Effect of the national integrated demonstration area for the prevention and control of noncommunicable diseases programme on behavioural risk factors in China: a synthetic difference-in-differences study

Wenlan Dong,[a,j] Xiao Zhang,[a,j] Shiwei Liu,[b,j] Xian Zhang,[c] Zhaosheng Li,[d] Ming Gu,[c,e] Yibing Yang,[f] Yichen E. Fang,[c] Han Li,[a] Jing Qian,[a] Mei Zhang,[a] Limin Wang,[a] Lijing L. Yan,[c,g,h,i,**] and Jing Wu[a,*]

[a]National Center for Chronic and Noncommunicable Disease Control and Prevention, Chinese Center for Disease Control and Prevention, Beijing, China
[b]Chinese Center for Disease Control and Prevention, Beijing, China
[c]Global Health Research Center, Duke Kunshan University, Kunshan, Jiangsu, China
[d]Economics Department, Duke University, NC, USA
[e]Division of Social Sciences, Duke Kunshan University, Kunshan, Jiangsu, China
[f]Office of Noncommunicable Disease and Aging Health Management, Chinese Center for Disease Control and Prevention, Beijing, China
[g]School of Public Health, Wuhan University, Wuhan, Hubei, China
[h]PKU Institute for Global Health and Development, Peking University, Beijing, China
[i]Department of Preventive Medicine, Feinberg School of Medicine, Northwestern University, Chicago, IL, USA

The Lancet Regional Health - Western Pacific 2024;50: 101167

Published Online xxx
https://doi.org/10.1016/j.lanwpc.2024.101167

Summary

Background The government-led community-based Chinese National Integrated Demonstration Areas for the Prevention and Control of Noncommunicable Diseases programme was launched in 2011, but no rigorous impact evaluation has been conducted to date. We aimed to evaluate the causal effects of this programme on behavioural risk factors.

Methods We used data from the latest five waves of the China Chronic Disease and Risk Factor Surveillance. The primary outcome is a behavioural risk score combining current smoking, passive smoking, drinking in last month, regular exercise, body mass index, and waist circumference. We applied the synthetic difference-in-differences method and constructed synthetic controls from the non-demonstration areas with the outcome. The average treatment effects on the treated were estimated for overall effect and by short- (1–2), medium- (3–4), and long-term (6–7 years) effects.

Findings We identified 26 demonstration areas (N = 72,193) and 100 non-demonstration areas (N = 275,397). Participants in the demonstration areas had higher education and income levels and different pre-implementation trends than non-demonstration areas. Using synthetic controls instead of non-demonstration areas reduced these pre-implementation differences. Compared to the synthetic controls, declines were observed in current smoking (−1.78% [−4.51%, 0.96%]), passive smoking (−8.09% [−14.27%, −1.90%]), and drinking in last month (−4.04% [−8.75%, 0.67%]) but not in the other factors. Behavioural risk score declined by 1.05 short-term (95% CI: −1.84, −0.26), 1.15 medium-term (95% CI: −2.08, −0.22), 2.82 long-term (95% CI: −4.79, −0.85), and 1.54 overall (95% CI: −2.51, −0.56).

Interpretation The programme improved behavioural risk scores, primarily through reductions in the prevalence of smoking and drinking, and the effect was long-lasting. Our findings provided empirical evidence for utilizing an integrated prevention and control strategy to fight against NCD in China and other countries facing similar challenges.

Abbreviations: ATT, Average treatment effect of the treated; DID, Difference-in-differences; GDP, Gross domestic product; NCD, Non-communicable disease; SC, Synthetic control; SDID, Synthetic differece-in-differences

*Corresponding author. No. 27 Nanwei Road, National Center for Chronic and Non-communicable Disease Control and Prevention, Chinese Center for Disease Control and Prevention Xicheng District, Beijing, 100050, China.
**Corresponding author. 8 Duke Avenue, Global Health Research Center, Duke Kunshan University, Kunshan, Jiangsu, 215316, China.
E-mail addresses: wujing@ncncd.chinacdc.cn (J. Wu), lijing.yan@duke.edu (L.L. Yan).
[j]These authors contributed equally to this work.

Funding The China National Key Research and Development Program (2018YFC1315304 and 2017YFC1310902); National Natural Science Foundation of China (81872721).

Keywords: Noncommunicable diseases; Behavioural risk factors; Synthetic difference-in-differences; Impact evaluation; Programme evaluation

Research in context

Evidence before this study

We searched PubMed and Chinese databases (China National Knowledge Infrastructure, Wan Fang database, and Baidu Scholar) for articles in English or Chinese up to June 2023, using a collection of terms including noncommunicable disease (“noncommunicable disease” or “chronic and noncommunicable disease” or “chronic disease” or “NCD”), demonstration area (“demonstration area” or “demonstration district”), integrated prevention and control (“integrated prevention and control” or “comprehensive prevention and control”), behavioural risk factor (“behavioural risk factor” or “tobacco” or “cigarette” or “smoke” or “passive smoke” or “secondhand smoke exposure” or “alcohol” or “drink” or “physical activity” or “exercise” or “body mass index” or “BMI” or “waist circumference” or “WC”), effect (“effect” or “impact”), and geographic locations (“China” or “Chinese”). We reviewed all studies without restrictions on the study design and found only one project with four articles published in Chinese journals. This project, conducted in 2016—five years after the initiation of the programme, evaluated the effect of the programme using qualitative research methods (focus group discussions and personal interviews) and a cross-sectional survey (N = 3891) from 10 demonstration areas without any control. Due to the lack of controls and the cross-sectional nature of self-reported data from demonstration areas only, no causal inference on the impact of the programme can be made from this evaluation attempt.

Added value of this study

We evaluated the potential causal effects of the programme on behavioural risk factors. To do so, we used the quasi-experimental synthetic difference-in-differences design to solve the selection bias and unobserved confounding introduced during the approval process of the demonstration areas. We used data (N = 347,590 from 126 areas) from the five most recent waves of the nationally representative China Chronic Disease and Risk Factor Surveillance, with the primary outcome being a composite behavioural risk score combining six behavioural factors, including current smoking, passive smoking, drinking in last month, regular exercise, body mass index, and waist circumference. Compared to synthetic controls, the programme led to a 1.78% decline in current smoking (95% CI: −4.51%, 0.96%), 8.09% decline in the passive smoking rate (95% CI: −14.27%, −1.90%), and 4.04% decline in drinking in last month (95% CI: −8.75%, 0.67%), but not significantly so in the other three factors. The composite behavioural risk score declined by 1.05 short-term (95% CI: −1.84, −0.26), 1.15 medium-term (95% CI: −2.08, −0.22), 2.82 long-term (95% CI: −4.79, −0.85), and 1.54 overall (95% CI: −2.51, −0.56). Our findings are consistent with the programme has led to significant and increasing improvements in behavioural risk scores.

Implications of all the available evidence

Our synthetic difference-in-differences quasi-experimental study provides evidence for the programme’s effectiveness on behavioural outcomes and evidence-based policy making for programme expansion and scaling up in China. It may also shed insight for other countries facing similar challenges in NCD prevention and control.

Introduction

Non-communicable diseases (NCDs) accounted for 73% of global deaths in 2017[1] and 74.36% in 2019.[2] Due to the high disease burden, NCD prevention and control are highly prioritized. In September 2011, the United Nations held the first High-Level Meeting on NCDs.[3] The Political Declaration on the Prevention and Control of NCDs, published by the United Nations in the same year, recognized “with profound concern” the global burden and threat of NCDs as a major “challenge of epidemic proportions for development” in the twenty-first century.[4] It also called for “a whole-of-government and a whole-of-society effort” to respond to the challenge.[5]

In China, the mobility burden and mortality caused by NCD are also alarmingly high and increasing. The share of NCD-caused mortality increased from 80.0% in 2002 to 86.6% in 2012 and to 88.5% in 2019, which is higher than the global average.[6] In 2011, the Chinese government had begun to pilot a programme called “the National Integrated Demonstration Area for the Prevention and Control of NCDs (NIDAN)”, led by the Ministry of Health (then called “National Health and Family Planning Commission”).[7] Only three months after the United Nations High-Level Meeting on NCDs, the first batch – a total of 39 areas (urban districts or rural counties) from 18 provinces were approved in December 2011 to become demonstration areas. Four more batches have been approved since then. In 2020, there were a total of 488 demonstration areas from all 31 provinces in China, covering 17.1% of all district-or county-level areas and more than 250 million people.

Among various sub-targets that this programme covered, changing behavioural factors are highly important as they are one of the key determinants for NCD prevention. Evaluating the impact of the programme on behavioural factors can identify priorities for improvement, provide evidence to guide its future expansion in China, and shed insight on launching similar programs in other countries. However, rigorous evaluation of this real-world programme is challenging. The gold standard of (cluster-) randomized controlled trials was not feasible as the areas could not be randomly assigned. Quasi-experimental methods such as difference-in-differences (DID) and the newer synthetic control (SC) method are reasonable substitutes as they evaluate the programme mimicking randomized controlled trials. However, in the demonstration programme, the assignment of demonstration areas are not random and must fulfill certain criteria to be approved and thus are inherently better than average in the prevention and control of NCDs. In a separate methodological paper, we have demonstrated that assumptions for the classical quasi-experimental method of DID or the newer SC method could not be satisfied, rendering them inappropriate as well.

In 2021, Arkhangelsky et al. introduced the innovative synthetic difference-in-differences (SDID) method.[8] It has been shown to outperform classical quasi-experimental methods in simultaneous adoption design by relaxing many of the required assumptions. In place of DID and SC, SDID has been applied in evaluating the effect of various programmes on health outcomes such as hospital lengths of stays,[9] obesity,[10] and diet-related health,[11] especially behavior factors such as consumption of fruits and vegetables.[12] In our methodological paper, we have shown that the SDID method is more suitable than DID or SC in evaluating programs implemented in a staggered setting (e.g., in batches as in the demonstration programme). Given the significance of this largest and longest-running national programme on NCDs, we aimed to use the SDID method to evaluate the effect of the programme on risk factors and health outcomes, with a particular focus on behavioural risk factors in this paper.

Methods

National NCD demonstration areas programme

The National NCD Demonstration Areas programme aims to improve area-level NCD prevention and control by establishing demonstration zones where local governments are required to form leadership teams. These teams enhance comprehensive control based on an action index system issued by the National Health Commission. The intervention includes promoting healthy lifestyles, physical fitness, tobacco control, creating supportive environments, advancing chronic disease screening, early diagnosis, and tiered medical services, etc.[13] The leadership team should distribute interventions among several government departments from seven key aspects, which also serve as the criteria for annually determining whether this area qualifies as a demonstration area: (1) governmental leadership; (2) supportive environment; (3) integrated system; (4) health education and health promotion; (5) whole-course management of NCDs; (6) surveillance and evaluation; and (7) innovation and dissemination[14,15] (see Panel 1 for more details). All implementation processes for each indicator in each area met the criteria was guided by government documents. The panel, appendix 1 and 2 provides the details of the intervention and the workflow of implemention progress.

The approval and initiation of demonstration areas has been carried out in five batches – 2011 (39 areas, reevaluated in 2016), 2012 (101 areas reevaluated in 2017), 2014 (125 areas, reevaluated in 2019), 2017 (103 areas, reevaluated in 2022), and 2020 (123 areas). A total of 488 areas were approved by 2020 with three areas lost due to changes in administrative divisions, covering 17.1% of all district-or county-level areas in China and more than 250 million people.

Synthetic difference-in-differences

We used the SDID method to estimate the causal effect of this programme on behavioural risk factors. In general, SDID combined the advantages of the DID and SC methods and relaxed corresponding assumptions. In traditional DID, parallel trend assumption is stickly required. However, SDID automatically satisfies the parallel trend assumption and applies a logic similar to SC by weighting both units and time periods. Instead of directing using non-demonstration areas as the counterfactual outcome for demonstration areas, Y^0, the SDID method involves choosing the optimal synthetic weight W^{sdid} to construct synthetic controls from the non-treated group which's outcome parallel with the average outcome for demonstration areas. It also looks for time weight $\Gamma = (\lambda_1, \lambda_2 ..., \lambda_{T_0})$ that balances pre-intervention time periods with post-intervention.[11] Therefore, it is relatively free from potential selection bias and confounding factors. The abstract of our methodological paper comparing the SDID with the DID and SC methods was included in Appendix 3. The effect of this programme was measured by the average treatment effect. Given the optimum area weight $\widehat{\omega}_N^{sdid}$ and time weight $\widehat{\lambda}_t$, we estimated average treatment effects of the treated (ATTs) as:

Panel 1. National NCD demonstration areas programme

To address the increasing burden of NCDs in China, the National Health Commission (formerly Ministry of Health) initiated the National Integrated Demonstration Areas for Prevention and Control of NCDs programme in 2011. The programme emphasizes the importance of integrated approaches for NCD prevention and control at the district (urban) or county (rural) level ("area" hereafter). The goal of the programme is to improve area-level NCD prevention and control and to promote effective integrated strategies at a larger scale.

The programme designates certain areas as demonstration zones and mandates local governments to form leadership teams within these zones. These teams are integral to the intervention, tasked with enhancing comprehensive control and intervention through an action index system issued by the National Health Commission. The intervention includes promoting healthy lifestyles, physical fitness, tobacco control, creating supportive environments, advancing chronic disease screening, early diagnosis, and tiered medical services, etc. The performance of these teams in NCD prevention is standardized and evaluated annually based on seven key aspects: (1) governmental leadership; (2) supportive environment; (3) integrated system; (4) health education and health promotion; (5) whole-course management of NCDs; (6) surveillance and evaluation; and (7) innovation and dissemination. Appendix 1 provides the detailed scoring system of criteria for evaluating whether an area's interventions meet the standards, including the 22 specific indicators and actions for the seven categories, along with their evaluation weights and the responsible government departments.

The workflow for the implementation progress of the programme, as well as the evaluation procedure, is explained in Appendix 2. In summary, one year after becoming provincial demonstration areas, the local area-level governments are eligible to apply to become national demonstration areas and, after approval by the National Health Commission-appointed expert panel, to continue to implement area-level multi-sectoral and community-based measures targeting health promotion and education, early diagnosis, and standardized management for preventing and controlling NCDs.

$$ATT^{sdid} = E[Y|t > T_0, D = 1] - E[\widehat{\lambda}_t Y|t > T_0, D = 1] - \sum_{n=1}^{N-K} \widehat{\omega_n^{sdid}} \left(E[Y|t > T_0, D = 0] - E[\widehat{\lambda}_t Y|t \leq T_0, D = 0] \right. \quad (1)$$

where Y is the outcome of interest, $t > T_0$ represents post-treatment period, $t \leq T_0$ represents pre-treatment period, D is a dummy variable indicates whether the area is a demographic area. Full details of the SDID model, optimal weight selection, and ATTs estimation can be found in Appendix 4.

Data sources

Individual-level behavioural outcomes were from China Chronic Disease and Risk Factor Surveillance (CCDRFS), a nationwide series of cross-sectional surveys conducted every three years since 2004, with the main aim of collecting information on major NCDs and related risk factors among Chinese adults. Study participants were selected based on a multistage stratified cluster sampling with the smallest primary sampling unit being urban districts or rural counties, just as the programme's implementation units, known as disease surveillance points in CCDRFS.[16] Up to the present, six surveys had been carried out in 2004 (covering 79 areas), 2007 (161 areas), 2010 (161 areas), 2013 (298 areas), 2015 (298 areas), and 2018 (298 areas).[17] We used data from all waves except 2004, as its questionnaire was incomparable with the rest.

Because each wave of CCDRFS did not necessarily include the same areas, there were 126 areas having data from all five waves from 2007 to 2018, with 26 areas becoming demonstration areas in three batches in 2011, 2012, or 2014 (see Appendix 5). The remaining 100 were non-demonstration areas and served as the donor pool to construct synthetic controls. We further combined the first and second batches (2011 and 2012) into one group because their implementation periods fell between two waves of CCDRFS (2010 and 2013).

In the SDID model, we included demographic and economic domains as covariates: sex ratio (male over female), the proportion of older adults aged 65 years or older (out of the total population), and the log-transformed (GDP) per capita. These data for the 126 areas were obtained from the local Bureau of Statistics yearbooks for each area in each survey year.

Outcome measurements

The focus of this evaluation was on behavioural risk factors. The following six variables from CCDRFS were included: current smoking, passive smoking, drinking in the last month, regular exercise, body mass index, and waist circumference (see Appendix 6 for detailed definitions or measurements). No dietary data were available in the 2007 wave. Body mass index and waist circumference as indicators of overall and central obesity were included owing to their close relationships with diet and physical activity.

The programme intervent from seven aspects, so a composite measure is needed to capture its effects succinctly and comprehensively, rather than focusing on individual outcomes. Therefore, we created a composite measure called 'behavioural risk score' to indicate the overall effect as the primary outcome for this programme evaluation. The six behavioural variables above were assigned equal weights. "Regular exercise" had a weight of −1 as its direction was opposite of the other variables. Then we added up six behavioural variables to generate the behavioural risk score. A lower behavioural risk score indicates a better overall outcome.

To obtain area-level outcomes, individual-level data were aggregated using weights accounting for sample design, non-response, and age-sex differences between selected and target population.

Statistical analysis

We performed descriptive analyses of outcomes by implementation status and year. The Shapiro–Wilk test was employed to assess the normality assumption, then Welch's t test and Wilcoxon rank sum test was separately applied for normally distributed variables and skewed variables. We used the SDID method to estimate the causal effect of this programme on behavioural risk factors. In this study, a synthetic control that had parallel trend in adjusted outcomes with each batch of demonstration areas was modelled as the optimal weighted combination of the donor pool (i.e., the 100 non-demonstration areas), with time weights being equal across pre-treatment period. Linear regression with identity link was performed to model the relationship between behavioural risk factors and predictors, including unit-fixed effects, time-fixed effects, and covariates (proportion of elderly, sex ratio, and log-transformed GDP per capita). The adjustment for covariates was fulfilled by subtracting the effect of covariates on outcomes. The linearity assumption was checked using penalized splines of covariates in generalized additive models (Appendix 7A, B, and C). Since log-transformed GDP per capita showed non-linear relationships with some outcomes, the natural cubic spline of log-transformed GDP per capita was instead incorporated into the model. The degree of freedom of the natural cubic term was set to three because it generated the best model fitting.

To estimate the effect of NIDAN, the average difference in the related risk factor between the demonstration areas and the synthetic control during the entire pre-treatment period was subtracted from the average difference during the entire post-treatment period. To determine the overall effect across multiple batches, SDID estimates were first obtained from the first two batches and the third batch, then a weighted average of these estimates was calculated by weighting to the fraction of treated unit/time–perod pairs. Considering that the effect of this programme may be transient and/or lagging and the availability of data, the short-term effect as 1–2 years post-treatment, medium-term effect as 3–4 years post-treatment, and long-term effect as 6–7 years post-treatment were further separately estimated. Clustered bootstrap of the corresponding counties/districts was performed 1000 times to obtain standard errors. As the asymptotic normality of obtained estimates had been proved by Arkhangelsky et al.,[8] which also held true in our study, the results were summarized as mean difference adding or subtracting 1.96 times its standard error. All statistical analyses were conducted in R (version 4.2.1) using the "synthdid"[8] and "xsynthdid"[18] (for time-varying covariates) packages. We reported the average treatment effects of the treated with 95% confidence interval.

Sensitivity analyses

Several sensitivity analyses were conducted to verify the robustness of our results. First, we repeated the analyses to use time weights that balance the pre-treatment and the post-treatment periods for the control group, i.e., to make the average pre-treatment change as similar as possible to the average post-treatment change in the control group. Second, in the donor pool, we excluded non-demonstration areas located in the same city with any demonstration area, to satisfy the "no interference across areas" assumption with more rigor. Third, areas with top 10% highest growth in GDP per capita during 2010–2018 were excluded to avoid the potential confounding from socioeconomic factors that may affect both health behaviors and selection of demonstration areas. Fourth, considering that the relatively short pre-treatment period might subject our study to bias due to overfitting to noise, we applied the weights derived from the behavioural risk score to each single behavioural variable.

Role of the funding source

The funder of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the report.

Results

The 26 demonstration and 100 non-demonstration areas in the CCDRFS over 2007–2018 included 72,193 and 275,397 participants, respectively. Table 1 reports the individual-level and area-level characteristics for two demonstration groups and all non-demonstration areas. Participants in the demonstration areas had higher education and household income per capita. Area-level GDP per capita was also much higher in demonstration areas than that in non-demonstration areas. The systematic pre-implementation differences between demonstration and non-demonstration areas indicated the need for a more compatible control group.

For the 2011–2012 batches, the pre-implementation trends for most behavioural factors were different between demonstration and non-demonstration areas (Fig. 1, left of the vertical grey bars and see Appendix 8 for detail results). However, when comparing demonstration areas with their synthetic controls, we found that the pre-implementation trends were parallel across all outcomes. This indicates that by applying the SDID

	Demonstration areas in the 2011/12 batch	Demonstration areas in the 2014 batch	Non-demonstration areas
Individual characteristics	–	–	–
N=	52,675	19,518	275,397
Age (years)	50.8 ± 15.0	51.6 ± 14.6	49.7 ± 14.6
Men (n, %)	23,486 (44.6)	8971 (46.0)	125,183 (45.5)
High school or above (n, %)	16,943 (32.2)	5016 (30.8)	54,852 (19.9)
Annual household income per capita (CNY)	12,500 (6400–25,000)	10,000 (4500–22,500)	8000 (3750–15,000)
Area-level characteristics	–	–	–
N=	19	7	100
Proportion of elderly (%)	9.9 ± 2.1	11.2 ± 2.3	9.3 ± 3.3
Sex ratio[a]	102.6 ± 4.3	102.6 ± 6.1	104.2 ± 4.1
Gross domestic products per capita (CNY)	56,367 (33,627–89,533)	35,423 (25,446–58,640)	29,304 (16,655–48,874)

BMI, body mass index; WC, waist circumference; GDP, gross domestic product. The data were presented as mean ± SD, n (%), or median (P_{25}, P_{75}). [a]The numer of men per 100 women.

Table 1: **Individual and area-level characteristics by implementation status of the demonstration programme, 2007–2018.**

method and using synthetic controls as the control group, the parallel trend assumption was satisfied (Fig. 1, left of the vertical grey bars). The three columns in Fig. 1 show the short-term (1–2 years), medium-term (3–4 years) and long-term (6–7 years) trends after initiation of the programme. Although the trends had different patterns across the risk factors, improvement for demonstration areas (solid red lines) over the SC (dash red lines) or non-demonstration areas (dash blue lines) was discernible for each time period, and the difference was increasing over time for most risk factors. For the composite measure of behavioural risk score, we observed clear, large, and increasing improvements over time.

The same trend was observed for the demonstration areas approved in 2014 (Fig. 2). The behavioural risk score decreased (improved) for the demonstration areas but increased for the two sets of controls. Most individual risk factors had similar trends favoring the demonstration areas but there was a noticeable decrease in regular exercise compared with non-demonstration areas in 2018.

The multivariable SDID estimates, adjusted for sex ratio, proportion of older adults, and GDP per capita, were calculated for the primary outcome of behavioural risk score. The results of behavioural risk score showed that the pooled effect of the programme was –1.54, with a 95% confidence interval (CI) ranging from –2.51 to –0.56. Specifically, the ATTs for the 2011–2012 batch and 2014 batch were –1.75 (95% CI: –3.01 to –0.48) and –0.69 (95% CI: –1.69 to 0.32), respectively (Table 2). As for each behavioural risk factor, the results indicate a pooled ATTs for the current smoking rate of –1.78%, with a 95% CI ranging from –4.51% to 0.96%. The ATTs for the 2011–2012 batch and 2014 batch were –1.76% (95% CI: –5.25% to 1.73%) and –1.85% (95% CI: –4.50% to 0.80%), respectively. For the passive smoking rate, the pooled ATT was –8.09%, with a 95% CI ranging from –14.27% to –1.90%. The subgroup ATTs for 2011–2012 batches and 2014 batch were –8.66% (95% CI: –16.73% to –0.59%) and –5.75% (95% CI: –14.05% to 2.55%), respectively. These findings suggest that the programme effectively reduced the prevalence of both current and passive smoking. As for the prevalence of drinking in the last month, the pooled ATTs were –4.04% (95% CI: –8.75% to 0.67%). For 2011–2012 batches and 2014 batch, the ATTs were –4.54% (95% CI: –10.54% to 1.45%) and –1.99% (95% CI: –4.60% to 0.61%), respectively. The results showed a reduction in the prevalence of drinking. As for regular exercise, the effect was not consistent across batches. Although the ATTs for body mass index and waist circumference were consistently negative across batches, the effect was too small to have clinical implications.

When examining the short-term, medium-term, and long-term effects separately, for the 2011–2012 batches, the effect sizes increased over time for the behavioural risk score as well as current smoking, and passive smoking (Table 3). Only short-term and medium-term results were available for the 2014 batch or pooled results. The effect sizes were generally smaller, with no clear increase over time except for smoking and drinking. Results from the four sets of sensitivity analyses (time weights, excluding same-city non-demonstration areas, excluding top 10% highest growth GDP, and applying weight from behavioural risk score to each single behavioural variables) were generally consistent, thus verifying the robustness of the SDID method and analyses (Appendices 9–13).

Discussion

The programme is the largest and longest-running national initiative on NCD prevention and control in China. With data from a series of independent nationally representative surveys (N = 347,590 for 126 areas), we adopted the SDID method to rigorously evaluate the impact of the programme on behavioural risk factors. We observed consistent and increasing improvements over time in the overall behavioural risk score for the demonstration areas compared to synthetic controls or non-demonstration areas. Out of the six behavioural risk factors examined, the programme reduced prevalence of current smoking, passive smoking, and drinking, with a trend toward increasing effects from short (1–2 years post-programme initiation), medium (3–4 years), to long-term (6–7 years). The results on regular exercise were mixed and the changes in body mass index and waist circumference were small.

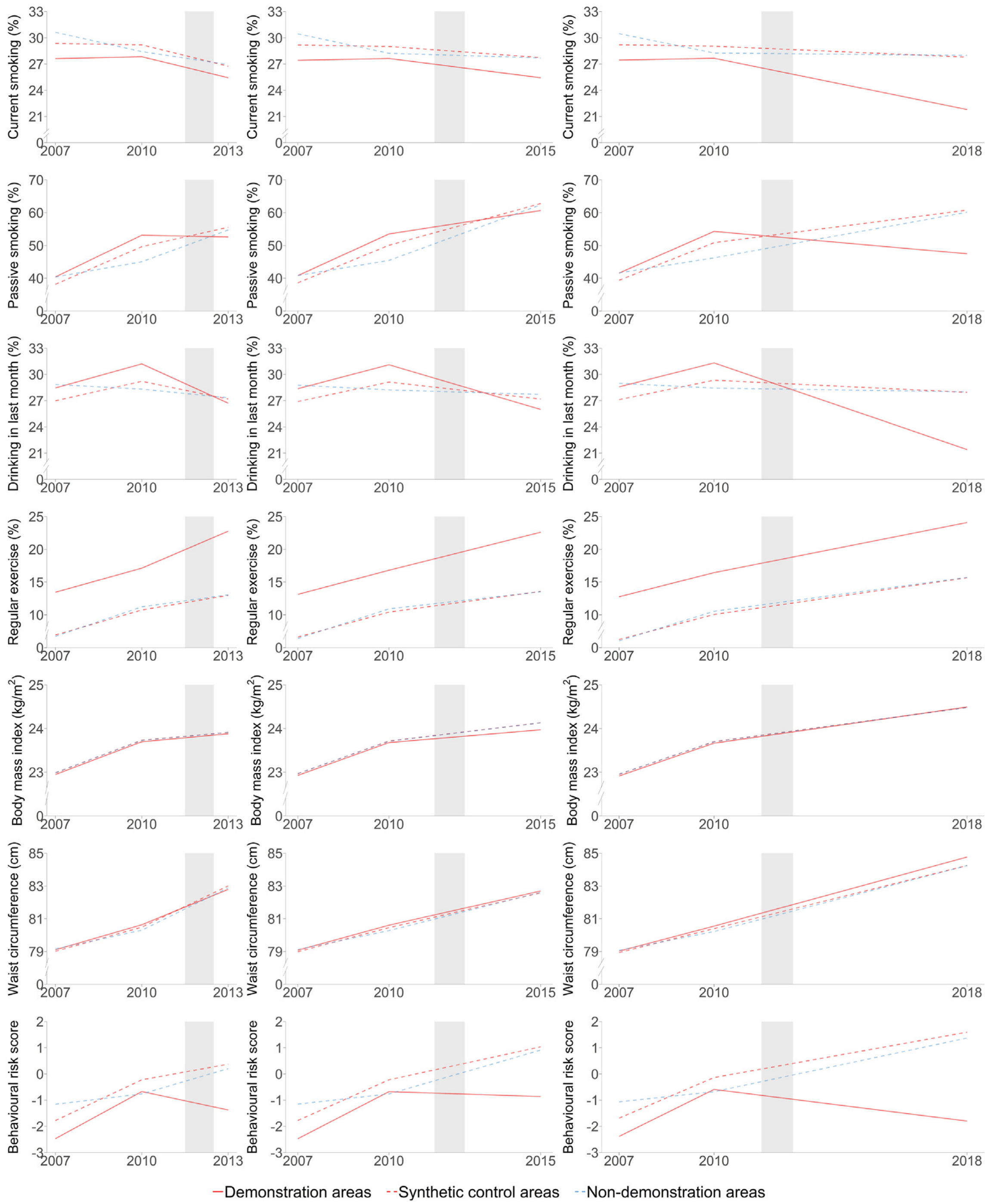

Fig. 1: Trend of health-related behaviors and factors in counties that had implemented NIDAN in 2011/12, their synthetic controls, and counties that had not implemented NIDAN before or in 2018, after 1–2 years (left), 3–4 years (middle), and 6–7 years (right) of implementation. The shaded area indicates the period of NIDAN implementation. NIDAN, non-communicable disease demonstration area; BMI, body mass index; WC, waist circumference; DA, demonstration area.

The etiology of NCDs are complex and multi-factorial, with behavioural factors playing a significant role.[2] To address the rapidly rising burden of NCDs, multi-pronged community-based strategies are needed and promoted by the United Nations and the World Health Organization.[19] This programme adheres to these principles and is characterized by government leadership and policy support, community engagement,

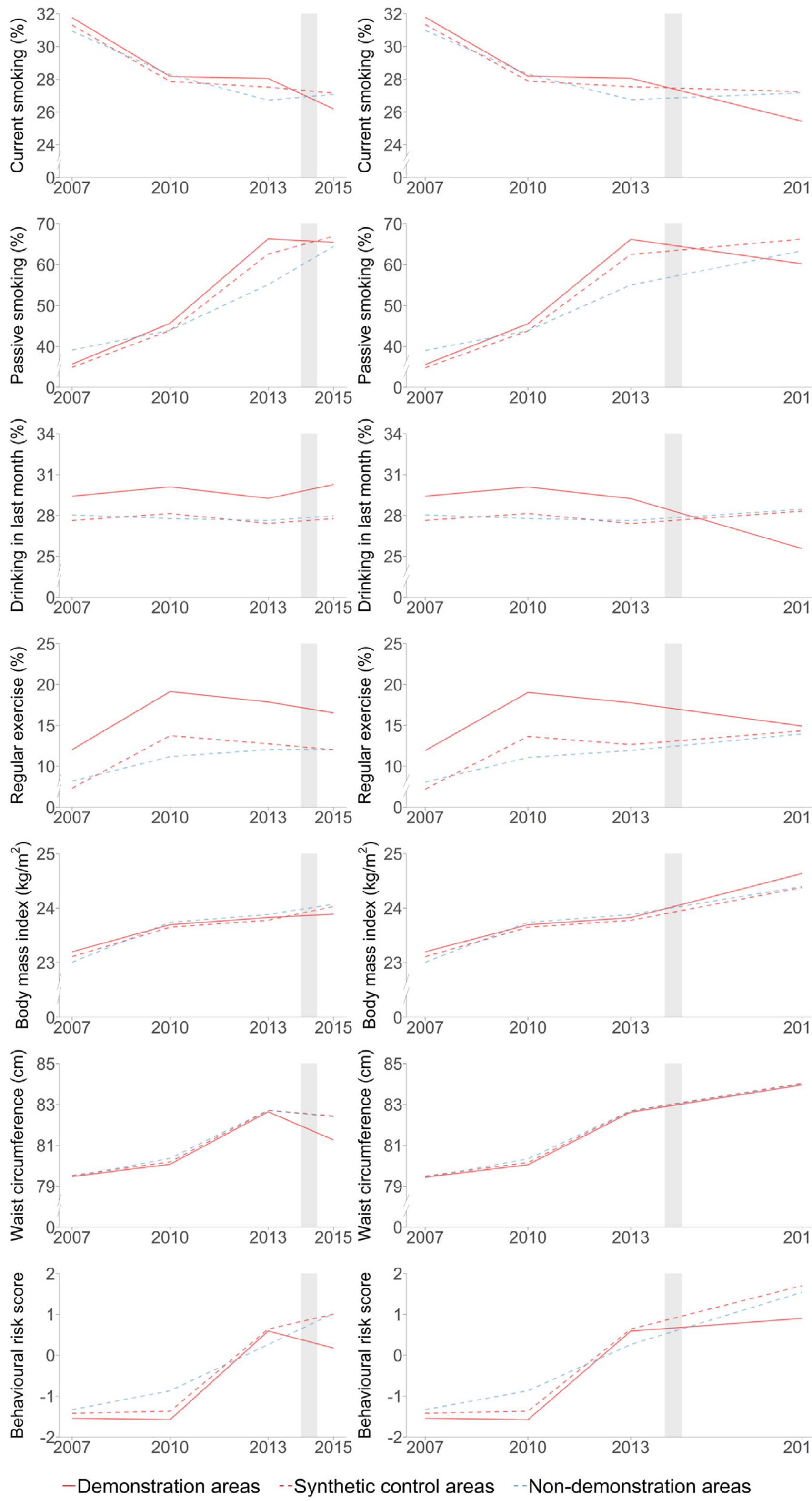

Fig. 2: Trend of health-related behaviors and factors in counties that had implemented NIDAN in 2014, their synthetic controls, and counties that had not implemented NIDAN before or in 2018, after 1 year (left) and 4 years (right) of implementation. The shaded area indicates the period of NIDAN implementation. NIDAN, non-communicable disease demonstration area; BMI, body mass index; WC, waist circumference; DA, demonstration area.

building a supportive (e.g., smoke-free) environment, health education and promotion, and surveillance and monitoring. It is more "integrated" or comprehensive than campaigns focusing on single or a limited set of risk factors such as a healthy diet or tobacco control. Therefore, the behavioural risk score comprised of six behavioural risk factors is a suitable composite measure to evaluate its effects. Our findings of improvements in behavioural risk score demonstrated the overall effectiveness of this programme. The increasing effects from short-, medium-, to long-term are consistent with the fact that it takes time for large regional programs to manifest their impacts.[20]

The effect of this programme on individual risk factors was generally reassuring with some mixed and non-consistent results. Tobacco control is a long-term process that requires the implementation of comprehensive policies and measures. Nationally, tobacco smoking among people aged 15 or above declined by 1.5 percentage points from 28.1% in 2010 to 26.6% in 2018, far from the 20% by 2030 goal set by the Healthy China Action Plan (2019–2030).[21] Against this backdrop, the demonstration areas had larger declines than control areas in the current smoking rate and passive smoking rate. This could be attributable to the programme's emphasis not only on smoking cessation but also on building smoke-free environments, mass media, and health education.

Drinking, associated with increased risks in more than 200 diseases and injuries,[22,23] is the third most important behavioural risk factor for NCDs after diet and tobacco use.[24] The results showed that interventions targeting drinking behaviors were effective. The programme's impact on reducing drinking behaviors accumulated over the middle and long-term, indicating a consistent and lasting effect.

Over the past 20 years, the prevalence of overweight and obesity has risen sharply in China.[25–27] Compared to control areas, the demonstration areas could not reverse the trend of increasing obesity but appeared to have a slower increase, especially waist circumference. For regular exercise, a behavior closely related to obesity, we observed increases in some years but decreases in others. Our results suggest that more intensive and comprehensive policies and measures need to be implemented to tackle the health and economic burdens of obesity.

Before this study, the only "systematic" evaluation of this programme was conducted in 2016. Besides qualitative interviews and site visits, the evaluation team surveyed 3891 residents in 10 selected demonstration areas from a total of 265 areas in 2016.[28] This study found 72.1% of participants reported consuming vegetables daily, 53.6% consuming fruits daily, and 86.9% walking at least 10 min per day.[29] Compared to one year ago, they self-reported decreasing salt and oil intake while increasing physical activity. Although individual-level data were analyzed and published in four papers in Chinese journals, these cross-sectional findings could only be compared with national averages to infer the effects of the programme on behavioural risk factors,[30,31] hypertension,[32] and diabetes[33] management. These studies failed to consider the systematic difference in pre-implementation between demonstration areas and non-demonstration areas, leading to potential overestimation of the results.

As the first causal evaluation of the programme, we didn't use cross-sectional analayes nor traditional quasi-experimental methods such as DID or SC, which have been shown to lead to biased policy effect estimates in evaluating such programme due to violating the key parallel trend assumption for DID or convex hull condition for SC caused by potential selection bias.[8,34–36] Rather, we adopted the innovative SDID method to overcome common biases and fully utilize the available data. The SDID method estimates ATTs by comparing the treatment group with a constructed synthetic control group. This approach is considered more reliable because SDID reduces reliance on the parallel trend assumption by constructing a synthetic control that ideally satisfies this assumption. To the best of our knowledge, the validation of SDID is based on the pre-treatment parallel trend between the treatment group

	Year of implementation	Absolute difference (95% CI)	*P*-value
Current smoking (%)	2011–2012	−1.76 (−5.25, 1.73)	0.32
	2014	−1.85 (−4.50, 0.80)	0.17
	Pooled	−1.78 (−4.51, 0.96)	0.20
Passive smoking (%)	2011–2012	−8.66 (−16.73, −0.59)	0.036
	2014	−5.75 (−14.05, 2.55)	0.18
	Pooled	−8.09 (−14.27, −1.90)	0.011
Drinking in last month (%)	2011–2012	−4.54 (−10.54, 1.45)	0.14
	2014	−1.99 (−4.60, 0.61)	0.13
	Pooled	−4.04 (−8.75, 0.67)	0.093
Regular exercise (%)	2011–2012	2.41 (−2.82, 7.64)	0.37
	2014	−2.58 (−5.83, 0.67)	0.12
	Pooled	1.43 (−2.59, 5.44)	0.49
Body mass index (kg/m^2)	2011–2012	−0.02 (−0.32, 0.29)	0.92
	2014	−0.00 (−0.21, 0.20)	0.99
	Pooled	−0.01 (−0.25, 0.22)	0.92
Waist circumference (cm)	2011–2012	−0.13 (−1.43, 1.17)	0.84
	2014	−0.53 (−1.62, 0.56)	0.34
	Pooled	−0.21 (−1.23, 0.80)	0.68
Behavioural risk score	2011–2012	−1.75 (−3.01, −0.48)	0.0071
	2014	−0.69 (−1.69, 0.32)	0.18
	Pooled	−1.54 (−2.51, −0.56)	0.0021

NIDAN, non-communicable disease demonstration area; BMI, body mass index; WC, waist circumference. [a]All estimates were adjusted for proportion of older adults aged 65 years or older, sex ratio, and natural cubic spline of log-transformed GDP per capita.

Table 2: **Overall effect[a] of the demonstration programme on behavioural risk factors by batch.**

	Year of implementation	1–2 years after implementation of the programme		3–4 years after implementation of the programme		6–7 years after implementation of the programme	
		Absolute difference (95% CI)	*P*-value	Absolute difference (95% CI)	*P*-value	Absolute difference (95% CI)	*P*-value
Current smoking (%)	2011–2012	0.23 (−3.24, 3.71)	0.90	−0.74 (−3.78, 2.29)	0.63	−4.44 (−9.69, 0.82)	0.099
	2014	−1.39 (−5.36, 2.58)	0.49	−2.21 (−5.81, 1.38)	0.23	–	–
	Pooled	−0.20 (−2.85, 2.44)	0.88	−1.14 (−3.48, 1.20)	0.34	–	–
Passive smoking (%)	2011–2012	−5.85 (−13.20, 1.51)	0.12	−4.95 (−11.92, 2.03)	0.17	−16.15 (−27.43, −4.86)	0.0055
	2014	−3.54 (−11.63, 4.55)	0.39	−8.14 (−20.19, 3.92)	0.19	–	–
	Pooled	−5.23 (−10.66, 0.21)	0.060	−5.80 (−11.75, 0.14)	0.056	–	–
Drinking in last month (%)	2011–2012	−2.20 (−7.30, 2.89)	0.40	−2.91 (−7.74, 1.92)	0.24	−8.26 (−16.80, 0.27)	0.059
	2014	0.64 (−2.55, 3.83)	0.69	−4.60 (−8.90, −0.30)	0.037	–	–
	Pooled	−1.44 (−5.17, 2.30)	0.45	−3.37 (−6.93, 0.20)	0.065	–	–
Regular exercise (%)	2011–2012	3.34 (−2.20, 8.88)	0.24	2.59 (−1.84, 7.02)	0.25	1.97 (−5.42, 9.36)	0.60
	2014	−0.55 (−3.33, 2.22)	0.70	−4.48 (−9.27, 0.30)	0.067	–	–
	Pooled	2.29 (−1.63, 6.21)	0.25	0.68 (−2.67, 4.04)	0.69	–	–
Body mass index (kg/m^2)	2011–2012	0.00 (−0.33, 0.34)	0.98	−0.12 (−0.39, 0.14)	0.35	0.05 (−0.40, 0.50)	0.84
	2014	−0.20 (−0.53, 0.13)	0.23	0.20 (−0.17, 0.56)	0.29	–	–
	Pooled	−0.05 (−0.29, 0.19)	0.68	−0.04 (−0.24, 0.17)	0.71	–	–
Waist circumference (cm)	2011–2012	−0.33 (−1.70, 1.03)	0.63	−0.01 (−1.39, 1.37)	0.99	0.40 (−1.61, 2.42)	0.70
	2014	−1.09 (−2.50, 0.32)	0.13	−0.00 (−1.11, 1.10)	0.99	–	–
	Pooled	−0.54 (−1.58, 0.50)	0.31	−0.01 (−1.02, 1.00)	0.99	–	–
Behavioural risk score	2011–2012	−1.18 (−2.25, −0.10)	0.033	−1.33 (−2.46, −0.19)	0.023	−2.82 (−4.79, −0.85)	0.0055
	2014	−0.70 (−1.66, 0.25)	0.15	−0.67 (−2.25, 0.90)	0.40	–	–
	Pooled	−1.05 (−1.84, −0.26)	0.0097	−1.15 (−2.08, −0.22)	0.015	–	–

[a]All estimates were adjusted for proportion of older adults aged 65 years or older, sex ratio, and natural cubic spline of log-transformed GDP per capita.

***Table 3:* Effect[a] of the demonstration programme on behavioural risk factors by batch and duration of implementation.**

and the synthetic control group, without requiring other testable statistical assumptions. Our methodological paper, using the outcome of past-month drinking and half of the CCDRFS sample as a case study, has shown that in a staggered implementation setting (in batches like this one), the SDID offers several advantages over DID or SC methods. Other studies have also found that the SDID method could improve the accuracy of policy effect estimates.[37]

Findings from our study demonstrated that the multi-pronged community-based programmes for NCD prevention and control are beneficial in promoting healthy behaviors in general. Compared to national or provincial-level NCD management programmes, the advantage of such programme lies in being controlled by area-level governments, with full departmental coordination. This promotes health integration into all policy implementations and continuously improves the policy environment for NCD prevention and control. Its effectiveness is worthy of rigorous evaluation to guide future evidence-based policy-making, implementation, and scaling up. We found that for environmental behaviour factors that the local government can systematically intervene in by restricting the sales and use,[38,39] such as smoking and drinking, this policy is effective; However, for individual behaviour factors, such as exercise and obesity, this programme has not shown consistent or large effect results with meaningful clinical implication within a short period, even if the demonstration areas provide guidance and support on healthy lifestyles. This result aligns with the characteristics of such programmes targeting environmental factor interventions. For the ineffectiveness of interventions targeting individual lifestyle factors, policymakers need to conduct an in-depth analysis of the causes and adjust relevant policies on this basis, such as providing more convenient public exercise venues, increasing investment in health education, developing personalized intervention tools, and encouraging the public to participate in exercise through economic incentives (e.g., fitness subsidies).

Our study has several limitations: 1) our analyses were constrained by the data available from CCDRFS. For the five waves from 2007 to 2018 (the most recent one conducted), the CCDRFS only contained 26 demonstration areas out of the 265 areas approved by then. Nevertheless, CCDRFS was nationally representative and the best data available for evaluation. We were also able to make use of data from 100 non-demonstration areas in the SDID analyses and evaluated short-term (1–2 years), medium-term (3–4 years), and long-term effects (6–7 years post-initiation); 2) the inclusion of evaluation indicators (six behavioural risk factors) relied on the variables available and compatible across multiple waves of CCDRFS, thus certain important indicators such as diet could not be evaluated. Instead, we included two physiologically measured indices (body mass index and waist circumference) closely related to both physical activity and diet in our evaluation; 3) The confidence intervals of our significant findings were relatively wide possibly due to the sample size, suggesting that our study might be underpowered to detect certain associations had they existed. To partially address this concern, we generated a composite measure – behavioural risk score to capture programme impacts more fully; 4) as the DID method, the SDID method also requires the "no anticipation" assumption, which means in our context no forward-looking areas reacted in advance of programme initiation.[18] Although the use of 2007 and 2010 as pre-implementation data did not suffer from this problem for the demonstration areas in the first two batches (2011–2012), the areas from the third batch (2014) might have reacted in advance after the official launch of the programme in 2011. However, since violating this assumption would under-estimate instead of over-estimate the treatment effect, it would make our results conservative and acceptable.

Future efforts should aim to systematically evaluate the impact of the programme using more temporal and spatial data on behavioural and other outcomes as they become available. A Cost-effectiveness analysis of the programme should be conducted to determine costs, budget impacts, and economic values. The implementation and evaluation of the NCD demonstration programme can provide evidence and valuable experience and lessons for future work in China as well as references and insight for other countries around the world facing similar challenges in NCD prevention and control.

Contributors
JW and LLY conceptualized the study and convened the full team of authors. WD, Xiao Z, and SL obtained the information on the National Integrated Demonstration Areas for Prevention and Control of NCDs programe and the data on CCDRFS. MZ and LW obtained the information and data for sensitivity analyses. YF and JQ collected data for the area-level covariates. WD, Xiao Z, SL, Xian Z, MG, HL, and LLY participated in weekly or biweekly meetings coordinated by HL and YY. Xiao Z, Xian Z, and ZL designed the study and conducted the analyses. WD, SL and MG provided important insights into the conceptual approach. WD, Xiao Z, and SL drafted the manuscript. Xian Z, MG, and LLY critically revised the manuscript. All authors read, edited, critically revised, and approved the final manuscript.

Data sharing statement
The data that support the findings of this study were used under license from the Chinese Centre for Disease Prevention and Control, and thus are not publicly available. However, these data may be available upon request when permission from the Chinese Centre for Disease Prevention and Control can be obtained.

Editor note
The Lancet Group takes a neutral position with respect to territorial claims in published maps and institutional affiliations.

Declaration of interests
We declare no competing interests.

Acknowledgements
We thank the research staffs from 31 provincial Centers for Disease Control and Prevention and 298 local Centers for Disease Control and Prevention for their collection of data and their efforts in quality control.

Appendix A. Supplementary data
Supplementary data related to this article can be found at https://doi.org/10.1016/j.lanwpc.2024.101167.

References
1 GBD 2017 Causes of Death Collaborators. Global, regional, and national age-sex-specific mortality for 282 causes of death in 195 countries and territories, 1980–2017: a systematic analysis for the global burden of disease study 2017. *Lancet*. 2018;392:1736–1788.
2 GBD 2019 Risk Factors Collaborators. Global burden of 87 risk factors in 204 countries and territories, 1990-2019: a systematic analysis for the global burden of disease study 2019. *Lancet*. 2020;396(10258):1223–1249.
3 Organization WH. United Nations high-level meeting on non-communicable disease prevention and control - NCD summit to shape the international agenda. https://www.who.int/news-room/events/detail/2011/09/19/default-calendar/united-nations-high-level-meeting-on-noncommunicable-disease-prevention-and-control—ncd-summit-to-shape-the-international-agenda; 2011. Accessed July 13, 2023.
4 Nations U. Political declaration of the high-level meeting of the general assembly on the prevention and control of non-communicable diseases 2012. https://undocs.org/zh/A/RES/66/2. Accessed November 21, 2023.
5 Programme UND. 'Whole of society' approach needed to tackle global epidemic of noncommunicable diseases. https://www.undp.org/asia-pacific/news/%E2%80%98whole-society%E2%80%99-approach-needed-tackle-global-epidemic-noncommunicable-diseases; 2019. Accessed November 21, 2023.
6 Peng W, Chen S, Chen X, et al. Trends in major non-communicable diseases and related risk factors in China 2002-2019: an analysis of nationally representative survey data. *Lancet Reg Health West Pac*. 2023;43:100809.
7 China TSCotPsRo. Notice of the general office of the ministry of health on the issuance of the management measures of the demonstration areas for comprehensive prevention and control of chronic non-communicable diseases. https://www.gov.cn/zwgk/2011-03/24/content_1830841.htm; 2011. Accessed June 9, 2023.
8 Arkhangelsky D, Athey S, Hirshberg DA, Imbens GW, Wager S. Synthetic difference-in-differences. *Am Econ Rev*. 2021;111(12):4088–4118.
9 Aragón MJ, Chalkley M, Kreif N. The long-run effects of diagnosis related group payment on hospital lengths of stay in a publicly funded health care system: evidence from 15 years of micro data. *Health Econ*. 2022;31(6):956–972.
10 Essays in applied economics linking policy, obesity, and health economics. https://ttu-ir.tdl.org/items/26f0fdef-3134-4cd2-9ce5-c9d0fcf4b6c0.
11 Caspi CE, Gombi-Vaca MF, Wolfson J, et al. Early results of a natural experiment evaluating the effects of a local minimum wage policy on the diet-related health of low-wage workers, 2018–2020. *Public Health Nutr*. 2023;26(11):2573–2585.
12 *A healthy start for healthy start vouchers? By hamideh mohtashami Borzadaran*and johannes Lohse†and emma frew‡*. 2023.
13 *Circular of the general office of the national health and family planning commission on the issuance of the administrative measures for the construction of national chronic disease comprehensive prevention and control demonstration areas*; 2016. http://www.nhc.gov.cn/jkj/s5878/201611/6d55c194a965460b9bc7ee9cb5cb4592.shtml. Accessed June 2, 2024.
14 Dong JMF, Dong W, Jiang Y, Liu S, Zhou M. *Creating national demonstration areas for the integrated prevention and control of*

noncommunicable diseases in China. Washington: Brookings Institution Press; 2020.
15 China NHCotPsRo. The workbook for assessment and evaluation of national integrated demonstration areas for non-communicable disease (NCD). https://www.gov.cn/xinwen/2016-11/21/content_5135074.htm; 2016. Accessed November 21, 2023.
16 Li Y, Zhang M, Jiang Y, Wu F. Co-variations and clustering of chronic disease behavioural risk factors in China: China chronic disease and risk factor surveillance, 2007. *PLoS One*. 2012;7(3):e33881.
17 Zhang M, Wang L, Wu J, et al. Data resource profile: China chronic disease and risk factor surveillance (CCDRFS). *Int J Epidemiol*. 2022;51(2):e1–e8.
18 Kranz S. Synthetic difference-in-differences with time-varying covariates: technical report. Available online at: https://github.com/skranz/xsynthdid; 2022.
19 Organization WH. Framework for action to implement the united Nations political declaration on noncommunicable diseases. https://www.emro.who.int/noncommunicable-diseases/ncd-infocus/declaration-ncds.html; 2011.
20 McLaren L, Ghali LM, Lorenzetti D, Rock M. Out of context? Translating evidence from the North Karelia project over place and time. *Health Educ Res*. 2007;22(3):414–424.
21 Goodchild M, Zheng R. Tobacco control and Healthy China 2030. *Tob Control*. 2019;28(4):409–413.
22 Rehm J, Gmel GE Sr, Gmel G, et al. The relationship between different dimensions of alcohol use and the burden of disease-an update. *Addiction*. 2017;112(6):968–1001.
23 Walsh B, Grant M. Public health implications of alcohol production and trade. *WHO Offset Publ*. 1985;88:1–55.
24 Li Y, Liu S, Wang L, Zhou M. Burden of disease attributable to main behavioural risk factor of chronic disease inactivity in China, 1990 and 2010. *Zhonghua Yu Fang Yi Xue Za Zhi*. 2015;49(4):303–308.
25 Evaluation IfHMa. Global health data exchange. GBD results tool. http://ghdx.healthdata.org/gbd-resultstool.
26 Wang L, Zhou B, Zhao Z, et al. Body-mass index and obesity in urban and rural China: findings from consecutive nationally representative surveys during 2004-18. *Lancet*. 2021;398(10294):53–63.
27 Wang Y, Zhao L, Gao L, Pan A, Xue H. Health policy and public health implications of obesity in China. *Lancet Diabetes Endocrinol*. 2021;9(7):446–461.
28 Li JJ, Li JL, Zhang J, et al. Study on the overall implementation status of the national demonstration areas for comprehensive prevention and control of non-communicable diseases. *Zhonghua Liu Xing Bing Xue Za Zhi*. 2018;39(4):417–421.
29 Zhang J, Jin RR, Li JJ, et al. Study on the effectiveness of implementation: the national demonstration areas for comprehensive prevention and control of non-communicable diseases. *Zhonghua Liu Xing Bing Xue Za Zhi*. 2018;39(4):394–400.
30 Jiang Y, Kong LZ, Li LM. Implementing the strategy of 'healthy China' and strengthening the setting-up of national demonstration areas, for comprehensive prevention and control of non-communicable diseases. *Zhonghua Liu Xing Bing Xue Za Zhi*. 2018;39(4):391–393.
31 Li JJ, Jin RR, Zhang J, et al. Study on consumption of vegetables and fruits and related influencing factors among residents from the national demonstration areas for comprehensive prevention and control of non-communicable diseases. *Zhonghua Liu Xing Bing Xue Za Zhi*. 2018;39(4):412–416.
32 Jin RR, Zhang J, Li JL, et al. Current status on prevalence, treatment and management of hypertension among Chinese adults in the national demonstration areas for comprehensive prevention and control of non-communicable diseases. *Zhonghua Liu Xing Bing Xue Za Zhi*. 2018;39(4):401–406.
33 Jin RR, Li JJ, Zhang J, et al. Management programs on diabetes among Chinese adults in the national demonstration areas for comprehensive prevention and control of non-communicable diseases. *Zhonghua Liu Xing Bing Xue Za Zhi*. 2018;39(4):407–411.
34 Abadie A. Semiparametric difference-in-differences estimators. *Rev Econ Stud*. 2005;72(1):1–19.
35 Imbens GW, Wooldridge JM. Recent developments in the econometrics of program evaluation. *J Econ Lit*. 2009;47(1):5–86.
36 Bhalotra S, Clarke D, Gomes JF, Venkataramani A. Maternal mortality and women's political power. *J Eur Econ Assoc*. 2023;21(5):2172–2208.
37 de Chaisemartin C, D'Haultfœuille X. Two-way fixed effects and differences-in-differences with heterogeneous treatment effects: a survey. The econometrics journal 2022; 26(3): C1-C30.42. abadie A. Using synthetic controls: feasibility, data requirements, and methodological aspects. *J Econ Lit*. 2021;59(2):391–425.
38 Regulations on the implementation of the tobacco monopoly law of the people's republic of China. https://flk.npc.gov.cn/detail2.html; 2021. Accessed May 24, 2024.
39 Advertising law of the people's republic of China. https://flk.npc.gov.cn/detail2.html; 2021. Accessed May 24, 2024.