Water Afar Off Quenches Not Thirst? The Health Benefits of the South–North Water Transfer Project in China

Huanxiu Guo¹, Ding Li², Shuang Ma³, and Peng Zhang⁴

¹Nanjing Audit University ²Southwestern University of Finance and Economics ³Guangzhou University, ⁴The Chinese University of Hong Kong, Shenzhen

January 10, 2024

Abstract

This paper is the first to estimate the health benefits of the largest inter-basin water diversion project in the world, namely, the Chinese South–North Water Transfer (SNWT) project. Using detailed household-level data, we exploit the natural experimental setting of the middle route of SNWT to implement a difference in discontinuity design. We find that SNWT reduced the incidences of waterborne diseases and cancer by more than 25% and 16%, respectively. The cost-benefit analysis shows that the health benefits outweigh the construction costs in 4–10 years, which highlights the importance of water transfer infrastructure for sustainable development in arid regions.

Keywords: Water scarcity, Water transfer, Clean drinking water, Public health, China **JEL codes:** O13, O18, I15, Q12, Q54

1 Introduction

Safe and clean drinking water, defined as water that does not represent any significant risk to health over a lifetime of consumption and is free of microbial pathogens, chemical, and radiological substances (WHO), is vital to human health and the sustainable development of human society (Uhlenbrook et al., 2022). In 2015, the UN General Assembly put forward a "human-centered" set of Sustainable Development Goals, of which Goal 6 aims to ensure safe drinking water and sanitation for all. However, given the uneven distribution of water resources across the globe, many countries are facing severe problems in the provision of clean drinking water. Approximately 2.1 billion people in the world lack safe and clean drinking water at home, of which over 1/10 live in northern China.¹ The acute water shortage, over-exploitation of groundwater, climate change, and water pollution all compound threats to drinking water and public health in the arid region (Tao and Xin, 2014).

Recognizing these threats, the Chinese government sought to quench the thirst of the north. Among other solutions, the South–North Water Transfer (SNWT) project is regarded as the largest and most ambitious one, aiming to divert tens of millions m³ of water from Yangtze River through three water routes that extend by over more than a thousand kilometers (Webber et al., 2017). As an ancient solution to imbalanced water resources, inter-basins water transfer had been widely used in countries, such as the United States, Canada, Australia, India, and South Africa (Faúndez et al., 2023).² Despite the worldwide prevalence of inter-basins water transfer projects, whether these projects effectively alleviate the water crisis remains unclear, and their controversial environmental and socio-economic impacts are still being debated (Berkoff, 2003; Barnett et al., 2015; Purvis and Dinar, 2020).

Proponents of these projects stressed on the benefits of water transfers for various uses, such as for domestic consumption, agricultural irrigation, energy production, ship transportation, and eco-environmental restoration, which contribute to the GDP and produce other economic benefits for the host countries (Berkoff, 2003; Crow-Miller, 2015). By contrast, critics of these projects argued that water transfers have complex socio-environmental consequences, such as relocation of

¹See WHO and UNICEF JMP, Progress on Drinking Water, Sanitation and Hygiene: Update and SDG Baselines (2017), available at https://data.unicef.org/resources/progress-drinking-water-sanitation-hygiene-2017-updatesdg-baselines.

²Inter-basins water transfer projects were discovered in ancient Babylon and the Egyptian civilization. Modern examples of these projects include the central Arizona project in the United States, the Quebec water transfer project in Canada, the national river linking project in India, and the Orange river development project in South Africa.

indigenous populations, changes in the natural cycle of river basins, loss in biodiversity of flora and fauna due to flooding, pollution spillovers during the transfer process, huge financial costs, and increased water prices (Purvis and Dinar, 2020; Sheng et al., 2020; Faúndez et al., 2023). However, the crucial health impacts of inter-basin water transfer projects have been largely overlooked in the literature. Nevertheless, academic debates and policymaking should be based on rigorous empirical evidence and holistic cost-benefit analysis, both of which are absent in the literature.

As the first to provide convincing evidence for the health benefits of inter-basin water transfer projects, this paper explores the causal impact of SNWT on public health and highlights the role of safe and clean drinking water to the success of this project. However, estimating causal relationships is empirically challenging in this context because of the residents' non-random access to SNWT. To address this problem, we exploit the specific regional topography and the middle route operation of SNWT. As shown in Figure S1, the middle route was constructed along the Taihang Mountain. The water level in the Danjiangkou Reservoir has been raised by 13 meters to allow water flows from the higher reservoir to the lower end of Beijing and Tianjin but not to the Loess Plateau in the west of Taihang Mountain due to its much higher altitude. Therefore, the middle route creates a natural spatial cutoff that divides the residents into treatment and control groups .³ We exploit this natural experimental setting to implement a difference in discontinuity design by comparing those residents located within narrow areas to the middle route before and after the operation of SNWT.

We collect comprehensive resident-level health data from the China Family Panel Studies (CFPS) and link them to the middle route of SNWT for the empirical test. This national representative survey data provides detailed information on the incidences of chronic diseases, such as waterborne diseases and cancer, which allow us to assess the causal impact of SNWT on residents' health outcomes. We also manually collect monthly monitoring data on the water quality of centralized drinking water sources in the studied areas from the official websites of local environmental governance authorities and compile them with other climatic and socio-economic data to investigate the related mechanism.

Our estimates confirm the substantial health benefits of SNWT, which has reduced the incidences of waterborne diseases and cancer by more than 25% and 16%, respectively. The health effects of this project were not felt until 4 years following its launch in 2014. With regard to its underlying mechanism, this project has significantly improved the water quality of centralized drinking water sources rather than changing other economic factors, such as income and consumption, or the

³Note that a small number of counties in the east of the middle route is not treated by SNWT, and we thus implement a fuzzy RD design in this context.

health behaviors of residents in the North China Plain. Further heterogeneous analysis shows that SNWT has more pronounced health effects on residents living in arid areas and have access to tap water. In terms of cost-effectiveness, SNWT can reduce households' health expenses by more than 60%, amounting to 59.50 billion yuan (8.29 billion US dollars) per year. While contrasting these health benefits with the estimated construction costs obtained from the literature, our simple back-of-the-envelope calculation suggests that the former outweighs the latter in 4–10 years.

This paper contributes to two strands of the literature. First, this paper contributes to the literature on the cost-benefit analysis of water infrastructure, which plays a crucial and indispensable role in agricultural production, energy production, economic development, and climate change mitigation (Duflo and Pande, 2007; Hansen et al., 2009; Dillon and Fishman, 2019; Alsan and Goldin, 2019; Jeuland, 2020). However, the huge public investment in water infrastructure requires a holistic analysis of its costs and benefits, which may vary greatly across different infrastructure types, locations, and socio-economic conditions (Strobl and Strobl, 2011; Olmstead and Sigman, 2015). While the agricultural and socio-economic effects of irrigation dams (Duflo and Pande, 2007; Sarsons, 2015; Mettetal, 2019; Gatti et al., 2021; Mary, 2022; Blakeslee et al., 2023), hydroelectric dams (Cole et al., 2014; De Faria et al., 2017; Severnini, 2023), and wells (Q. Huang et al., 2013; Sekhri, 2014; Sayre and Taraz, 2019; Blakeslee et al., 2020) have been subjected to numerous economic evaluations, only a few studies have focused on the health benefits of water infrastructure (Zhang, 2012; Zhang and Xu, 2016). To advance this important research stream, we examine a new type of water infrastructure for long-distance water transfers and quantify its health benefits. To the best of our knowledge, this study is the first to estimate the health benefits of the largest inter-basin water diversion project in the world.⁴

Second, this paper contributes to the literature on water pollution and public health. Water pollution poses a significant concern for developing and developed countries, where it has triggered various public health issues, such as child mortality (Galiani et al., 2005; Greenstone and Hanna, 2014; Fan and He, 2023), cognitive and physical capacity impairments (Pitt et al., 2021), digestive cancer (Ebenstein, 2012), diarrheal disease (Bhalotra et al., 2021), gastrointestinal illness (Marcus, 2022) and dental health (Aggeborn and Öhman, 2021). However, how to solve water pollution and effectively improve the quality of drinking water in arid and semi-arid regions remain unclear. In addition to the existing solutions, such as privatization of water services (Galiani et al., 2005),

⁴There are fruitful policy debates on the impacts of SNWT, but most of these discussions were grounded on the political and institutional implications of this project in the absence of systematic empirical evidence (Berkoff, 2003; Moore, 2014; Barnett et al., 2015; Pohlner, 2016; Webber et al., 2017; Sheng et al., 2018).

government environmental regulations (Greenstone and Hanna, 2014), provision of piped water (Fan and He, 2023), water disinfection (Bhalotra et al., 2021), and public notification of water violations (Marcus, 2022), we propose long-distance water transfer as a new solution. Although this solution seems controversial and costly, our calculations suggest that the health benefits of water transfer projects outweigh their investment in the short or mid term.

2 Background

2.1 Acute water shortage and health consequences in northern China

The distribution of water resources in China is highly uneven, with 80% of these resources distributed in the south, depriving the northern region of access to this resource. By 2050, the population in northern China is projected to increase to 340 million, while per capita water resources in the region is projected to decline to below 347 m³, with only about 100 m³ of drinking water (Liu and Yang, 2012). Given the limited surface water resources in the country, groundwater has been overexploited. The groundwater table in northern China has been rapidly decreasing by 1 to 3 meters a year, thus causing serious environmental problems (Gleeson et al., 2010). Consequently, groundwater in northern China shows increased bicarbonate concentrations and overall hardness (Currell et al., 2012).

This water problem is further exacerbated by severe water pollution. According to the Ministry of Water Resources, the surface water quality in major northern river systems, such as the Yellow River, Huai River, and Hai River, was continuously classified as unfit for domestic and agricultural uses (superior to Class IV).⁵ Meanwhile, drinkable groundwater (Class I–III) occupied only 22.1% of shallow groundwater and 26.4% of deep groundwater in northern China (Han et al., 2016). This acute water crisis has exposed millions of people in northern China to unsafe drinking water. The elevated concentrations of arsenic, fluoride, and nitrate can be dispersed and accumulated in humans via long-term consumption of groundwater and lead to endemic fluorosis and arsenicosis. Toxic contaminants in water, such as nitrate, are also linked to chronic illnesses in the digestive system and increase the incidence of digestive cancers (Ebenstein, 2012). Many "cancer villages" have been documented in northern China, where water scarcity and water pollution are acute (Han et al., 2016). China's Center for Disease Control also confirmed that the mortality rates due to digestive tract tumors were significantly higher in the tributary areas of the Huai River Basin than

⁵Available at: http://www.mwr.gov.cn/zwzc/hygb/dxsdtyb/201604/P020160405539942030096.pdf.

in the more distant control areas (Yang and Zhuang, 2014).

2.2 Solution of SNWT

The SNWT project is regarded as the most ambitious inter-basin water diversion project aiming to solve the water crisis in northern China (Webber et al., 2017). SNWT consists of three routes, namely, the eastern, middle, and western routes, to connect the Yangtze, Huai, Yellow, and Hai rivers and make up the three verticals of the national water network (see Figure S1 for the planning of the three routes of SNWT). After 10 years of construction, the eastern route started transferring water at the end of 2013, while the middle route began its operation a year later at the end of 2014.⁶

Thus far, SNWT has transferred over 50 billion m³ of water from the south to northern China, i.e., 44.7 billion via the middle route and 5.3 billion via the eastern route, which are equivalent to about 10% of the annual flow of the Yellow River. These massive water transfers have benefited more than 40 major northern cities and 140 million residents.⁷ For example, Tianjin is now completely using south water, while the shares of south water in the domestic water supply of Zhengzhou and Beijing are 90% and 70%, respectively. The Chinese government claims that the quality of water provided by SNWT is consistently exceeding the national standard, making it safe for agricultural, industrial, and domestic uses.⁸

Although the long-distance water transfer and complicated environmental consequences of SNWT have been fiercely debated, the potential health effects of this project have been widely overlooked. This gap thus calls for a rigorous empirical assessment and holistic cost-benefit analysis of SNWT. However, these efforts present a challenge due to the non-random access of residents to SNWT. Regional heterogeneities may also confound with SNWT and impede the identification of its health effects. In this study, the specific regional topography and the middle route operation of SNWT offer a natural experimental setting for investigating this problem. As shown in Figure S1, the middle route was constructed along the Taihang Mountain. The water level in the Danjiangkou Reservoir has been raised by 13 meters to allow water flows from the higher reservoir to the lower end of Beijing and Tianjin but not to the Loess Plateau due to its much higher altitude. Consequently, the middle route creates a natural spatial cutoff that divides the residents into treatment and control groups. We exploit this natural experimental setting to implement a difference in discontinuity

⁶The western route is at the planning stage and is yet to be built.

⁷According to a news report published by Xinhua Net on January 8, 2022. Accessed via: http://www.gov.cn/xinwen/2022-01/08/content_5667043.htm

⁸The Chinese government stated that the water quality was above degree III for the eastern route and above degree II for the middle route. More details can be found in http://www.gov.cn/xinwen/2021-12/12/content_5660275.htm.

design by comparing those residents located within narrow areas to the middle route before and after the operation of SNWT.

3 Empirical strategy

To quantify the health benefits of water transfers to northern China, we estimate the causal relationship between SNWT and residents' health outcomes. The non-random access of residents to SNWT gives rise to a potential endogeneity problem that may hinder the identification of the causal impact of this project. We address this problem by using the natural experimental setting of the middle route of SNWT to implement a fuzzy regression discontinuity design (RDD). We estimate the following econometric model:

$$Health_{ijt} = \alpha SNWT_{ijt} + \beta f(Dist_j) + \rho_j + \iota_i + \tau_t + \epsilon_{ijt}$$

s.t. $-h \le Dist_j \le h$, (1)

where $Health_{ijt}$ is the health outcome (measured by the incidence of waterborne diseases or cancer) of resident *i* located in county *j* in year *t*, *SNWT* denotes whether a household has access to SNWT, and $f(Dist_j)$ is a polynomial function of the forcing variable, which is the shortest distance between the county administrative center and the middle route of SNWT. Note that we identified the counties with access to SNWT according to the official document, "Master Plan of the South to North Water Transfers Project". We control for the one-degree-latitude fixed effect ρ_j , the resident fixed effect ι_i , and the year fixed effect τ_t to capture any geographical factors, time-invariant resident characteristics, and year-specific shocks. ϵ_{ijt} denotes the error term, which is assumed to be independent and identically distributed (i.i.d).

Eq. (1) can be estimated using a non-parametric or parametric approach. According to Gelman and Imbens, 2019, the parametric estimation of RDD tends to generate sensitive results related to the order of polynomial function of the forcing variable and have some undesirable statistical properties. Therefore, we use the non-parametric local linear approach for the main analysis and implement parametric estimation for robustness checks. To determine the appropriate bandwidth for estimation, we adopt the method proposed by Calonico et al., 2014 to choose the optimal bandwidth h.

Another challenge in non-parametric estimation is accounting for the fixed effects. To address this problem, we follow the two-step approach of Lee and Lemieux, 2010. In the first step, we regress the resident health outcomes on the set of fixed effects and predict the residualized health outcomes. In the second step, using the predicted value, we re-estimate Eq. (1) using the traditional non-

parametric local linear approach. This two-step approach can improve the precision of estimation without causing bias.

$$Health_{ijt} = \rho_j + \iota_i + \tau_t + \epsilon_{ijt} \tag{2}$$

$$\begin{aligned} H\hat{ealth}_{ijt} &= \alpha SNWT_{ijt} + \beta f(Dist_j) + \epsilon_{ijt} \\ \text{s.t.} &- h \leq Dist_j \leq h. \end{aligned} \tag{3}$$

To explore the panel structure of the data, we estimate Eq. (3) separately for the pre- and post-2014 periods (i.e., the middle route of SNWT commenced its operations in 2014). To compare the results before and after the water transfers, we follow the literature and adopt the following difference-in-discontinuity (Dif-in-Disc) model to estimate the treatment effect of SNWT (Grembi et al., 2016; He et al., 2020):

$$Health_{ijt} = \alpha SNWT_{ijt} + \beta Dist_j + \gamma Post_t + \theta SNWT_{ijt} \times Post_t + \rho SNWT_{ijt} \times Dist_j + \delta Post_t \times Dist_j + \varphi SNWT_{ijt} \times Post_t \times Dist_j + \rho_j + \iota_i + \tau_t + \epsilon_{ijt} s.t. - h < Dist_i < h.$$

$$(4)$$

On the basis of the traditional RDD model, we introduce a dummy $Post_t$ to indicate if the time period is post 2014. The coefficient θ before the interaction term of *SNWT* and *Post* captures the treatment effect of SNWT as the change in discontinuity before and after 2014. In this way, we combine the cross-sectional variation of RDD and the temporal variation of the panel data to clearly identify the health benefits of SNWT.

4 Data

This section presents our multi-source datasets for empirically investigating the health benefits of SNWT. These datasets include the individual-level survey data from CFPS, station-level monitoring data on drinking water quality, county-level characteristics derived from the China County Statistical Yearbook, and county-level climatic data from the China Meteorological History Dataset.

4.1 CFPS

We collect our health data from the CFPS, which is a comprehensive and longitudinal social survey administrated by Peking University in China that covers about 16,000 households across the 162 administrative counties of 25 provinces or municipalities and autonomous regions. This dataset stores information about various aspects of Chinese families, such as their economic activities, education outcomes, family dynamics and relationships, migration, and health. With a coverage accounting for 94.5% of the total population of China, the CFPS has been widely used for studying the society and households of contemporary China (Cao et al., 2022; Deng et al., 2022; W. Huang and Liu, 2023).

With six waves (2010, 2012, 2014, 2016, and 2018) conducted thus far, the CFPS allows us to compile a panel dataset for empirically investigating the health outcomes of households before and after SNWT. We select 1,970 households in 34 counties of 4 provinces (Hebei, Henan, Shanxi and Shaanxi) where crossed the middle route of SNWT. We derive detailed resident-level information on the incidence of chronic diseases (e.g., infectious diseases, parasitic diseases, and cancers), and resident health behaviors, such as smoking, drinking, and doing exercise. We also collect household-level information, such as health expenses, access to tap water, and cost of water consumption. To control for household heterogeneity, we derive variables on households' socio-economic characteristics, such as income, consumption, family size, total asset, and proximity to hospitals.

4.2 Monitoring data on centralized drinking water sources

To explore the mechanism by which SNWT provides clean and safe drinking water, we compile station monitoring data on drinking water quality. China started monitoring the water quality of its centralized drinking water sources in 2012, but only until 2016 did the Ministry of Ecology and Environment (MEE) issue the measures for disclosing information on the national water quality monitoring of centralized drinking water sources, which ensured the full disclosure of comprehensive monitoring data at the local level. Since 2016, an increasing number of cities have regularly disclosed centralized drinking water quality data from monitoring sources managed at the national, provincial, and municipal city levels.

For each sample city, we manually collect monthly monitoring data of 97 city level centralized drinking water sources in 2016 and 2018 from reports on the water quality of centralized drinking

water sources published by local environmental authorities.⁹ These data provide information on the general grade of drinking water quality (I–V) and details about water hardness and water contaminants (i.e., sulfate, nitrate, phosphorus, and manganese). This comprehensive information allows us to accurately measure the effects of SNWT on water quality of major drinking water sources.

4.3 Other complementary data

One important condition for the internal validity of RDD is to ensure the continuity of other socioeconomic characteristics at the cutoff. To address the systematic differences between counties with and without access to SNWT, we collect comprehensive data on county characteristics, such as GDP, area, population, government expenditure, industrial production, and number of hospital beds, from the China County Statistical Yearbook. We also collect data on various climatic conditions, including near surface air temperature, sunshine duration, wind speed, air specific humidity, near surface air pressure, and precipitation, from the China Meteorological History Dataset (V3.0) published by the China Meteorological Administration. We initially interpolate station-level meteorological data to grids of $500 \text{ m} \times 500 \text{ m}$ using the Barnes interpolation method (Barnes, 1964) and then aggregate the grid data to the county level using GIS. The definition and data sources of all the variables used in this paper are presented in Table S1, and the summary statistics are presented in Table S2 in the appendices.

4.4 Balance check

In our sample, 12 counties have access to SNWT, and 22 counties have no access. To ensure that the treatment and control groups are comparable, we restrict our analysis to a smaller bandwidth and then perform a balance check for the number of socio-economic characteristics of the counties. The results of the balance check with the RD estimation are presented in Table S3 in the appendices.

After restricting the sample within 200 km of the middle route, most of the differences in the characteristics between the SNWT and non-SNWT areas become insignificant. We also change the bandwidth to 100 km or 300 km for another balance check, and the results are presented in Tables S3 and S5. After changing the bandwidth to 100 km or 300 km, we note some differences in several county and household characteristics, but the statistical significances are very weak and inconsistent. Therefore, with a smaller bandwidth, those counties with and without access to SNWT

⁹Comprehensive monitoring data on centralized drinking water sources became publicly available only in 2016.

are well balanced in most socio-economic characteristics and climatic conditions, thus allowing us to continue with the RD analysis.

5 Results

5.1 Effects of SNWT on individuals' health outcomes

We start by testing the impact of SNWT on residents' health outcomes as measured by the incidence of waterborne disease and cancer.¹⁰ Figure 1 presents the RD graphs before and after SNWT. Interestingly, the incidences of waterborne diseases and cancer are higher at the regional cutoff, but we observe no significant discontinuity before SNWT. This observation is consistent with findings from the literature on trans-boundary water pollution, which shows that water pollution is generally severe at administrative borders (Sigman, 2002; Duvivier and Xiong, 2013; Cai et al., 2016; Lipscomb and Mobarak, 2016). After the operation of the project, residents living to the east of the middle route (i.e., with access to SNWT) show significant improvement in terms of incidences of waterborne diseases and cancer, which demonstrate clear discontinuity at the cutoff.

We then estimate the treatment effect of SNWT to quantify its health benefits, and Table 1 reports the results. We present the non-parametric estimates of the RD model for the subsamples before and after the operation of SNWT. Afterward, we present the treatment effect estimated using the Dif-in-Disc model as in Grembi et al., 2016. Consistent with the figure, SNWT has no significant effects on individual health outcomes before its operation, but these effects become significant and quite robust after the operation of the project. The magnitudes of the treatment effect are -0.25 (exp(-0.282)-1) ~ -0.39 (exp(-0.501)-1) for waterborne diseases and -0.16 (exp(-0.172)-1) ~ -0.20 (exp(-0.221)-1) for cancer.¹¹ The point estimates suggest that SNWT reduced waterborne diseases and cancer by at least 25% and 16%, respectively, for residents living in areas close to the middle route.

We then investigate the dynamic effects of SNWT. We estimate the treatment effect of the project year by year, and the results are presented in Figure 2 and Table S6. As shown in the RD graph and the estimation results, the health effects of SNWT are insignificant prior to the operation of SNWT. Only in 2018 (i.e., 4 years after its operation) did the project start to reveal its health benefits.

¹⁰The waterborne diseases, including infectious, parasitic, and digestive system diseases, are defined according to the WHO. Cancer is defined as malignant and benign tumors in the CFPS.

¹¹To ensure that the estimates are within [-1,1], we rescale the point estimates using $e^x - 1$ as in He et al., 2020.

Combined with previous estimates of health benefits, our dynamic results provide strong evidence supporting the substantial yet lagged health benefits of SNWT to residents living in the region.

5.2 Robustness checks

We subject our results to a battery of robustness checks. First, to ensure the comparability of the treatment and control groups, we further implement the balance check with alternative kernel. The results are reported in Table S7. As noted, all of the checked characteristics remain comparable between the treated and control groups. This balance check also ensures the internal validity of the spatial RD estimation.

Second, we use an alternative bandwidth to check the RD estimation of health benefits. Instead of selecting the optimal bandwidth using the CCT method, we use the alternative bandwidths of 100, 200, and 300 km to re-estimate the RD model. The results are reported in Table S8. We observe greater statistical significance when we enlarge the bandwidth of the RD estimation because enlarging such bandwidth includes more observations for the analysis, thus increasing the statistical power for the estimation. The direction and magnitude of the estimated coefficients remain consistent with the main results.

Third, we estimate the RD model using a parametric approach and estimate the treatment effect using a classical difference-in-difference (DID) model. The results are reported in Table S9. The parametric RD estimations yield consistent results with smaller magnitude and significance that support our previous findings. The DID model also supports our conclusion on the health benefits of SNWT.

Fourth, we further check the health benefits of SNWT using other non-water relevant diseases such as diseases of respiratory system and endocrine system as placebo test. If the health benefits are due to other changing socio-economic factors related to SNWT, then the effect on non-relevant diseases should also be significant. The estimation results are reported in Table S10, and the RD results are insignificant with other diseases as the dependent variable.

5.3 Mechanisms

We then explore the mechanism underlying the health benefits of SNWT. First, we check the impact of SNWT on major economic factors that also affect individual health outcomes. For instance, the increasing income and consumption and changing health behaviors, such as smoking, drinking, and doing exercise, may impact residents' health outcome. To rule out these competing explanations, we estimate the Dif-in-Disc model using these economic factors as dependent variables, and the results are presented in Figure S2 and Table S11. Results show that none of these economic factors

are significantly affected by SNWT because the water transferred by the project is used primarily for resident consumption and secondly for industrial production and agricultural irrigation.

Second, to evaluate the important role of clean drinking water, we analyze the impact of SNWT on drinking water quality. We use different indicators of drinking water quality as dependent variables to check the change of drinking water quality after operation of the project.¹² The estimation results are reported in Table 2 and the corresponding RD graph is presented in Figure S3 in the appendices.

As shown in Table 2, SNWT has significantly reduced the amount of water contaminants that exceed the national safety standard for indicators of water hardness, sulfate, nitrate, and Total Phosphorus (TP). The general degree of drinking water quality is 0.13–0.15 (4.56%–5.26%) better in sources supplied by SNWT. In other words, the transferred water meets the high quality standards set by the government. The replacement of over-exploited groundwater with clean and safe water significantly improves the drinking water quality for residents. These estimation results provide direct evidence and sound explanation to the substantial health benefits of SNWT.

5.4 Heterogeneous effect of SNWT

The SNWT project aims to transfer water to the North China Plain via new water canals and tap water pipelines. Therefore, access to tap water should be a critical condition for households' consumption of south water. We use the question in the CFPS asking residents about their water sources for cooking to determine those residents who actually consume tap water for daily use. We then split our sample according to residents' access to tap water and test the heterogeneity of SNWT.

Another interesting heterogeneity to explore is by arid and humid areas, because the project can bring greater health benefits by relieving water scarcity in arid areas than in humid areas. To measure aridity, we use the climatic aridity index as proposed by the Food and Agriculture Organization (FAO):

$$Aridity_{it} = p_{it}/ET_{jt} \tag{5}$$

, where p_{jt} is the precipitation of county j in year t, and ET_{jt} is the potential evapotranspiration

¹²Due to the very small variation of quality in time for a given drinking water source, we are unable to control for the source-specific fixed effect in the model. However, we control for the fixed effects of one-degree latitude and year as in the previous analysis.

calculated using the FAO–Penman–Monteith method.¹³ Based on the aridity index, the areas are delineated into hyper-arid (*aridindex* < 0.03), arid (0.03 < aridindex < 0.20), semi-arid (0.20 < aridindex < 0.50), sub-humid (0.50 < aridindex < 0.75), and humid areas (*aridindex* > 0.75). Given that sub-humid areas also have arid conditions, we define "arid areas" to collectively represent the hyper-arid, arid, semi-arid, and sub-humid areas (*aridindex* < 0.75). We then split our sample of households into arid versus humid areas.

The results for the heterogeneity of SNWT are reported in Table S12. As expected, the health benefits of SNWT are significant only for residents living in arid areas and with access to tap water. These results show interesting heterogeneous effects of the SNWT and provide important policy implications for social planners. In arid areas where water resources are scarce, comprehensive planning and investment are needed to build local drinking water infrastructure and utilities, such as tap water pipe networks, in order to deliver water to vulnerable residents who truly need such resource and to ensure the success and inclusivity of SNWT.

5.5 Cost-benefit analysis

The final step is to conduct a cost-benefit analysis of SNWT. To calculate the monetary value of the health benefits of SNWT, we estimate the impact of this project on households' health expenses (logged) as in the previous analysis. The results are reported in Table 3 and the corresponding RD graph is presented in Figure S4 in the appendices. The RD estimates are insignificant before SNWT and become significant and quite robust after the project. The magnitudes of the treatment effect are -0.62 (exp(-0.978)-1) ~ -0.66 (exp(-1.077)-1), which suggest that SNWT has reduced households' health expenses by more than 60%.

Given that the average health expenses of households in our sample is 4,672 yuan (650 US dollars) per year, SNWT thus allows these households to save more than 2,800 yuan (390 US dollars) per year without taking into account the value of lives. According to the Chinese government, approximately 85 million people or 21.25 million households have directly benefited from the middle route of SNWT.¹⁴ Therefore, the total health expenses saved by the project amount to 59.50 billion yuan (8.29 billion US dollars) per year.

The financial costs of SNWT are also enormous. While conservative estimates from the Chinese government suggest that the construction of the middle route of SNWT cost more than 200 billion

¹³More details about the FAO–Penman–Monteith method can be found in the appendices.

¹⁴More details can be found in http://nsbd.mwr.gov.cn/zx/zxdt/202207/t20220722_1586569.html.

yuan (27 billion US dollars), previous studies have estimated a much higher cost of 580 billion yuan (79.4 billion US dollars) (Webber et al., 2017).¹⁵ Many other estimates fall within this scope. By contrasting the health benefits of SNWT to its construction costs, we easily find that the former exceeds the latter in 4–10 years. This simple calculation justifies the cost-effectiveness of SNWT and supports the economic rationale for the investment decision of the Chinese government.

However, this back-of-the-envelope calculation is preliminary because these estimates do not include the costs for resettlement of about 300,000 people and the costs of shutting down factories and farms for environmental protection in water source areas along the middle route. These enormous costs need to be further justified in future studies by conducting a more comprehensive evaluation of the value of lives and other environmental and socio-economic benefits of the project.

6 Conclusion

SNWT has diverted massive amounts of water from the Yangtze River to quench the thirst of residents in northern China. To evaluate its cost-effectiveness from the public health perspective, this paper estimates the causal impact of SNWT on residents' health outcomes using a differencein-discontinuity design. SNWT has reduced the incidence of waterborne diseases and cancer in the region by more than 25% and 16%, respectively, by significantly improving the quality of centralized drinking water sources for residents with access to the transferred water. The project also produces more pronounced health benefits for residents living in arid regions and have better access to tap water.

Based on these findings, we suggest that cost-benefit analyses of SNWT should take into account the substantial health benefits of the project. Water diversion projects typically build large-scale water infrastructures to optimize the allocation of water resources, but these infrastructures often entail complex environmental consequences and require a huge public investment. Therefore, a more thorough and rigorous evaluation of the potential benefits of these infrastructures are needed to support and direct policy-making. As demonstrated by the case of SNWT, given that residents in arid regions are very sensitive to safe and clean water, the water transfer project will produce significant health benefits for these residents that overweigh the costs. This case offers important lessons regarding the economic valuation of safe and clean water, which is vital for people's adaptation to the global water crisis.

¹⁵More details can be found in http://hb.ifeng.com/news/focus/detail_2014_11/16/3156564_0.shtml.

References

- Aggeborn, L., & Öhman, M. (2021). The effects of fluoride in drinking water. *Journal of Political Economy*, *129*(2), 465–491.
- Allen, R. G., Pereira, L. S., Raes, D., Smith, M., et al. (1998). Crop evapotranspiration-guidelines for computing crop water requirements-fao irrigation and drainage paper 56. *Fao, Rome*, 300(9), D05109.
- Alsan, M., & Goldin, C. (2019). Watersheds in child mortality: The role of effective water and sewerage infrastructure, 1880–1920. *Journal of Political Economy*, *127*(2), 586–638.
- Barnes, S. L. (1964). A technique for maximizing details in numerical weather map analysis. *Journal of Applied Meteorology and Climatology*, 3(4), 396–409.
- Barnett, J., Rogers, S., Webber, M., Finlayson, B., & Wang, M. (2015). Sustainability: Transfer project cannot meet china's water needs. *Nature*, *527*(7578), 295–297.
- Berkoff, J. (2003). China: The south–north water transfer project—is it justified? *Water Policy*, 5(1), 1–28.
- Bhalotra, S. R., Diaz-Cayeros, A., Miller, G., Miranda, A., & Venkataramani, A. S. (2021). Urban water disinfection and mortality decline in lower-income countries. *American Economic Journal: Economic Policy*, 13(4), 490–520.
- Blakeslee, D., Dar, A., Fishman, R., Malik, S., Pellegrina, H. S., & Bagavathinathan, K. S. (2023). Irrigation and the spatial pattern of local economic development in india. *Journal of Development Economics*, 161, 102997.
- Blakeslee, D., Fishman, R., & Srinivasan, V. (2020). Way down in the hole: Adaptation to long-term water loss in rural india. *American Economic Review*, *110*(1), 200–224.
- Cai, H., Chen, Y., & Gong, Q. (2016). Polluting thy neighbor: Unintended consequences of china s pollution reduction mandates. *Journal of Environmental Economics and Management*, 76, 86–104.
- Calonico, S., Cattaneo, M. D., & Titiunik, R. (2014). Robust nonparametric confidence intervals for regression-discontinuity designs. *Econometrica*, 82(6), 2295–2326.
- Cao, D., Zhou, Z., Liu, G., Shen, C., Ren, Y., Zhao, D., Zhao, Y., Deng, Q., & Zhai, X. (2022).
 Does social capital buffer or exacerbate mental health inequality? evidence from the china family panel study (cfps). *International Journal for Equity in Health*, 21(1), 1–17.
- Cole, M. A., Elliott, R. J., & Strobl, E. (2014). Climate change, hydro-dependency, and the african dam boom. *World Development*, *60*, 84–98.
- Crow-Miller, B. (2015). Discourses of deflection: The politics of framing china's south-north water transfer project. *Water Alternatives*, 8(2).

- Currell, M. J., Han, D., Chen, Z., & Cartwright, I. (2012). Sustainability of groundwater usage in northern china: Dependence on palaeowaters and effects on water quality, quantity and ecosystem health. *Hydrological Processes*, *26*(26), 4050–4066.
- De Faria, F. A., Davis, A., Severnini, E., & Jaramillo, P. (2017). The local socio-economic impacts of large hydropower plant development in a developing country. *Energy Economics*, 67, 533–544.
- Deng, Y., Liao, L., Yu, J., & Zhang, Y. (2022). Capital spillover, house prices, and consumer spending: Quasi-experimental evidence from house purchase restrictions. *The Review of Financial Studies*, 35(6), 3060–3099.
- Dillon, A., & Fishman, R. (2019). Dams: Effects of hydrological infrastructure on development. Annual Review of Resource Economics, 11, 125–148.
- Duflo, E., & Pande, R. (2007). Dams. The Quarterly Journal of Economics, 122(2), 601-646.
- Duvivier, C., & Xiong, H. (2013). Transboundary pollution in china: A study of polluting firms' location choices in hebei province. *Environment and Development Economics*, 18(4), 459– 483.
- Ebenstein, A. (2012). The consequences of industrialization: Evidence from water pollution and digestive cancers in china. *Review of Economics and Statistics*, 94(1), 186–201.
- Fan, M., & He, G. (2023). Clean water and infant health: Evidence from piped water provision in china. *Journal of the Association of Environmental and Resource Economists*, 10(1), 159–193.
- Faúndez, M., Alcayaga, H., Walters, J., Pizarro, A., & Soto-Alvarez, M. (2023). Sustainability of water transfer projects: A systematic review. *Science of The Total Environment*, 860, 160500.
- Galiani, S., Gertler, P., & Schargrodsky, E. (2005). Water for life: The impact of the privatization of water services on child mortality. *Journal of Political Economy*, *113*(1), 83–120.
- Gatti, N., Baylis, K., & Crost, B. (2021). Can irrigation infrastructure mitigate the effect of rainfall shocks on conflict? evidence from indonesia. *American Journal of Agricultural Economics*, *103*(1), 211–231.
- Gelman, A., & Imbens, G. (2019). Why high-order polynomials should not be used in regression discontinuity designs. *Journal of Business & Economic Statistics*, *37*(3), 447–456.
- Gleeson, T., VanderSteen, J., Sophocleous, M. A., Taniguchi, M., Alley, W. M., Allen, D. M., & Zhou, Y. (2010). Groundwater sustainability strategies. *Nature Geoscience*, *3*(6), 378–379.
- Greenstone, M., & Hanna, R. (2014). Environmental regulations, air and water pollution, and infant mortality in india. *American Economic Review*, *104*(10), 3038–3072.
- Grembi, V., Nannicini, T., & Troiano, U. (2016). Do fiscal rules matter? *American Economic Journal: Applied Economics*, 1–30.

- Han, D., Currell, M. J., & Cao, G. (2016). Deep challenges for china's war on water pollution. *Environmental Pollution*, 218, 1222–1233.
- Hansen, Z. K., Libecap, G. D., & Lowe, S. E. (2009). Climate variability and water infrastructure: Historical experience in the western united states (tech. rep.). National Bureau of Economic Research.
- He, G., Wang, S., & Zhang, B. (2020). Watering down environmental regulation in china. *The Quarterly Journal of Economics*, *135*(4), 2135–2185.
- Huang, Q., Wang, J., Rozelle, S., Polasky, S., & Liu, Y. (2013). The effects of well management and the nature of the aquifer on groundwater resources. *American Journal of Agricultural Economics*, 95(1), 94–116.
- Huang, W., & Liu, H. (2023). Early childhood exposure to health insurance and adolescent outcomes: Evidence from rural china. *Journal of Development Economics*, *160*, 102925.
- Jeuland, M. (2020). The economics of dams. Oxford Review of Economic Policy, 36(1), 45-68.
- Lee, D. S., & Lemieux, T. (2010). Regression discontinuity designs in economics. *Journal of Economic Literature*, 48(2), 281–355.
- Lipscomb, M., & Mobarak, A. M. (2016). Decentralization and pollution spillovers: Evidence from the re-drawing of county borders in brazil. *The Review of Economic Studies*, 84(1), 464–502.
- Liu, J., & Yang, W. (2012). Water sustainability for china and beyond. *Science*, 337(6095), 649–650.
- Marcus, M. (2022). Testing the water: Drinking water quality, public notification, and child outcomes. *Review of Economics and Statistics*, *104*(6), 1289–1303.
- Mary, S. (2022). Dams mitigate the effect of rainfall shocks on hindus-muslims riots. *World Development*, 150, 105731.
- Mettetal, E. (2019). Irrigation dams, water and infant mortality: Evidence from south africa. *Journal* of Development Economics, 138, 17–40.
- Monteith, J. L. (1965). Evaporation and environment. *Symposia of the Society for Experimental Biology*, *19*, 205–234.
- Moore, S. M. (2014). Modernisation, authoritarianism, and the environment: The politics of china's south–north water transfer project. *Environmental Politics*, 23(6), 947–964.
- Olmstead, S. M., & Sigman, H. (2015). Damming the commons: An empirical analysis of international cooperation and conflict in dam location. *Journal of the Association of Environmental and Resource Economists*, 2(4), 497–526.
- Penman, H. L. (1948). Natural evaporation from open water, bare soil and grass. Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences, 193(1032), 120–145.

- Pitt, M. M., Rosenzweig, M. R., & Hassan, M. N. (2021). Identifying the costs of a public health success: Arsenic well water contamination and productivity in bangladesh. *The Review of Economic Studies*, 88(5), 2479–2526.
- Pohlner, H. (2016). Institutional change and the political economy of water megaprojects: China's south-north water transfer. *Global Environmental Change*, *38*, 205–216.
- Purvis, L., & Dinar, A. (2020). Are intra-and inter-basin water transfers a sustainable policy intervention for addressing water scarcity? *Water Security*, *9*, 100058.
- Sarsons, H. (2015). Rainfall and conflict: A cautionary tale. *Journal of Development Economics*, *115*, 62–72.
- Sayre, S. S., & Taraz, V. (2019). Groundwater depletion in india: Social losses from costly well deepening. *Journal of Environmental Economics and Management*, 93, 85–100.
- Sekhri, S. (2014). Wells, water, and welfare: The impact of access to groundwater on rural poverty and conflict. *American Economic Journal: Applied Economics*, *6*(3), 76–102.
- Severnini, E. (2023). The power of hydroelectric dams: Historical evidence from the united states over the twentieth century. *The Economic Journal*, *133*(649), 420–459.
- Sheng, J., Tang, W., & Webber, M. (2020). Can interbasin water transfer affect water consumption and pollution? lessons from china's south–north water transfer project. *Environmental Policy* and Governance, 30(6), 345–358.
- Sheng, J., Webber, M., & Han, X. (2018). Governmentality within china's south-north water transfer project: Tournaments, markets and water pollution. *Journal of Environmental Policy & Planning*, 20(4), 533–549.
- Sigman, H. (2002). International spillovers and water quality in rivers: Do countries free ride? *American Economic Review*, 92(4), 1152–1159.
- Strobl, E., & Strobl, R. O. (2011). The distributional impact of large dams: Evidence from cropland productivity in africa. *Journal of Development Economics*, *96*(2), 432–450.
- Tao, T., & Xin, K. (2014). Public health: A sustainable plan for china's drinking water. *Nature*, *511*(7511), 527–528.
- Uhlenbrook, S., Yu, W., Schmitter, P., & Smith, D. M. (2022). Optimising the water we eat—rethinking policy to enhance productive and sustainable use of water in agri-food systems across scales. *The Lancet Planetary Health*, 6(1), e59–e65.
- Webber, M., Crow-Miller, B., & Rogers, S. (2017). The south–north water transfer project: Remaking the geography of china. *Regional Studies*, *51*(3), 370–382.
- Yang, G., & Zhuang, D. (2014). Atlas of the huai river basin water environment: Digestive cancer mortality. Springer Science & Business Media.
- Zhang, J. (2012). The impact of water quality on health: Evidence from the drinking water infrastructure program in rural china. *Journal of Health Economics*, *31*(1), 122–134.

Zhang, J., & Xu, L. C. (2016). The long-run effects of treated water on education: The rural drinking water program in china. *Journal of Development Economics*, *122*, 1–15.





Panel A: Before SNWT

Note: This figure presents the RD graph of the incidences of waterborne diseases and cancer for residents live next to the middle route of SNWT. Panel A is the graph before operation of SNWT and Panel B is the graph after operation of SNWT. The red vertical line indicates the middle route and the hollow points are the bin average of incidences of waterborne diseases and cancer.





Note: This figure presents the non-parametric RD estimates of the impact of SNWT on incidences of waterborne diseases and cancer over year. The red vertical dash line indicates the year of SNWT operation. The green vertical line is 95% confidence interval.

	(1)	(2) Waterborne	(3) e	(4)	(5) Cancer	(6)
Dep. Var.	Before	After	Dif-in-Disc	Before	After	Dif-in-Disc
Conventional	0.0070	-0.0369*	-0.2822**	0.0038*	-0.0111**	-0.1718***
	(0.0053)	(0.0198)	(0.1105)	(0.0022)	(0.0049)	(0.0582)
Bias-corrected	0.0087	-0.0474**	-0.5008***	0.0035	-0.0115**	-0.2207***
	(0.0053)	(0.0198)	(0.1105)	(0.0022)	(0.0049)	(0.0582)
Robust	0.0087	-0.0474*	-0.5008***	0.0035	-0.0115**	-0.2207**
	(0.0065)	(0.0248)	(0.1691)	(0.0026)	(0.0059)	(0.0868)
Obs.	18,599	11,811	21,955	18,592	11,830	21,967
Bandwidth	157.95	189.91	182.18	179.30	200.88	179.49
Year FE	Y	Y	Y	Y	Y	Y
Latitude FE	Y	Y	Y	Y	Y	Y
Individual FE	Y	Y	Y	Y	Y	Y

Table 1. The impact of SNWT on individual health outcomes

Note: This table reports the non-parametric RD estimates of the impact of SNWT on the incidence of waterborne diseases and cancer for individual located next to the SNWT middle line. Column 1 and 4 report the discontinuity before operation of the SNWT middle line. Column 2 and 5 report the discontinuity after operation of the SNWT middle line. Column 3 and 6 estimate the treatment effect of the SNWT using a Difference-in-Discontinuity model as in Grembi (2016). For all models, the bandwidth is selected with method by Calonico, Cattaneo and Titiunik (2014), and we report the results with three different procedures, i.e., conventional RD estimates with a conventional variance estimator; bias-corrected RD estimates with a conventional variance estimator; and bias-corrected RD estimates with a robust variance estimator. The default kernel function is triangular. We also control for the year fixed effects, 1-degree latitude fixed effects, and individual fixed effects using the two step methods suggested by Lee and Lemieux (2010). Robust standard errors are in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

	(1)	(2)	(3)	(4)	(5)	(6)
Dep. Var.	Grade	Hardness	Sulfate	Nitrate	TP	Manganese
Conventional	-0.1474***	-0.0194***	-0.0254***	-0.0148*	-0.0003***	-0.0025
	(0.0258)	(0.0029)	(0.0024)	(0.0088)	(0.0001)	(0.0052)
Bias-corrected	-0.1335***	-0.0206***	-0.0262***	-0.0173**	-0.0004***	-0.0020
	(0.0258)	(0.0029)	(0.0024)	(0.0088)	(0.0001)	(0.0052)
Robust	-0.1335***	-0.0206***	-0.0262***	-0.0173*	-0.0004***	-0.0020
	(0.0303)	(0.0030)	(0.0024)	(0.0103)	(0.0001)	(0.0056)
Obs.	1,901	1,991	1,991	1,991	1,991	1,988
Bandwidth	131.20	136.11	111.27	182.29	56.989	139.92
Year FE	Y	Y	Y	Y	Y	Y
Latitude FE	Y	Y	Y	Y	Y	Y

Table 2. The impact of SNWT on drinking water quality

Note: This table reports the non-parametric RD estimates of the impact of SNWT on the grade of water quality and the contaminants above the national standards. See Table S1 for variable definition. For all models, the bandwidth is selected with method by Calonico, Cattaneo and Titiunik(2014), and we report the results with three different procedures, i.e., conventional RD estimates with a conventional variance estimator; bias-corrected RD estimates with a conventional variance estimator. The default kernel function is triangular. We also control for the month fixed effects and 1-degree latitude fixed effects using the two step methods suggested by Lee and Lemieux (2010). Robust standard errors are in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

	(1)	(2)	(3)
		ln(Health_expense)	
Dep. Var.	Before	After	Dif-in-Disc
Conventional	0.1809	-0.4203*	-0.9778**
	(0.1414)	(0.2318)	(0.4774)
Bias-corrected	0.1434	-0.5342**	-1.0774**
	(0.1414)	(0.2318)	(0.4774)
Robust	0.1434	-0.5342*	-1.0774*
	(0.1914)	(0.2971)	(0.6411)
Obs.	6,456	4,517	7,879
Bandwidth	49.87	54.11	180.75
Year FE	Y	Y	Y
Latitude FE	Y	Y	Y
Household FE	Y	Y	Y

Table 3. The impact of SNWT on households' health expense

Note: This table reports the non-parametric RD estimates of the impact of SNWT on logged health expenses for household located next to the SNWT middle line. Column 1 reports the discontinuity before operation of the SNWT middle line. Column 2 reports the discontinuity after operation of the SNWT middle line. Column 3 estimates the treatment effect of the SNWT using a Difference-in-Discontinuity model as in Grembi (2016). For all models, the bandwidth is selected with method by Calonico, Cattaneo and Titiunik (2014), and we report the results with three different procedures, i.e., conventional RD estimates with a conventional variance estimator; bias-corrected RD estimates with a conventional variance estimator; and bias-corrected RD estimates with a robust variance estimator. The default kernel function is triangular. We also control for the year fixed effects, 1-degree latitude fixed effects, and individual fixed effects using the two step methods suggested by Lee and Lemieux (2010). Robust standard errors are in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

A Supplementary figures and tables



Figure S1. Water transfer routes of the SNWT project

Note: The map shows two water routes of the SNWT project, i.e., the middle route and the eastern route. The middle route was constructed to realize artesian water transfer along the Taihang mountains. The treament and control counties are selected based on the official document, "Master Plan of the South to North Water Transfers Project".





Note: This figure presents the RD graph of various economic factors for residents live next to the middle route of SNWT. Panel A is the graph before operation of SNWT and Panel B is the graph after operation of SNWT. The red vertical line indicates the middle route and the hollow points are the bin average of the economic factors.

-400 -200 0 200 400 Distance to the SNWT mid-line

-400 -200 0 200 400 Distance to the SNWT mid-line

-400 -200 0 200 400 Distance to the SNWT mid-line

-400 -200 0 200 400 Distance to the SNWT mid-line

-400 -200 0 200 400 Distance to the SNWT mid-line

-400 -200 0 200 400 Distance to the SNWT mid-line

-400 -200 0 200 400 Distance to the SNWT mid-line

6



Figure S3. Quality of centralized drinking water sources after the SNWT

Note: This figure presents the RD graph of water quality for centralized drinking water sources located next to the middle route of SNWT. The red vertical line indicates the middle route and the hollow points are the bin average of water quality indicators.



Figure S4. The impact of SNWT on household health expense

Note: This figure presents the RD graph of the health expenses (logged) for residents live next to the middle route of SNWT. Panel A is the graph before operation of SNWT and Panel B is the graph after operation of SNWT. The red vertical line indicates the middle route and the hollow points are the bin average of health expenses.

Var	Definition	Unit	Data source
	Panel A. County level	ariables	
	The shortest geographical distance from		
Distance	the county administrative center to the	km	Calculated by authors
	middle route of the SNWT.		
dub	The Gross Domestic Production of the	unillinu anu an	China statistical vaarbook (county laval)
100	county.	unnfullun	Cillia statistical feat over (county iever)
Area	The administration area of the county	km^2	China statistical yearbook (county level)
Pop	The population of the county	10000 people	China statistical yearbook (county level)
Riccal Evn	General public budget expenditure of the		China statistical vasibook (connew laval)
riscal_Exp	county government.	10000 yuan	CIIIIIA SIALISLICAL YEALDOUR (COUTILY IEVEL)
Industial_output	Total industrial output.	millionyuan	China statistical yearbook (county level)
Hospital_bed	Number of beds in hospital of the county.	N/A	China statistical yearbook (county level)
Temp	Annual average near surface (2m) air temperature of the county.	K	China Meteorological Administration
	famos em to emandante		
Suns_h	Annual average of sun shine duration of the county.	hour	China Meteorological Administration
Wind_speed	Annual average of wind speed of the county.	ms^{-1}	China Meteorological Administration
Humiditv	Annual average of air specific humidity of	keke-1	China Meteorological Administration
Comments of the second s	the county.	00	
Dressing	Annual average near surface (2m) air	P_{A}	China Meteorological Administration
	pressure of the county.	5	
Dracinitation	Annual average precipitation of	111 IVI	China Meteorological Administration
F I cc i pitation	the county.	11111	CIIIIa Microlological Auminima anun

Table S1. Definition and data sources of variables

	Panel B. Household l	vel variables		
II.a.14b arris	Household's annual expenditure on		Searce	
ncautexp	medical care.	унап	UFF3	
Food_exp	Household's annual expenditure on food.	yuan	CFPS	
Dress_exp	Household's annual expenditure on dress.	yuan	CFPS	
	Indicator of household's access to tap			
Tap_water	water. With", 1" for have access and "0"	N/A	CFPS	
	for no access.			
Income	Annual net income of the household.	yuan	CFPS	
Wage	Annual wage income of the household.	yuan	CFPS	
Family_size	Family size of the household.	N/A	CFPS	
Total_asset	Annual net asset of the household	yuan	CFPS	
Prox_hospital	Shortest distance to the nearest hospital	km	CFPS	
Water_fare	Annual fare for irrigation.	yuan	CFPS	
Irrig_fare	Annual fare for water consumption.	yuan	CFPS	
	Panel C. Individual le	vel variables		
	Indicator of household's chronical illness			
	in the past six months. "1" for any of the			
Watarharna dicanca	following diseases: infectious diseases;	VIV		
	parasitic disease; malignant and benign	W M		
	tumors; digestive system diseases.			
	"0" for none.			
	Indicator of household's chronical illness			
Cancer	in the past six months. "1" for malignant	N/A	CFPS	
	and benign tumors. "0" for none.			

	Indicator of household's chronical illness in the		
	past six months. "1" for any diseases except for the		
Other_diseases	following diseases: infectious diseases; parasitic	N/A	CFPS
	disease; malignant and benign tumors; digestive		
	system diseases. "0" for none.		
Cuncline	Indicator of smoke in the past month.	NIA	Steps
SIIINUIIIC	"1" for yes, "0" for no.	NIA	CFF3
	Indicator of drinking alcohol for more		
Alcoholic	than three times in the past month.	N/A	CFPS
	"1" for yes, "1" for no.		
Exercise	Frequency of doing sport in the past month.	N/A	CFPS
	Panel D. Water monitoring station	ı level variable	s
C. mode	Grade of drinking water quality, I is the best	NIA	Report on Water Quality of Centralized
Olauc	and V is the worst.	W M	Drinking Water Source
Hardneee	Drinking water hordness evoeds multinles		Report on Water Quality of Centralized
1141 011055	DIMINING WARD MANNINGSS CACCOLS IMMULTICS.		Drinking Water Source
Cultara	Sulfata avoaade multinlae		Report on Water Quality of Centralized
Эшпасс			Drinking Water Source
Nitrata	Nitrata avoaade multinlae		Report on Water Quality of Centralized
INILIAIC	minary varyed miningers		Drinking Water Source
đT	Total nhocnhorns avraads multinlas		Report on Water Quality of Centralized
11	rotat priception as exercas manufaces		Drinking Water Source
Mananaca	Mananaca avraade multinlae		Report on Water Quality of Centralized
IVIALIZATIVOS	Mailgancee caveras manipues		Drinking Water Source
Note: This table presents	the definition and data sources of key variables used in the paper.		

	~ ~ ~				
	Obs	Mean	Std.Dev.	Min	Max
]	Panel A. County	level variables		
Distance	170	126.04	83.70	15.93	324.26
GDP	122	15,124.77	10,395.33	2,260.00	50,962.86
Area	134	1,327.35	695.90	437.00	3,510.00
Рор	114	67.74	40.84	16.00	181.00
Fiscal_Exp	126	259,008.70	166,358.80	65,313.00	1,017,018.00
Industial_output	54	1,644,170.00	1,542,577.00	22,342.00	6,972,897.00
Hospital_bed	57	1,705.05	1,107.28	368.00	5,486.00
Temp	170	13.31	2.53	6.44	16.51
Suns_h	170	5.77	0.87	3.92	7.65
Wind_speed	170	2.10	0.41	1.13	3.41
Humidity	170	61.68	5.67	50.57	76.11
Pressure	170	972.28	40.18	887.08	1,015.31
Precipitation	170	631.72	183.14	366.18	1,235.23
	Pa	nel B. Househo	ld level variable	es	
Health_exp	15,318	4,672.17	13,619.36	0.00	500,000.00
Foodexp	15,318	11,453.95	11,351.15	0.00	144,000.00
Dressexp	15,318	1,995.90	3,135.86	0.00	120,000.00
Tap_water	15,196	0.60	0.49	0.00	1.00
Income	14,564	37,110.53	40,166.74	0.00	1,100,000.00
Wage	15,286	27,517.08	36,916.79	0.00	1,203,800.00
Family_size	15,318	3.95	1.87	0.00	15.00
Total_asset	11,779	266,070.10	538,852.70	-1959200.00	30600000.00
Prox_hospital	6,121	360.24	1,273.47	0.00	60,000.00
Water_fare	11,870	31.23	56.80	0.00	2,000.00
Irrig_fare	5,081	536.39	1,318.12	0.00	40,000.00
	Pa	anel C. Individu	al level variable	S	
Waterborne_disease	37,384	0.03	0.17	0.00	1.00
Cancer	37,389	0.00	0.07	0.00	1.00
Other_diseases	39,995	0.35	0.48	0.00	1.00
Smoke	37,915	0.28	0.45	0.00	1.00
Alcoholic	37,449	0.12	0.33	0.00	1.00
Exercise	37,278	2.48	2.95	0.00	50.00

Table S2. Summary statistics

	Panel D. W	Vater monitori	ng station level	variables	
Grade	1,901	2.8532	0.4769	1.0000	5.0000
Hardness	1,991	0.0038	0.0319	0.0000	0.5400
Sulfate	1,991	0.0048	0.0439	0.0000	0.5300
Nitrate	1,991	0.0029	0.0494	0.0000	1.1600
ТР	1,991	0.0001	0.0054	0.0000	0.2400
Manganese	1,988	0.0036	0.0448	0.0000	0.7000

Note: This table presents the summary statistics of key variables used in the paper.

	(1)	(2)	(3)	(4)	(5)	(6)
		Panel A. C	County socio-	economic chara	cteristics	
Dep. Var.	GDP	Area	Pop	Fiscal_Exp	Industial_output	Hospital_bed
RD	-0.0244	-0.0026*	-0.0100	-0.0281	-0.1589	-0.0035
	(0.0740)	(0.0014)	(0.0089)	(0.0735)	(0.1339)	(0.0722)
Obs.	69	81	61	75	71	37
		Panel B. C	County meteo	rological charac	eteristics	
Dep. Var.	Temp	Suns_h	Wind	Humidity	Pressure	Precipitation
RD	0.0045	0.0154	0.0119	-0.0075	0.0003	0.0484
	(0.0101)	(0.0624)	(0.0415)	(0.0153)	(0.0009)	(0.1035)
Obs.	102	102	102	102	102	102
		Pan	el C. Househo	old characteristi	cs	
Dep. Var.	Income	Family_size	Total_asset	Prox_hospital	Water_fare	Irriga_fare
RD	-0.0559	0.0159	-0.0840	0.0032	-0.1109	0.2454
	(0.0763)	(0.0116)	(0.0527)	(0.0593)	(0.0732)	(0.1766)
Obs.	6,929	7,407	6,897	4,488	4,583	1,133

Table S3. Balance check of characteristics in Non-SNWT and SNWT areas

Note: This table reports the balance check of socio-economic and meteorological characteristics at county and household level for the non-SNWT and SNWT areas before operation of the SNWT middle line (2010-2014). See Table S1 for variable definition. It performs the non-parametric RD estimation with 200 km as the default bandwidth. The default kernel function is triangular. Robust standard errors are in parentheses. *** p < 0.01, ** p < 0.05, * p < 0.1.

(1)	(2)	(3)	(4)	(5)	(6)
	Panel A. C	County socio-	economic chara	cteristics	
GDP	Area	Pop	Fiscal_Exp	Industial_output	Hospital_bed
0.0662	0.0001	0.0181	0.0691	-0.0416	0.0148
(0.1064)	(0.0009)	(0.0158)	(0.1090)	(0.1683)	(0.1209)
69	81	61	75	71	37
	Panel B. C	County meteo	rological charac	eteristics	
Temp	Suns_h	Wind	Humidity	Pressure	Precipitation
-0.0002	0.0280	-0.0463	-0.0061	0.0011	0.0675
(0.0127)	(0.1182)	(0.0702)	(0.0254)	(0.0018)	(0.1227)
102	102	102	102	102	102
	Pan	el C. Househo	old characteristi	cs	
Income	Family_size	Total_asset	Prox_hospital	Water_fare	Irrig_fare
-0.0505	0.0095	-0.1320*	0.0054	0.2030**	0.1467
(0.0997)	(0.0141)	(0.0685)	(0.0747)	(0.0914)	(0.1563)
6,929	7,407	6,897	4,488	4,583	1,133
	(1) GDP 0.0662 (0.1064) 69 Temp -0.0002 (0.0127) 102 Income -0.0505 (0.0997) 6,929	(1) (2) Panel A. C GDP Area 0.0662 0.0001 (0.1064) (0.0009) 69 81 Panel B. C Temp Suns_h -0.0002 0.0280 (0.0127) (0.1182) 102 102 Panel B. C Panel B. C .0.0505 0.0280 (0.0127) (0.1182) 102 102 .00505 0.0095 (0.0997) (0.0141) 6,929 7,407	(1) (2) (3) Panel A. Curry socio-o Panel A. Curry socio-o GDP Area Pop 0.0662 0.0001 0.0181 (0.1064) (0.0009) (0.0158) 69 81 61 Panel B. Curry meteo Panel B. Curry meteo Temp Suns_h Wind -0.0002 0.0280 -0.0463 (0.0127) (0.1182) (0.0702) 102 102 102 Income Family_size Total_asset -0.0505 0.0095 -0.1320* (0.0997) (0.0141) (0.0685) 6,929 7,407 6,897	(1) (2) (3) (4) Panel A. County socio-consic charae GDP Area Pop Fiscal_Exp 0.0662 0.0001 0.0181 0.0691 (0.1064) (0.0009) (0.0158) (0.1090) 69 81 61 75 Panel B. County meteorological charae Temp Suns_h Wind Humidity -0.0002 0.0280 -0.0463 -0.0061 (0.0127) (0.1182) (0.0702) (0.0254) 102 102 102 102 Panel C. Househol characteristi Income Family_size Total_asset Prox_hospital -0.0505 0.0095 -0.1320* 0.0054 (0.0997) (0.0141) (0.0685) (0.0747) 6,929 7,407 6,897 4,488	(1)(2)(3)(4)(5)Panel A. County socio-economic characteristicsGDPAreaPopFiscal ExpIndustial output0.06620.00010.01810.0691-0.0416(0.1064)(0.0009)(0.0158)(0.1090)(0.1683)6981617571Panel B. County meteorolical characteristicsTempSuns_hWindHumidityPressure-0.00020.0280-0.0463-0.00610.0011(0.0127)(0.1182)(0.0702)(0.0254)(0.0018)102102102102102Panel E. C. Household characteristicsFamily_sizeTotal assetProx_hospitalMater_fare-0.05050.0095-0.1320*0.00540.2030**(0.0997)(0.0141)(0.0685)(0.0747)(0.0914)6,9297,4076,8974,4884,583

Table S4. Balance check of characteristics in Non-SNWT and SNWT areas (100km)

Note: This table reports the balance check of socio-economic and meteorological characteristics at county and household level for the non-SNWT and SNWT areas before operation of the SNWT middle line (2010-2014). See Table S1 for variable definition. It performs the non-parametric RD estimation with 100 km as the default bandwidth. The default kernel function is triangular. Robust standard errors are in parentheses. *** p < 0.01, ** p < 0.05, * p < 0.1.

	(1)	(2)	(3)	(4)	(5)	(6)
		Panel A. C	County socio-	economic chara	cteristics	
Dep. Var.	GDP	Area	Pop	Fiscal_Exp	Industial_output	Hospital_bed
RD	-0.0087	-0.00414**	-0.0119	-0.0497	-0.1452	-0.0605
	(0.0722)	(0.0019)	(0.0098)	(0.0725)	(0.1393)	(0.0627)
Obs.	69	81	61	75	71	37
		Panel B. C	County meteo	rological charac	teristics	
Dep. Var.	Temp	Suns_h	Wind	Humidity	Pressure	Precipitation
RD	0.0054	0.0098	0.0164	-0.0053	0.0002	0.0424
	(0.0082)	(0.0457)	(0.0353)	(0.0120)	(0.0007)	(0.0857)
Obs.	102	102	102	102	102	102
		Pan	el C. Househo	old characteristi	cs	
Dep. Var.	Income	Family_size	Total_asset	Prox_hospital	Water_fare	Irrig_fare
RD	-0.0451	0.0073	-0.0463	0.0016	-0.0503	0.0778
	(0.0570)	(0.0090)	(0.0412)	(0.0440)	(0.0583)	(0.1668)
Obs.	6,929	7,407	6,897	4,488	4,583	1,133

Table S5. Balance check of characteristics in Non-SNWT and SNWT areas (300km)

Note: This table reports the balance check of socio-economic and meteorological characteristics at county and household level for the non-SNWT and SNWT areas before operation of the SNWT middle line (2010-2014). See Table S1 for variable definition. It performs the non-parametric RD estimation with 300 km as the default bandwidth. The default kernel function is triangular. Robust standard errors are in parentheses. *** p < 0.01, ** p < 0.05, * p < 0.1.

Dep. Var.	(1)	(2)	(3) Waterborr	le (4)	(2)	(9)	(7)	(8) Cancer	(6)	(10)
Year	2010	2012	2014	2016	2018	2010	2012	2014	2016	2018
Conventional	-0.0104 (0.0243)	0.0051 (0.022)	0.0081 (0.0104)	0.0157* (0.0091)	-0.1321** (0.0537)	0.0097 (0.0120)	0.0014 (0.0029)	0.0134*** (0.0038)	-0.0043 (0.0046)	-0.0827*** (0.0227)
Bias-corrected	-0.0113	0.0026	0.0114	0.0146	-0.1811***	0.0123	0.0018	0.0148***	-0.0028	-0.1118^{***}
Robust	(0.0279)	(0.0279) (0.0279)	(0.0129) (0.0129)	0.0146 (0.0111)	(0.0679)	0.0123 0.0123 (0.0145)	0.0018 0.0018 (0.0034)	(0.0044) (0.0044)	-0.0028 (0.0055)	-0.1118^{***} (0.0291)
Obs.	5,638	6,017	6,944	6,417	5,394	5,637	5,994	6,961	6,437	5,393
Bandwidth Year FE	70.09 Y	62.55 Y	136.33 Y	122.18 Y	137.25 Y	63.73 Y	9.60 Y	134.47 Y	131.39 Y	138.51 Y
Latitude FE Individual FE	Ч	YY	YY	YY	ΥΥ	YY	YY	YY	YY	ХХ

Table S6. The dynamic health effects of SNWT

models, the bandwidth is selected with method by Calonico, Cattaneo and Titiunik(2014), and we report the results with three different procedures, i.e., conventional RD estimates with a conventional variance estimator; bias-corrected RD estimates with a conventional variance estimator; and bias-corrected RD estimates with a robust variance estimator. The default kernel function is triangular. We also control for the year fixed effects, 1-degree latitude fixed effects, and individual fixed effects using the two step methods suggested by Lee and Lemieux (2010). Robust standard errors are in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

	(1)	(2)	(3)	(4)	(5)	(6)
		Panel A. Co	ounty socio-ec	onomic characte	ristics	
Dep. Var.	GDP	Area	Рор	Fiscal_Exp	Indus_output	Hospital_bed
RD	-0.0257	-0.0009	-0.0127	-0.0666	-0.1581	-0.0401
	(0.0758)	(0.0011)	(0.0096)	(0.0813)	(0.1518)	(0.0688)
Obs.	69	81	61	75	71	37
		Panel B. Co	ounty meteoro	ological characte	ristics	
Dep. Var.	Temp	Suns_h	Wind	Humidity	Pressure	Precipitation
RD	0.0054	0.0095	0.0194	-0.0067	0.0002	0.0497
	(0.0105)	(0.0524)	(0.0417)	(0.0141)	(0.0008)	(0.1023)
Obs.	102	102	102	102	102	102
		Panel	C. Househol	d characteristics		
Dep. Var.	Income	Family_size	Total_asset	Prox_hospital	Water_fare	Irriga_fare
RD	-0.0063	0.0151	-0.0419	0.0013	-0.0703	-0.1681
	(0.0619)	(0.0104)	(0.0476)	(0.0490)	(0.0675)	(0.1929)
Obs.	6,929	7,407	6,897	4,488	4,583	1,133

Table S7. Balance check of characteristics in Non-SNWT and SNWT areas (Uniform kernel)

Note: This table reports the balance check of socio-economic and meteorological characteristics at county and household level for the non-SNWT and SNWT areas before operation of the SNWT middle line (2010-2014). See Table S1 for variable definition. It performs the non-parametric RD estimation with 200 km as the default bandwidth. The kernel function is uniform. Robust standard errors are in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

	(1)	(2)	(3)	(4)	(5)	(6)
Dep. Var.		Waterborne	2		Cancer	
Conventional	-0.5161*	-0.2738**	-0.1656**	-0.3934**	-0.1691***	-0.1732***
	(0.3012)	(0.1064)	(0.0794)	(0.1667)	(0.0561)	(0.0420)
Bias-corrected	0.4643	-0.2633**	-0.5429***	0.1750	-0.2102***	-0.2079***
	(0.3012)	(0.1064)	(0.0794)	(0.1667)	(0.0561)	(0.0420)
Robust	0.4643	-0.2633	-0.5429***	0.1750	-0.2102*	-0.2079***
	(0.7044)	(0.2221)	(0.1453)	(0.3553)	(0.1132)	(0.0764)
Obs.	12,295	23,776	28,837	12,297	23,788	28,849
Bandwidth	100	200	300	100	200	300
Year FE	Y	Y	Y	Y	Y	Y
Latitude FE	Y	Y	Y	Y	Y	Y
Individual FE	Y	Y	Y	Y	Y	Y

Table S8. Robustness check with alternative bandwidth

Note: This table reports the non-parametric RD estimates of the impact of SNWT on the incidence of waterborne diseases and cancer for individual located next to the SNWT middle line using alternative bandwidths, i.e., 100km, 200km and 300km. We report the results with three different procedures, i.e., conventional RD estimates with a conventional variance estimator; bias-corrected RD estimates with a conventional variance estimator; and bias-corrected RD estimates with a convention is triangular. We also control for the year fixed effects, 1-degree latitude fixed effects, and individual fixed effects using the two step methods suggested by Lee and Lemieux (2010). Robust standard errors are in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

	(1)	(2) Waterborne	(3)	(4)	(5) Cancer	(6)
Dep. Var.	Before	After	DID	Before	After	DID
Distance	-3.77e-05	-9.04e-06	-3.82e-05*	1.63e-06	-4.30e-05**	-2.20e-05**
	(2.66e-05)	(2.85e-05)	(2.29e-05)	(7.26e-06)	(1.76e-05)	(9.09e-06)
SNWT	0.0059	-0.0212*	0.0068*	0.0021	-0.0093*	0.0057***
	(0.0078)	(0.0108)	(0.0039)	(0.0028)	(0.0051)	(0.0015)
SNWT*Post			-0.0087*			-0.0085***
			(0.0049)			(0.0027)
Obs.	18,670	11,523	30,193	18,664	11,534	30,198
R^2	0.002	0.002	0.002	0.002	0.005	0.003
Year FE	Y	Y	Y	Y	Y	Y
Latitude FE	Y	Y	Y	Y	Y	Y

Table S9. Robustness check with parametric estimation

Note: This table reports the parametric RD and Difference-in-Difference estimates of the impact of SNWT on the incidence of waterborne diseases and cancer for individual located next to the SNWT middle line. Column 1 reports the discontinuity before operation of the SNWT middle line. Column 2 reports the discontinuity after operation of the SNWT middle line. Column 3 estimates the treatment effect of the SNWT using a Difference-in-Difference model. For all models, we set a bandwidth of 200km by default. We also control for the year fixed effects and 1-degree latitude fixed effects. Robust standard errors are clustered at community level and reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

	(1)	(2) Other_diseases	(3)
Dep. Var.	Before	After	Dif-in-Disc
Conventional	-0.0084	0.0917	0.1599
	(0.0100)	(0.0608)	(0.1913)
Bias-corrected	-0.0093	0.1032*	0.2230
	(0.0100)	(0.0608)	(0.1913)
Robust	-0.0093	0.1032	0.2230
	(0.0120)	(0.0722)	(0.2901)
Obs.	20,500	12,557	23,741
Bandwidth	130.87	143.83	192.87
Year FE	Y	Y	Y
Latitude FE	Y	Y	Y
Household FE	Y	Y	Y

Table S10. Placebo test with other_diseases

Note: This table reports the non-parametric RD estimates of the impact of SNWT on the incidence of other diseases for individual located next to the SNWT middle line. Column 1 reports the discontinuity before operation of the SNWT middle line. Column 2 reports the discontinuity after operation of the SNWT middle line. Column 3 estimates the treatment effect of the SNWT using a Difference-in-Discontinuity model as in Grembi (2016). For all models, the bandwidth is selected with method by Calonico, Cattaneo and Titiunik (2014), and we report the results with three different procedures, i.e., conventional RD estimates with a conventional variance estimator; bias-corrected RD estimates with a conventional variance estimator. The default kernel function is triangular. We also control for the year fixed effects, 1-degree latitude fixed effects, and individual fixed effects using the two step methods suggested by Lee and Lemieux (2010). *** p<0.01, ** p<0.05, * p<0.1.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Dep. Var.	Income	Wage	Food	Dress	Smoking	Alcoholic	Exercise
Conventional	0.7421	0.7892	0.6171	-0.0360	-0.0545	-0.1654	0.2963
	(0.4894)	(0.4876)	(0.3954)	(0.4078)	(0.1139)	(0.1393)	(0.5329)
Bias-corrected	0.6894	0.4323	0.3789	-0.2830	-0.1767	-0.0724	-0.4979
	(0.4894)	(0.4876)	(0.3954)	(0.4078)	(0.1139)	(0.1393)	(0.5329)
Robust	0.6894	0.4323	0.3789	-0.2830	-0.1767	-0.0724	-0.4979
	(0.7021)	(0.7114)	(0.5536)	(0.5389)	(0.2277)	(0.2185)	(0.7423)
Obs.	8,906	6,300	9,109	7,932	28,911	24,279	18,926
Bandwidth	205.21	190.49	209.88	179.48	277.02	215.05	150.83
Year FE	Y	Y	Y	Y	Y	Y	Y
Latitude FE	Y	Y	Y	Y	Y	Y	Y

Table S11. The economic effects of SNWT

Note: This table reports the non-parametric RD estimates of the impact of SNWT on different economic factors influencing residents' health outcome using a Difference-in-Discontinuity model as in Grembi (2016). See Table S1 for variable definition. For all models, the bandwidth is selected with method by Calonico, Cattaneo and Titiunik(2014), and we report the results with three different procedures, i.e., conventional RD estimates with a conventional variance estimator; bias-corrected RD estimates with a conventional variance estimator; and bias-corrected RD estimates with a robust variance estimator. The default kernel function is triangular. We also control for the year fixed effects, 1-degree latitude fixed effects, and household fixed effects using the two step methods suggested by Lee and Lemieux (2010). Robust standard errors are in parentheses. *** p < 0.01, ** p < 0.05, * p < 0.1.

	(1)	(2)	(3)	(4)	(5)	(9)	(2)	(8)
Dep. Var.	~	Water	borne	s 2	~	Can	cer	× ,
Subsample	Tap=0	Tap=1	Arid=0	Arid=1	Tap=0	Tap=1	Arid=0	Arid=1
Conventional	0.0477	-0.4330***	-0.9720	-0.1656***	-0.0261	-0.2532***	-0.0413	-0.0647***
	(0.0755)	(0.1414)	(0.7678)	(0.0472)	(0.0384)	(0.0787)	(0.3396)	(0.0246)
Bias-corrected	0.0790	-0.9504***	-0.6153	-0.1461***	0.0110	-0.4786***	-0.0523	-0.0638***
	(0.0755)	(0.1414)	(0.7678)	(0.0472)	(0.0384)	(0.0787)	(0.3396)	(0.0246)
Robust	0.0790	-0.9504***	-0.6153	-0.1461^{**}	0.0110	-0.4786***	-0.0523	-0.0638*
	(0.1273)	(0.1999)	(0.9105)	(0.0684)	(0.0611)	(0.1077)	(0.3807)	(0.0340)
Obs.	7,653	13,493	5,129	5,528	9,582	14,303	5,160	5,082
Bandwidth	187.07	168.75	95.13	73.46	222.88	194.76	93.37	73.26
Year FE	Υ	Υ	Υ	Υ	Υ	Υ	Υ	Υ
Latitude FE	Υ	Υ	Υ	Υ	Υ	Υ	Υ	Υ
Household FE	Υ	Υ	Υ	Υ	Υ	Υ	Υ	Υ
Vote This table ren	orts the non-nar	ametric RD estimate	as of the imnact	of SNWT on the in	dividual inciden	res of waterborne d	iseases and canc	er hy their access

Ľ
Ę
S
the
of
effects
geneous
Heterog
S 12.]
Table ?

ч Ч 44

cess **Note:** This table reports the non-parametric RD estimates of the impact of SNWT on the individual incidences of waterborne diseases and cancer by uneur access to tap water and regional aridity using Difference-in-Discontinuity model as in Grembi (2016). For all models, the bandwidth is selected with method by Calonico, mator; bias-corrected RD estimates with a conventional variance estimator; and bias-corrected RD estimates with a robust variance estimator. The default kernel function is triangular. We also control for the year fixed effects, 1-degree latitude fixed effects, and household fixed effects using the two step methods suggested by Lee and Lemieux (2010). Robust standard errors are in parentheses. *** p<0.01, ** p<0.05, * p<0.11. Cattaneo and Titiunik (CCT). (2014), and we report the results with three different procedures, i.e., conventional RD estimates with a conventional variance esti-

B The FAO-Penman-Monteith Method

This section presents the FAO-Penman-Monteith Method to calculate the potential evapotranspiration for the aridity index. In 1948, Penman, 1948 combined the energy balance with the mass transfer method and derived an equation to compute the evaporation from an open water surface from standard climatological records of sunshine, temperature, humidity and wind speed. This so-called combination method was further developed by many researchers and extended to cropped surfaces by introducing a bulk surface resistance term (Monteith, 1965). The original Penman-Monteith equation is:

$$\lambda ET = \Delta (R_n - G) + \rho_a c_p \frac{(e_s - e_a)}{r_a} / \Delta + \gamma (1 + \frac{r_s}{r_a})$$

where λET is the evaporative latent heat flux, Δ represents the slope of the saturation vapour pressure temperature relationship, R_n is the net radiation, G is the soil heat flux, $(e_s - e_a)$ represents the vapour pressure deficit of the air, ρ_a is the mean air density at constant pressure, c_p is the specific heat of the air, γ is the psychrometric constant, and r_s and r_a are the (bulk) surface and aerodynamic resistances.

The Penman-Monteith method as formulated above includes all parameters that govern energy exchange and corresponding latent heat flux (evapotranspiration) from uniform expanses of vegetation. The equation can be utilized for the direct calculation of any crop evapotranspiration as the surface and aerodynamic resistances are crop specific. For our study, we adopt an updated Penman-Monteith approach recommended by FAO (Allen et al., 1998) utilizing some assumed constant parameters for a clipped grass reference crop. The new equation is as follows:

$$ET_0 = 0.408\Delta(R_n - G) + \gamma \frac{900u_2(e_s - e_a)}{T + 273} / \Delta + \gamma (1 + 0.34u_2)$$

where ET_0 is the reference evapotranspiration rate, T is the mean air temperature and u_2 is the wind speed at 2 meters above the ground. The ET_0 is calculated using the county level climatic data and then we calculate the aridity index.