C_1 * DN + C_2 * DN^2$$

The intercalibration coefficients, based on just under 40,000 data points in the chosen districts for each observation, are derived from regressions in Stata and are reported in Appendix Table A2. These intercalibration coefficients are applied throughout Kazakhstan and allow us to estimate both national and local patterns. Figure 4 provides Kazakhstan's mean luminosity over time based on both calibrated and uncalibrated data. The post-Soviet collapse, recovery, 2015-16 recession, and subsequent recovery are picked up by both calibrated and uncalibrated measures, but the patterns in the latter measure exhibit far larger swings. The large divergence between the two series from 1995-2007 should serve as a caution against merging uncalibrated data. It is also worth noting that mean luminosity values are extremely low, reflecting Kazakhstan's vast size and low population density. Patterns restricting observations to data points with luminosity values greater than zero (or 4 or 6, as is common in the literature) are similar to those in Figure 4c.

Figure 4c. "Kazakhstan" mean luminosity since Independence



75 https://ru.wikipedia.org/wiki/%D0%96%D0%B0%D0%BC%D0%B1%D1%8B%D0%BB%D1%81%D0%BA%D0%B0%D1%8F_%D0%BE%D0%B1%D0%BB%D0%B0%D1%81%D1%82%D1%8C

Mapping each data point into specific *raions* and cities is not a trivial matter, since boundaries are periodically changed. The shapefile on administrative boundaries of Kazakhstan was taken from The Humanitarian Data Exchange website.[76] This shapefile contains information on country/oblast/raion-level divisions of Kazakhstan as of 2019. Because the MedInform health dataset covers 2000-2018, some of the districts in the health dataset are outdated and have different boundaries in 2019. For this reason, the 2019 administrative division shapefile was modified to match MedInform data. Specifically, the following changes were made:

- Kegen district, which was formed in 2018, was merged into Rayimbek district in our 2019 shapefile, because it was part of Rayimbek district in the MedInform dataset.
- For the same reason, Zhetisay and Keles districts from Turkestan oblast were merged, respectively, into Maktaaral and Saryagash districts.
- The MedInform dataset has no data on Baikonur city, which is part of Karmakshy district in the Kyzylorda oblast. For this reason, Baikonur city was added to the Karmakshy district in the modified shapefile.
- In the 2019 shapefile, Baizak, Zhualy and Jambyl districts in Jambyl oblast were incorrectly represented and their boundaries were redefined to match their actual boundaries from administrative map of **Jambyl oblast.**[77]

6. Findings

6.1. Propensity Score Matching (PSM)

In the absence of a well-defined control group, we adopt a propensity score matching method to define respective controls for extreme and maximal, high and minimal radiation risk zones. This method allows us to match treatment groups to observations as similar as possible based on a list of socio-economic characteristics. Given the limited number of districts and the underlying problem that much of Kazakhstan was "treated" in one or more ways, finding comparable, untreated matches is extremely difficult. In order to alleviate these issues and improve matching quality we have chosen kernel matching and imposed a common support for treatment and control groups. The advantage of this specific matching technique is that more information is used in estimation of counterfactual outcomes because kernel matching utilizes weighted averages from all observations in the control group.[78]

Having selected the matching technique, we turn to the terms on which we match. Such characteristics are: percent males aged 0-15, percent females aged 0-15, percent male 16-62, percent male 63+, percent female 58+; mean, standard deviation and variety of night lights of a district, percent Kazakh and percent Russian of a district, and nurses per 100,000 population. Unlike the regressions, the list of controls is different in two ways. Firstly, PSM is sensitive to the number of variables used in the matching process. The more variables we control for, the harder it is for the algorithm to find adequate matches. For this reason, we use fewer controls compared to regressions. Secondly, we want to match based on the healthcare quality of the districts and are less concerned with endogeneity issues in PSM compared to the regressions. Hence, we add nurses per 100 000 as a proxy for healthcare quality of the district. Despite our best efforts, heterogeneity of the districts of Kazakhstan makes

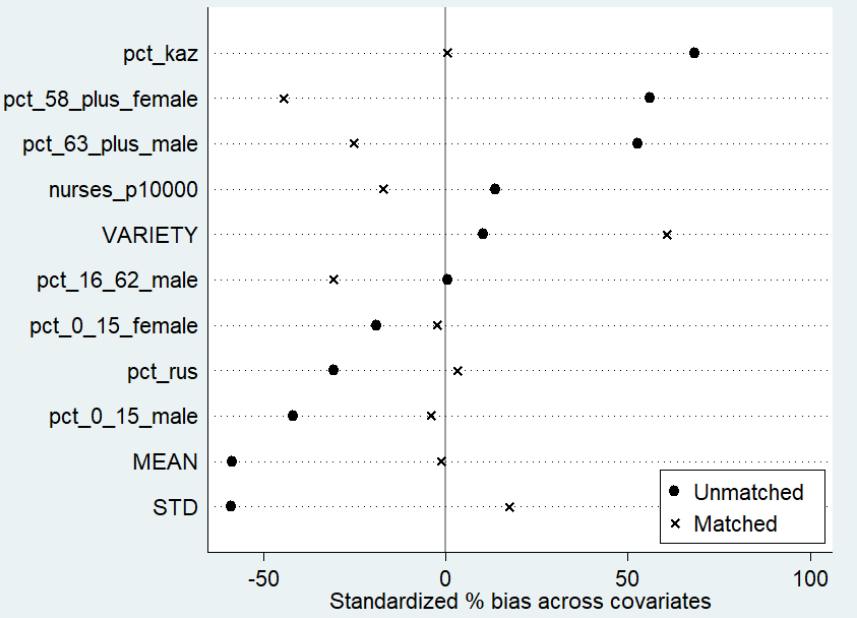
76 https://data.humdata.org/dataset/kazakhstan-administrative-boundaries-taxonomy
77 https://ru.wikipedia.org/wiki/%D0%96%D0%B0%D0%BC%D0%B1%D1%8B%D0%BB%D1%81%D0%BA%D0%B0%D1%8F_%D0%BE%D0%B1%D0%BB%D0%B0%D1%81%D1%82%D1%8C
78 Caliendo & Kopeinig, 2008

the matches imperfect. The quality of the matching as well as potential reasons for differences across groups are discussed below in greater detail.

Zone of Extreme and Maximal Radiation Risk

In this treatment group, 4 out of 57 observations were out of common support after PSM procedure. Overall, zone of extreme and maximal radiation has the lowest quality of matching with mean and median bias around 19% and 17%, respectively. Rubin's B is 29% and Rubin's R is below 0.5. We believe that heterogeneity of the matched observations in age-gender structure and luminosity patterns is the primary reason for the poor quality of the matches. Matching results are summarized below in Figure 5a.

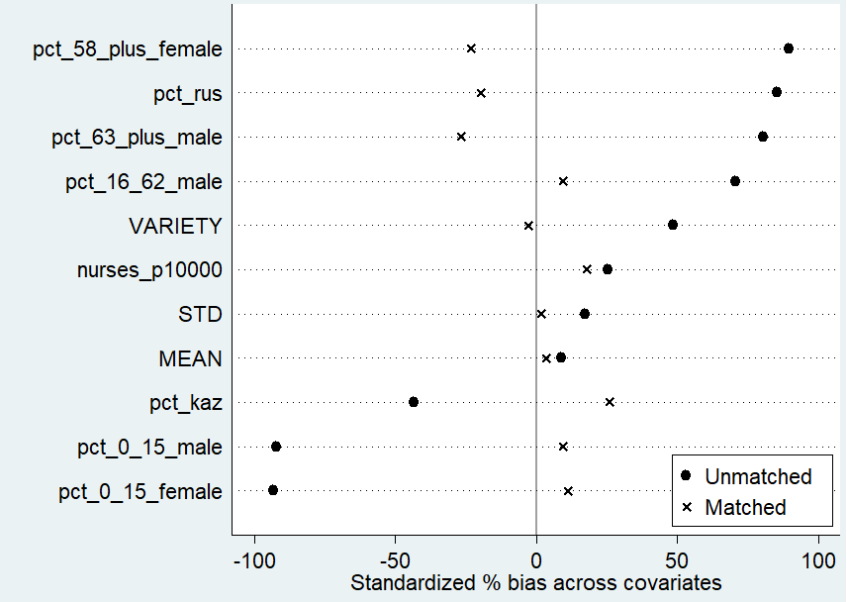
Figure 5a. ZEMR matching results



Zone of High Radiation Risk

The zone of high radiation risk presents most difficulty for matching procedure. In this treatment group, 71 out of 189 observations were dropped because of the lack of common support among treatment and control groups. Dropped observations belong to four districts: Oskemen and Ridder cities, and Glubokoe and Shemonaiha raions. These observations have specific luminosity patterns and atypical ethnic composition with dominant Slavic population, which makes it difficult to find adequate matches across the rest of the Kazakhstan. For the rest of the observations, analysis shows Rubin's B at 53% and Rubin's R of 0.5-2. Mean bias is at 14% while median bias is 11%. Results of the matching for each covariate are presented below in Figure 5b.

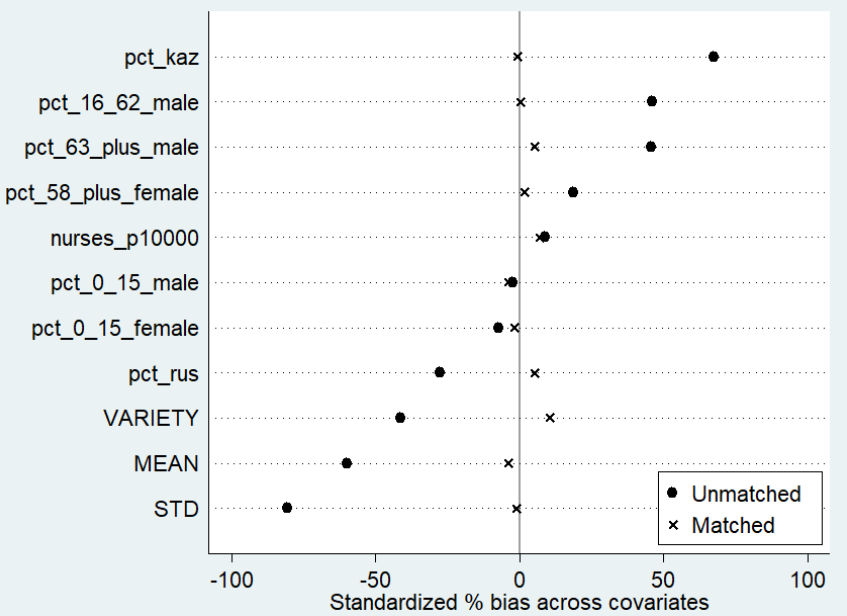
Figure 5b. ZHR matching results



Zone of Minimal Radiation Risk

The best matching among the three treatment groups is achieved in minimal radiation risk zone. Despite the fact that Rubin's R is below 0.5, Rubin's B is less than 25%. Moreover, median and mean bias are below 4%. After the matching procedure, 2 out of 133 observations were dropped due to poor match quality. The results of the PSM for zone of minimal radiation risk are shown below in Figure 5c. This zone does not contain big cities and is less ethnically diverse than the other two zones. We believe that these are the reasons why matching was most successful in this zone.

Figure 5c. ZMR matching results



6.2. PSM results

This section contains propensity score matching results for each variable of interest in more detail than in the preceding comparative section. The comparison is made across treatment, all-control and "hypothetical control" groups. "Hypothetical controls" are calculated in the process of matching and represent the values of treated observations as if they were untreated. The means of "hypothetical controls" are then compared with the means of treatment groups and used in calculation of average treatment effect on the treated (ATT). ATTs and their t-stat values after matching are presented in Tables 2a-4c. Unless otherwise noted, all the results are significant at the 5% level or less. Note also that the control group is the same for all treatment groups as the kernel matching technique relies on weighing all observations in the control group. Hence, means of the control groups before matching coincide.

Prevalence and incidence of all types of cancer

The effects of nuclear testing at the SNP on cancer morbidity and mortality have been extensively researched. In their analysis of solid cancer mortality patterns among the Semipalatinsk Historical Cohort, Bauer et al. (2005) find a strong relationship between exposure level and cancer morbidity. Similar findings emerge in Vakulchuk et al. (2014), who report a dramatic increase in cancer mortality in 1970s, and a sharp rise in incidence of cancer in 1980s among exposed population.

Our analysis of cancer prevalence and incidence among treated and control population is less conclusive. Despite the fact that means of cancer prevalence in extreme/maximal, high risk zones differ from the mean of the all-control group (0.97% and 1.08% versus 0.73%), and these differences become less prominent in kernel matching with common support. After matching, means of extreme/maximal and high risk zone drop to 0.96% and 0.75%, respectively, while means of these hypothetical control groups increase to 0.85%. As a result, we observe a "positive" effect of the radiation on cancer prevalence in high radiation risk zone but no statistically significant results are observed for the minimal radiation risk zone.

Similar to cancer prevalence, we observe huge mean differences in cancer incidence between extreme/maximal, high risk treatment zones and all-control group (0.24% and 0.26% versus 0.18%). However, after kernel matching only the extreme/maximal radiation risk zone has statistically significant effect (0.24% versus 0.20% in the hypothetical control). Possible reasons for the loss of significance after matching are provided in the "general notes" section below.

Prevalence and incidence of all forms of Tuberculosis

There is limited amount of academic literature on the effects of long exposure to ionizing radiation on prevalence or incidence of tuberculosis. Belozerov et al. (2008) report that the incidence of active forms of tuberculosis depends on the level of radiation exposure. The authors conclude that districts with radiation exposure over three times the natural radiation background have 1.4 times and 2.9 times higher incidence of active forms and extrapulmonary forms of TB, respectively.

Our findings also suggest that exposure to radiation may have elevated TB levels for the treated population. We do not observe statistically significant mean differences for TB prevalence by simply comparing means of treatment regions with the mean for the all-control group. However, kernel PSM reveals that means of the treatment groups after matching statistically differ from the means of their hypothetical controls. More precisely, these differences are 0.28% (treatment) versus almost 0.18% (hypothetical control) in extreme/maximal, 0.24% versus nearly 0.20% in high, and 0.24% versus 0.20% in minimal radiation risk zone. In other words, prevalence of TB is 1.2-1.6 times higher in treated areas than in their respective controls. Please note that results for high risk zone are slightly below 5% significance level.

Similar to prevalence, statistically significant mean differences in incidence of TB are only observed after matching. Specifically, incidence of TB is nearly 0.13% (treatment) versus 0.09% (hypothetical control) in extreme/maximal, 0.11% versus 0.09% in high, and 0.1% versus 0.09% in minimal radiation risk zone. In short, the incidence of TB is 1.13-1.37 times higher in treated areas than in their respective controls.

Crude Death Rate (CDR): Deaths per 1000

The existing literature on mortality and exposure to radiation at SNP finds that radiation levels are associated with cancer mortality.[79] These studies report dramatic increases in cancer mortality in treated areas. In contrast, our results on the relation between radiation exposure and elevated mortality as measure by the CDR are inconclusive. For extreme/maximal radiation risk zone, we do not observe statistically significant differences both before and after matching. Those who live in high radiation risk zone have significantly different means than the all-control comparator – 1.2% versus 0.9% (25% more). However, those differences flip after kernel matching procedure – mean mortality becomes 9% higher in hypothetical control than in treatment group. We also observe significant differences after matching in minimal radiation risk zone. Those exposed to radiation in this group have 9% higher death rates than their hypothetical control. Our dataset does not contain mortality by diseases, but rather presents a total measure from all possible causes. We believe that it is the reason why we notice discrepancies between our findings and academic literature.

Incidence of diseases of musculoskeletal system and connective tissue

The effect of radiation exposure at SNP on diseases of the bone, muscles and connective tissue has been explored amply in the academic literature. A study of veterans of special risk subdivisions who participated in nuclear testing at SNP revealed a moderate but statistically significant increase in prevalence of musculoskeletal and connecting tissue diseases.[80] Elevated prevalence of bone, muscles and connective tissue diseases has also been noted in villages near the SNP. The number of patients with rheumatism and polyarthritis of rheumatoid and other origin as well as number of patients with deforming osteoarthritis was reported to be two-three times greater in treated villages than in control.[81] Moreover, the literature on the long-run effects of radiation indicates that effects on the musculoskeletal system are long-run and typically observed month to years after the exposure.[82]

Our analysis of incidence of musculoskeletal and connecting tissue diseases aligns with the existing findings. Means of musculoskeletal and connecting tissue diseases in treatment groups are statistically different from all-control group. Incidence of diseases of bone, muscles and connective tissue is almost 2.7% in extreme and maximal, 3.1% in high risk, and nearly 1.9% in minimal radiation risk zones. This is much greater than the incidence in the all-control group – about 1.5%. After kernel matching, we still observe statistically significant differences. We can conclude that treated areas have 1.34-2.24 times higher incidence than their respective controls.

Incidence of endocrine, nutritional and metabolic diseases, immunity disorders

The relation between SNP exposure and morbidity from endocrine-metabolic-immune disorders also has been studied. Alishev et al., (2012) report a twofold increase in the prevalence of endocrine diseases among veterans of special risk subdivisions who participated in nuclear testing compared to a control group. Apart from participants in nuclear testing, there is evidence of higher morbidity from EMB disorders in the civilian population near the SNP. Belozerov et al. (2008) con-

79 Vakulchuk et al., 2014; Belozerov et al., 2008
80 Alishev et al., 2012
81 Balmukhanov et al., 2006
82 ICRP, 2012

clude that nearly half of the population in the nuclear polygon region has an incomplete immune system and the population as a whole has developed a secondary immune deficiency.

Elevated morbidity levels of EMB disorders can also been noticed from our data. The average incidence of endocrine-metabolic-immune disorders in two most treated areas is significantly different from all-control group. Those living in extreme/maximal and high risk zones have mean incidence of 1.3% and about 1%, respectively – all far greater than the 0.7% observed in the all-control districts. Kernel matching also reveals positive statistically significant differences in incidence of EMB diseases for individuals in extreme/maximal and high radiation risk zones. Population from these treatment groups is 1.24-1.49 times more likely suffer from EMB disorders than their hypothetical controls. However, differences are negative for the zone of minimal radiation risk. This may seem counterintuitive, but it aligns well with the existing studies of the long-term effects of radiation. It has been established that high doses of chronic irradiation are likely to be immunosuppressive, while exposure at low doses actually may boost immune responses.[83]

Incidence of diseases of the circulatory system

There are several studies relating SNP and circulatory system diseases. Grosche et al. (2011) analyze mortality patterns from cardiovascular diseases among the Semipalatisk Historical Cohort, concluding that there is no discernable risk of radiation-related mortality from diseases of circulatory system. Another approach was taken by Markabayeva et al. (2018), who explores hypertension prevalence in some of the low and intermediate radiation risk territiories. The results of their study suggest that higher levels of radiation exposure lead to higher risks of cardiovascular system diseases. In contrast to previous studies, we find significant mean differences in incidence of circulatory system diseases in extreme/maximal and minimal radiation risk zones. Those residing in extreme/maximal risk zone have incidence of 3.2%, those in minimal risk zone nearly 2.3%, while incidence in the all-control group is close to 2.0%. After matching, we observe significant differences only in extreme/maximal risk zone. Residents of the most treated areas are 1.29 times more likely suffer from circulatory system diseases than those from hypothetical control group.

Prevalence of congenital anomalies

We do not observe statistically significant differences before and after matching for extreme/maximal and high radiation risk zones. These results coincide with findings of Vakulchuk et al. (2014) who conclude that exposed population and control group do not differ in the number congential malformations after 1985. As for minimal radiation risk zone, prevalence of congenital anmalies is 0.09%, which is around 30% less than 0.13% observed in all-control and hypothetical control groups.

Hospital discharges, diseases of respiratory system

Results from existing studies suggest that exposure to ionizing radiation is associated with elevated risk of developing respiratory diseases. Alsihev et al. (2012) report a moderate yet statistically significant increase in incidence of respiratory diseases among veterans of special risk subdivisions who participated in nuclear testing at SNP. Increased incidence of respiratory diseases has also been observed among local population in the vicinity of SNP. Belozerov et al. (2008) note that districts adjacent to SNP have incidence of ARVI 21% above the national level. It has also been established that exposure to radiation considerably elevates the risk of death from respiratory diseases.[84] Our findings support these results. Those living in extreme/maximal risk zone have mean prevalence of about 26%, those in high risk zone – 26.5%, those in minimal risk zone –

83 ICRP, 2012
84 ICRP, 2012

23.4%. Mean prevalence values observed in all treatment groups are significantly greater than 17.8% for the all-control group. After kernel matching, statistically significant differences prevail. We conclude that treated areas have 1.5-2.0 times higher prevalence of respiratory diseases than their respective hypothetical controls.

Incidence of ischemic heart disease

Our results suggest that incidence of ischemic heart diseases is positively correlated with the degree of treatment. We observe statistically significant mean differences for extreme/maximal and high risk zones. Those living in extreme/maximal radiation risk zone have incidence of 0.70%, those in high radiation risk zone – 0.42%, while incidence in all-control group is around 0.36%. Similar to the incidence of circulatory system diseases, kernel matching produces statistically significant differences only in extreme/maximal areas of treatment. We find that those in most treated areas are 2.27 times more likely to experience conditions with ischemic heart diseases than their hypothetical controls.

Incidence of diseases of the nervous system and sense organs

It has been established that exposure to radiation at SNP is associated with elevated likelihood of nervous system diseases. Belozerov et al. (2008) state that the treated population is 1.6-10.5 times more likely to have nervous system diseases than controls. Similar results are reported by Balmukhanov et al. (2006), who find that rates of neurocirculatory dystonia in treated villages twice exceeded the rates of controls. Our findings align with the existing studies. Those living in extreme/maximal risk zone have an incidence of 2.8%, in high risk zone – 2%, and 1.7% in minimal radiation risk zone. This is 16-89% more than the 1.5% incidence in the all-control group. After matching, these differences still prevail and treatment groups have 1.4-2.8 times more incidence than their respective hypothetical controls. Academic literature suggests that children's brains are especially vulnerable to ionizing radiation and long-run negative impacts on nervous system, such as cognitive or behavioral defects, may occur even from a low dose.[85] Therefore, the effects we observe today may reflect the exposure of children during the testing. Unfortunately, we do not possess the age-specific incidences; hence, this question requires further investigation.

Incidence of narcologic disorders

The relationship between incidence of narcological disorders and exposure to radiation has not been addressed in previous studies. Before matching, we note significant mean differences between treatment and the all-control groups. The mean value of narcological disorders in extreme/maximal exposure zone is nearly 0.50%, in high exposure 0.24%, and in minimal exposure zone 0.20% - all groups have significantly greater mean incidence than the 0.16% observed in all-control group. Statistical differences prevail in extreme/maximal and minimal risk zones after matching. Those living in extreme/maximal risk zone are 3.4 times, and those living in minimal risk zone are 1.8 times more likely to suffer from narcological disorders than their hypothetical controls.

Incidence of mental disorders

A previous study by Belozerov et al. (2008) concludes that those exposed to radiation are 2.1-6.0 times more likely to suffer from psychiatric disorders. Our findings are consistent with the existing results. Although we do not observe statistically significant mean differences between treatment groups and all-control, differences are noticed in extreme/maximal and minimal radiation risk zones after kernel matching. Specifically, those in extreme/maximal risk zone are 2.4, and those in minimal risk zone are 1.4 times more likely to experience mental disorders.

85 ICRP, 2012

Incidence of diseases of the digestive system

One of the ways through which people were exposed to radiation is consumption of irradiated products that enter the food chain.[86] Ingestion of irradiated foods thus may result in higher risks of alimentary canal-related disease among exposed groups. Vakulchuk et al. (2014) report that prevalent types of cancer among the SNP treated population include esophageal, stomach and small intestine cancers, while Balmukhanov et al. (2006) note that diseases of alimentary tract are common in exposed Kainar, Sarzhal, and Kokpekty villages. Our findings support existing results. Those in extreme/maximal and minimal radiation risk zones have significantly greater means of incidence of diseases of digestive system – 3.8% and 3.7% versus 3.1% in all-control. After matching, we lose significance in extreme/maximal risk zone, but observe significant differences in high and minimal risk zones: people from these treatment groups are 1.25-1.27 times more likely to experience diseases of the digestive system.

Incidence of diseases of genitourinary system

The relationship between exposure to radiation at SNP and incidence of genitourinary system diseases has not been explored at the population level. Existing studies tend to utilize sample analysis. For instance, Alsihev et al. (2012) report a 4 times higher incidence of genitourinary system diseases among veterans of special risk subdivisions who participated in nuclear testing at the SNP. A study by Balmukhanov et al. (2006), which focuses on treated villages of Sarzhal and Kainar, notes that 33%-54% of all villagers have urinary tract morbidities and that a majority (!) of women are affected by gynecological diseases. Our findings indicate that higher incidence of genitourinary diseases is observed at population level as well. Those in extreme/maximal risk zone have mean incidences of 5.0%, 4.5% in high, and 3.8% in minimal risk zones- all significantly greater than the 3.2% incidence observed in all-control (10% significance level for minimal risk zone). After kernel matching, we observe statistically significant differences in high and minimal risk zones. People from these areas are 1.19-1.42 times more likely to suffer from diseases of genitourinary system than their hypothetical controls. These results tend to align with the existing literature on the long-term effects of radiation on genitourinary system. Kidney damage due to radiation tends to show up and progress more than 10 years after the low dose exposure, while even one time intake of certain radioactive elements may cause long-term suppressive effects on gonads.[87]

Incidence of diseases of the skin and subcutaneous

It has been reported that areas exposed to radiation have higher incidence of skin diseases. Balmukhanov et al. (2006) note that morbidity rates of general dermapathology are about 2-3 times, and rates congenital and autoimmune skin diseases are 7-9 times higher in treated villages than in control. Scholars from ICRP have found that exposure to radiation can cause long-run responses of the skin that can occur months and years after irradiation. Such skin responses are dermal erythematous reactions, atrophy, induration, telangiectasia, necrosis and fibrosis.[88] Our findings support existing study results. We observe incidences of about 5.7% in extreme/maximal, 4.4% in high, and 3.4% in minimal radiation risk zones, respectively – all of the rates are very different from 2.6% observed in all-control group. The results are still significant after kernel matching: treated areas have 1.35-2.81 times higher incidence of skin and subcutaneous diseases than their respective controls.

86 Balmukhanov et al., 2006
87 Balmukhanov et al., 2006
88 ICRP, 2012

Incidence of injuries and poisoning registered at out-patient departments

The effect of exposure to ionizing radiation from SNP on incidence of poisoning and injuries has not been explored, and there is no obvious reason to suspect a connection Our findings partially confirm this hypothesis – no statistically significant difference before and after matching are observed in extreme/maximal and minimal zones. However, differences do appear in the high risk zone – 4.7% versus 3.2% incidence for the all-control population. The effect diminishes, but retains significance after kernel matching. Those in high risk zone are 1.15 times more likely to experience injuries or poisoning than their hypothetical counterparts.

Incidence of pregnancy complications, childbirth and puerperium

The relation between exposure to ionizing radiation at SNP and pregnancy complications has been explored in academic literature. Belozerov et al. (2008) report that people from exposed areas are 1.7-16.6 times more likely to experience pregnancy related complications or complications at birth and postnatal period. Our findings align with the existing literature results. We observe huge mean differences for all treatment groups. The incidence of pregnancy complications is around 9% in extreme/maximal, about 8% in high, and 5.5% in minimal radiation risk zones, while the incidence of pregnancy complications in all-control group is only 3.3%. These differences do not vanish after kernel matching and we observe that on average treated areas are 1.5-2.3 more likely to have complications in pregnancy than hypothetical controls.

Infant mortality, per 1000 live births

Infant mortality is a variable of key interest. Previous studies have reported a huge effect of radiation exposure and nuclear testing on levels of infant mortality during years of above ground testing.[89] These studies also report that infant mortality in treated areas returned to pre-testing levels around the 1970s. Our results suggest that effects of radiation exposure on infant mortality may still be present. Mean values of infant mortality in extreme/maximal, high and minimal radiation risk zones are 15.6, 16.9 and 16.2, respectively. All of these values are 19-29% greater than mean infant mortality of 13.1 observed in the all-control group. After matching, we still observe statistically significant differences. Infant mortality in treatment groups is 1.28-1.49 times greater than infant mortality in hypothetical controls.

Incidence of cerebro-vascular diseases

It appears that the relation between exposure to ionizing radiation at the SNP and cerebro-vascular diseases has not been fully addressed in academic literature. Our results on this subject are inconclusive. We see significant mean differences in incidence of cerebrovascular diseases for extreme/maximal and high risk groups before matching – 0.31% in extreme/maximal risk zone, 0.26% in high risk zone versus 0.18% in the all-control group. After kernel matching, the mean incidence in the hypothetical control for the extreme/maximal zone increases to 0.30% - for this reason, statistically significant differences are no longer observed. As for high radiation risk zone, the mean of the treatment group drops to 0.18%, while mean of the hypothetical control increases to 0.25%. As a result, we see a "positive" effect of radiation on incidence of cerebrovascular diseases.

Incidence of syphilis

The relationship between exposure to radiation at SNP and incidence of syphilis has not been studied in academic literature. Our findings suggest that there are no statistically significant mean differences between treatment and control groups both before and after matching.

89 Vakulchuk et al., 2014; Balmukhanov et al., 2006

Hospital discharges, certain conditions originating in perinatal period, per 100 000 live births

Before matching, we observe statistically significant mean differences for all treatment groups. The mean value of certain conditions originating in perinatal period is about 2.1% in extreme/maximal, around 2.4% in high, and 1.6% in minimal radiation risk zone. These values are 23-77% higher than the 1.3% observed in the all-control group. After matching, we lose significance in extreme/maximal risk zone, but results are robust in the other two zones. Hospital discharges related to perinatal conditions are 1.37-1.40 times more likely to occur in these treatment areas than in hypothetical controls.

6.3. Additional variables with fewer observations

In general, variables presented below contain fewer observations from both treatment and control groups.

Incidence of brucellosis

Belozerov et al. (2008) found that areas in the vicinity of the nuclear test site have 23.4% higher incidence of brucellosis than the national level. Our analysis of the incidence of brucellosis is inconclusive. Mean incidence of brucellosis is 11 in extreme/maximal, 9 in high, and 32 in minimal radiation risk zone, while the mean value in all-control group is around 17. As a result, we observe both positive and negative statistically significant differences between treated and control groups. The nature of such relationship remains unclear: an obvious explanation is randomness driven by the limited amount of observations. After matching, we observe statistically significant differences only in minimal radiation risk zone. Those in minimally exposed group have 2.2 times higher incidence of brucellosis than their hypothetical control.

Prevalence of diabetes mellitus

Studies analyzing the effect of exposure to radiation at SNP on prevalence of diabetes also seem to be absent in academic literature. Our findings do not reveal statistical differences in extreme/maximal and minimal treatment groups both before and after matching. Those in high radiation risk zone have prevalence of 1.6% while those in the all-control group - 1.2%. However, mean differences vanish after kernel matching.

Incidence of gonococcal infections

It appears that the relation between exposure to radiation at SNP and incidence of gonococcal infections has not been studied. Our results suggest that there are no statistically significant mean differences in extreme/maximal and minimal radiation risk zones before matching. However, mean incidence of gonococcal infections is significantly different in high risk zone, with an estimated 46 cases/district/year versus 30 cases in the all-control group. After kernel matching, significant differences are only present in minimal radiation risk zone. PSM results suggest that incidence of gonococcal infections in zone of minimal exposure is twice as higher than in the hypothetical control.

Incidence of viral hepatitis

The exposure to radiation at SNP has been associated with elevated incidence of viral hepatitis. Belozerov et al. (2008) report that areas adjacent to the test site exceed the national level of viral hepatitis by 110%. Such differences are not observed from our analysis. On the contrary, mean incidence in all treated groups is lower than in the all-control group. The incidence of viral hepatitis is 28 in extreme/maximal, 22 in high and 23 in minimal risk zone – all below 36 cases observed in the all-control group. However, after kernel matching statistically significant differences vanish.

Perinatal mortality rate and Fetal death rate, per 1000 births

Studies on the effect of radiation at SNP on a young population tend to focus more on infant mortality, congenital anomalies or pregnancy complications, while the effect on perinatal and fetal mortality remains unexplored. We find statistically significant mean differences in perinatal mortality across all treatment groups. Perinatal mortality is 1.7% in extreme/maximal, 1.6% in high and minimal radiation risk zones – all far greater than 1.1% observed in the all-control group. After matching, significant differences prevail only in the high risk zone. This treatment group has 1.41 times higher perinatal mortality than its hypothetical control.

Findings on fetal mortality are similar – we observe statistically significant mean differences in all risk zones. The mean fetal death rate is 1.1% in extreme/maximal risk zone, 0.9% in high and minimal risk zones, and around 0.8% in the all-control group. After kernel matching, positive significant differences prevail only in high risk zone. The level of fetal mortality in this treatment group is 1.33 higher than its hypothetical control. Matching also reveals negative significant differences for zone of minimal radiation risk – those living in this area are predicted to have about a 20% lower fetal death rate than their matched controls.

Number of abortions per hundred births (including stillbirths)

To our knowledge, there are no studies linking ionizing radiation exposure to the number of abortions in a region. However, it is possible that there will be a higher level of abortions in regions with greater fetal abnormalities or problems in pregnancy. This link seems likely to be especially acute in recent years, as the Government of Kazakhstan now offers free genetic screening for those pregnant women whose children have confirmed risk of developing birth defects or disorders, and those whose fetuses appear to suffer defects are encouraged to undergo an emergency pregnancy termination.[90] To the extent that pregnant women to undertake genetic testing, elevated levels of abortions imply that our findings above of the impacts of radiation on birth defects and anomalies are understatements, because those with most pronounced effects of radiation are simply not born. We would also expect this effect to be most prominent from the mid-2000s onward when genetic testing became more widespread.

Our findings do indicate that zone differences are present. The mean number of abortions is 51 in extreme/maximal, around 108 in high, and about 52 in minimal radiation risk zones. All of these values are significantly greater than 39 in all control group. Significant differences prevail in high and minimal risk zones after propensity score matching. The number of abortions for women living in these areas is predicted to be 1.35-3.10 times higher than in respective hypothetical controls.

Two additional points need to be mentioned. First, the zone differences we observe align with existing findings on abortion patterns in Kazakhstan. It has been established that the ethnically Slavic population of Kazakhstan has higher abortion rates than the Kazakh population.[91] This aligns with observed greater differences in zone of high radiation risk, where more ethnically Slavic and European population is present. Second, the quality of the data on the number of abortions is questionable. Apart from the fact that it is difficult to track the abortions precisely and we expect undercounts, there were nine observations where the number of abortions were suspiciously elevated (with 6 of them belonging to the Ulan district of East Kazakhstan Oblast - ZHR). These observations were greater 1000 and were excluded from PSM and regression estimations. Thus, the results we report should be regarded as tentative, but consistent with a downward bias in our estimates of the effects of radiation exposure on birth outcomes.

90 Butts, 2010
91 Olds & Westoff, 2004

General notes on matching results

Two points need to be mentioned after the propensity score matching. Firstly, given the omission of 71 observations from four districts in the zone of high radiation risk due to their socio-economic differences, means of several variables have significantly dropped after propensity scores were calculated. Variables like prevalence, incidence of cancer and death per 1000 are good examples. As a result, the differences between the treatment and matched control and corresponding t-stats for ZHR may be lower than in would be the case with more complete data. For such variables, regression analysis may convey more information than PSM. Second, due to the socio-economic differences within treatment groups and the small number of observations (especially for zone of extreme/maximal risk), standard errors after the estimation of propensity scores are much larger than in the regressions, doubling for some variables. This problem is especially important for prevalence and incidence of cancer, and the crude death rate in the zone of extreme/maximal radiation risk. Again, this may result in lower t-stats for some of the variables. An alternative method potentially would be to regroup the treated districts by socio-economic characteristics and not by the degree of reported radiation exposure but we have not done so here.

Table 2a. PSM results for zone of extreme/maximal radiation: most number of observations

Kernel PSM results for Zone of extreme and maximal radiation: Table 2a						
Probit Regression				Number of Obs	=	2698
				LR chi2(11)	=	400.14
Log likelihood = -76.18525				Prob > chi2	=	0.0000
				Pseudo R2	=	0.7242
Variable	Sample	Treated	Controls	Difference	S.E.	T-stat
Prevalence of all types of cancer	Unmatched	968.63	730.90	237.73	56.98	4.17
	ATT	976.47	858.36	118.11	115.51	1.02
Incidence of all types of cancer	Unmatched	242.34	183.44	58.91	10.40	5.67
	ATT	243.48	195.57	47.91	20.38	2.35
Incidence of all forms of TB	Unmatched	124.42	111.56	12.86	8.21	1.57
	ATT	126.31	91.73	34.57	17.51	1.97
Prevalence of all forms of TB	Unmatched	276.00	262.69	13.31	24.74	0.54
	ATT	281.97	176.15	105.82	50.96	2.08
Death per 1000	Unmatched	10.46	9.83	0.62	0.37	1.66
	ATT	10.55	10.23	0.31	0.77	0.41
Incidence of diseases of the musculoskeletal system and connective tissue	Unmatched	2650.72	1476.95	1173.76	151.76	7.73
	ATT	2696.25	1201.96	1494.29	325.22	4.59
Incidence of endocrine, nutritional and metabolic diseases, immunity disorders	Unmatched	1298.99	686.71	612.28	74.83	8.18
	ATT	1308.54	876.70	431.84	179.11	2.41
Incidence of diseases of the circulatory system	Unmatched	3209.91	1980.83	1229.07	159.16	7.72
	ATT	3193.51	2467.79	725.72	368.22	1.97
Prevalence of congenital anomalies	Unmatched	148.91	138.89	10.03	23.00	0.44
	ATT	151.42	141.11	10.31	43.75	0.24
Hospital discharges, diseases of respiratory system	Unmatched	25965.22	17867.52	8097.70	1260.52	6.42
	ATT	25910.46	12717.60	13192.85	2753.11	4.79
Incidence of ischaemic heart disease	Unmatched	704.36	364.18	340.18	44.07	7.72
	ATT	690.54	303.84	386.70	99.45	3.89
Incidence of diseases of the nervous system and sense organs	Unmatched	2815.73	1486.41	1329.32	150.08	8.86
	ATT	2833.21	1009.93	1823.28	337.84	5.40
Incidence of narcologic disorders	Unmatched	497.49	163.04	334.45	31.98	10.46
	ATT	494.31	145.30	349.01	96.39	3.62
Incidence of mental disorders	Unmatched	130.56	108.69	21.86	13.54	1.62
	ATT	134.85	56.5	78.32	29.39	2.66
Incidence of diseases of the digestive system	Unmatched	3820.73	3076.66	744.07	340.02	2.19
	ATT	3825.47	3256.33	569.14	686.05	0.83
Incidence of diseases of the genitourinary system	Unmatched	5010.41	3172.65	1837.76	532.67	3.45
	ATT	5068.15	4113.75	954.40	1026.71	0.93
Incidence of diseases of the skin and subcutaneous	Unmatched	5675.01	2647.88	3027.13	287.12	10.54
	ATT	5630.62	2003.43	3627.19	710.21	5.11
Incidence of injuries and poisoning registered in out-patient departments	Unmatched	3476.68	3226.40	250.28	284.16	0.88
	ATT	3558.64	4334.40	-775.76	541.02	-1.43

Table 2b. PSM results for zone of extreme/maximal radiation: intermediate number of observations

Kernel PSM results for Zone of extreme and maximal radiation: Table 2b						
Probit Regression				Number of Obs	=	2619
				LR chi2(11)	=	391.07
Log likelihood = -75.194857				Prob > chi2	=	0.0000
				Pseudo R2	=	0.7222
Variable	Sample	Treated	Controls	Difference	S.E.	T-stat
Incidence of pregnancy complications, childbirth and puerperium	Unmatched	8933.86	3385.57	5548.29	467.78	11.86
	ATT	9069.10	5006.01	4063.09	970.15	4.19
Infant mortality, per 1000 live births	Unmatched	15.66	13.12	2.54	0.89	2.87
	ATT	15.88	10.63	5.25	1.83	2.87
Incidence of cerebrovascular diseases	Unmatched	309.26	184.87	124.39	28.91	4.30
	ATT	303.83	301.84	1.99	61.64	0.03
Incidence of syphilis	Unmatched	42.87	50.90	-8.03	7.40	-1.09
	ATT	44.65	25.46	19.19	14.70	1.31
Hospital discharges, certain conditions originating in perinatal period	Unmatched	2079.95	1331.78	748.17	200.79	3.73
	ATT	2098.74	1583.74	515.00	380.78	1.35

Table 2c. PSM results for zone of extreme/maximal radiation: least number of observations

Kernel PSM results for Zone of extreme and maximal radiation: Table 2c						
Variable	Sample	Treated	Controls	Difference	S.E.	T-stat
Incidence of brucellosis	Unmatched	10.89	16.86	-5.97	3.91	-1.53
	ATT	10.10	2.33	7.77	7.36	1.06
Prevalence of diabetes mellitus	Unmatched	1197.28	1150.10	47.17	297.77	0.16
	ATT	1173.73	1744.46	-570.73	526.49	-1.08
Incidence of gonococcal infections	Unmatched	30.79	29.91	0.88	6.53	0.13
	ATT	30.80	10.59	20.20	13.71	1.47
Incidence of viral hepatitis	Unmatched	28.09	36.72	-8.63	10.29	-0.84
	ATT	35.87	14.73	21.14	26.03	0.81
Fetal death rate per 1000 births	Unmatched	10.92	7.61	3.31	1.00	3.30
	ATT	10.84	11.28	-0.44	2.20	-0.20
Perinatal mortality rate per 1000 births	Unmatched	16.56	11.07	5.49	1.21	4.55
	ATT	16.82	15.75	1.07	2.72	0.39
Number of abortions on 100 born alive and dead	Unmatched	51.02	38.64	12.38	5.69	2.17
	ATT	53.80	33.56	20.23	12.99	1.56

Table 3a. PSM results for zone of high radiation: most number of observations

Kernel PSM results for Zone of high radiation: Table 3a						
Probit Regression				Number of Obs	=	2830
				LR chi2(11)	=	683.77
Log likelihood = -352.14482				Prob > chi2	=	0.0000
				Pseudo R2	=	0.4926
Variable	Sample	Treated	Controls	Difference	S.E.	T-stat
Prevalence of all types of cancer	Unmatched	1082.79	730.90	351.89	32.71	10.76
	ATT	756.42	854.99	-98.57	44.61	-2.21
Incidence of all types of cancer	Unmatched	267.30	183.44	83.87	6.01	13.95
	ATT	203.25	210.69	-7.44	8.16	-0.91
Incidence of all forms of TB	Unmatched	107.28	111.56	-4.28	4.57	-0.94
	ATT	107.39	94.54	12.85	6.79	1.89
Prevalence of all forms of TB	Unmatched	233.47	262.69	-29.22	13.88	-2.10
	ATT	239.77	196.49	43.29	22.25	1.95
Death per 1000	Unmatched	12.22	9.83	2.39	0.22	10.93
	ATT	9.58	10.41	-0.83	0.25	-3.31
Incidence of diseases of the musculoskeletal system and connective tissue	Unmatched	3119.01	1476.95	1642.05	91.11	18.02
	ATT	3160.75	1524.97	1635.78	228.81	7.15
Incidence of endocrine, nutritional and metabolic diseases, immunity disorders	Unmatched	981.62	686.71	294.91	41.56	7.10
	ATT	923.49	743.35	180.14	58.62	3.07
Incidence of diseases of the circulatory system	Unmatched	2121.35	1980.83	140.52	87.60	1.60
	ATT	2035.27	2109.52	-74.25	121.32	-0.61
Prevalence of congenital anomalies	Unmatched	138.39	138.89	-0.49	12.84	-0.04
	ATT	118.18	122.78	-4.60	14.97	-0.31
Hospital discharges, diseases of respiratory system	Unmatched	28365.45	17867.52	10497.93	708.11	14.83
	ATT	26506.32	15600.89	10905.42	1121.44	9.72
Incidence of ischaemic heart disease	Unmatched	422.73	364.18	58.55	24.04	2.44
	ATT	410.02	386.31	23.71	29.72	0.80
Incidence of diseases of the nervous system and sense organs	Unmatched	1993.35	1486.41	506.94	83.33	6.08
	ATT	2263.44	1177.83	1085.62	121.06	8.97
Incidence of narcologic disorders	Unmatched	238.80	163.04	75.76	17.07	4.44
	ATT	130.15	141.48	-11.33	19.00	-0.60
Incidence of mental disorders	Unmatched	108.11	108.69	-0.58	7.54	-0.08
	ATT	76.70	92.13	-15.42	9.70	-1.59
Incidence of diseases of the digestive system	Unmatched	3057.74	3076.66	-18.92	187.00	-0.10
	ATT	3355.87	2632.74	723.13	218.44	3.31
Incidence of diseases of the genitourinary system	Unmatched	4531.23	3172.65	1358.58	293.40	4.63
	ATT	4639.58	3246.87	1392.72	307.81	4.52
Incidence of diseases of the skin and subcutaneous	Unmatched	4435.22	2647.88	1787.34	160.21	11.16
	ATT	3982.62	2236.57	1746.06	215.90	8.09
Incidence of injuries and poisoning registered in out-patient departments	Unmatched	4755.65	3226.40	1529.25	162.59	9.41
	ATT	4016.76	3475.44	541.32	215.29	2.51

Table 3b. PSM results for zone of high radiation: intermediate number of observations

Kernel PSM results for Zone of high radiation: Table 3b						
Probit Regression				Number of Obs	=	2748
				LR chi2(11)	=	686.18
Log likelihood = -334.72011				Prob > chi2	=	0.0000
				Pseudo R2	=	0.5062
Variable	Sample	Treated	Controls	Difference	S.E.	T-stat
Incidence of pregnancy complications, childbirth and puerperium	Unmatched	7991.47	3385.57	4605.90	272.87	16.88
	ATT	8894.87	3867.31	5027.56	577.95	8.70
Infant mortality, per 1000 live births	Unmatched	16.91	13.12	3.79	0.51	7.36
	ATT	16.55	12.71	3.84	1.05	3.65
Incidence of cerebrovascular diseases	Unmatched	260.39	184.87	75.52	16.52	4.57
	ATT	178.67	243.40	-64.72	22.61	-2.86
Incidence of syphilis	Unmatched	58.44	50.90	7.55	4.22	1.79
	ATT	42.51	39.32	3.19	5.23	0.61
Hospital discharges, certain conditions originating in perinatal period	Unmatched	2356.33	1331.78	1024.56	114.25	8.97
	ATT	2049.86	1490.80	559.07	145.93	3.83

Table 3c. PSM results for zone of high radiation: least number of observations

Kernel PSM results for Zone of high radiation: Table 3c						
Variable	Sample	Treated	Controls	Difference	S.E.	T-stat
Incidence of brucellosis	Unmatched	9.14	16.86	-7.72	2.26	-3.42
	ATT	13.22	12.42	0.80	2.76	0.29
Prevalence of diabetes mellitus	Unmatched	1651.35	1150.10	501.24	163.49	3.07
	ATT	1381.67	1387.22	-5.55	168.54	-0.03
Incidence of gonococcal infections	Unmatched	45.66	29.91	15.75	3.71	4.24
	ATT	14.98	19.96	-4.99	4.12	-1.21
Incidence of viral hepatitis	Unmatched	22.51	36.72	-14.21	5.66	-2.51
	ATT	19.03	15.88	3.15	6.22	0.51
Fetal death rate per 1000 births	Unmatched	9.29	7.61	1.68	0.59	2.87
	ATT	10.20	7.67	2.54	1.46	1.74
Perinatal mortality rate per 1000 births	Unmatched	15.60	11.07	4.52	0.70	6.48
	ATT	15.53	11.05	4.49	1.50	2.98
Number of abortions on 100 born alive and dead	Unmatched	107.84	38.64	69.2	3.86	17.92
	ATT	101.92	32.89	69.02	13.51	5.11

Table 4a. PSM results for zone of minimal radiation: most number of observations

Kernel PSM results for Zone of minimal radiation: Table 4a						
Probit Regression				Number of Obs	=	2774
				LR chi2(11)	=	470.98
Log likelihood = -298.28395				Prob > chi2	=	0.0000
				Pseudo R2	=	0.4412
Variable	Sample	Treated	Controls	Difference	S.E.	T-stat
Prevalence of all types of cancer	Unmatched	695.17	730.90	-35.73	37.87	-0.94
	ATT	687.73	670.04	17.69	45.14	0.39
Incidence of all types of cancer	Unmatched	190.65	183.44	7.21	6.97	1.03
	ATT	188.98	175.10	13.89	8.90	1.56
Incidence of all forms of TB	Unmatched	102.14	111.56	-9.42	5.42	-1.74
	ATT	102.69	89.56	13.13	6.31	2.08
Prevalence of all forms of TB	Unmatched	238.49	262.69	-24.20	16.45	-1.47
	ATT	239.30	198.10	41.21	20.25	2.03
Death per 1000	Unmatched	10.33	9.83	0.50	0.25	1.98
	ATT	10.30	9.45	0.84	0.35	2.39
Incidence of diseases of the musculoskeletal system and connective tissue	Unmatched	1867.62	1476.95	390.67	98.81	3.95
	ATT	1867.83	1390.49	477.34	91.78	5.20
Incidence of endocrine, nutritional and metabolic diseases, immunity disorders	Unmatched	724.16	686.71	37.45	48.43	0.77
	ATT	727.23	844.37	-117.13	49.90	-2.35
Incidence of diseases of the circulatory system	Unmatched	2256.72	1980.83	275.89	103.99	2.65
	ATT	2244.47	2110.58	133.89	119.90	1.12
Prevalence of congenital anomalies	Unmatched	92.07	138.89	-46.81	15.09	-3.10
	ATT	92.03	132.24	-40.21	13.33	-3.02
Hospital discharges, diseases of respiratory system	Unmatched	23427.14	17867.52	5559.62	946.61	5.87
	ATT	23446.30	15698.85	7747.44	2283.16	3.39
Incidence of ischaemic heart disease	Unmatched	400.29	364.18	36.11	28.55	1.26
	ATT	400.09	381.91	18.18	27.67	0.66
Incidence of diseases of the nervous system and sense organs	Unmatched	1728.27	1486.41	241.86	97.43	2.48
	ATT	1732.98	1241.46	491.52	94.96	5.18
Incidence of narcologic disorders	Unmatched	201.75	163.04	38.71	20.08	1.93
	ATT	201.42	109.80	91.63	25.10	3.65
Incidence of mental disorders	Unmatched	99.24	108.69	-9.45	8.82	-1.07
	ATT	99.02	69.83	29.19	8.45	3.45
Incidence of diseases of the digestive system	Unmatched	3709.50	3076.66	632.84	222.60	2.84
	ATT	3698.38	3021.93	676.45	211.86	3.19
Incidence of diseases of the genitourinary system	Unmatched	3795.47	3172.65	622.82	349.08	1.78
	ATT	3789.32	3162.76	626.56	304.94	2.05
Incidence of diseases of the skin and subcutaneous	Unmatched	3412.23	2647.88	764.35	184.18	4.15
	ATT	3399.64	2510.59	889.05	177.61	5.01
Incidence of injuries and poisoning registered in out-patient departments	Unmatched	3174.29	3226.40	-52.12	187.31	-0.28
	ATT	3165.84	2947.64	218.20	181.77	1.20

Table 4b. PSM results for zone of minimal radiation: intermediate number of observations

Kernel PSM results for Zone of minimal radiation: Table 4b						
Probit Regression				Number of Obs	=	2693
				LR chi2(11)	=	466.79
Log likelihood = -287.42852				Prob > chi2	=	0.0000
				Pseudo R2	=	0.4481
Variable	Sample	Treated	Controls	Difference	S.E.	T-stat
Incidence of pregnancy complications, childbirth and puerperium	Unmatched	5478.95	3385.57	2093.38	314.34	6.66
	ATT	5522.63	3732.54	1790.09	434.92	4.12
Infant mortality, per 1000 live births	Unmatched	16.23	13.12	3.11	0.60	5.18
	ATT	15.90	12.34	3.55	0.85	4.18
Incidence of cerebrovascular diseases	Unmatched	183.58	184.87	-1.29	18.89	-0.07
	ATT	182.11	205.86	-23.75	18.57	-1.28
Incidence of syphilis	Unmatched	44.26	50.90	-6.63	4.95	-1.34
	ATT	42.46	33.63	8.83	6.02	1.47
Hospital discharges, certain conditions originating in perinatal period	Unmatched	1646.48	1331.78	314.71	133.30	2.36
	ATT	1634.02	1164.46	469.56	148.87	3.15

Table 4c. PSM results for zone of minimal radiation: least number of observations

Kernel PSM results for Zone of minimal radiation: Table 4c						
Variable	Sample	Treated	Controls	Difference	S.E.	T-stat
Incidence of brucellosis	Unmatched	32.44	16.86	15.58	2.72	5.73
	ATT	32.08	14.81	17.26	3.68	4.69
Prevalence of diabetes mellitus	Unmatched	1082.08	1150.10	-68.02	193.83	-0.35
	ATT	1077.77	1126.62	-48.85	162.53	-0.30
Incidence of gonococcal infections	Unmatched	26.68	29.91	-3.23	4.16	-0.78
	ATT	26.82	12.98	13.83	4.40	3.15
Incidence of viral hepatitis	Unmatched	23.17	36.72	-13.55	6.73	-2.01
	ATT	23.17	15.64	7.53	6.74	1.12
Fetal death rate per 1000 births	Unmatched	9.18	7.61	1.57	0.66	2.38
	ATT	9.23	11.59	-2.35	0.95	-2.47
Perinatal mortality rate per 1000 births	Unmatched	15.89	11.07	4.82	0.81	5.96
	ATT	15.84	14.96	0.87	1.24	0.71
Number of abortions on 100 born alive and dead	Unmatched	51.70	38.63	13.06	3.80	3.44
	ATT	51.81	38.34	13.47	5.13	2.62

7. Regression Analysis

In order to explore another perspective on the effects of ionizing radiation from SNP and perform robustness checks, we also include a comprehensive regression analysis. We start with simple random effects model described in section V. Given that we have 30 variables of interest and we run model 1 for each of those, we only report radiation or environmental zone dummies. The results are presented in Table 5.

Key findings of other regression models available in the Online Appendix are discussed below. Given the large number of health characteristic variables, we categorize them into the following seven groups: Birth and maternal health related variables, cancer prevalence and incidence, TB and dis

eases of respiratory system, circulatory system diseases and related variables, other systemic diseases, infectious diseases, and other health conditions. We present the results in a form of bullet points and provide regression output in the 200+-page Online Appendix. The year fixed effects model with zone-year interactions for every treatment dummy were excluded from the Online Appendix but are available upon request.

Table 5. Random Effects regression model coefficients.

Table 5. Random effects regression table: coefficients and std errors of treatment dummies

Variable	zemr	zhr	zmr	eco_cat	eco_crs	eco_pcr	eco_fnd	constant	Mean of all control
Prevalence of all types of cancer	91.35 (69.95)	-19.72 (41.22)	-44.55 (46.67)	108.70 (70.23)	297.95*** (50.83)	63.26 (38.62)	-18.48 (31.59)	630.24*** (143.59)	730.90 (8.31)
Incidence of all types of cancer	42.34*** (13.82)	17.24** (8.18)	10.28 (9.20)	19.51 (13.81)	64.16*** (10.02)	1.82 (7.62)	-0.98 (6.20)	185.93*** (37.30)	183.44 (1.52)
Incidence of all forms of TB	71.33*** (16.53)	16.98* (9.77)	28.31** (11.02)	23.69 (16.56)	1.74 (12.00)	-30.61*** (9.12)	-3.42 (7.44)	-6.27 (38.79)	111.56 (1.19)
Prevalence of all forms of TB	180.66*** (43.95)	35.77 (26.02)	51.96* (29.27)	71.28 (43.93)	-3.24 (31.86)	-85.23*** (24.23)	32.03 (19.71)	2.11 (115.82)	262.69 (3.60)
Death per 1000	1.49*** (0.43)	0.22 (0.26)	1.24*** (0.29)	0.02 (0.43)	-0.87*** (0.31)	-0.04 (0.24)	0.41** (0.19)	10.84*** (1.16)	9.83 (0.05)
Incidence of diseases of the musculoskeletal system and connective tissue	1211.08*** (403.07)	1361.67*** (237.48)	376.80 (268.97)	238.87 (404.71)	411.40 (292.93)	-406.47* (222.55)	-255.72 (182.08)	2234.45*** (823.59)	1476.95 (22.02)
Incidence of endocrine, nutritional and metabolic diseases, immunity disorders	624.91*** (193.03)	239.51** (113.93)	30.57 (128.78)	80.89 (193.66)	559.67*** (140.22)	-158.91 (106.56)	0.33 (87.08)	740.01* (418.60)	686.71 (10.74)
Incidence of diseases of the circulatory system	737.53* (410.20)	19.24 (241.73)	-165.20 (273.72)	226.36 (411.82)	588.48** (298.10)	-410.43* (226.48)	-272.82 (185.27)	3433.94*** (844.24)	1980.83 (22.90)
Prevalence of congenital anomalies	-0.22 (57.33)	18.97 (33.78)	-63.82* (38.25)	-25.47 (57.57)	56.98 (41.67)	-22.08 (31.65)	-25.48 (25.90)	287.25** (117.05)	138.89 (3.37)
Hospital discharges, diseases of respiratory system	7355.62* (3776.38)	7702.51*** (2217.72)	5091.97** (2520.49)	1994.25 (3795.42)	3670.11 (2745.76)	-6723.98*** (2084.99)	-3015.99* (1708.71)	23574.93*** (7066.69)	17867.52 (182.57)
Incidence of ischaemic heart disease	263.08*** (100.98)	49.86 (59.70)	-24.74 (67.32)	-42.40 (101.18)	-44.57 (73.30)	-88.54 (55.73)	-69.04 (45.46)	671.84*** (235.90)	364.18 (6.35)
Incidence of diseases of the nervous system and sense organs	1665.72*** (415.85)	544.68** (243.80)	419.94 (277.55)	1771.62*** (418.09)	1028.05*** (302.39)	-727.02*** (229.57)	-473.44** (188.27)	3142.80*** (747.90)	1486.41 (21.67)
Incidence of narcologic disorders	447.02*** (59.95)	14.14 (35.46)	131.17*** (39.96)	127.99** (60.02)	103.92** (43.49)	-20.56 (33.09)	-21.00 (26.96)	-1.65 (145.81)	163.04 (4.40)
Incidence of mental disorders	106.63*** (25.38)	21.76 (15.01)	58.18*** (16.92)	41.15 (25.43)	35.00* (18.42)	-21.61 (14.01)	-0.97 (11.42)	-57.10 (60.11)	108.69 (1.96)

CONTINUED...

Table 5 continued. Random Effects regression model coefficients.

Table 5 continued. Random effects regression table: coefficients and std errors of treatment dummies

Variable	zemr	zhr	zmr	eco_cat	eco_crs	eco_pcr	eco_fnd	constant	Mean of all control
Incidence of diseases of the digestive system	824.64 (1109.60)	488.83 (650.39)	414.92 (740.60)	5259.53*** (1115.65)	2394.14*** (806.88)	-692.29 (612.55)	904.70* (502.42)	1772.72 (1985.03)	3076.66 (49.56)
Incidence of diseases of the genitourinary system	2296.60*** (851.51)	1483.12*** (503.85)	636.75 (565.54)	1863.24** (847.64)	1323.78** (615.91)	-864.25* (468.86)	-761.13** (379.64)	4639.81* (2626.70)	3172.65 (78.00)
Incidence of diseases of the skin and subcutaneous	3870.17*** (794.02)	1915.39*** (466.99)	1319.60** (529.92)	1626.80** (797.71)	1234.64** (577.23)	-842.92* (438.41)	40.63 (359.03)	3640.71** (1543.28)	2647.88 (40.97)
Incidence of injuries and poisoning registered in out-patient departments	893.35 (662.39)	541.70 (389.10)	316.71 (442.09)	1567.31** (665.68)	470.94 (481.60)	-631.85* (365.72)	-408.92 (299.68)	4803.14*** (1247.52)	3226.40 (41.64)
Incidence of pregnancy complications, childbirth and puerperium	5185.39*** (1389.53)	3939.31*** (817.80)	1591.51* (927.27)	-266.02 (1395.60)	288.03 (1009.98)	-299.35 (767.30)	-63.77 (628.26)	-447.17 (2753.84)	3385.57 (68.46)
Infant mortality, per 1000 live births	5.59*** (1.79)	5.29*** (1.06)	5.25*** (1.19)	1.82 (1.77)	2.37* (1.29)	3.82*** (0.98)	0.64 (0.79)	1.80 (6.84)	13.12 (0.13)
Incidence of cerebrovascular diseases	79.29 (65.99)	13.74 (39.03)	-40.58 (44.00)	-5.51 (66.12)	43.66 (47.90)	-38.14 (36.43)	7.06 (29.73)	338.04** (154.24)	184.87 (4.22)
Incidence of syphilis	33.16*** (10.84)	8.31 (6.41)	29.09*** (7.20)	-1.50 (10.77)	3.21 (7.82)	-23.67*** (5.97)	-1.91 (4.83)	-28.46 (33.16)	50.90 (1.09)
Hospital discharges, certain conditions originating in perinatal period	1052.14** (534.65)	839.11*** (314.52)	642.36* (356.81)	-245.95 (537.10)	389.58 (388.65)	-374.96 (295.31)	186.30 (241.78)	1843.28* (1044.88)	1331.78 (29.54)
Incidence of brucellosis	-6.40 (10.68)	-4.00 (6.34)	12.76* (7.14)	-10.19 (10.84)	20.35*** (7.75)	2.81 (5.93)	-12.93** (4.89)	14.82 (23.37)	16.86 (0.64)
Prevalence of diabetes mellitus	-132.79 (322.47)	24.22 (189.02)	-95.93 (211.26)	35.92 (312.81)	374.87 (229.00)	52.32 (174.27)	-183.12 (139.62)	-573.40 (1256.2)	1150.10 (42.46)
Incidence of gonococcal infections	33.04** (14.78)	6.28 (8.66)	28.89*** (9.84)	16.47 (14.84)	13.40 (10.71)	6.57 (8.17)	10.11 (6.68)	-7.05 (27.42)	29.91 (0.93)
Incidence of viral hepatitis	10.45 (14.59)	3.67 (8.64)	15.33 (9.61)	20.25 (14.12)	63.47*** (10.17)	29.20*** (7.96)	-9.52 (6.52)	143.28*** (54.62)	36.72 (1.49)
Fetal death rate per 1000 births	0.12 (1.69)	1.63 (1.00)	0.67 (1.11)	1.45 (1.66)	0.73 (1.20)	0.50 (0.92)	0.32 (0.75)	5.23 (5.21)	7.61 (0.14)
Perinatal mortality rate per 1000 births	4.46** (2.02)	5.21*** (1.20)	5.18*** (1.34)	2.85 (1.99)	5.10*** (1.45)	-0.20 (1.10)	0.57 (0.89)	11.49* (6.42)	11.07 (0.17)
Number of abortions on 100 born alive and dead	23.49 (17.82)	68.47*** (10.50)	23.06* (11.90)	1.34 (17.91)	-1.87 (12.96)	-4.86 (9.84)	3.09 (8.06)	61.95* (34.04)	38.67 (0.83)

Standard errors in parenthesis: *** p<0.01, ** p<0.05, * p<0.1

7.1. Birth and maternal health related variables

Incidence of pregnancy complications, childbirth and puerperium

- Models 6-8 with fixed year effects suggest an increase in pregnancy complications after 2002 for all-control groups. A similar trend can also be noticed from models 4-5 with time variables (positive time and negative time squared coefficients).
- Dummies for radiation risk zones are statistically significant in models 4-6. Model 7 also indicates that zones of radiation risk have statistically different levels. Model 6 predicts 5161, 4248 and 2011 more cases of pregnancy complications for extreme/maximal, high and minimal radiation risk zones, respectively. Observed mean incidence of pregnancy complications, childbirth and puerperium is around 3386 – those living in extreme and maximal regions (ZEMR) have 2.52, those in high risk regions (ZHR) have 2.25, and those in "minimal" excess risk regions (ZMR) have 1.59 times higher incidence.
- Treatment zones do not converge to the control group over time and have their own unique trend.
- Our regression results support the conclusions from kernel matching – PSM suggests that treated areas have 1.5-2.3 times higher incidence than the hypothetical controls.

Infant mortality, per 1000 live births

- In comparison with the base year, models with fixed year effects (6-8) suggest that national level of infant mortality has been on a decline since 2002. Models with time variables (4 and 5) also confirm this decreasing trend.
- Results on differences between radiation exposed areas and all-control group are contradictory. Models 4 and 6 indicate significant differences for each of the treated areas, while models 5, 7 and 8 suggest that areas exposed to radiation have insignificant level differences from all-control group. Based on model 6, extreme/ maximal, high and minimal radiation risk zones have additional 3, 4.5 and 2.9 infant deaths per 1000 live births on average. Given that the observed mean infant mortality in all-control groups is around 13, areas exposed to radiation have 22-34% more infant deaths.
- The exposed areas do not seem to follow the national level trend over time.
- PSM predicts that areas exposed to radiation have 1.28-1.49 times higher infant mortality than all-control group. This result is partially confirmed by regressions.

Prevalence of congenital anomalies

- Both models with fixed year effects (6-8) and models with time variables (4-5) indicate the growth of the prevalence of the congenital anomalies at national level starting from 2002. As an example, models 7 and 8 predict a peak in the prevalence that results in an additional 191 cases of congenital abnormalities registered in the year 2017 compared to the base year 2000.
- Models 4-6 predict significant differences for ZEMR and ZHR. Model 7 predicts statistically significant differences for zones of radiation risk (ZRR) – on average, districts that belong to the exposed areas have 90 cases of congenital anomalies more. Somewhat similar results are noted from model 8: 140 cases more for zone of high radiation risk. Given the fact that observed mean for the control group is around 139, it can be said that on average prevalence of congenital anomalies in ZHR is twice the prevalence of all-control group.

Zones of radiation risk tend to converge to the national pattern over time.

- Regression results contradict the PSM findings, where no statistically significant differences were observed after matching for ZEMR and ZHR. These differences are likely to arise because observations from three districts in ZHR are off-support and not included in estimation.

Hospital discharges, certain conditions originating in perinatal period

- Broadly speaking, all of the models predict an "inverse-U" shaped trend of the variable with the peak in the year 2007 (around 1000 cases more compared to the base year 2000).
- Regression analysis provides contradictory results on the statistical differences of areas exposed to radiation. Models without interaction terms (4 and 6) predict significant differences for all three treatment groups, while models with year interactions (7 and 8) indicate that differences are insignificant. Such discrepancies can come from the fact that trends of treated areas are very similar to the trend of the all-control group; for this reason treatment dummies lose significance when the trends (treatment interactions) are calculated in models 7 and 8.
- Models with year fixed effects indicate that exposed areas tend to follow the trend of the all-control group.
- PSM results predict that certain conditions originating in perinatal period are 1.37-1.40 times more likely in ZHR and ZMR. Regression results from model 6 align with PSM findings – certain conditions originating in perinatal period are 1.43-1.62 times higher than the observed mean of all-control group.

Fetal death rate per 1000 births

- All of the models predict a "U shaped" trend of the standardized fetal death rate for all-control group. Compared to the year 2000, there was a drop in fetal death rate around 2003-2007, followed by an increase 2008.
- Regression results are inconclusive. Models without interaction terms (4 and 6) indicate that zones of high and minimal radiation risk are significantly different from the all-control group; model 5 indicates such differences only for the zone of minimal risk, while models 7 and 8 suggest that none of the exposed areas are different.
- There were spikes of fetal death rate of unknown origin in ZEMR and ZHR in 2009 and 2013-2014 that are better "explained" by year fixed-effects rather than zone differences. Omitting these acute rises in fetal death rate in ZEMR and ZHR, it can be said that all treatment groups follow the trend of the all-control group over time.
- PSM predicts positive significant differences only for zone of high radiation risk. Differences in regressions and PSM results can be attributed to the smaller number of observations of fetal death rate in treated areas, which may have affected the quality of matching.

Perinatal mortality rate per 1000 births

- Perinatal mortality in all-control group decreased in 2004-2007, spiked in 2008, and then gradually decreased to the year 2000 level over the 2009-2018 period. It is unclear whether the observed trends reflected actual events or changes in measurement quality.
- Regression results are inconclusive. Models without interaction terms (4 and 6) indicate that all zones of radiation risk are different from the all-control group, while models with time or year interactions (5 and 8) suggest that only high and minimal radiation risk zones are different. Based on model 8, those living in zone of high radiation risk have 5.5 and those living in minimal radiation risk zone have 6.9 cases more on average, respectively. Given that the observed mean of all-control group is around 11 cases, the regression model predicts that ZHR has a 1.5 and ZMR has a 1.63 times higher perinatal mortality rate.
- Omitting occasional spikes in perinatal mortality in treated areas, it can be said that all treatment groups follow the trend of the all-control group over time.
- Similarly to fetal death rate, PSM predicts positive significant differences only for zone of high radiation risk – those living in this area have 1.33 times higher perinatal mortality rate than matched controls. Differences in regressions and PSM results can be attributed to the smaller number of observations of perinatal mortality rate in treated areas, which may have affected the quality of matching.

Number of abortions per hundred births (including stillbirths)

- All of the models with time or year variables predict a decline in a number of abortions throughout the 2000-2018 period. Models 6-8 predict a decline of 21-23 cases in the number of abortions in 2018. Compared to the mean of all control group, which is around 39 cases, the predicted decrease in the number of abortions is around 56%.
- Regression analysis provides contradictory results on the statistical differences of areas exposed to radiation. Models 4, 6 and 7 predict significant differences for all treatment groups, while models with time/year-zone interactions (5 and 8) indicate significant differences only for zone of high radiation risk. Based on model 6, those living in zone of extreme/maximal radiation risk have 15, those in high risk zones have 51, and those in minimal radiation risk zone have 14 abortions more on average. Given that the observed mean of all-control group is around 39 cases, the regression model predicts that ZEMR has 1.38, ZHR has 2.30, and ZMR has a 1.36 times higher number of abortions.
- The time trends in ZEMR and ZMR resemble the trend of all control group, while the number of abortions in ZHR does not seem to follow that trend.
- Regressions partially align with the PSM results. PSM predicts 3.10 and 1.35 times higher abortions in ZHR and ZMR, respectively. Given certain data issues on the number of abortions mentioned in PSM section, more conclusive results require better data.

7.2. Cancer prevalence and incidence

Prevalence of all types of cancer

- All regression models indicate that cancer prevalence has been steadily increasing over 2000-2018 period. Year fixed effect models indicate that prevalence of cancer has increased by around 200 cases per raion over the period of 19 years.
- From the regressions alone, it is difficult to state that there are significant differences in prevalence in districts exposed to radiation. Models 4, 5 and 6 indicate that all zones of radiation risk have significantly elevated prevalence of cancer, while models 7 and 8 suggest that such differences are insignificant. These differences are likely to arise because of the increase in standard errors in the latter models.
- All models with time or year interactions indicate that areas exposed to radiation tend to follow the trend of the all-control group.
- PSM findings suggest that there are no positive significant differences in cancer prevalence in treated areas. Similar conclusion can be drawn from regression models 7 and 8. However, differences do emerge in graphs for ZEMR and ZHR, which implies that results from PSM and abovementioned regression specifications may be incorrect. There are two potential reasons for this: increased of the standard errors for treatment groups and off-support observations from ZHR in kernel matching.

Incidence of all types of cancer

- All year fixed effect models indicate that incidence of all types of cancer in the all-control group has risen between 2002-2005 and 2012-2016, likely because of population aging. Based on models 7 and 8, it can be said that standardized incidence increased by 7-15 cases in 2002-2005, and by 11-24 cases in 2012-2016. Given that the observed mean of all-control group is around 183 cases, the incidence has risen by 4-8% and 6-13% within these periods.
- The coefficients on all treatment groups are statistically significant in models 4-7. Based on model 6 results, the incidence is 22% higher in ZEMR, 12% higher in ZHR and 8% higher in ZMR.
- Trends of exposed areas are not very different from the trend of all-control group.
- Both PSM and regression results predict statistically significant differences in incidence of cancer for ZEMR. However, PSM and regressions do not align on ZHR and ZMR. These differences

are likely to arise because observations from three districts in ZHR were off-support and not included in PSM estimation.

7.3. TB and diseases of respiratory system

Prevalence of all forms of TB

- All year fixed effects models suggest that prevalence of TB in all-control districts was on the rise until 2006, sharply declined around 2007-2008 and then steadily decreased throughout 2009-2018.
- All regression models predict insignificant positive differences for treated areas. Moreover, models 4 and 6 predict that zones with high and minimal radiation risk have statistically negative coefficients. Based on model 6, on average both ZHR and ZMR have about fewer 30 cases of TB than the all-control group.
- Year fixed effects models predict that trends of exposed areas are not very different from the trend of the all-control group throughout 2000-2018.
- PSM results depict positive significant differences for all treated groups. Admittedly, this may contradict the regression findings, which suggest that exposed areas are not very different from the all-control group. However, PSM results indicate that after matching on age-gender, ethnic composition and economic variables the prevalence in the exposed areas would have been lower than observed in our data had there not been radiation exposure.

Incidence of all forms of TB

- The trend of incidence of all forms of TB in all-control group is similar to the trend of prevalence of TB. All year fixed effects models suggest that incidence of TB increased over 2001-2005, and then steadily decreased throughout 2006-2018.
- All regression models predict either negative or insignificant positive differences for treated areas. As an example, model 6 predicts that zones with high and minimal radiation risk have 7 and 10 fewer cases, respectively.
- Year fixed effects models predict that trends of exposed areas are not very different from the trend of the all-control group throughout 2000-2018. Occasional differences for zones of extreme/ maximal and high-risk zones are observed.
- PSM results yield positive, significant differences for all treated groups. This contradicts the regression findings, which suggest that exposed areas are not very different from all-control group. However, we tend to favor the PSM results which indicate that, after matching on demographic composition and economic variables, the prevalence in the exposed areas is elevated.

Hospital discharges, diseases of respiratory system

- Year FE models (6,7,8) suggest that after 2004 there was an increase in the incidence of respiratory system diseases at national level. Based on the calculations from model 7, the number of respiratory diseases was 1.07-1.24 times higher than in the base year.
- All models predict significant differences. Model 8 predicts 17,307, 9,944 and 8,421 more cases of diseases of respiratory system for extreme/maximal, high and minimal radiation risk zones, respectively. Observed mean incidence of diseases of respiratory system is around 17,868. By implication, those living in ZEMR have 97%, those in ZHR have 56% , and those in ZMT have 47% higher incidences.
- Despite some converging trends, differences between zones exposed to radiation and all-control group prevail in 2018.
- Regression findings align with PSM results: after matching, ZEMR has 2.03, ZHR has 1.69, and ZMR has 1.49 times higher incidence, respectively, than their matched controls. These are among the most elevated conditions we find.

7.4. Circulatory system diseases and related variables

Incidence of diseases of the circulatory system

- Starting from 2002, all year FE models (6, 7 and 8) predict significant yearly differences in incidence compared to the base year. This result suggests that incidence of circulatory diseases has been on a rise in all-control group.
- Model 7, 8 predicts insignificant differences for radiation exposed zones. Models 4, 5 and 6 predict significant differences for zones of extreme/maximal and high radiation risk. Based on model 6, treatment dummies for ZEMR and ZHR are 1251 and 179, respectively. Given that the mean incidence of diseases of circulatory system is 1981 in the all-control group, we can conclude that those living in ZEMR have 1.63 and those living in ZHR have 1.09 times higher incidence.
- Trends over time of radiation treated areas are not very different from the trend of the all-control group.
- These results partially align with PSM results, where significant differences after matching were noted in zone of extreme and maximal radiation risk. PSM results suggest that residents of the most treated zone have 1.29 times higher incidence of diseases of circulatory system.

Incidence of ischemic heart disease

- Similar to the incidence of circulatory system diseases, incidence of ischemic heart diseases was on the rise from 2002.
- Regression models suggest different results. Models without interaction terms (4 and 6) predict positive significant differences for zone of extreme/maximal and high radiation risk. More precisely, treatment dummies are around 325 and 45 cases for ZEMR and ZHR, respectively. Given that the observed mean in all-control group is around 364 cases, models 4 and 6 suggest that ZEMR has 1.89 and ZHR has 1.12 times higher incidence of ischemic heart diseases. Models 5 and 8 suggest that such differences are rather insignificant (ZHR is significant at only 10% in model 5). Such differences in regression outcomes can be attributed to high variability of observations within treatment groups.
- Regression results indicate that time trends of areas exposed to radiation are similar to the trend of all-control group.
- PSM results partially align with regression results from models 4 and 6. Both techniques indicate that incidence of ischemic heart disease in ZEMR is statistically different than incidence in all-control group– 2.27 times more based on PSM and 1.89 times more based on regressions. As for ZHR, PSM predicts the differences to be insignificant.

Incidence of cerebrovascular diseases

- Models with time dummies indicate that there is no clear time trend over 2000-2018 period. Year FE models suggest that incidence of cerebrovascular diseases for all-control group followed a “U shape”. There were two spikes in incidence in 2002 and 2017-2018. During these years incidence were elevated by around 90-110 cases (models 6, 7 and 8). Given that observed mean of all-control group is close to 185, incidence in all-control group was 1.49-1.59 times higher during these years. For the years between 2002 and 2017, incidence was elevated by 20-60 cases compared to the base year of 2000. In general, regressions indicate that incidence of cerebrovascular diseases has risen over the period of 19 years.
- Models 4 and 5 suggest that both ZEMR and ZHR are statistically different from the all-control group, while models 6 and 8 indicate that only ZEMR is statistically different. Such differences are likely to arise due to higher standard errors in the year fixed-effects models.
- Regression results indicate that time trends of areas exposed to the radiation are not very different from the trend of the all-control group.
- Regression findings contradict the PSM results, where no positive significant differences for

exposed areas are observed. In essence, once we control for age-gender, ethnic and economic composition, and restrict comparison to similar districts, zone differences disappear.

7.5. Other systemic diseases

Incidence of diseases of the musculoskeletal system and connective tissue

- Incidence of diseases of musculoskeletal system and connective tissue in all-control group follows a "U shape". Based on year fixed-effects models, after a sharp increase in incidence around 2002-2006 incidence stabilized, and increased once again after 2016. As with other time trends, it is possible that changes in reporting standards and measurement account for some or all of the patterns: what matters to us above all are differences across comparable regions.
- All year and time effects regression models predict positive significant differences for all treated areas. Based on model 8, the treatment dummy for zone of extreme/maximal risk is 2,394 (implying that many additional cases), for the zone of high radiation risk is 2,893, and 698 for zone of minimal radiation risk. Given observed mean around 1,477 cases, regression indicates a 2.62, 2.96, and 1.47 times higher incidence for ZEMR, ZHR and ZMR, respectively.
- Overall, the differences between treated groups and all-control tend to shrink over time, but do not disappear completely. The biggest differences between treated and control are observed in zone of high radiation risk. A likely explanation is that ZHR has a higher proportion of people of retirement age stemming from ethnic composition differences in this treatment group (Russian and Ukrainian populations tend to be older and less healthy).
- PSM and regression findings align well for diseases of musculoskeletal system and connective tissue diseases. PSM predicts 1.34-2.24 times higher incidence, while regressions predict 1.47-2.96 times higher incidence for exposed areas. Both PSM and regressions predict highest differences for ZHR and lowest differences for ZMR.

Incidence of endocrine, nutritional and metabolic diseases, immunity disorders

- The incidence in all-control group follows an inverse "U" shape. All year FE models predict a rise in incidence from 2002-2010 after which incidence gradually decreases.
- Most of the regression models predict positive significant differences for extreme/maximal and high radiation risk zones (insignificant difference for ZEMR in model 8). Based on model 6, those in extreme/maximal risk zone have additional 551 cases, while those in high risk zone have 241. Observed mean incidence of EMB diseases and immunity disorders is around 687, which implies that those in ZEMR and ZHR have 1.80 and 1.35 times higher incidence, respectively.
- Based on the year fixed effects models, time trends of the treated and all-control groups seem to be similar: convergence of the gaps between ZEMR, ZHR and all-control group is not observed.
- Both PSM and regressions predict positive significant differences for ZEMR and ZHR. While regressions predict 1.35-1.80 times higher incidence, PSM predicts that incidence is 1.24-1.49 times higher.

Incidence of diseases of the nervous system and sense organs

- The overall trend in all-control group follows the inverse "U" shape. All year FE models predict a rise in incidence from 2002-2009 after which incidence gradually decreases to the base year level.
- All of the year fixed effects or time dummy models predict that incidence of diseases of the nervous system and sense organs in areas exposed to radiation is significantly different from the all-control group. Based on model 8, those living in extreme/maximal risk zone have 2,416, those in high have 1,016, and those in minimal 692 cases more on average. Compared to 1,486 – the observed mean of all-control group, treated areas have 1.47-2.62 times higher incidence.
- Trends of zone of high and minimal radiation risk are similar to all-control group, but zone of

extreme/maximal radiation is different. After 2010, it converges to the trend of all-control group, for this reason the gap between ZEMR and all-control decreases.
- PSM and regression results support each other. After kernel matching, we find that treated areas have 1.4-2.8 times higher incidence than their matched controls.

Incidence of diseases of the digestive system
- Incidence of diseases of diseases of the digestive system in all-control group follows a "U" shape. Based on year FE models, incidence sharply increases in 2002-2003, gradually declines, and sharply increases again in 2016-2017.
- Regression models suggest different results. Models without interaction terms (4 and 6) predict positive significant differences for zone of extreme/maximal and minimal radiation risk. Based on model 6, treatment dummies are around 845 and 483 cases for ZEMR and ZMR, respectively. Given that the observed mean in all-control group is around 3,077 cases, model 6 results suggest that ZEMR has 1.27 and ZMR has 1.16 times higher incidence of diseases of digestive system. However, models 5, 7 and 8 suggest that zone differences are rather insignificant. Such differences in regression outcomes can be attributed to high variability of observations within the treatment groups.
- Based on fixed effects model with zone-year interactions, time trends of the exposed areas are similar to the trend of the all-control group.
- Kernel PSM predicts positive significant differences for zones of high and minimal radiation risk. The incidence is 1.27 times higher in ZHR and 1.25 higher in ZMR. Due to higher standard errors, PSM predicts insignificant differences for ZEMR.

Incidence of diseases of the genitourinary system
- There is no clear time trend for the diseases of genitourinary system. Incidence remains relatively stable over the entire period of the analysis, with the exception of a spike of unknown origin in 2005.
- By and large, models 4-8 predict positive significant differences for areas exposed to radiation (model 5 predicts insignificant differences for ZMR, while model 8 predicts insignificant differences for ZMR and ZHR). Based on model 6, treatment dummies are 1,979, 1,367 and 638 for ZEMR, ZHR and ZMR, respectively. Given that the observed mean in all-control group is around 3,173 cases, model 6 results predict that ZEMR has 1.62, ZHR has 1.43 and ZMR has 1.20 times higher incidence of diseases of genitourinary system, respectively.
- Over time, trends in extreme/maximal and minimal radiation risk zones tend to converge to the trend of all-control group. The difference in the zone of high radiation risk does not converge over time.
- Kernel PSM predicts positive significant differences for zones of high and minimal radiation risk. Incidence is 1.42 times higher in ZHR and 1.19 times higher in ZMR. Higher standard errors are the likely reason why significant differences for ZEMR are not observed after kernel matching.

Prevalence of diabetes mellitus
- Year FE models demonstrate that the prevalence of diabetes has been increasing steadily over 2000-2018. Given that coefficient for the year 2018 is around 1,300 in models in 7 and 8, and the observed all-control group coefficient is around 1,150, it appears that prevalence more than doubled over the period of 19 years, perhaps reflecting an aging population.
- None of the year or time dummy models predict significant differences for any of the treated groups.
- Year FE models indicate that time trends of areas exposed to radiation are very similar to the trend of all-control group.
- Similar to regressions, PSM does not establish any statistical differences between treatment and the all-control groups.

7.6. Infectious diseases

Incidence of syphilis

- The incidence of syphilis in the all-control group declines sharply within 2001-2005 period and then decreases at a slower rate. Based on year fixed effect models with interactions, the year 2018 coefficient is around -114, while the constant term is around 150-155. As a result, we can conclude that models 7 and 8 predict an overall decline in incidence of syphilis around 73-76%.
- Regression results on the effects of radiation diverge. Model 4 and 6 predict negative significant differences for zone of high radiation risk. Both models predict that those living in the area of high radiation risk tend to have around 8-9 fewer cases of syphilis on average. Contrary to model 4 and 6 results, model 5 predicts positive significant differences for ZHR and negative significant differences for ZEMR. Model 5 predicts that treatment dummies are around -33 and 16 for ZEMR and ZHR, respectively. Lastly, model 8 predicts negative significant differences for ZEMR only with 42 cases less than in all-control group.
- Despite differences in starting points, incidence of syphilis in treated groups tends to converge to the trend of all-control group over time.
- PSM establishes no significant differences between treatment and all-control groups.

Incidence of brucellosis

- Based on year FE models, incidence of brucellosis in all-control group increases during 2001-2010 and declines thereafter.
- Regression models suggest different results. Models 4-6 predict negative significant differences for the zone of extreme/maximal and positive significant differences for zone of minimal radiation risk. Models 4 and 6 predict treatment dummies of around -13 for ZEMR and +10 for ZMR. Given that observed mean of all-control is around 17, these models suggest that incidence is 76% less in ZEMR and 59% more in ZMR. Model 5 predicts negative significant differences for both ZEMR and ZMR. A negative coefficient for treatment dummy only captures the fact that incidence of brucellosis grows over time in ZMR. Models 7 and 8 capture no significant differences in areas exposed to radiation.
- Trends of extreme/maximal and high radiation risk zones are similar to the trend of all-control group. The trend of zone of minimal radiation risk is different from other groups. These differences are captured by model 5 and model 8. Model 5 suggests that incidence in this group follows an inverse “U: shape, while model 8 indicates that peak in incidence occurs throughout 2006-2011.
- PSM results partially align with regression findings. Kernel matching predicts that those living in minimal risk zone have 2.2 times higher incidence of brucellosis than their controls.

Incidence of gonococcal infections

- Year FE and time dummy models predict a steady decrease in incidence of gonococcal infections over 2000-2018 period. Based on model 8, coefficient on 2018 year is around -51 and the constant is close to 170. Therefore, the regression predicts that ceteris paribus incidence in 2018 has dropped by 30% when compared to the base year 2000.
- Regression models suggest different results. Models 4 and 6 predict positive significant differences for zones of high and minimal radiation risk. These models predict treatment dummies around 6.3 for ZHR and 11 for ZMR. Given that observed mean of all-control is around 30, models 4 and 6 suggest that incidence of gonococcal infections is 21% and 36% higher in ZHR and ZMR, respectively. Models 5, 7 and 8 predicts no significant differences for all treatment group dummies.
- Based on models 7 and 8, it can be concluded that trends of extreme/maximal and minimal radiation risk zones are similar to the trend of all-control group. The trend for the zone of high radiation risk differs from that of other groups. These differences are captured by model 5 and

model 8. Model 5 suggests that incidence in this group follows an inverse "U" shape; while model 8 indicates that incidence in zone of high radiation risk was elevated throughout 2004-2007 period.

- PSM results partially align with regression findings. Kernel matching predicts that those living in minimal risk zone have 2 times higher incidence of gonococcal infections than matched controls. PSM predicts insignificant differences in zone of high radiation risk, likely reflecting a loss of observations that are off-support during the matching procedure.

Incidence of viral hepatitis

- All year fixed effects and time dummy models predict that the incidence of viral hepatitis in all-control group decreases at decreasing rate throughout 2000-2018. Based on model 8, coefficient on 2018 year is around -106 and constant is close to 178. Therefore, the regression predicts that ceteris paribus incidence of viral hepatitis in 2018 has dropped by about 60% when compared to the base year 2000.
- Regression models suggest different results. Models 4 and 6 establish that areas exposed to radiation are not statistically different from the all treatment group. Model 7 predicts negative significant differences for zones of radiation risk, while model 8 indicates that such differences are present only in zone of high radiation risk. Based on model 7, zones of radiation risk on average have 45 cases less.
- Based on year FE models, we can conclude that all treatment groups converge to the trend of all-control group over time.
- PSM predicts no significant differences for all treatment groups. The PSM results could reflect a smaller number of observations of incidence of viral hepatitis compared to other variables in the dataset, but inconsistent regression results also may reflect little or no true effect.

7.7. Other

Incidence of diseases of the skin and subcutaneous tissue

- There is a decreasing trend of diseases of skin and subcutaneous at country level. Based on model 7, there is an 11-14% decrease from 2012 onward in the incidence compared to the base year.
- All models predict significant treatment group differences. Model 8 predicts 4,453, 1,478 and 1,752 more cases of diseases of skin and subcutaneous for extreme/maximal, high and minimal radiation risk zones, respectively. Observed mean incidence of diseases of skin and subcutaneous is around 2,648 – those living in ZEMR have 2.68, those in ZHR have 1.55, and those in ZMR have 1.66 times higher incidence.
- Despite some converging trends, large differences between zones exposed to radiation and the all-control group prevail in 2018.
- Regression findings align with results of PSM , indicating 1.35-2.81 times higher incidence of diseases of skin and subcutaneous in treated areas after matching. The exceptionally high values for those living in ZEMR districts in both PSM and a variety of regression specifications points to a future research focus.

Incidence of narcological disorders

- Incidence of narcological disorders was on the rise from 2002 until 2006 but has since gradually decreased. Based on model 8, incidence in 2018 is 95 cases fewer than the incidence/district in the year 2000, which is around 490 cases.
- By and large, models 4-8 predict positive significant differences for zones of extreme/maximal and minimal radiation (model 8 predicts insignificant differences for ZMR, model 7 predicts significant differences for zones of radiation risk). Based on model 6, treatment dummies are

around 339 and 104 cases for ZEMR and ZMR, respectively. Given that the observed mean in the all-control group is around 163 cases, model 6 results suggest that ZEMR has 3.08 and ZMR has 1.64 times higher incidence of narcological disorders.

- Based on fixed effects model with zone-year interactions, time trends of the exposed areas are not very different from the trend of all-control group. The trend in the zone of extreme/maximal radiation risk slightly diverges from the trend of all-control in 2009-2013 period.
- PSM yields very similar results those from regressions. PSM predicts that zone of extreme/maximal radiation risk has 3.4 times and zone of minimal radiation risk has 1.8 times higher incidence than matched control group. The causal mechanism underlying this stark finding – whether it reflects decreased resistance to controlled substances or an endogenous reaction caused by higher stress – remains unclear to us.

Incidence of mental disorders

- After a short period of rise in incidence of mental disorders in 2002-2004, Kazakhstan has experienced a steady decrease. Year fixed effects models suggest that incidence of mental disorders has dropped by approximately 100 cases/district over the period of 19 years.
- Regression models suggest different results. Models without interaction terms (4 and 6) predict positive significant differences for zone of extreme/maximal and minimal radiation risk. Based on model 6, treatment dummies are around 50 and 19 cases for ZEMR and ZMR, respectively. Given that the observed mean in all-control group is around 109 cases, model 6 results suggest that ZEMR has 1.46 and ZMR has 1.17 times higher incidence of mental disorders. Models 5 and 8 suggest that significant zone differences are only observed in zone of extreme/maximal radiation risk. In these specifications, the ZEMR treatment dummy is around 100.
- Based on fixed effects model with zone-year interactions, time trends of the zones of high and minimal radiation risk are not very different from the trend of the all-control group. The trend of zone of extreme/maximal radiation risk converges to the trend of all-control group after 2014.
- PSM results are similar to the predictions of regression models 4 and 6. PSM predicts that zone of extreme/maximal radiation risk has 2.4 times and zone of minimal radiation risk has 1.4 times higher incidence than matched control group. As with narcological disorders, the causal mechanism remains unclear, though the differences are large indeed.

Crude death rate

- Based on year FE models, standardized death rates in the all-control group rose from 2003-2008 and declined afterwards. These models imply that by the end of 2018 the death rate has fallen by approximately 2.6 per thousand. Given that the observed mean of all-control group is around 9.8, we can conclude that death rate has decreased by 26.5% over the period of 19 years. This reflects a remarkable increase in the health of Kazakhstan's population, especially as the age variable we include do not fully reflect overall population aging.
- Regression models suggest varied results. Models without interaction terms (4 and 6) predict positive significant differences for the minimal radiation risk zone. Based on model 6, the ZMR treatment dummy is 0.49. Given the observed mean of all-control group this represents a 5% increase in death rates. Models 5, 7 and 8 suggest that no significant differences are observed for any exposed group.
- Time trends of the exposed areas are similar to the time trend of all-control group.
- SM findings for ZEMR and ZMR align with the regression results – PSM predicts that those in minimal radiation risk zone have 9% higher death rates than matched control, while those in extreme/maximal radiation risk zone are not statistically different. For those in high radiation risk zone, PSM predicts lower death rates – this is likely caused by the loss of observations that are off-support during matching procedure.

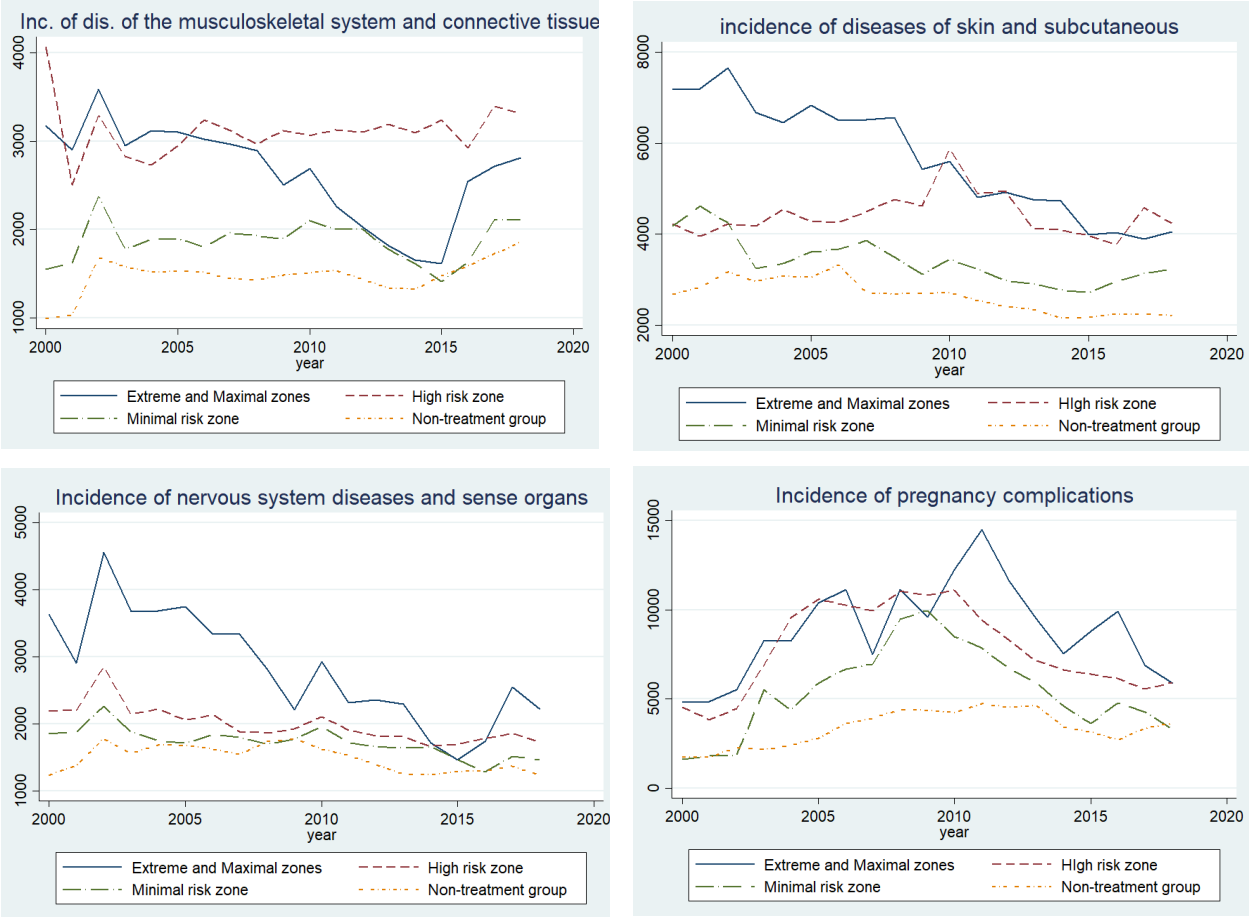
Incidence of injuries and poisoning registered at out-patient departments

- Incidence of injuries and poisoning in all-control group remains more or less stable over the period of 19 years. Based on FE models, we can tell that there was an increase in incidence in 2002-2005, and decrease in 2010-2014 periods.
- All year fixed effects and time dummy regression models suggest that all treatment zones have significantly greater incidence of injuries and poisoning. Based on model 8, treatment dummies are around 1,931, 1,111 and 1,070 cases for ZEMR, ZHR and ZMR, respectively. The observed mean incidence in all-control group is around 3,226 cases, which implies that incidence ZEMR has 1.60, ZHR has 1.34 and ZMR has 1.33 times higher incidence than all-control group.
- Based on model 8, time trend of the zone of high radiation risk is similar to the trend for the all-control group. Trends of zones of extreme/maximal and minimal radiation risk converge to the trend of all-control group after 2008, while differences in trends for ZHR remain over time.
- PSM and regressions provide diverging results. PSM predicts a 1.15 times higher incidence in ZHR only, while regressions predict significant differences for all treatment groups. The nature of such differences is unclear and requires further investigation as there are insignificant differences before matching in ZEMR and ZMR. As with elevated narcological and mental disorders, the causal mechanism remains unclear.

7.8. Time Trends

As noted above, some outcome variables appear to exhibit different time trends for treated and control groups in both regression and PSM analysis. Specifically, these variables are incidence of diseases of the musculoskeletal system and connective tissue, incidence of diseases of the nervous system and sense organs, incidence of diseases of the skin and subcutaneous, and incidence of pregnancy compilations, childbirth and puerperium. As can be seen in Figure 6, group differences can be easily observed. In general, the degree of the difference depends on the level of exposure to radiation. We can also notice that these differences do not diminish over time for incidence of musculoskeletal diseases and pregnancy complications, while differences decline, but do not disappear completely for incidence of nervous system and skin diseases. Ultimately, it is important to distinguish between outcomes for which differences are disappearing over time and those that remain large, and to focus on the latter group.

Figure 6. Time trends of "key" variables over 2000-18 period



7.9. Zones of ecological disasters and peaceful nuclear explosions

Apart from the areas exposed to ionizing radiation from SNP, Kazakhstan has several areas with officially-designated "ecologically disastrous" conditions. Most of these areas are impacted by the desiccation of the Aral Sea; however, some are affected by a history of "peaceful nuclear explosions" and/or uranium mining. Despite the fact that these areas were excluded from the PSM estimations and were assigned separate treatment dummies in regression analysis, it is reasonable to question how these areas are different from the others. Running PSM on the ecologically disastrous areas is not the focus of this paper, but the Regression Analysis section, and especially the Online Appendix, can shed some light on the issue. Preliminary findings suggest that these areas have elevated levels of prevalence and incidence of all types of cancers, prevalence and incidence of TB, incidence of viral hepatitis, incidence of digestive, genitourinary, nervous and musculoskeletal system diseases. These results tend to align with the existing studies on the ecologically problematic areas of Kazakhstan.[92] The presence of a substantial number of regions suffering from other treatments and with observably worse health outcomes serves as an important caution to not simply take all regions outside the SNP as a control.

92 Whish-Wilson, 2002; Ataniyazova, 2003

8. Conclusion

In this paper, we analyze the long-term effects of radiation exposure for regions in and near the Semipalatinsk Nuclear Polygon. The uniqueness of our approach is in the use of district level population data and propensity score matching. Utilizing the broad set of socio-demographic controls and using satellite-night light data as a proxy for district level economic activity, we are able to generate matches for districts exposed to radiation fallout from SNP and analyze their "treatment effects" 11-29 years after the last nuclear test (38-56 years after the last atmospheric explosion).

To paraphrase Faulkner, the past is neither past nor over. Our findings make it clear that the atomic tests have a very long shadow. Using 30 health characteristic variables from MedInfo, we find a large subset of variables, where significant differences were observed in all radiation exposure zones for both PSM and regressions. Such health variables include incidence of diseases of the musculoskeletal system and connective tissue, incidence of diseases of the nervous system and sense organs, incidence of diseases of the skin and subcutaneous tissue, diseases of respiratory system, and incidence of pregnancy complications, childbirth and puerperium. We also establish a list of health indicators where significant differences were predicted for most of the "treatment" zones or regression models. These variables include incidence of endocrine, nutritional and metabolic diseases, immune disorders; standardized infant mortality, incidence of diseases of genitourinary system, incidence of diseases of the digestive system, certain conditions originating in perinatal period, and incidence of narcological disorders. In general, our results are broadly consistent with existing academic literature on long-term effects of exposure to ionizing radiation. Our results strengthen these past studies by defining larger treatment groups and consistent comparisons across a very wide array of outcomes.

The time trend analysis of the variables above suggests exposure effects are persistent and intergenerational in nature. Ongoing research that builds on what has been done here will be needed to determine how rapidly the risks are diminishing, both in the SNP and in other parts of Kazakhstan that have been "treated" in a myriad of ways.

Sadly, SNP is not a unique case in a world history of atomic testing and radioactive pollution. Exposure to ionizing radiation has been observed in the USA at the Nevada Test Site, the Rocky Flats Plant in Colorado, and the Hanford Site in Washington State, in Russia at the Mayak Plant in Chelyabinsk Region, and in many other places. What makes a SNP case special is that (a) the exposure levels were exceptionally high, (b) the "treated" population had severely restricted mobility and therefore could not leave, and suffered prolonged exposure, and (c) it has been uniquely well-documented. For this latter reason, we conclude with an expression of recognition and admiration to the administrators and statisticians at Kazakhstan's Ministry of Health. We also would like to acknowledge Government's recognition of those hundreds of thousands of persons exposed to radiation, and to its efforts to uncover and fully detail the test site area, and to make major strides in cleaning it up. These combined efforts are what make this research possible.

References

Abishev, A.Kh., Yakovenko, A.M., Mackey, K.G. (2016). Instrumental precision of peaceful nuclear explosions conducting sites on the territory of Kazakhstan. Research and Technology Review National Nuclear Center of the Republic of Kazakhstan, 4, 92-97.

Akleyev, A.V. (2012). Chronic Radiation Syndrome. Springer.

Apsalikov, K. N., A. Lipikhina, B. Grosche, T. Belikhina, E. Ostroumova, S. Shinkarev, V. Stepanenko et al. "The State Scientific Automated Medical Registry, Kazakhstan: An important resource for low-dose radiation health research." Radiation and environmental biophysics 58, no. 1 (2019): 1-11.

Ataniyazova, O.A. (2003, March 18). Health and Ecological Consequences of the Aral Sea Crisis. The 3rd World Water Forum Regional Cooperation in Shared Water Resources in Central Asia, Kyoto, Japan.

Balmukhanov, S.B., Abdrakhmanov, J.N., Balmukhanov, T., Gusev, B.I., Kurakina, .N.N., & Raisov, T.G. (2006). Medical Effects and Dosimetric Data from Nuclear Tests at the Semipalatinsk Test Site. https://www.researchgate.net/publication/235195996_Medical_Effects_and_Dosimetric_Data_from_Nuclear_Tests_at_the_Semipalatinsk_Test_Site.

Bauer, Susanne, Boris I. Gusev, Ludmila M. Pivina, Kazbek N. Apsalikov, and Bernd Grosche. "Radiation exposure due to local fallout from Soviet atmospheric nuclear weapons testing in Kazakhstan: solid cancer mortality in the Semipalatinsk historical cohort, 1960–1999." Radiation research 164, no. 4 (2005): 409-419.

Bauer, Susanne, Boris Gusev, Tatyana Belikhina, Timur Moldagaliev, and Kazbek Apsalikov. "The legacies of Soviet nuclear testing in Kazakhstan: fallout, public health and societal issues." Radioactivity in the Environment, vol. 19, pp. 241-258. Elsevier, 2013.

Becker, Charles M., and Dina S. Urzhumova. "Mortality recovery and stabilization in Kazakhstan, 1995–2001." Economics & Human Biology 3.1 (2005): 97-122.

Brunn, Stanley D. “Fifty Years of Soviet Nuclear Testing in Semipalatinsk, Kazakhstan: Juxtaposed Worlds of Blasts and Silences, Security and Risks, Denials and Memory.” Engineering Earth, 2011, pp. 1789–1818.

Bul’bulian MA, Tokareva GD. Descriptive epidemiological study of malignant neoplasms in the Semipalatinsk district of the Kazakh S. S. R. Vestnik Akad Med Nauk SSSR. 1991(7):59–63.

Butts, A. (Director). (2010). After the Apocalypse. [Film]. Tigerlily Films.

Caliendo, M., & Kopeinig, S. (2008). Some Practical Guidance for the Implementation of Propensity Score Matching. Journal of Economic Surveys, 22(1), 31-72.

Demoscope (n.d.) Retrieved from http://www.demoscope.ru/weekly/ssp/census.php?cy=6.

Elvidge, C. D., Baugh, K. E., Kihn, E. A., Kroehl, H. W., & Davis, E. R. (1997). Mapping city lights with nighttime data from the DMSP Operational Linescan System. Photogrammetric Engineering and Remote Sensing, 63(6), 727-734.

Elvidge, Christopher D., Daniel Ziskin, Kimberly E. Baugh, Benjamin T. Tuttle, Tilottama Ghosh, Dee W. Pack, Edward H. Erwin, and Mikhail Zhizhin. "A fifteen year record of global natural gas flaring derived from satellite data." Energies 2, no. 3 (2009): 595-622.

Gilbert, Ethel S. "Ionizing Radiation and Cancer Risks: What Have We Learned from Epidemiology?" International Journal of Radiation Biology, U.S. National Library of Medicine, June 2009, www.ncbi.nlm.nih.gov/pmc/articles/PMC2859619/.

Grosche, Bernd, Daniel T. Lackland, Charles E. Land, Steven L. Simon, Kazbek N. Apsalikov, Ludmilla M. Pivina, Susanne Bauer, and Boris I. Gusev. "Mortality from cardiovascular diseases in the Semipalatinsk historical cohort, 1960–1999, and its relationship to radiation exposure." Radiation research 176, no. 5 (2011): 660-669.

Grosche, B., Zhunussova, T., Apsalikov, K. and Kesminiene, A., 2015. Studies of health effects from nuclear testing near the Semipalatinsk Nuclear Test Site, Kazakhstan. Central Asian journal of global health, 4(1).

Gusev, Boris Ivanovich, Zhanat N. Abylkassimova, and Kazbek Negmatovich Apsalikov. "The Semipalatinsk nuclear test site: a first assessment of the radiological situation and the test-related radiation doses in the surrounding territories." Radiation and environmental biophysics 36, no. 3 (1997): 201-204.

Gusev, B. I., et al. "The Semipalatinsk Nuclear Nuclear Polygon: A First Analysis of Solid Cancer Incidence (Selected Sites) Due to Test-Related Radiation." Radiation and Environmental Biophysics, vol. 37, no. 3, 1998, pp. 209–214.

ICRP. (2012). ICRP Statement on Tissue Reactions / Early and Late Effects of Radiation in Normal Tissues and Organs – Threshold Doses for Tissue Reactions in a Radiation Protection Context. ICRP Publication 118.

Kamiya, K., Ozasa, K., Akiba, S., Niwa, O., Kodama, K., Takamura, N., Zaharieva, E.K., Kimura, Y., & Wakeford, R. (2015). Long-term effects of radiation exposure on health. Lancet, 386, 469-478.

Kassenova, T. (2022). Atomic Steppe: How Kazakhstan gave up the bomb. Stanford University Press.

Katayama, Hiroaki, Kazbek N. Apsalikov, Boris I. Gusev, Boris Galich, Madina Madieva, Gulsum Koshpessova, Asel Abdikarimova, and Masaharu Hoshi. "An attempt to develop a database for epidemiological research in Semipalatinsk." Journal of radiation research 47, no. Suppl_A (2006): A189-A197.

Land, Charles E., Zhaksybaj Zhumadilov, Boris I. Gusev, M. H. Hartshorne, P. W. Wiest, P. W. Woodward, L. A. Crooks et al. "Ultrasound-detected thyroid nodule prevalence and radiation dose from fallout." Radiation research 169, no. 4 (2008): 373-383.

Markabayeva, Akbayan, Susanne Bauer, Lyudmila Pivina, Geir Bjørklund, Salvatore Chirumbolo, Aiman Kerimkulova, Yulia Semenova, and Tatyana Belikhina. "Increased prevalence of essential hypertension in areas previously exposed to fallout due to nuclear weapons testing at the Semipalatinsk Nuclear Polygon, Kazakhstan." Environmental research 167 (2018): 129-135.

MedInform (n.d.) Retrieved from http://www.medinfo.kz/#/dpsraion.

Mikhailov, V. N., ed. (1996). USSR Nuclear Weapons Tests and Peaceful Nuclear Explosions: 1949 through 1990. Saratov, Russia: The Ministry of the Russian Federation for Atomic Energy, and Ministry of Defense of the Russian Federation.

Mudie, Nadejda Y., et al. “Sex Ratio in the Offspring of Parents with Chronic Radiation Exposure from Nuclear Testing in Kazakhstan.” Radiation Research, vol. 168, no. 5, 2007, pp. 600–607.

Mudie, Nadejda Y., et al. “Twinning in the Offspring of Parents with Chronic Radiation Exposure from Nuclear Testing in Kazakhstan.” Radiation Research, vol. 173, no. 6, 2010, pp. 829–836.

Olds, S., & Westoff, C. (2004). Abortion and contraception in Georgia and Kazakhstan. Open Society Institute.

Shkolnik, V.S. (2002). The Semipalatinsk Test Site: Creation, Operation, and Conversion. (P.B. Gallagher, Trans.). Small World Associates.

Stawkowski, Magdalena E. "Everyday Radioactive Goods? Economic Development at Semipalatinsk, Kazakhstan." Journal of Asian Studies 76.2 (2017): 423-436.

Stubblefield, M.D. (2011). Radiation Fibrosis Syndrome: Neuromuscular and Musculoskeletal Complications in Cancer Survivors. American Academy of Physical Medicine and Rehabilitation, 3(11), 1041-1054.

Teleuov, Murat K. "Health conditions of the population and some outcomes of the international collaborative research programs in the Semipalatinsk region of Kazakhstan." International Congress Series. Vol. 1299. Elsevier, 2007.

Ургушбаева, Г. М., Кибатаев, К. М., & Мамырбаев, А. А. (2015). Экологические последствия ядерных взрывов в Западном Казахстане (обзор). Международный медицинский журнал MEDICUS (Волгоград), (2), 52.

Vakulchuk, Roman, Kristian Lundby Gjerde, Tatiana Belikhina, and Kazbek Apsalikov. "Semipalatinsk nuclear testing: the humanitarian consequences: Report prepared for the Second Conference on Humanitarian Impact of Nuclear Weapons in Nayarit, Mexico, 13–14 February 2014." (2014).

Whish-Wilson, P. (2002). The Aral Sea environmental health crisis. Journal of Rural and Remote Environmental Health, 1(2), 29-34.

Алишев, Н. В., Драбкин, Б. А., Шубик, В. М., Николаева, Н. А., & Пучкова, Е. И. (2012). Последствия ядерных испытаний на Семипалатинском полигоне. Медицина экстремальных ситуаций, (1 (39)).

Белозеров, Е. С., Киселева, Л. М., Макаров, М. А., & Игнатьев, С. Б. (2008). Индуцированные радиацией факторы, определяющие высокую инфекционную заболеваемость. Сибирский медицинский журнал (Иркутск), 82.

Логачев, В.А. (1997). Семипалатинский Полигон: Обеспечение общей и радиационной безопасности ядерных испытаний. Вторая типография ФУ «Медбиоэкстрем».

Логачев, В.А. (2001). Мирные ядерные взрывы: обеспечение общей и радиационной безопасности при их проведении. Изд.АТ.

Логачев, В.А. (2002). Ядерные Испытания СССР: современное радиоэкологическое состояние полигонов. Изд.АТ.

Логачев, В.А., & Логачев, Л.А. (2004). Основы обеспечения радиационной безопасности населения и радиоэкологических требований при проведении на Семипалатинском полигоне ядерных испытаний в атмосфере. Research and Technology Review National Nuclear Center of the Republic of Kazakhstan, 1(17). 5-14.

Логинова, О. (н.д.). Саим Балмуханов и Бахия Атчабаров. Врачи ядерного полигона. Власть. https://polygon.vlast.kz/doctors.

Лукашенко С.Н., Стрильчук Ю.Г., Умаров М.А., Кривицкий П.Е., Айдарханов А.О., Паницкий А.В., Актаев М.Р., Айдарханова А.К., Ларионова Н.В., Яковенко Ю.Ю., Субботин С.Б., Турченко Д.В., Ляхова О.Н., Романенко В.В., Койгельдинова М.Т., & Байгазинов Ж.А. (2017). Семипалатинский испытательный полигон. Современное состояние. Национальный ядерный центр Республики Казахстан Институт Радиационной Безопасности и Экологии, 52 с.

Назарбаев, Н.А. (2001). Эпицентр мира. Елорда.

Шепель, В.Н. (2007). Из истории Семипалатинского полигона. 1951-1992 гг. Архив Президента Республики Казахстан.

Судариков, Андрей Михайлович. "Проблемы обеспечения военной безопасности Советского Союза на начальном этапе" холодной войны"(ЛП Берия: политическая судьба руководителя отечественного ВПК)." Геополитика и безопасность 2 (2017): 74-82.

Appendix

Table A1. Pension Benefits for SNP Radiation-exposed Individuals in Kazakhstan and Russia

	Kazakhstan	**Russia**
Benefits cover 1949-1963	yes	yes
Benefits cover 1963-1990	yes	no
Eligibility from 5+ rem	yes	yes
Eligibility from 0.1+ rem	yes	no
Lump sum Benefits	Yes, for all five types of territories for every year lived throughout 1949-1990 (payments vary by territory)	No lump sum benefits offered
Monthly Benefits	In form of supplementary pensions for those who retired before 01/01/98 and received 35+ rems In form of additional labor payment for all five types of territories (payments vary by territory)	Yes (payments vary by territory)
Ability to "retire earlier"	If you lived in zones of emergency and maximal radiation risk (35+ rems) on the period from August 29 1949 to July 5, 1963 for no less than 5 years: men – upon reaching the age of 50 years with total work experience no less than 25 years; women – upon reaching the age of 45 years with total work experience no less than twenty years.[93]	If you resided in 25+ rems territories on the period from August 29 1949 to 1963 you can retire 10 years earlier: men – with total work experience no less than 25 years; women – with total work experience no less than 20 years.[94]

Note: It seems that monthly benefits for those who were exposed to 25+ rems are higher in Russia (as they get around 2300 rubles, compared to around 6000 tenge for those who were exposed to 35+ rems of radiation).

93. https://egov.kz/cms/en/articles/ecological_disaster_zones
94. https://docs.cntd.ru/document/9012767

Table A2. Intercalibration satellite matching coefficients

Selected Satellite	Year	Reference Satellite-Year	c_0	c_1	c_2	r_squared	n_points
F10	1992	F18 -2010	0.5720801	1.189029	-0.0026269	0.8964	39945
F10	1993	F18 -2010	0.1425793	1.153042	-0.0025901	0.8888	39945
F10	1994	F18 -2010	0.3509641	1.233192	-0.0039215	0.8928	39945
F12	1994	F18 -2010	0.6379268	1.270426	-0.0035664	0.8729	39945
F12	1995	F18 -2010	0.5527987	1.421314	-0.0066622	0.8757	39945
F12	1996	F18 -2010	0.5604528	1.652799	-0.0105613	0.8762	39945
F12	1997	F18 -2010	0.8678802	1.685768	-0.0113331	0.8785	39945
F12	1998	F18 -2010	0.8235058	1.635441	-0.0104544	0.8778	39945
F12	1999	F18 -2010	0.9501305	1.838845	-0.0143084	0.874	39945
F14	1997	F18 -2010	0.8115268	2.102258	-0.0186778	0.8848	39945
F14	1998	F18 -2010	0.9884518	2.22266	-0.0209698	0.8674	39945
F14	1999	F18 -2010	0.7485338	2.356055	-0.0233576	0.8912	39945
F14	2000	F18 -2010	0.8717967	2.043029	-0.0177786	0.897	39945
F14	2001	F18 -2010	0.9572389	2.28445	-0.0222661	0.8839	39945
F14	2002	F18 -2010	0.7626567	1.94393	-0.0161206	0.9001	39945
F14	2003	F18 -2010	0.8104049	2.0602	-0.0183067	0.8874	39945
F15	2000	F18 -2010	0.6560252	1.622598	-0.0103985	0.8915	39945
F15	2001	F18 -2010	0.7484002	1.885753	-0.0149526	0.8806	39945
F15	2002	F18 -2010	0.5409654	1.818938	-0.0134981	0.8936	39945
F15	2003	F18 -2010	0.8616156	2.246474	-0.0213641	0.8942	39945
F15	2004	F18 -2010	0.4508075	2.469991	-0.0252682	0.9144	39945
F15	2005	F18 -2010	0.3910836	2.475016	-0.0254249	0.9175	39945
F15	2006	F18 -2010	0.3289179	2.218945	-0.0206249	0.9492	39945
F15	2007	F18 -2010	0.3288791	1.970973	-0.0162126	0.9671	39945
F15	2014	F18 -2010	0.2232898	1.264555	-0.0048301	0.9413	39945
F15	2015	F18 -2010	0.2329244	1.42831	-0.007669	0.9509	39945
F15	2016	F18 -2010	0.1800695	1.375328	-0.0061646	0.9542	39945
F15	2017	F18 -2010	0.3042964	1.546636	-0.0084804	0.935	39945
F15	2018	F18 -2010	0.3749043	1.319941	-0.0058063	0.9182	39945
F15	2019	F18 -2010	0.3326902	1.21889	-0.00443	0.9315	39945
F16	2004	F18 -2010	0.4906877	2.083186	-0.0180146	0.9121	39945
F16	2005	F18 -2010	0.5056196	2.567493	-0.0269949	0.9134	39945
F16	2006	F18 -2010	0.8455413	2.055692	-0.0177791	0.9301	39945
F16	2007	F18 -2010	0.3181397	1.641349	-0.0106362	0.9635	39945
F16	2008	F18 -2010	0.4542078	1.578038	-0.0100462	0.9592	39945
F16	2009	F18 -2010	0.380155	1.49236	-0.0084157	0.9673	39945
F16	2016	F18 -2010	0.5806708	1.657959	-0.0115159	0.926	39945
F16	2017	F18 -2010	0.7829743	1.623976	-0.0111712	0.8943	39945
F16	2018	F18 -2010	0.5828748	1.284443	-0.0056751	0.9246	39945
F16	2019	F18 -2010	0.4911256	1.209641	-0.0040003	0.9152	39945
F18	2010	F18 -2010	0	1	0	1	39945
F18	2011	F18 -2010	0.1907832	1.170019	-0.0039924	0.9341	39945
F18	2012	F18 -2010	0.4213364	1.05665	-0.0020118	0.9355	39945
F18	2013	F18 -2010	0.3455815	1.052602	-0.0012669	0.9499	39945

Table A3. Population of East Kazakhstan Oblast by raion/district.

	1959	1970	1979	1989	1999	2000	2001	2002
East KZ Oblast	1255153	1559078	1646039	1771769	1532949	1516755	1499033	1482429
Oskemen	207707	230340	274287	325020	320608	316910	312472	309763
Kurchatov					9301	9327	9316	9314
Leninogorsk	83862	82342	78077	78749	65030	64472	63822	63061
Semei	197018	282658	325548	375828	298266	295696	294236	294187
Abai	13563	21969	25632	27080	17938	17870	17625	17441
Ayagoz	62562	91629	97449	85595	82044	81643	80906	80136
Beskaragai	32456	36733	33930	32959	28150	27884	27417	26892
Boroduliha	37183	52478	52849	55586	49155	48445	47575	46444
Glubokoe	45512	77294	74406	77305	67327	66648	66237	65877
Zharma	69553	73705	75268	74598	60411	59394	58066	56494
Zaisan	27081	33615	36364	37127	39572	39330	39032	38703
Zyryan	127326	118453	110544	111685	93916	92750	91708	90538
Kokpekty	54943	62398	59541	57521	45862	45169	44161	42891
Korshim	47433	57227	57499	53880	45130	44525	43725	42669
Katonkaragai	52400	56834	52627	49116	45178	44776	44132	43486
Tarbagatai	36323	62472	71463	74379	65701	65318	64703	63979
Ulan	44519	61690	63119	67010	45886	44939	44189	43348
Urzhar	64298	93089	93653	88705	95497	94735	93820	92353
Shemonaiha	51414	64152	63783	67561	57977	56924	55891	54853

Table A3 continued	2003	2004	2005	2006	2007	2008	2009	2010
East KZ Oblast	1465961	1455412	1442097	1431180	1424513	1417384	1417764	1398073
Oskemen	307104	305053	302244	299400	298827	298136	298866	316369
Kurchatov	9483	9866	9940	10169	10359	10569	10747	10406
Leninogorsk	62156	61952	61382	60758	60057	59269	58727	58511
Semei	294903	297282	301984	305473	308129	310257	313503	326965
Abai	17127	16870	16344	16101	15937	15833	15689	15314
Ayagoz	79153	78277	76605	75351	75001	74640	74653	74603
Beskaragai	26296	25851	25185	24728	24417	23755	23396	21997
Boroduliha	45293	44318	43253	42239	41554	40839	40093	39664
Glubokoe	65751	65882	65907	65962	65488	65467	65400	63642
Zharma	54840	53273	51417	50197	49368	48854	49058	44835
Zaisan	38511	38362	38072	37876	37966	38204	39040	35808
Zyryan	88963	87899	86612	85694	84941	84152	83005	76485
Kokpekty	41676	41057	40306	39598	39122	38528	38014	34241
Korshim	41650	40824	40157	39529	38924	38269	37591	31535
Katonkaragai	42651	41863	40976	40262	39537	38657	38129	29799
Tarbagatai	63111	62243	61185	60405	59737	59009	58325	46561
Ulan	42751	42619	42101	41601	41206	40770	42363	39964
Urzhar	90669	88994	86721	85134	83840	82630	82261	82613
Shemonaiha	53873	52927	51706	51706	50103	49546	48904	48761

Table A3 continued	2011	2012	2013	2014	2015	2016	2017	2018	2018 as % of 1999
East KZ Oblast	1398078	1395060	1394068	1394388	1395466	1396019	1389568	1383745	90.27
Oskemen	318818	321240	321326	325877	328848	333077	335683	341064	106.38
Kurchatov	11033	11308	11510	11673	12095	12281	12390	12383	133.14
Leninogorsk	58127	57832	58092	58007	58071	58045	57840	57541	88.48
Semei	329046	331439	335524	337733	339312	343638	344376	347284	116.43
Abai	15376	15400	15235	15361	15375	15286	15011	14548	81.10
Ayagoz	74418	73814	73434	73602	73992	74139	73931	73022	89.00
Beskaragai	21648	21387	21054	20832	20603	20133	19563	19057	67.70
Boroduliha	39087	38514	37896	37336	37206	36812	36568	36314	73.88
Glubokoe	63564	63532	63907	64083	64243	64037	63883	62870	93.38
Zharma	44597	44027	43180	42293	41920	41032	39717	38701	64.06
Zaisan	36773	37041	37924	37924	38119	37940	37480	37196	94.00
Zyryan	75234	74074	73314	72238	71234	70089	68917	67724	72.11
Kokpekty	33435	32316	32081	31573	31417	30754	29949	28950	63.12
Korshim	30783	29976	29483	28719	27931	27098	26382	25126	55.67
Katonkaragai	29223	28539	28021	27279	26627	25748	25009	23935	52.98
Tarbagatai	46111	44988	44141	43482	42663	41763	40715	39592	60.26
Ulan	40378	40310	40486	40150	40424	40524	40352	39373	85.81
Urzhar	82195	81980	81020	80094	79656	78405	77031	74766	78.29
Shemonaiha	48232	47343	46440	46132	45730	45218	44771	44299	76.41

Notes:

1. 1959-1989 data are taken USSR censuses from Demoskop.[95] During Soviet era, Kurchatov was a closed (secret, military) city. For this reason, no population information prior to 1999 is available.
2. 1999-2018 data are taken from Bureau of National Statistics of Republic of Kazakhstan.[96]

95 http://www.demoscope.ru/weekly/ssp/census.php
96 https://stat.gov.kz/official/industry/66/statistic/5

Table A4. Population of Kazakhstan's Oblasts by nationality/ethnicity, 1970-2019

Total population	**1970**	**1979**	**1989**	**1999**	**2004**	**2010**	**2018**	**1999 as % of 1989**	**2018 as % of 1970**	**2018 as % of 1999**
Akmola oblast	959729	994932	1064406	836271	748930	735134	738942	78.57	76.99	88.36
Kazakhs	195296	229024	266831	313498	310291	349330	376750	117.49	192.91	120.18
Russians	424421	442506	459348	329454	281991	262167	244786	71.72	57.68	74.30
Ukranians	102829	92557	91236	62228	52274	37760	32475	68.21	31.58	52.19
Germans	124906	127948	139032	52334	33106	26193	26090	37.64	20.89	49.85
Tatars	21064	22190	22332	17272	15445	13702	13247	77.34	62.89	76.70
Total population	**1970**	**1979**	**1989**	**1999**	**2004**	**2010**	**2018**	**1999 as % of 1989**	**2018 as % of 1970**	**2018 as % of 1999**
Aktobe oblast	550582	630383	732653	682558	671812	763591	857711	93.16	155.78	125.66
Kazakhs	261632	328392	407222	482285	504798	607983	706917	118.43	270.20	146.58
Russians	145218	158298	173281	114416	96011	102692	99589	66.03	68.58	87.04
Ukranians	77090	74794	74547	46848	38227	25112	22115	62.84	28.69	47.21
Germans	31473	30084	31628	10721	7204	5518	5691	33.90	18.08	53.08
Tatars	14747	15567	16924	11675	10186	9460	9275	68.98	62.89	79.44
Total population	**1970**	**1979**	**1989**	**1999**	**2004**	**2010**	**2018**	**1999 as % of 1989**	**2018 as % of 1970**	**2018 as % of 1999**
Almaty oblast	1272492	1453401	1642917	1558534	1571194	1836162	2017277	94.86	158.53	129.43
Kazakhs	486948	602215	741737	926137	970970	1248636	1448292	124.86	297.42	156.38
Russians	481944	514011	518315	339984	308994	305579	274027	65.59	56.86	80.60
Ukranians	28803	27376	29971	13512	10905	6221	3788	45.08	13.15	28.03
Germans	84364	93170	94123	18927	13280	8855	8592	20.11	10.18	45.40
Tatars	16501	18972	19551	15647	14812	13569	12530	80.03	75.93	80.08
Total population	**1970**	**1979**	**1989**	**1999**	**2004**	**2010**	**2018**	**1999 as % of 1989**	**2018 as % of 1970**	**2018 as % of 1999**
Atyrau oblast	340343	373706	424708	440286	457215	520988	620684	103.67	182.37	140.97
Kazakhs	239972	285081	338998	391672	413123	475519	573723	115.54	239.08	146.48
Russians	76316	67957	63673	38013	33905	33613	33521	59.70	43.92	88.18
Ukranians	5960	3913	3749	1444	1262	803	769	38.52	12.90	53.25
Germans	1872	1694	1401	687	552	467	478	49.04	25.53	69.58
Tatars	5252	4979	4913	2728	2517	2308	2325	55.53	44.27	85.23
Total population	**1970**	**1979**	**1989**	**1999**	**2004**	**2010**	**2018**	**1999 as % of 1989**	**2018 as % of 1970**	**2018 as % of 1999**
East Kazakhstan oblast	1563874	1657403	1767225	1531024	1455412	1398073	1383745	86.63	88.48	90.38
Kazakhs	509809	597217	687879	743098	745984	789722	830815	108.03	162.97	111.80
Russians	881608	899047	914424	694705	631957	555286	505010	75.97	57.28	72.69
Ukranians	34896	36152	35702	15696	12952	6754	4104	43.96	11.76	26.15
Germans	69136	65610	66924	32141	22147	13943	13069	48.03	18.90	40.66
Tatars	27964	27483	27982	24506	22981	17651	16463	87.58	58.87	67.18
Total population	**1970**	**1979**	**1989**	**1999**	**2004**	**2010**	**2018**	**1999 as % of 1989**	**2018 as % of 1970**	**2018 as % of 1999**
Jambyl oblast	792321	929983	1038667	988840	985552	1034487	1117220	95.20	141.01	112.98
Kazakhs	322762	410040	507302	640346	667510	740571	811323	126.23	251.37	126.70
Russians	256267	282403	275424	179258	153079	121415	111203	65.08	43.39	62.04
Ukranians	36454	36003	33903	10013	6283	5094	2961	29.53	8.12	29.57
Germans	66356	69427	70150	11394	7130	4431	4346	16.24	6.55	38.14
Tatars	14214	15998	16618	12576	11132	9239	8881	75.68	62.48	70.62
Total population	**1970**	**1979**	**1989**	**1999**	**2004**	**2010**	**2018**	**1999 as % of 1989**	**2018 as % of 1970**	**2018 as % of 1999**

West Kazakhstan oblast	513077	585501	629494	616800	603832	603858	646927	97.98	126.09	104.88
Kazakhs	253127	301622	351123	399030	410985	438762	490924	113.64	193.94	123.03
Russians	197171	217743	216514	174018	152152	134771	127102	80.37	64.46	73.04
Ukranians	32018	32141	28092	19634	16872	11338	9454	69.89	29.53	48.15
Germans	4135	4722	4550	2434	1641	1274	1296	53.49	31.34	53.25
Tatars	12454	12854	12703	10104	9364	8611	8444	79.54	67.80	83.57
Total population	**1970**	**1979**	**1989**	**1999**	**2004**	**2010**	**2018**	**1999 as % of 1989**	**2018 as % of 1970**	**2018 as % of 1999**
Karaganda oblast	1560468	1713208	1745448	1410218	1330927	1346822	1380538	80.79	88.47	97.90
Kazakhs	294719	373408	449837	529478	542448	631031	708337	117.70	240.34	133.78
Russians	788777	859363	817900	614416	558384	527263	494663	75.12	62.71	80.51
Ukranians	153543	143566	128547	78755	68847	49166	41201	61.27	26.83	52.32
Germans	147233	154602	159208	57229	41306	32915	32322	35.95	21.95	56.48
Tatars	47899	52603	52769	39313	35982	32540	30975	74.50	64.67	78.79
Total population	**1970**	**1979**	**1989**	**1999**	**2004**	**2010**	**2018**	**1999 as % of 1989**	**2018 as % of 1970**	**2018 as % of 1999**
Kostanay oblast	985571	1089068	1223844	1017729	913435	883379	875616	83.16	88.84	86.04
Kazakhs	188267	230054	279787	314801	304750	330118	352610	112.51	187.29	112.01
Russians	432109	483260	535100	430242	382535	378605	361435	80.40	83.64	84.01
Ukranians	179376	174975	178140	130449	113478	83717	73274	73.23	40.85	56.17
Germans	91456	97284	110440	57410	36902	28042	27538	51.98	30.11	47.97
Tatars	21115	24992	27812	20070	17701	16912	16077	72.16	76.14	80.10
Total population	**1970**	**1979**	**1989**	**1999**	**2004**	**2010**	**2018**	**1999 as % of 1989**	**2018 as % of 1970**	**2018 as % of 1999**
Kyzylorda oblast	494352	562191	574464	596215	607491	689011	783156	103.79	158.42	131.35
Kazakhs	346362	428042	504126	561630	578430	657017	753027	111.41	217.41	134.08
Russians	91797	86084	37960	17155	13660	15977	14783	45.19	16.10	86.17
Ukranians	11032	12613	3139	844	618	417	285	26.89	2.58	33.77
Germans	3116	2236	1960	376	195	177	149	19.18	4.78	39.63
Tatars	7159	6087	4538	2309	1947	1775	1569	50.88	21.92	67.95
Table A4 Continued										
Total population	**1970**	**1979**	**1989**	**1999**	**2004**	**2010**	**2018**	**1999 as % of 1989**	**2018 as % of 1970**	**2018 as % of 1999**
Mangystau oblast	159234	248842	324243	314669	349668	503241	660317	97.05	414.68	209.84
Kazakhs	72106	108290	165043	247644	289751	445610	599065	150.05	830.81	241.91
Russians	60008	99923	106801	46630	40100	39845	37355	43.66	62.25	80.11
Ukranians	7525	9348	10159	4124	3629	2205	1880	40.59	24.98	45.59
Germans	618	1081	1136	554	436	284	279	48.77	45.15	50.36
Tatars	3170	4706	5193	2490	2198	1680	1666	47.95	52.56	66.91
Total population	**1970**	**1979**	**1989**	**1999**	**2004**	**2010**	**2018**	**1999 as % of 1989**	**2018 as % of 1970**	**2018 as % of 1999**
Pavlodar oblast	697947	807224	942313	806983	745238	744395	754854	85.64	108.15	93.54
Kazakhs	175691	216113	268512	311862	319902	357415	390754	116.14	222.41	125.30
Russians	310004	370916	427658	337924	299449	286677	270721	79.02	87.33	80.11
Ukranians	85839	83185	86651	62585	53375	39550	33244	72.23	38.73	53.12
Germans	73614	81487	95342	43835	26590	20864	20554	45.98	27.92	46.89
Tatars	13972	16801	20152	17064	15655	14217	14040	84.68	100.49	82.28
Total population	**1970**	**1979**	**1989**	**1999**	**2004**	**2010**	**2018**	**1999 as % of 1989**	**2018 as % of 1970**	**2018 as % of 1999**
North Kazakhstan oblast	874986	884345	912065	725980	674497	592791	558584	79.60	63.84	76.94
Kazakhs	153733	179671	206060	214697	211900	198440	193992	104.19	126.19	90.36
Russians	458783	463114	469636	361461	336682	298362	277807	76.97	60.55	76.86

Ukranians	88902	77350	70525	46980	41713	29135	23523	66.61	26.46	50.07
Germans	86028	84465	86716	41157	27164	20818	19805	47.46	23.02	48.12
Tatars	18938	19846	19977	16472	15450	12952	12182	82.45	64.33	73.96
Total population	**1970**	**1979**	**1989**	**1999**	**2004**	**2010**	**2018**	**1999 as % of 1989**	**2018 as % of 1970**	**2018 as % of 1999**
South Kazakhstan oblast	1289088	1568985	1823528	1978339	2150256	2511580	2929196	108.49	227.23	148.06
Kazakhs	609175	801862	1017470	1340889	1483005	1818807	2135364	131.79	350.53	159.25
Russians	282553	300365	278473	162098	153798	135726	128895	58.21	45.62	79.52
Ukranians	37364	34830	33033	13039	10139	5548	3688	39.47	9.87	28.28
Germans	76382	50742	44526	5261	3844	2503	2525	11.82	3.31	47.99
Tatars	34426	37039	34615	23672	22469	19029	18349	68.39	53.30	77.51
Total population	**1970**	**1979**	**1989**	**1999**	**2004**	**2010**	**2018**	**1999 as % of 1989**	**2018 as % of 1970**	**2018 as % of 1999**
Astana city	181322	232322	281252	319324	510533	649152	1030577	113.54	568.37	322.74
Kazakhs	23068	32870	49798	133585	285054	457569	805718	268.25	3492.80	603.15
Russians	104010	133432	152147	129480	157685	124034	138175	85.10	132.85	106.72
Ukranians	19820	23197	26054	18070	19423	12978	14176	69.36	71.52	78.45
Germans	10564	14436	18913	9591	9199	7268	9243	50.71	87.50	96.37
Tatars	6554	7866	9339	8286	10559	9259	11624	88.72	177.36	140.28
Total population	**1970**	**1979**	**1989**	**1999**	**2004**	**2010**	**2018**	**1999 as % of 1989**	**2018 as % of 1970**	**2018 as % of 1999**
Almaty city	777569	956817	1071927	1129356	1175208	1390610	1801993	105.36	231.75	159.56
Kazakhs	105728	169476	255133	434397	512085	745712	1072694	170.26	1014.58	246.94
Russians	530931	612783	615365	510366	471955	452105	469614	82.94	88.45	92.02
Ukranians	32010	35974	42243	22835	19400	11662	10577	54.06	33.04	46.32
Germans	16824	21219	20806	9390	6947	5856	7499	45.13	44.57	79.86
Tatars	20260	24643	25329	24770	24337	21090	24474	97.79	120.80	98.81

Notes:

1. 1970-1999 Data on Oblast population by ethnicity are taken from “Национальный Состав Республики Казахстан. Том 1[97].

2. 2004-2018 data are taken from “population of Republic of Kazakhstan by ethnicity” from Bureau of National Statistics of Republic of Kazakhstan[98].

97 http://www.demoscope.ru/weekly/ssp/census.php
98 https://stat.gov.kz/official/industry/66/statistic/5

Table A5. Total number of "polygon pensioners" by oblast, 2002-2006[99].

Number SNTS pensioners	1 Jan 2002	1 July of 2002 (1st half of 2002)	1 Jan 2003 (2nd half of 2002)	1st July 2003 (1 half of 2003)	1 Jan of 2004 (2 half of 2003)	1st July 2004 (1 half of 2004)	1 Jan 2005 (2 nd half of 2004)	1st April 2005	1st July of 2005 (1 half of 2005)	1st October 2005	1 Jan 2006 (2 half of 2005)	1 July 2006
Kazakhstan	15565	15947	15952	15931	15934	16030	13459	12329	12261	11688	11619	12462
Akmola	86	90	91	79	85	84	85	86	85	52	51	52
Aktobe	5	5	4	4	4	4	4	5	5	87	84	118
Almaty	603	616	608	627	632	769	769	781	781	574	577	566
Atyrau	7	7	7	7	7	11	8	8	8	0	0	0
West Kazakhstan	65	65	69	69	68	69	69	71	60	8	8	9
Zhambyl	82	57	57	57	54	55	49	35	44	13	13	13
Karagandy	134	135	134	139	138	137	135	134	132	200	199	198
Kostanay	71	99	99	84	84	107	103	101	99	17	15	14
Kyzylorda	11	11	15	15	15	15	14	14	14	4	4	4
Mangystau	6	8	8	9	8	7	7	7	7	41	40	37
Turkistan	45	31	29	28	26	22	15	15	16	11	11	11
Pavlodar	4161	4197	4197	4137	4055	4017	1503	1493	1482	1475	1466	1444
North Kazakhstan	30	29	31	31	30	29	28	28	28	5	7	8
East Kazakhstan	10012	10337	10345	10385	10464	10432	10402	9287	9235	8962	8901	9746
city of Astana	35	36	36	35	34	38	40	38	36	8	10	10
city of Almaty	212	224	222	225	230	234	228	226	229	231	233	232

99 https://stat.gov.kz/official/industry/66/statistic/5