From Both Sides, But Not for All: Heterogeneous Effects of a Nationwide Mental Health Policy in China

Jan 2026

Abstract

We evaluate a nationwide policy in China combining mental healthcare supply expansion with stigma reduction. Using a difference-in-differences design, we find significant average improvements in mental health. However, benefits are concentrated among low- and medium-risk population, with no detectable gains for the high-risk. We attribute this heterogeneity to an accessibility-privacy trade-off: community-based care expansion reduces travel time but increases social visibility that disproportionately deter high-risk individuals. A discrete choice experiment confirms that high-risk individuals respond weakly to accessibility improvements due to privacy concerns. We conclude that accessibility-focused interventions can inadvertently exclude the vulnerable, necessitating privacy-protected delivery channels for high-risk groups.

Keywords: Mental Health; Difference-in-Difference; Discrete Choice Experiment; Policy Evaluation

JEL Codes: I12; I18; I38; O15

1 Introduction

Mental health disorders affect nearly one billion people worldwide, yet treatment gaps remain even in high-income countries (World Health Organization, 2025; Patel et al., 2018). In response, governments are increasingly adopting integrated interventions that work from both sides of the market: expanding service supply while simultaneously reducing demand-side barriers through stigma reduction campaigns. Recent examples include India's National Mental Health Policy and Singapore's National Mental Health and Well-being Strategy, both combining infrastructure investment with public awareness initiatives. Whether this two-pronged approach effectively improves population mental health, and whether its benefits reach those most in need, remains an open empirical question with direct policy implications.

We address this gap by evaluating China's National Social Psychological Service System Construction Pilot Work Plan (hereafter, the Plan), launched in November 2018 by the National Health Commission and nine other agencies. The Plan exemplifies the both-sides strategy: designated pilot cities received additional financial and institutional support over three years (2019–2021) to expand mental healthcare providers, while local governments organized media campaigns and public advocacy to reduce stigma and inform citizens of available services. We focus on two core policy questions. First, does this integrated intervention improve average mental health outcomes in treated cities? Second, are the benefits distributed equitably across the mental health distribution, or do they bypass high-risk individuals most in need of care?

Empirically, we employ a canonical Difference-in-Differences (DID) design with nationally representative data from the China Family Panel Studies (CFPS), comparing the mental health trajectories in treated (pilot) cities and control (regular) cities.[1] Our findings reveal that the policy significantly improved mental health in treated cities, evi-

[1] As we show in Table 3, using data from the China City Statistical Yearbook, we find that treated cities are significantly better off in terms of GDP per capita and baseline mental health resources compared to their control counterparts. This confirms the findings in Wang and Yang (2025) that policy experimentation is not random in China. We address selection bias by controlling for interactions between time dummies and baseline city characteristics in section 6, yielding consistent results.

denced by a 0.068 standard deviation reduction in standardized mental health scores[2] and a 4 percentage point decrease in the probability of being at high risk of severe disorder.

However, this average effect conceals a striking policy failure: all improvements concentrate among individuals with relatively better mental health, while high-risk individuals—the policy's stated priority—experience no detectable benefit. This distributional pattern contradicts the intervention's explicit targeting objective and raises a critical question: why does improved accessibility bypass those most in need?

Understanding this distributional failure requires examining the institutional details of the supply expansion. Pilot cities experienced significant growth in community-based, non-medical counseling organizations—which offer high accessibility but limited anonymity—while showing no change in specialized psychiatric hospitals. This compositional shift creates a fundamental trade-off: community-based providers reduce travel costs but increase visibility within local social networks. In stigmatized contexts, this trade-off may differentially affect individuals across the mental health distribution, potentially explaining why accessibility improvements failed to reach the most vulnerable.

Motivated by this observation, we consider two potential mechanisms to explain this failure. The first is a structural mismatch in supply expansion: expanding counseling organizations without expanding psychiatric hospitals may fail to serve individuals requiring medication-based treatment. However, this channel offers only a partial explanation. Severe psychotic disorders requiring exclusive pharmacological management have a relatively low prevalence in the general population in China (Huang et al., 2019), yet our high-risk sample, identified via psychometric screening, likely consists predominantly of individuals with common mental disorders for which counseling is clinically effective (Singla et al., 2017). This compositional reality points to a second, more fundamental mechanism.

We then propose and test a time-privacy trade-off mechanism rooted in the privacy costs of community-based care. Our conceptual framework models care-seeking as dependent on three costs: monetary, time, and psychological. A key insight emerges: proximity

[2] Mental health scores are constructed based on Kessler-6 and CES-D scales. A higher score indicates a worse mental health condition.

to mental health services generates a dual effect. While closer providers reduce commuting time, they simultaneously increase the risk of being recognized by acquaintances, raising privacy costs in stigmatized contexts. For vulnerable individuals with severe mental health needs, this privacy penalty can outweigh convenience gains, resulting in null effects despite expanded supply. Conversely, individuals with milder conditions, who face lower psychological barriers and less fear of social judgment, capture the benefits of improved accessibility.

To test this time-privacy trade-off mechanism independently of the policy context, we conduct a Discrete Choice Experiment (DCE) eliciting preferences over service attributes (price, commuting time, stigma level).[3] The DCE reveals striking heterogeneity: individuals with poorer baseline mental health exhibit high price sensitivity but low responsiveness to reduced commuting time. This is consistent with our dual-effect framework: for vulnerable populations, extreme proximity heightens privacy concerns, making them less sensitive to travel distance reductions. In contrast, low-risk individuals respond strongly to accessibility improvements. This pattern provides robust, independent evidence that the policy's emphasis on visible, community-based expansion inadvertently yields muted gains for the most vulnerable by imposing uncompensated privacy costs.[4]

Our study contributes to the literature in three distinct ways. First, we provide causal, population-level evidence on the distributional effects of a large-scale mental health intervention in a developing country. Existing research on mental healthcare access focuses primarily on targeted groups in developed economies (Costantini, 2024; Shafer et al., 2024) or localized experiments (Baranov et al., 2020; Barker et al., 2022; Vlassopoulos et al., 2024), limiting its generalizability. While prior work documents that healthcare infrastructure expansion increases utilization and reduces risk (Feyman et al., 2023; Ding, 2023; Harrell et al., 2023; Dias and Fontes, 2024; Cassidy et al., 2025), evidence on who benefits from broad supply expansions remains scarce. We fill this gap by leveraging a nationwide

[3]In the DCE, commuting time is expressed in minutes, with longer times indicating reduced accessibility. Stigma is represented by the percentage of individuals in society holding discriminatory attitudes toward those with mental illness.

[4]To address sample composition concerns, we employ entropy balancing to reweight CFPS data to match DCE demographics (section 6), confirming consistent estimates.

policy variation and documenting a critical distributional failure: accessibility-focused interventions can inadvertently exclude the most vulnerable, potentially exacerbating inequalities in mental healthcare utilization.

Second, we contribute to understanding demand-side barriers in stigmatized healthcare by proposing and experimentally validating a dual effect of proximity. Prior research has separately analyzed financial barriers through insurance expansions (Finkelstein et al., 2012; Andersen, 2015; Solomon and Dasgupta, 2022) and non-monetary barriers such as stigma and wealth (Angelucci and Bennett, 2024; Schwandt, 2018; Thornicroft, 2008; Bharadwaj et al., 2017; Brewer et al., 2024). We advance this literature by demonstrating that accessibility is not merely a function of time costs but fundamentally incorporates psychological privacy costs. Through our DCE, we show that for vulnerable individuals, proximity acts as a double-edged sword—a finding with broad implications for policy design in other stigmatized services (e.g., HIV testing, addiction treatment, reproductive health).

Third, we identify policy design lessons for mental health interventions. Our findings reveal limitations of one-size-fits-all accessibility strategies. While integrated interventions addressing both supply and demand barriers are essential complements to insurance reforms, accessibility-focused designs can deliver limited benefits for the most vulnerable. Our results suggest a two-pronged strategy: enhance physical accessibility for low-risk, convenience-sensitive populations through community-based care, while providing privacy-protected, specialized services for high-risk, privacy-sensitive populations, potentially through spatially or institutionally separated facilities. Ultimately, the optimal allocation of resources between physical accessibility and specialized care depends on the specific policy objective.

The remainder of this paper proceeds as follows. section 2 describes China's mental healthcare system and the policy intervention. section 3 presents the data and section 4 reports the DID results. section 5 develops the theoretical framework and presents the DCE design and findings. section 6 provides robustness checks. section 7 concludes with policy implications.

2 Institutional Background

In November 2018, China's National Health Commission, in collaboration with the Central Political and Legal Affairs Commission and eight other central departments,[5] launched the "National Social Psychological Service System Construction Pilot Work Plan" (hereafter, the Plan). This three-year initiative (2019–2021) was designed to establish and enhance the national mental health service framework in response to the growing recognition of mental health importance and the persistent shortage of available resources.

Consistent with many national policy experiments in China that foster innovation through regional pilots, the Plan adopts an incentivized framework through the strategic designation of pilot cities.[6] Applicant city governments were required to submit comprehensive work plans detailing their implementation goals and measures. Selected pilot cities are expected to lead in the enhancement of public mental health systems and serve as regional role models. Two cohorts of pilot cities were announced: 50 cities commenced implementation in 2019, with Wuhan city in Hubei province joining in 2020.[7]

To facilitate achievement of the Plan's objectives, pilot cities receive support through a two-tiered framework. The first tier provides financial and institutional support from city government agencies to alleviate resource constraints and streamline implementation. The second tier involves strategic oversight and expert guidance from provincial and central authorities to ensure alignment with national goals and the dissemination of effective practices. The Plan operates through two distinct channels that target both supply-side and demand-side barriers to mental healthcare access. First, on the supply

[5] The other eight departments are: Publicity Department of the CPC Central Committee, Ministry of Education, Ministry of Public Security, Ministry of Civil Affairs, Ministry of Justice, Ministry of Finance, State Bureau of Letters and Calls, China Disabled Persons' Federation.

[6] There are no specific, publicly available criteria showing how pilot cities were selected. However, as is common in policy experimentation in China, treated (pilot) cities are often better across various dimensions at baseline, such as GDP, a phenomenon documented by Wang and Yang (2025). As shown in Table 3, we confirm this pattern in our sample. We specifically address this issue of non-random selection of treated cities in section 6.

[7] We present the geographical distribution of pilot and regular cities across China in Supplementary Materials. The distribution of pilot cities is purposefully dispersed and non-clustered across various regions, encompassing not only economically developed Eastern and Central provinces (e.g., surrounding the Bohai Sea and the Yangtze River Delta) but also Western and Northeastern provinces (e.g., Xinjiang, Tibet, and Inner Mongolia). For further details regarding the policy's specific directives and the complete list of pilot cities, the translated policy document is provided in Supplementary Materials. In our analysis, we exclude Wuhan city as it was severely impacted by COVID-19 in 2020, making it a potential outlier.

side, the Plan significantly increased service availability in pilot cities. This expansion is most visible in expanded accessibility of mental healthcare resources through community-based service providers.[8] For instance, a news report from Longyan city in Fujian province highlights that mental healthcare services have become widely accessible to residents, effectively addressing community mental health needs.[9]

Second, on the demand side, the Plan targets a crucial barrier to utilization: stigma associated with mental illness. A government directive issued in 2019 mandated that pilot city governments implement public awareness campaigns strategically designed to inform the public about service availability and actively reduce stigma to foster a more supportive environment for help-seeking.[10] Local governments were encouraged to utilize diverse media channels—including radio, television, and websites—to maximize outreach. For example, Longyan City, an early pilot, organized over 400 activities to promote mental healthcare and educate the public on appropriate service access, aiming to correct stigmatized attitudes toward mental illness.[11] These campaigns represent a deliberate effort to shift public perceptions and encourage timely help-seeking behavior, complementing the supply-side expansion with demand-side interventions.

3 Empirical Strategy

3.1 Data and Measurement

To evaluate the Plan's impact, we use two primary data sources. The first dataset comes from official policy documents, which document the list of designated pilot cities and their respective implementation timelines. The Plan was officially announced in November 2018, with implementation spanning 2019 to 2021. Specifically, 50 pilot cities began implementation in 2019, and Wuhan was the sole additional city to join the pilot program

[8] Empirically, as shown in Figure 3, the Plan primarily increased the supply of general counseling organizations (Panel b). Conversely, the Plan had no significant effect on the supply of professional psychiatric hospitals (Panel a). This pattern indicates that the policy-induced market response originated mainly from easily established, non-medical providers who face significantly lower regulatory and financial entry barriers compared to medical institutions.

[9] Source: https://m.thepaper.cn/newsDetail_forward_30122067

[10] Source: https://ncmhc.org.cn/channel/newsinfo/6263

[11] Source: https://m.thepaper.cn/newsDetail_forward_8902694

in 2020. The second, and central, dataset comprises individual and household-level panel data from the China Family Panel Studies (CFPS). CFPS is a nationally representative, biennial longitudinal survey initiated in 2010 by the Institute of Social Science Survey (ISSS) at Peking University (Xie and Hu, 2014). This comprehensive survey provides rich information on respondents' demographics, physical and mental health conditions, social ideologies, and crucial family-related characteristics such as family size and household income.

For mental health measurement, CFPS employs different scales across survey waves. The 6-item Kessler (K6) scale, a well-established and validated instrument for assessing psychological distress over the past 30 days, is used in the first and third waves (2010 and 2014). The K6 scale comprises six questions addressing feelings of nervousness, hopelessness, restlessness, depression, effort, and worthlessness. Each item is scored on a 5-point Likert scale, ranging from "never" (0 points) to "almost every day" (4 points), yielding a total score from 0 to 24, where higher scores indicate greater psychological distress. For the remaining five waves (2012, 2016, 2018, 2020, and 2022), CFPS uses the Center for Epidemiologic Studies Depression Scale (CES-D), which assesses various depressive symptoms experienced in the past week. Notably, CFPS employed two versions of the CES-D: the 20-item version in 2012 and 2016, and a condensed 8-item version for the 2018, 2020, and 2022 waves. In both versions, respondents provide frequency responses on a 4-point scale: "not at all" (0 points), "a little" (1 point), "some" (2 points), and "a lot" (3 points). To ensure consistency, items phrased to reflect positive affect or behavior are reverse-scored. Consequently, the total score for the CES-D-20 ranges from 0 to 60, while the CES-D-8 ranges from 0 to 24. The specific questions for both the K6 and CES-D scales are shown in **??**.

To ensure comparability across survey waves using different scales, we standardize all mental health scores obtained from the K6, CES-D-20, and CES-D-8 scales, following Luo et al. (2023). Specifically, we normalize raw scores to a mean of zero and a standard deviation of one within each survey wave. Consequently, these standardized metrics capture an individual's relative position within the distribution of symptom severity, where higher values denote worse mental health outcomes. To capture the prevalence of clinical distress, we complement the continuous measure with a binary indicator for "caseness." We define caseness as a binary indicator for individuals scoring above established thresholds,[12] signifying a level of symptom severity that clinically warrants further professional investigation. Table 1 presents descriptive statistics for these key variables.[13]

Table 1: Summary statistics

Variable	Obs	Mean	Std. Dev.	Min	Max
Mental health					
Standardized mental health score	127,529	0.000	1.000	-1.363	6.617
Caseness	127,529	0.189	0.391	0	1
Ideological outcomes					
Relative income (0-5)	122,711	2.555	1.056	1	5
Relative status (0-5)	127,021	2.923	1.047	1	5
Physical health					
Self-rated health status	127,524	3.156	1.284	1	5
Self-rated health status relative to the last year	127,528	1.775	0.616	1	3
Addictive behaviors					
Smoked in the last month	127,528	0.308	0.462	0	1
Drank 3 times a week last month	127,522	0.157	0.363	0	1
Demographics					
Birth year	127,529	1,966	14.42	1,930	1,992
Gender (male = 1)	127,529	0.498	0.500	0	1
Urban residence	127,529	0.481	0.500	0	1
Employment status (employed = 1)	127,529	0.728	0.445	0	1
Years of education	127,529	7.610	4.951	0	19
Married	127,529	0.871	0.335	0	1
Family size	127,529	4.259	1.927	1	26
Low income households	127,529	0.314	0.464	0	1
Middle income households	127,529	0.383	0.486	0	1
High income households	127,529	0.303	0.460	0	1

Note: This table reports the summary statistics (observation count, mean, standard deviation, minimum, and maximum) for the main variables used in the analysis.

3.2 Identification

To quantify the causal effects of the Plan on individual mental health, we leverage across-city variation in exposure to the Plan within a canonical difference-in-differences framework. While the Plan was officially initiated in June 2019, we account for the inherent policy implementation lag—the time required for increased resources to be in place and for new mental healthcare supply to become accessible. Therefore, we conservatively define the effective onset of treatment as 2020. This strategic choice of treatment timing,

[12] Cutoffs are defined as follows: score of 12 or higher for the K6 (Kessler et al., 2003), 16 or higher for the CES-D-20 (Radloff, 1977), and 7 or higher for the CES-D-8 (Bi et al., 2023).

[13] To ensure identification, we restrict the analytic sample to individuals who remained in their original city throughout the study period, excluding observations associated with cross-city migration.

coupled with the 2-year observational gap in the CFPS dataset, positions us to assess the immediate and early-stage consequences of the Plan following its implementation. Our primary focus is to assess the overall impact of the Plan on residents in the treated cities,[14] and to investigate whether the Plan disproportionately benefited those most in need of help. Empirically, we estimate the following baseline specification to measure the Plan's overall effects on the outcomes of interest:

$$y_{ict} = \alpha + \beta Plan_{ct} + \gamma_i + \lambda_t + \eta Envi_{ct} + \epsilon_{ict} \quad (1)$$

where y_{ict} is the outcome of interest for individual i living in city c at time t.[15] γ_i denotes individual fixed effects and λ_t denotes year fixed effects. $Plan_{ct}$ is an indicator variable that takes the value of 1 if city c started to implement the Plan at survey wave t. $Envi'_{ct}$ includes a set of environmental factors of city c at time t that might influence residents.[16] ϵ_{ict} is the error term. We estimate Equation 1 using OLS with standard errors clustered at the city level.

The identifying assumption in our setup is that, in the absence of the Plan, the mental health conditions of residents in treated and control cities would have evolved similarly. When no pre-trend is found, i.e., the common trend assumption holds, the coefficient β captures the intention-to-treat effect (ITT) of the Plan on residents' mental health. To test the common trend assumption and examine the dynamics of the Plan's effects, we estimate the following event study specification:

$$y_{ict} = \alpha + \sum_{n=-10}^{2} \beta_n Plan^n_{ct} + \gamma_i + \lambda_t + \eta Envi_{ct} + \epsilon_{ict} \quad (2)$$

$Plan^n_{ct}$ is a set of indicator variables that take the value of 1 if for treated city c in survey wave t, the Plan was n years away. The coefficients β_n capture the dynamic

[14] We refer to pilot cities as treated cities and regular cities as control cities.

[15] When y_{ict} is a dummy variable indicating whether the individual has high risk of caseness or not, we estimate a linear probability model.

[16] Environmental controls are included as critical determinants of individual mental health (Chen et al., 2024); furthermore, these factors are plausibly exogenous to the implementation of the Plan. Environmental controls include average PM2.5 concentration per year ($\mu g/m^3$), surface layer height (m), surface pressure (Pa), surface specific humidity (g/g), surface wind speed (m/s), and surface air temperature ($Kelvin$).

differences in mental health outcomes between pilot and control cities n years relative to the policy implementation, compared to the reference period. 2018 serves as the reference year in our identification. Standard errors are clustered at the city level.

Another objective of this study is to examine whether the Plan disproportionately benefited those most in need of help. We define "people most in need of help" based on the concept of endogenous stratification, a frequently used method to explore whether the policy helped those who would otherwise obtain the worst outcomes in the absence of treatment. We then divide the sample into three subgroups based on these predicted outcomes (low, middle, and high risk of caseness) and estimate the Plan's impact on the outcomes of interest for the three subgroups, respectively. Following Abadie et al. (2018), we estimate the predicted probability of an individual having severe mental disorder in the absence of the Plan, using the leave-one-out (LOO) approach with 100 bootstrap repetitions to predict the probability of caseness without the treatment.[17]

4 Results

This section presents the causal impact of the Plan on mental health outcomes. We begin by estimating the overall treatment effect on the full sample, followed by an assessment of heterogeneous effects across subgroups stratified by predicted probability of severe mental disorder. We then examine potential downstream impacts on physical health and addictive behaviors.

4.1 Overall impact on mental health

Table 2 presents the estimated average treatment effect of the Plan on mental health outcomes for the full sample. Column (1) shows that the Plan leads to a statistically significant reduction of 0.076 standard deviations in the mental health score. This effect remains robust when controlling for environmental factors in Column (2), with a slightly

[17] LOO approach generates a predicted outcome for each observation i using coefficients estimated from the control group excluding observation i. The set of predictors includes: birth year, gender, type of residence (urban/rural), employment status, years of education, marital status, self-rated health status, family size, and family income level. We include more details of this process in Supplementary Materials.

Table 2: Effects of the Plan on mental health: Full sample

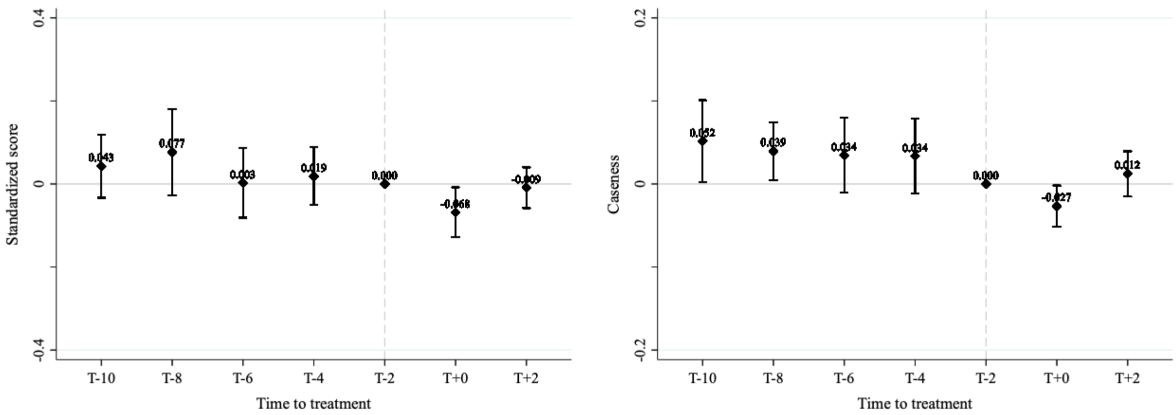
	Standardized score		High risk of caseness	
	(1)	(2)	(3)	(4)
Treat × Post	-0.076***	-0.068**	-0.051***	-0.040***
	(0.028)	(0.031)	(0.016)	(0.015)
Observations	127,529	127,529	127,529	127,529
R-squared	0.548	0.549	0.474	0.475
Individual FE	YES	YES	YES	YES
Survey Wave FE	YES	YES	YES	YES
Environmental Control	NO	YES	NO	YES
Control mean	0.032		0.199	
Std Dev	1.013		0.399	

Notes: Standard errors are clustered at city level and are reported in parentheses. The table presents DID estimation results on the effects of the Plan. The dependent variables are the standardized mental health score (columns 1 and 2) and a dummy variable indicating a high risk of severe mental disorder (columns 3 and 4). All regressions include individual and survey wave fixed effects, as indicated. Columns (2) and (4) include a set of environmental control variables. Standardized scores for mental health (based on Kessler 6 and CESD scales) are normalized to a mean of zero and a standard deviation of one by survey year. The reported Control mean and SD reflect the average and dispersion, respectively, of the control group across the entire sample period. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

attenuated magnitude of 0.068 standard deviations. Columns (3) and (4) demonstrate that the Plan significantly reduces the probability of caseness by approximately 4 percentage points, with small reduction in magnitude after controlling for environmental covariates.

To assess the validity of our identification strategy, we estimate the dynamic treatment effects using the event study specification in Equation 2. Figure 1 presents the event study coefficients for both mental health outcomes. The pre-treatment coefficients are consistently close to zero and statistically insignificant for both outcomes, providing strong support for the parallel trends assumption. This indicates that, in the absence of the Plan, mental health trajectories in piloting cities would have evolved similarly to those in control cities.

Upon policy initiation at $T+0$, we observe an immediate and statistically significant reduction in both the standardized mental health score and the probability of severe mental disorder. However, this initial improvement does not persist over time. At $T+2$, the point estimates for both outcomes turn positive and lose statistical significance,



(a) Standardized score

(b) Probability of caseness

Figure 1: Dynamic effects of the Plan on mental health: Full sample

Note: The figure presents the estimated coefficients (β_k) from an event study model derived from Equation 2, with 95% confidence intervals based on standard errors clustered at the city level. . The coefficients show the effect of the Plan in event time k, relative to the year before implementation (the reference period, $k = -2$). The panel title represents the outcome of interest in the regression.

suggesting that the positive effects of the Plan on mental health may fade within the limited observation window of our analysis.[18]

To contextualize the magnitude of these estimates, we note that the reduction of 0.068 standard deviations in the standardized mental health score corresponds to a Cohen's d of approximately 0.07. Most notably, our observed 4 percentage point reduction in the probability of caseness constitutes a 20% decrease relative to the control mean. This demonstrates that despite a modest shift in the average score, the intervention generates a reduction in caseness outcome that is comparable in relative terms to major health policy expansions.[19]

[18] It is important to highlight that although the Plan officially began in mid-2019, we classify all individuals as treated starting from 2020 due to data availability constraints and the typical time lag in policy implementation. Consequently, when interpreting the "immediate impact," one should consider this temporal offset. Though it is evident that any observable effects are relatively short-lived, a straightforward two-tailed t-test reveals that lockdown stringency in 2022, as measured by the Baidu Migration Index (we include more details of this index in section 6), was significantly higher in cities that received the treatment. This finding suggests that the temporary impact observed could be attributed to the more stringent lockdown measures implemented in these treated cities.

[19] For instance, Finkelstein et al. (2012) document that Medicaid coverage in the Oregon Health Insurance Experiment led to broad improvements in measures of self-reported physical and mental health, with effect sizes averaging 0.2 standard deviations. Notably, the study finds that Medicaid coverage resulted in a 32% increase in self-reported overall happiness.

4.2 Heterogeneous treatment effects

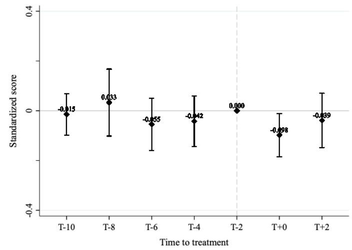
A central question is whether the Plan effectively reached individuals most in need of mental health support. To investigate this, we estimate heterogeneous treatment effects across subgroups stratified by predicted probability of caseness. Figure 2 presents event study estimates for three groups: low, middle, and high risk of caseness groups.

The results reveal a striking pattern of heterogeneity. For both the low and middle risk of caseness groups, the pre-treatment coefficients are close to zero and statistically insignificant, supporting the parallel trends assumption. At policy initiation ($T + 0$), both groups experience statistically significant reductions in mental health scores and in the probability of severe mental disorder. Specifically, for the low risk of caseness group, the Plan reduces the standardized mental health score by approximately 0.098 (Cohen's d $\approx$ 0.122) and the probability of severe disorder by 3.7 percentage points (representing roughly 50% of the control group mean). For the middle risk of caseness group, the corresponding reductions are 0.095 in the standardized score (Cohen's d $\approx$ 0.105) and 5 percentage points in the probability of severe disorder (approximately 25.68% of the control group mean).

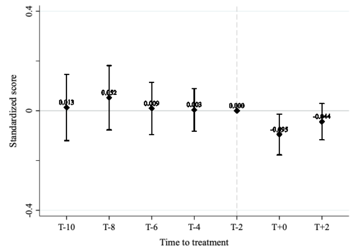
However, these initial improvements are not sustained over time. By $T + 2$, the point estimates remain negative but lose statistical significance and diminish in magnitude, indicating a gradual fading of the treatment effect. This temporal pattern is consistent with the full-sample findings and suggests that the Plan's beneficial impact on mental health may attenuate within approximately two years post-implementation.

In contrast, the high risk of caseness group, comprising individuals who would otherwise obtain the worse outcomes without treatment, exhibits neither statistically significant nor economically meaningful treatment effects. For this subgroup, the point estimate at $T + 0$ for the probability of severe disorder is a negligible -0.015, and the coefficient for the standardized mental health score, while negative, is also statistically insignificant. This finding suggests that the Plan, in its current implementation, did not effectively reach or substantially benefit the most vulnerable individuals.[20]

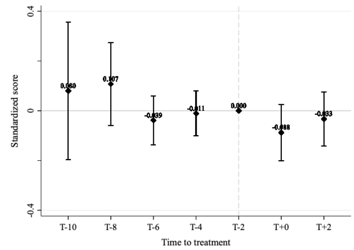
[20] We explicitly consider whether this null result stems from the pandemic disproportionately exac-



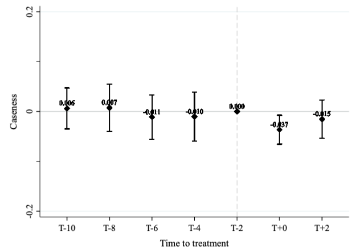
(a) Low risk of caseness



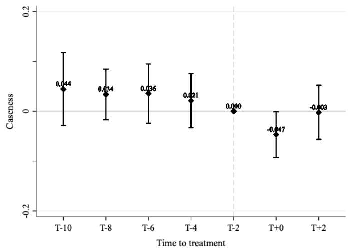
(b) Middle risk of caseness



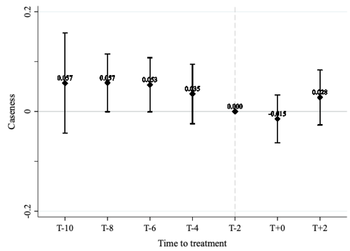
(c) High risk of caseness



(d) Low risk of caseness



(e) Middle risk of caseness



(f) High risk of caseness

Figure 2: Heterogeneous treatment effects: by risk of caseness

Notes: The figure presents the estimated coefficients (β_k) for low to high risk of caseness groups from an event study model derived from Equation 2, with 95% confidence intervals based on standard errors clustered at the city level. The title on the left side of each plot in each panel represents the outcome of interest in the regression. The coefficients show the effect of the Plan on the mental health outcome in event time k, relative to the reference year (the reference period, $k = -2$). The panel title represents the corresponding subgroup created by LOO.

4.3 Downstream impacts on physical health and addictive behaviors

Given the well-documented interrelationships between mental health, physical health, and addictive behaviors, we examine whether the Plan's effects on mental health translate into improvements in downstream outcomes. We focus on two sets of outcomes: (1) self-rated physical health measures, including self-rated health status and self-rated health relative to the previous year; and (2) addictive behaviors, specifically smoking and alcohol consumption, which are often used as coping mechanisms for psychological distress.

As reported in Table 1, we find that the Plan slightly reduced the probability of having smoked in the last month on average. Although the magnitude is relatively small at approximately 1 percentage point, this result suggests a beneficial spillover effect from the mental health intervention into the domain of behavioral health risk factors. We do not observe statistically significant effects on other physical health or drinking outcomes, which may reflect either insufficient statistical power, or longer lag times required for these outcomes to respond.

In summary, we find that the Plan, on average, improved population mental well-being and generated modest spillover benefits in reducing smoking behaviors. The policy produced significant and immediate improvements in mental health outcomes for the general population and for individuals at low to moderate risk of severe mental disorder. However, the Plan appears to have fallen short in adequately addressing the needs of the most vulnerable individuals. Specifically, we observe no statistically significant or economically meaningful impacts on the mental health outcomes of individuals with a high predicted probability of severe mental disorder. This suggests that a broad, two-sided intervention—while effective for a large segment of the population—may not constitute a sufficient solution for those most in need of help. For these individuals, improved well-being may require more sustained, intensive, and clinically tailored support. Furthermore, while the Plan's positive effects on the low and middle-risk groups were substantial and

erbating distress for high-risk individuals, thereby masking potential policy benefits. However, since pandemic severity was balanced across treatment and control cities, such a cancellation effect would largely be absorbed by the control group trend, as detailed in Section 6.

immediate, they were not persistent, fading within approximately two years of implementation. These findings underscore the importance of both targeting and sustainability in the design of large-scale mental health interventions.

5 Mechanism

Having established that the Plan generated better mental health outcomes for low- and medium-risk individuals while having no detectable impact on high-risk individuals, we now investigate the mechanisms underlying this pattern. We propose two potential mechanisms: a structural mismatch in provider type and a time-privacy trade-off mechanism. The former mechanism posits that the Plan prioritized the expansion of general counseling services, which may be a clinically imperfect substitute for the specialized psychiatric care required by severe cases. The latter mechanism suggests that the heterogeneous effect is also driven by behavioral frictions in the adoption process of general counseling services: high-risk individuals may face higher psychological costs, particularly regarding privacy and social stigma, that attenuate their uptake of the community-based services promoted by the Plan.

In this section, we first leverage organization registration data to characterize the precise nature of the supply shock, shown in subsection 5.1. Then, to test the time-privacy trade-off mechanism, we proceed in two steps. First, we develop a conceptual framework to formalize how the interaction between specific service attributes and patient preferences determines equilibrium adoption (subsubsection 5.2.1). Second, to directly identify the friction preventing uptake, we deploy a discrete choice experiment (DCE) that estimates preference heterogeneity for these specific attributes—most notably privacy and cost—across mental health severity profiles (subsubsection 5.2.3).

5.1 Structural mismatch in provider type

The first channel is a structural mismatch between the type of care supplied and the clinical needs of the population. As shown in section 2, the Plan prioritized the expansion of general counseling institutions rather than specialized psychiatric hospitals. This distinction is critical: only the latter are authorized to administer pharmacological interventions. Consequently, for individuals suffering from severe mental health disorders that require medication, an expansion of non-medical counseling may be a clinically ineffective substitute for psychiatric care. To validate this channel, we first formally estimate the Plan's impact on the entry of mental counseling organizations.

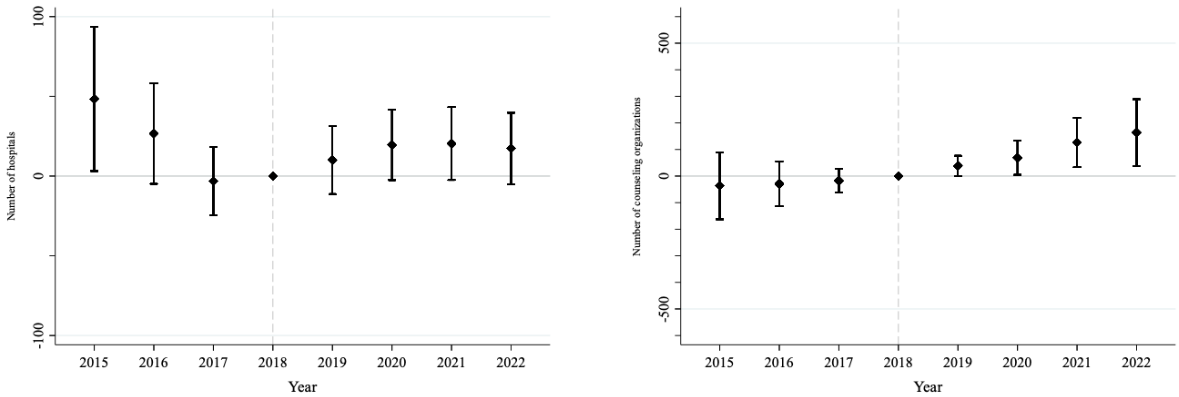
We use organization registration data from Tianyancha and Qichacha for 2015–2022, which aggregate publicly available government records of entity establishment and registration. We identify mental healthcare providers using a three-step approach.[21] Figure 3 presents event-study estimates of the Plan's impact on the number of each provider type, using the specification:

$$y_{ct} = \alpha + \gamma_c + \lambda_t + \sum_{n=-J}^{J} \beta_n \text{Plan}_{ct}^n + \mathbf{X'}_{c0} \times \lambda_t \boldsymbol{\psi} + \epsilon_{ct} \tag{3}$$

where y_{ct} is the number of providers (hospitals or counseling organizations) in city c at year t; Plan_{ct}^n are indicators for years relative to treatment; γ_c and λ_t are city and year fixed effects; and $\mathbf{X'}_{c0} \times \lambda_t$ controls for city-specific trends based on 2018 baseline characteristics (population density, GDP per capita, unemployment rate, number of medical and commercial mental counseling service providers, and internet penetration rate).

Two key supply-side facts emerge from this analysis. First, while the Plan induced a substantial, immediate, and persistent entry of *general counseling organizations* (Panel b), we find no statistically significant expansion of *specialized psychiatric hospitals* (Panel a). Controlling for the supply of general counseling institutions makes the baseline DID estimates for both mental health outcomes statistically insignificant (Table 11), confirming that counseling service expansion is the primary channel. However, this supply-side

[21] Briefly: (1) we filter entities whose business scope contains the keyword "mental healthcare," excluding those with disclaimers indicating no consultation services; (2) we employ a large language model (Deepseek-chat, August 2025 version, trained on 9 examples) to determine whether the remaining 102,345 records offer specific mental healthcare services; (3) we classify organizations into two types (Deepseek-chat, trained on 3 examples): *Professional Hospitals* (requiring integration with City Statistical Yearbook data for full count due to varied naming conventions) and *General Counseling Organizations* (commercial entities offering non-medical talk therapy, without diagnostic or prescriptive rights). In Supplementary materials, we provide a detailed explanation for the major differences between these provider types.



(a) Number of hospitals

(b) Number of counseling organizations

Figure 3: Dynamic effects of the Plan on mental healthcare supply composition

Notes: The figure plots event-study coefficients β_n from equation (3), with 95% confidence intervals based on standard errors clustered at the city level. The omitted category is $t = -1$ (the survey year before treatment). Panel (a) shows no detectable increase in professional hospitals, while Panel (b) reveals a sustained increase in general counseling organizations beginning in 2019 (the year the Plan was implemented). This suggests the Plan's supply expansion was heavily skewed toward community-based, lower-privacy providers.

structural mismatch offers only a partial explanation for our main results. Crucially, our high-risk subsample in CFPS is identified via psychometric screening scales rather than clinical diagnoses. Consequently, the majority of these individuals likely suffer from common mental disorders amenable to counseling, rather than severe psychotic disorders that necessitate pharmacological management. Moreover, given the relatively low base rate of severe psychotic disorders in the general Chinese population[22], such cases constitute a small fraction of our sample. It is therefore unlikely that the heterogeneous effects in the broader CFPS population are primarily driven by limited access to psychiatric medication.

5.2 Time-privacy trade-off

Given the limitations of the structural mismatch explanation discussed above, we propose a second, plausibly dominant mechanism: a time-privacy trade-off mechanism arising from heterogeneous preferences over price, commuting time, and public stigma, rather than clinical unavailability. To isolate this mechanism, we analyze the decision to utilize

[22]For example, prevalence rate of schizophrenia and other psychotic disorders that require medical treatments are roughly 0.7% of the population (Huang et al., 2019).

counseling services under the assumption that, for the relevant subpopulation, the clinical utility of care does not vary across provider types.[23] Our analysis proceeds in two steps. First, we present a conceptual framework to formalize how the Plan generates overall and heterogeneous responses we observe (subsubsection 5.2.1). Second, we deploy a discrete choice experiment (DCE) to directly measure preference heterogeneity for service attributes across severity groups (subsubsection 5.2.3).

5.2.1 Conceptual framework

To structure our mechanism analysis, we outline a simple conceptual framework.[24] As illustrated in Figure 4 and Section 2, the Plan operates through two channels: a supply-side expansion of provider networks and a demand-side reduction in public stigma. Our framework incorporates both channels and analyzes their joint effects on service uptake. We model the individual's decision to seek care as a function of monetary cost, time cost, and psychological cost. We also derive testable predictions about individual preferences that the DCE can assess.

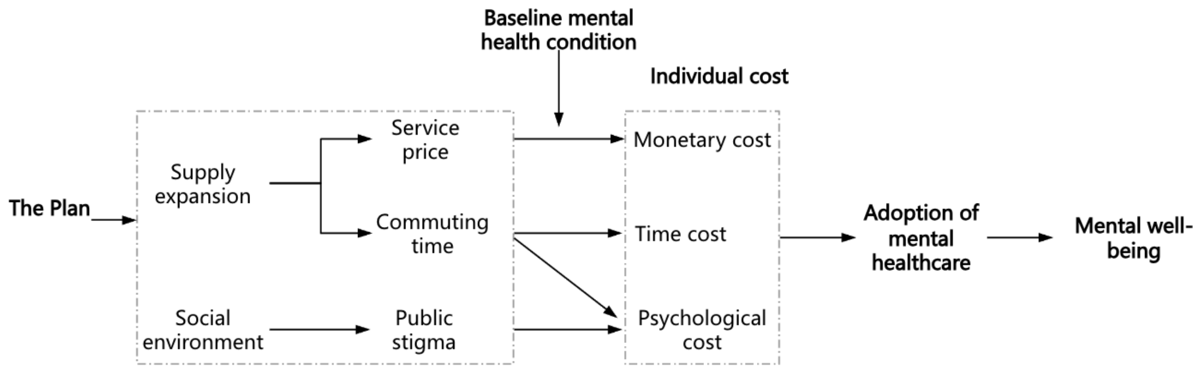


Figure 4: Graphical illustration of the conceptual framework

We summarize the main elements and predictions here; formal derivations are presented in Supplementary Materials.

[23]This assumption of substitutability between provider types, although simplified, is grounded in recent literature. For example, Singla et al. (2017) demonstrate that while non-specialist providers (NSPs) differ from professionals in training and setting, the magnitude of their impact on common mental disorders is similar to that of specialists.

[24]We do not intend to estimate a structural model of service demand, but rather to formalize the mechanisms generating heterogeneous treatment effects.

The framework features heterogeneous populations. Individuals differ in their baseline mental health status. Baseline severity affects both the need for healthcare services and preferences over service attributes, where we assume a concave utility function in mental health, implying that individuals with worse baseline mental health have higher needs for care.[25]

We model the Plan's effects through two channels. First, the Plan expands the supply of mental counseling providers, which reduces both the monetary price of services through increased competition and the commuting time of accessing care through improved geographic coverage. Second, the Plan reduce the public stigma. In equilibrium, the quantity of mental health services adopted is jointly determined by the interaction of these two channels.

Individuals face three distinct types of costs when seeking mental healthcare: (1) *monetary costs*, determined by the price of services; (2) *time costs*, driven by commuting time to providers; and (3) *psychological costs*, which capture stigma-related concerns about social visibility and reputational consequences, depending on both proximity to providers[26] and public stigma. Therefore, improved accessibility generates a *dual effect*: Closer providers reduce time costs but may increase psychological costs by making healthcare-seeking behavior more visible to one's social network.

The model implies two sets of key predictions at the policy level. First, supply expansion increases service uptake by lowering prices and reducing commuting time, thereby raising mental health outcomes in the population. Crucially, while commuting-time savings apply broadly, stigma-related psychological costs are more salient for individuals with worse baseline mental health. As a result, improved accessibility yields a weaker net

[25]We acknowledge that the relationship between baseline mental health and marginal benefits of adoption could exhibit non-monotonicity in extreme cases. For individuals with extremely severe psychopathology, treatment resistance may limit marginal gains. However, empirical evidence supports the assumption of diminishing marginal returns within general populations: individuals with worse baseline mental health typically demonstrate larger treatment effects due to greater "room for improvement" (Bower et al., 2013). Since our sample is drawn from a general population rather than a refractory clinical cohort, the concave assumption serves as the most parsimonious approximation.

[26]This is because the social distance of potential observers matters: individuals may be more concerned about stigma from acquaintances or community members than from strangers (Rüsch et al., 2014). The perceived risk of reputational damage is significantly higher when seeking care from nearby, visible providers.

gain in adoption for those most in need when psychological costs offset time-cost savings. We formalize these insights below, distinguishing average from distributional effects.

Hypothesis 1 (Supply expansion improves mental health). Supply expansion increases service uptake and thereby improves mental health outcomes on average.

Hypothesis 2 (Heterogeneous impact of supply expansion). The health gains from supply expansion are smaller for individuals with worse baseline mental health.

Second, turning to the demand-side intervention, reductions in public stigma directly lower the psychological costs of seeking care. Since psychological costs are higher for individuals with worse baseline mental health, reducing public stigma delivers larger net gains in uptake for them, translating into greater improvements in mental health outcomes for that group. The demand-side intervention thus generates parallel hypotheses with contrasting distributional implications.

Hypothesis 3 (Stigma reduction improves mental health). Reducing public stigma lowers psychological costs, increases service uptake, and thereby improves mental health outcomes on average.

Hypothesis 4 (Heterogeneous impact of stigma reduction). The health gains from stigma reduction are larger for individuals with worse baseline mental health.

These hypothesis map directly to our empirical results. Hypothesis 1 and 3 together predict the main difference-in-differences findings on health outcomes: the Plan improves population mental health on average. Critically, Hypothesis 2 and 4 yield opposing distributional predictions: supply expansion delivers smaller gains to high-risk individuals (privacy costs offset accessibility), while stigma reduction delivers larger gains (alleviating psychological barriers). Our DID results—benefits concentrated among lower-risk individuals, null effects for high-risk—align with Hypothesis 2 but contradict Hypothesis 4, indicating that supply expansion dominated the policy mix while stigma reduction was insufficient to offset privacy loss associated with improved accessibility.

To test the underlying mechanism directly, we state two additional hypothesis at the preference level about how individuals value service attributes. The model presumes that adoption decreases with higher prices, improves with better geographic access, and increases as public stigma falls. Importantly, these responses vary with baseline mental health in ways that mirror the mechanism above: individuals with worse baseline mental health face stronger stigma-related costs, shaping their sensitivities to price, access, and stigma.

Hypothesis 5 (Preference-level effects of attributes). Holding other factors fixed, adoption decreases with higher prices, increases with shorter commuting time, and increases with lower stigma.

Hypothesis 6 (Heterogeneity in attribute sensitivities). Relative to individuals with better baseline mental health, those with worse baseline mental health exhibit higher price sensitivity and lower responsiveness to commuting-time reductions.

We evaluate Hypothesis 5 and 6 using a DCE that directly elicits preference over price, access, and public stigma and tests whether attribute sensitivities vary with baseline mental health as implied by the framework. Taken together, the theory and these preference-level tests help explain why the Plan can produce overall improvements in health while delivering muted benefits for the most in need when stigma remains high—thereby motivating the descriptive evidence and the DCE that follow.

5.2.2 Suggestive evidence on service adoption

Before turning to direct preference elicitation, we present descriptive evidence on two key empirical patterns that motivate our DCE design: (1) the nature of the Plan's supply expansion, and (2) suggestive patterns of service adoption.

Given the lack of administrative data on mental health service utilization, we provide indirect, suggestive evidence using online consumer engagement data from Dazhongdianping, China's largest lifestyle and online-to-offline (O2O) platform (analogous to Yelp or TripAdvisor).[27]

We collected user comment data for three mental counseling organizations in Chongqing, one of the pilot cities. These organizations have been active on the platform for at least six years (available since 2019). Figure 5 plots the cumulative number of comments over

[27] Dazhong Dianping is one of China's largest platforms for user-generated reviews and ratings of local businesses, covering dining, entertainment, leisure activities, and other services.

time, from the earliest available date for each organization through December 2022. Two

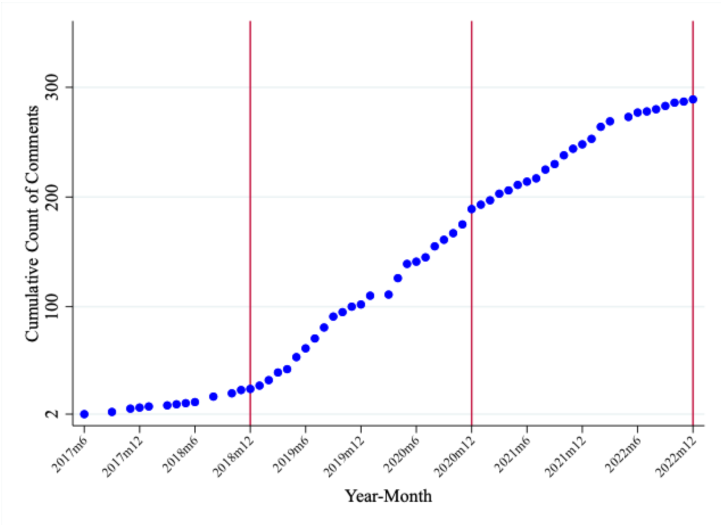


Figure 5: Cumulative user comments on mental counseling organizations in Chongqing

Notes: The figure plots the cumulative number of user comments on three mental counseling organizations in Chongqing, obtained from Dazhongdianping. The slope of each line represents the rate of new comments. The steepest growth occurs between late 2018 (when the Plan was announced) and December 2020, suggesting increased consumer engagement following the Plan's implementation.

patterns are noteworthy. First, there is a clear upward trend in the cumulative count of comments since June 2017. Second, the slope of the lines—representing the rate of new comments—seems to be the steepest and most sustained between late 2018 and December 2020, precisely the period when the Plan was announced and implemented.

While we acknowledge that this increase could partially reflect the growing popularity of the Dazhongdianping platform itself, the timing of the acceleration, which coincides with the Plan's implementation, combined with the documented supply-side expansion (Figure 3) suggests that the observed surge in consumer engagement at least partly reflects an underlying increase in mental healthcare adoption in treated cities. This provides supportive, albeit indirect, evidence linking the policy's shock to greater public utilization, consistent with Hypothesis 1 and 3.

Taken together, the evidence in this subsection establishes two key empirical patterns: (1) the Plan primarily expanded community-based, low-privacy providers, and (2) this expansion coincided with increased consumer engagement. These patterns set the stage for our direct preference elicitation in the next subsection, where we test whether the observed heterogeneous treatment effects in section 4 can be explained by differential preferences across mental health severity groups.

5.2.3 Discrete choice experiment

The descriptive evidence shown in the previous section documented that the Plan expanded community-based commercial counseling organizations, i.e., providers characterized by high accessibility but potentially low privacy. Our heterogeneous treatment effects (section 4) show that this supply expansion improved outcomes for low-risk and medium-risk individuals, but not for high-risk individuals. Our conceptual framework (subsubsection 5.2.1) attributes this pattern to heterogeneous preferences: high-risk individuals may respond weakly to accessibility improvements if stigma-related psychological costs dominate. To test this mechanism directly, we conduct a discrete choice experiment (DCE) to elicit preferences for mental health service attributes across mental health severity groups.

Design and implementation A DCE is a stated preference method that presents respondents with hypothetical choice scenarios to elicit their preferences for services based on varying attributes and levels (de Bekker-Grob et al., 2012). Building on our conceptual framework, we selected three core attributes:

1. Price: The monetary cost per hour of service. Four levels: ¥500, ¥300, ¥100, ¥0.[28]
2. Access: Commuting time to the provider. Three levels: 100 minutes, 30 minutes, 0 minutes.
3. Stigma: Following Jackson et al. (2021), we operationalize stigma as the percentage of individuals in the respondent's immediate social environment who tend to discriminate against mentally ill individuals. Three levels: 90%, 40%, 10%.

To isolate preferences for specific attributes and rule out confounding inferences regarding service quality or provider type, we presented respondents with a standardized decision context. At the start of the DCE, respondents received the following instructions:

[28]These price points were determined by examining two major online mental counseling platforms (Alibaba-affiliated) with the highest sales volume as of February 2025. For a certified "senior" counselor (typically with 5–6 years of experience), the average hourly rate was approximately ¥500.

"In the following section, you will be presented with a series of hypothetical mental health counseling service scenarios. Each scenario includes multiple service options. For each option, you will be asked to consider and evaluate three key features: 1) Cost per session (e.g., RMB 200/session or RMB 500/session); 2) Travel time to and from the location (e.g., 20 minutes round trip or 1 hour round trip); 3) Perceived level of social stigma (i.e., the estimated percentage of people around you who may hold negative attitudes toward individuals with mental health conditions, such as 10% or 90%). All options offer the same treatment effectiveness.[29] These factors are mutually independent—for instance, a higher fee does not imply a shorter travel distance, and shorter travel time does not necessarily correspond to lower stigma. Please make your selections purely based on your personal preferences. There are no correct or incorrect answers. Your authentic responses are critical to our research."

Figure 2 demonstrates an example of the choice set in this DCE.

We used a D-efficient fractional factorial design to construct choice sets, generated using Stata 17. The final design resulted in 10 choice sets, each comprising two alternative mental healthcare service options and an opt-out option ("I would not seek care").

The formal survey was conducted in April 2025, in partnership with Wenjuanxing, a prominent online survey company in China.[30] The survey consisted of three parts: (1) demographics and personal information, (2) mental healthcare use and attitudes, and (3) the DCE choice tasks. We collected responses from 308 individuals, yielding 9,240 choice observations.

[29] We carefully considered the clinical mismatch hypothesis—that high-risk individuals respond weakly to accessibility improvements not due to stigma, but because they associate community-based (proximate) care with lower clinical capability (e.g., lack of prescribing rights). In the experimental protocol, respondents were instructed to evaluate scenarios under the binding condition that "All options offer the same treatment effectiveness." By holding clinical utility constant for everyone by design, we decoupled the technological quality of care from its physical accessibility. Consequently, if high-risk respondents were avoiding nearby options solely due to clinical mismatch, this constraint should have neutralized that avoidance.

[30] A pilot survey was distributed in March 2025 to validate question clarity and reasonableness. Revisions were made based on feedback before the formal data collection in April.

Empirical specification We estimate a mixed logit (random parameters logit) model to capture unobserved preference heterogeneity across respondents:

$$U_{ij} = \boldsymbol{\beta}'\mathbf{X}_{ij} + \boldsymbol{\eta}_i'\mathbf{X}_{ij} + \epsilon_{ij} \tag{4}$$

where U_{ij} is the utility of individual i choosing alternative j; $\mathbf{X}_{ij}$ is the vector of attribute levels in alternative j; $\boldsymbol{\beta}$ is the population mean of the preference parameters; $\boldsymbol{\eta}_i$ is individual i's deviation from the population mean (assumed to follow a normal distribution); and ϵ_{ij} is an i.i.d. Type I extreme value error term. We use 1,000 Halton draws for simulation.

To test for preference heterogeneity by mental health severity, we estimate an extended specification with interaction terms:

$$U_{ij} = \boldsymbol{\beta}'\mathbf{X}_{ij} + \boldsymbol{\delta}'(\mathbf{X}_{ij} \times \text{Bad}_i) + \boldsymbol{\eta}_i'\mathbf{X}_{ij} + \epsilon_{ij} \tag{5}$$

where Bad_i is an indicator for "poor mental health," defined as a CES-D score above the 50th percentile. The coefficients $\boldsymbol{\delta}$ capture how preferences differ for individuals with worse baseline mental health.

Baseline preferences Table 4 presents the baseline mixed logit estimates (equation (4)). Consistent with our theoretical predictions (P3 and P5), respondents exhibit strong preferences for lower prices, shorter commuting times, and lower stigma levels. Specifically, it shows that reducing the price from ¥500 to ¥300, ¥100 or ¥0 significantly increases utility (coefficients: 0.346, 1.478 and 1.514, $p < 0.001$). Similarly, reducing commuting time from 100 minutes to 0 minutes significantly increases utility (coefficient: 0.681, $p < 0.001$), though the effect of 30 minutes is not statistically significant. Reducing stigma from 90% to 40% or 10% also significantly increases utility (coefficients: 0.852 and 1.882, both $p < 0.001$).

The large and significant standard deviations of the random parameters reveal substantial unobserved heterogeneity in preferences across respondents, motivating our investigation of heterogeneity by mental health severity.

Preference heterogeneity by mental health severity Table 3 presents results from the interaction specification (equation (5)). The findings strongly support our theoretical predictions.

Price sensitivity: Individuals with worse mental health exhibit *stronger* responses to price reductions. The interaction coefficients for Price_100 $\times$ Bad and Price_0 $\times$ Bad are positive and significant (0.769, $p < 0.001$; 0.574, $p = 0.028$), implying that reducing the price from ¥500 to ¥100 or ¥0 increases odds of adoption by 115.8% and 77.5% *more* for the "bad" group compared to the "good" group. This is consistent with our framework: individuals with worse mental health require more intensive services, so price reductions generate larger total cost savings.

Accessibility sensitivity: Individuals with worse mental health exhibit *weaker* responses to accessibility improvements. While the "good" group significantly values zero commuting time (coefficient: 0.904, $p < 0.001$), the "bad" group's preference is attenuated by 0.412 units (interaction coefficient: -0.412, $p = 0.024$), representing a 33.7% decrease in the odds of the "bad" group choosing the zero-commuting time option. This pattern is precisely what our framework predicts: improved accessibility reduces time costs for everyone, but it *increases* social visibility, which disproportionately burdens high-need individuals who require intensive, repeated service use.

Stigma: Both groups exhibit strong preferences for lower stigma environments, with no significant difference in stigma sensitivity (interaction coefficients for Stigma_40 $\times$ Bad and Stigma_10 $\times$ Bad are not statistically significant).

Interpretation and policy implications The DCE results provide direct evidence for the mechanisms underlying our heterogeneous treatment effects. Individuals with worse mental health conditions—who require more intensive mental health services—face a distinct cost structure: they are highly price-sensitive (because they consume more services) but respond weakly to accessibility improvements (because proximity increases social visibility). This cost structure makes them less responsive to the Plan's supply

Table 3: Mixed logit estimates: Heterogeneity by mental health severity

Attribute and Levels	**Coef.**	**SE**	**p**	**SD**	**SE**	**p**
Price (ref: ¥500)						
¥300	0.117	0.194	0.548	0.993	0.166	0.000
¥100	1.017	0.178	0.000	−0.683	0.227	0.003
¥0	1.199	0.207	0.000	1.609	0.177	0.000
Price_300 × Bad	0.391	0.241	0.104	0.368	0.422	0.384
Price_100 × Bad	0.769	0.219	0.000	−0.499	0.476	0.295
Price_0 × Bad	0.574	0.262	0.028	0.773	0.422	0.067
Access (ref: 100 min)						
30 minutes	0.172	0.153	0.261	0.913	0.149	0.000
0 minutes	0.904	0.147	0.000	−0.484	0.202	0.017
Access_30 × Bad	−0.070	0.191	0.716	−0.508	0.259	0.050
Access_0 × Bad	−0.412	0.182	0.024	0.531	0.302	0.079
Stigma (ref: 90%)						
40%	1.012	0.182	0.000	1.259	0.164	0.000
10%	2.099	0.265	0.000	2.348	0.185	0.000
Stigma_40 × Bad	−0.215	0.234	0.358	0.703	0.333	0.035
Stigma_10 × Bad	−0.232	0.334	0.487	−0.849	0.423	0.045
N				9,240		
Log likelihood				−640.9		
Pseudo R^2				0.113		

Notes: The table reports mixed logit estimates with interactions. The "× Bad" terms measure how preferences differ for respondents with CES-D scores above the median. Positive (negative) coefficients indicate that the "bad" group values that attribute level more (less) than the "good" group.

expansion, which primarily added community-based, low-privacy providers.

Empirically, we explore the interaction between accessibility and stigma further in Table 5, where we include three-way interactions ($\text{Access}_0 \times \text{Stigma}_{\text{level}} \times \text{Bad}$) using a conditional logit model.[31] The results confirm that even under low-stigma environments, individuals with worse mental health conditions exhibit lower adoption odds when services are extremely accessible (zero commuting time), suggesting that privacy concerns—such as fear of being recognized by neighbors or acquaintances—can suppress uptake even when general stigma is low. This underscores the importance of designing services that balance accessibility with anonymity.

6 Robustness Checks

This section addresses several potential threats to the validity of our baseline findings. We organize the robustness checks into two categories. First, we address core threats to causal identification, including non-random treatment assignment (subsection 6.1), confounding from the COVID-19 pandemic (subsection 6.2), and the potential influence

[31]The use of conditional logit is simply to facilitate the interpretation of the results.

of a concurrent competing policy (Long term Care Insurance Pilot Plan) (subsection 6.3). Second, we examine methodological concerns related to measurement validity and sample representativeness (subsection 6.4 and subsection 6.5).

6.1 Non-random assignment of treatment status

A primary concern for the internal validity of our baseline findings stems from the non-random selection of piloting cities. As illustrated in section 2, the application-selection process for piloting cities, a common feature of policy experimentation in China, is evident and often exhibits positive selection in terms of economic development and administrative capacity (Wang and Yang, 2025). This raises concerns about the representativeness of treated cities and the potential for upwardly biased treatment effect estimates.

Indeed, as presented in Table 3, we confirm that piloting cities were significantly "better off" in several baseline dimensions. For instance, they are on average more affluent and possess more mental healthcare resources, as measured by the number of mental healthcare organizations. These pre-existing differences in baseline characteristics could plausibly be associated with differential time trends in mental health outcomes, thereby confounding our estimates.

To address this potential source of bias, we follow Gruber et al. (2023) and explicitly control for differential time trends associated with observed baseline city characteristics. ITT estimates from this augmented specification are presented in columns (1) and (2) of Table 9. Reassuringly, the Plan continues to lead to statistically significant reductions in both the standardized mental health score (by 0.05 standard deviations) and the probability of having a high risk of severe disorder (by 3 percentage points). While the magnitudes are slightly smaller than our primary estimates, their statistical significance and sign reinforce the robustness of our core finding. The dynamic effects, illustrated in Figure 3 and Figure 4, further confirm the absence of pre-trends and the consistency of the treatment effect pattern across subgroups, supporting the validity of our baseline results even after accounting for differential trends driven by observed city characteristics.

6.2 The impacts of COVID-19

A significant concern for causal interpretation arises from the temporal coincidence of the Plan's implementation and the COVID-19 pandemic. The year 2020, which marks the first full year of policy exposure in our analysis, also coincides with the outbreak of COVID-19, raising the possibility that pandemic-related shocks could confound our estimated treatment effects.

We offer three main lines of evidence to demonstrate that our estimates are not materially biased by COVID-19. First, our baseline model incorporates both individual and survey wave fixed effects. Importantly, survey wave (year) fixed effects absorb common macroeconomic shocks and uniform temporal trends across all cities, including the nationwide impact of the COVID-19 pandemic. This design feature ensures that our treatment effect is identified from differential changes between treated and control cities, net of any common pandemic-related shock.

Second, we provide direct empirical evidence that the severity of the COVID-19 shock did not differ significantly between treated and control cities. Using daily COVID-19 data from Dingxiangyuan, we calculate each city's cumulative COVID-19 mortality rate by the end of 2020.[32] As shown in Table 6, there is no statistically significant difference in mortality rates between treatment and control groups. The mortality rates in both groups were remarkably small and highly similar, indicating that the fundamental severity of the immediate health shock did not differentially affect our treated and control cities.

To address concerns that differential local containment policies might have imposed varying levels of mobility restrictions and psychological distress on individuals, we employ a triple-difference (DDD) specification to examine the heterogeneous treatment effects of the Plan on individuals residing in cities with varying stringency of COVID-19 lockdown measures. For this analysis, we utilize data from the Baidu Qianxi Platform, which provides a Migration Scale Index reflecting population migration levels. Although the index is dimensionless and its construction method remains undisclosed, we construct an annual total migration index by summing the in-migration and out-migration indices to

[32] $Mortality\ rate_{2020} = \frac{COVID\text{-}19\ death\ count}{Population}$

assess the lockdown stringency of each city. Cities are then classified into two subgroups based on their total migration levels in 2020: those with high stringency (above the 50th percentile) and those with low stringency (below the 50th percentile).

The results, as detailed in Table 10, indicate that the interaction term Treat $\times$ Post $\times$ Lockdown does not exhibit a significant additional impact on either mental health outcome. Specifically, the coefficients for standardized score and caseness are (-0.007) and (0.012), respectively, suggesting that lockdown stringency does not significantly alter the treatment effects. Conversely, the Treat$\times$Post term shows a statistically significant reduction in both standardized scores -0.062^{*} and caseness -0.051^{***} following the treatment implementation, aligning closely with our baseline estimates. These findings reinforce the robustness of our primary results and highlight the Plan's efficacy in improving mental health outcomes, irrespective of local lockdown stringency.

6.3 Competing policies

In estimating the causal effect of the Plan, one concern is whether the observed changes in mental health outcomes in treated cities relative to control cities are attributable solely to the Plan, and not to other concurrent policy interventions. Specifically, the expansion of the Long-Term Care Insurance (LTCI) Pilot Plan, which began in 2016 and was formally expanded across multiple cities in 2020, presents a potential confounding factor. The LTCI system's primary goal is to provide financial and service guarantees for the severely disabled population, covering basic daily living and related medical care. While LTCI does not directly target mental health, it might generate a significant positive psychological externality. By alleviating the substantial economic burden and caregiving stress on households, the LTCI system can indirectly improve the psychological well-being of both the disabled individuals and their family caregivers.

If LTCI pilot cities were imperfectly allocated across our Plan's treatment and control groups, this positive spillover could potentially bias our baseline DID estimates, likely leading to an overestimation of the Plan's true causal effect. To account for this concurrent policy, we incorporate a control for the LTCI policy's impact by including an

LTCI-specific treatment interaction term with survey year (LTCI_{DID}) into our baseline regression framework. Columns (3) and (4) in Table 9 report the results of this augmented model. The estimated coefficients for the Plan's core DID term remain statistically significant at the 5% and 1% levels, respectively, and, notably, does not drop dramatically in magnitude after controlling for the LTCI policy's effects. For the Standardized Score (Column 3), the estimated treatment effect is -0.065 ($p < 0.05$). For the High Risk of Caseness binary outcome (Column 4), the estimated effect is -0.036 ($p < 0.01$). This consistent and increased magnitude suggests that our main findings are robust and are not driven by the concurrent LTCI Pilot Plan.

6.4 Comparability of CES-D and K6 scales

A methodological concern in our baseline analysis pertains to the comparability of the two mental health measurement instruments employed across different survey waves. While both the K6 and CES-D scales are widely recognized and validated instruments for assessing psychological well-being, they capture conceptually distinct constructs: the CES-D primarily measures depressive symptoms in the general population (Radloff, 1977), whereas the K6 captures broader psychological distress (Furukawa et al., 2003). Although our main analysis employs standardized scores and validated cutoffs to construct the indicator for caseness to mitigate the direct comparability issues between the raw scales, a more convincing demonstration of robustness would utilize a consistent outcome measure throughout. To address this concern, we conduct an additional robustness check using only CFPS survey waves that employ the CES-D scale (2012, 2016, 2018, 2020, and 2022). While this results in a comparatively shorter pre-treatment period (2012 and 2016), it offers valuable insights into the consistency of our findings using a single, homogeneous measurement instrument.

ITT estimates using the CES-D-only data are presented in columns (5) and (6) of Table 9. Overall, these estimates indicate that the Plan leads to a significant reduction of 0.076 standard deviations in the normalized CES-D score and a 3.5 percentage point reduction in the probability of caseness. The magnitudes of these point estimates are highly similar to our primary baseline estimates. Visual inspection of the event study plots in Figure 5 confirms the absence of pre-trends, supporting the parallel trends assumption. Figure 6 reports event study estimates for the three subgroups based on predicted probabilities of severe mental disorder. The findings are largely consistent with our main results: the Plan significantly improves mental health outcomes for the low and middle risk of caseness groups, while showing no meaningful impact on the high risk of caseness group. Although the high-probability group exhibits a statistically significant drop in the standardized CES-D score at $T + 0$, this does not translate into a significant reduction in the probability of severe mental disorder (Panel f), suggesting that any improvement in this subgroup is not substantively meaningful. However, it is important to note that the limited number of pre-treatment time points necessitates cautious interpretation of the visual event study plots.

6.5 Sample representativeness

A potential concern arises from the reliance on two distinct datasets in our analysis: the broader CFPS Sample (used for the main policy analysis and serving as a national benchmark) and the DCE Survey Sample (used for mechanism analysis). Online surveys are prone to convenience sampling bias and may over-represent educated, urban respondents with specific interest in the topic, thereby threatening the generalizability of our mechanism findings.[33] To ensure the main policy findings are consistent and comparable to the mechanism findings, and specifically, to assess the policy effect within the specific demographic profile captured by our DCE sample, we employ Entropy Balancing (EB) (Hainmueller, 2012) to reweight the larger CFPS data. EB is a flexible reweighting method designed to generate a set of optimal weights that satisfies a predefined set of

[33]To assess the extent of this concern, we conduct a detailed comparison of six key baseline characteristics across the two samples: age, years of education, location, marital status, gender, and the share of respondents with a high risk of caseness. These comparisons are included in Supplementary Materials. We observe non-trivial differences between the two samples. The Survey Sample does not fully reflect the regional distribution observed in the CFPS. For example, respondents from Henan province are under-represented in the Survey Sample, and there is a disproportionately greater share of respondents from treated cities. Additionally, the Survey Sample has higher risk of caseness, and is younger, more educated and slightly less likely to be male or married compared to the pooled CFPS sample.

moment conditions (i.e., balance constraints) between the two samples.[34] Specifically, we generate EB weights for the CFPS Sample, forcing the weighted mean (the first moment) of five key covariates—age, education, gender, marital status, and severe mental health status—to precisely match the unweighted mean of these covariates in the DCE Survey Sample (our target distribution).

The last two columns in Table 9 present the re-estimated results. After accounting for compositional differences using EB weights, our results remain consistent in sign and magnitude with the baseline estimates. Specifically, the coefficient of the main policy variable retains the expected negative sign across both mental health outcomes. While the effect on the standardized score becomes statistically insignificant and the effect on the caseness indicator is only marginally significant, the magnitude of the coefficients remains very similar to our primary findings.

We further address the concern that our mechanism derived from the younger, more educated DCE sample may not generalize to the broader CFPS population. As shown in Table 7, the high-risk group in our survey sample is indeed significantly younger and more educated than respondents in the national CFPS data. This raises a critical question: is the null treatment effect observed in the national data driven by the specific demographic profile (e.g., age) of the high-risk group, rather than the channels identified in the DCE? To test this, we re-estimate the heterogeneous treatment effects using Entropy Balancing weights, while retaining the original Leave-One-Out (LOO) risk classification from Section 3. This approach allows us to observe whether the policy's failure persists within a reweighted national sample that demographically mimics our DCE respondents. The results in Table 8 confirm our baseline findings: even after adjusting the CFPS sample to match the younger, more educated profile of the survey respondents, the treatment effect for the high-risk group remains statistically indistinguishable from zero. In contrast, the low-risk group continues to exhibit significant improvements in mental well-being. This robustness suggests that the barriers to adoption, likely the privacy costs identified in the

[34]Mathematically, the EB weight ω_j for each observation j in the CFPS Sample is chosen to minimize the entropy metric: $\min_{\omega_j} \sum_{j=1}^{N_{\text{CFPS}}} \omega_j \log(\omega_j/q_j)$ subject to a set of M moment-matching constraints. The first-moment constraint for a covariate X is defined as: $\sum_{j:\text{CFPS}} \omega_j X_j = \sum_{i:\text{Survey}} \frac{1}{N_{\text{Survey}}} X_i$

DCE, are structural constraints that bind across demographic groups, rather than being unique to the older or less educated CFPS population.

7 Discussion

In this study, we examine the impact of a nationwide policy intervention—the National Social Psychological Service System Construction Pilot Work Plan—designed to simultaneously expand mental healthcare capacity and reduce public stigma around mental illness. We document a significant overall improvement in population mental health following the Plan's implementation, but these improvements are primarily driven by individuals with better mental health conditions. To understand the mechanisms driving this heterogeneous response, we investigate the underlying adoption dynamics through two complementary approaches.

First, we provide descriptive evidence consistent with increased service engagement, using user-generated comments on mental health facilities from Dazhongdianping. The data reveal accelerated comment growth between late 2018 and 2020, temporally aligning with the Plan's announcement and initial implementation. We acknowledge important caveats: Dazhongdianping users skew young, urban, and tech-savvy, likely over-representing individuals with mild-to-moderate conditions who are willing to publicly share their experiences. We therefore interpret these patterns cautiously as suggestive evidence of increased public engagement, not definitive proof of adoption rates.

Second, we conducted a discrete choice experiment (DCE) to directly estimate preference parameters for three key attributes of mental healthcare services: price, accessibility (commuting time), and anticipated stigma. The DCE results confirm our conceptual framework: individuals exhibit higher utility for lower prices, reduced commuting times, and lower perceived stigma. Crucially, the DCE reveals heterogeneous preferences by baseline mental health severity. Individuals with worse mental health are significantly more price-sensitive (77% stronger response to price reductions; Table 3) but less responsive to accessibility improvements (46% weaker response to reduced commuting time).

We attribute this counterintuitive finding to proximity-induced privacy concerns: highly accessible, community-based services increase the risk of involuntary disclosure to neighbors or acquaintances, disproportionately burdening high-risk individuals who require more intensive, frequent treatment.

This preference heterogeneity is key to explaining the observed treatment effect heterogeneity in our main analysis (section 4). Since the Plan's operational channel primarily expanded supply and accessibility rather than direct financial subsidies or anonymity protections, it aligned optimally with the preferences of individuals with better baseline mental health—the group most responsive to reduced commuting time and least concerned about privacy risks. Conversely, high-risk individuals, who prioritize affordability and anonymity over physical proximity, derived limited benefit from the Plan's accessibility-focused interventions.

Our findings align with existing literature documenting that supply expansions can improve mental health outcomes (Feyman et al., 2023; Costantini, 2024), but we contribute novel evidence that the composition and modality of expanded capacity matter critically. Simply increasing the number of community-based facilities is insufficient—and may even be counterproductive—for the most vulnerable populations. We move beyond aggregate supply-demand analyses to identify preference heterogeneity across mental health severity levels, with direct implications for policy design.

7.1 Policy Implications: Severity-Contingent Prescriptions

Our findings yield severity-contingent policy prescriptions that challenge a one-size-fits-all accessibility mandates in mental health policy. The optimal policy approach differs fundamentally depending on the target population's mental health status.

For the general population with low-to-moderate mental health concerns, policymakers should continue expanding accessible community-based services. Our DCE estimates suggest that each 10-minute reduction in commuting time increases adoption probability by approximately 7% for this group (Table 3, Access_0 main effect). The Plan's community integration strategy that expands the supply of general counseling services, appears well-suited to this population's preferences. Policy priorities should include establishing geographic coverage targets, such as ensuring at least one mental health facility within short public transit in all urban districts. Integration with primary care systems can reduce referral frictions and normalize help-seeking behavior. Public awareness campaigns that emphasize the commonality and treatability of mild-to-moderate mental health concerns can further reduce ambient stigma levels and encourage early intervention.

For high-risk individuals with worse mental health conditions, however, a fundamentally different approach is required. Rather than emphasizing physical proximity, policymakers should prioritize anonymity-preserving delivery mechanisms. Our DCE results indicate this group exhibits 77% stronger price sensitivity but 46% weaker responsiveness to accessibility improvements compared to those with better mental health. Three specific policy interventions emerge as particularly promising for this population.

First, telehealth platforms with robust privacy protections can eliminate proximity-induced disclosure concerns while maintaining professional quality of care. China's rapid expansion of internet infrastructure and widespread smartphone adoption make telehealth particularly feasible. Policy priorities should include establishing insurance reimbursement parity for in-person and telehealth consultations, implementing regulatory standards for data security and patient anonymity, and developing training programs to certify psychiatrists in remote diagnosis and therapy delivery. Telehealth has the additional advantage of reducing geographic barriers for patients in rural or underserved areas where specialized psychiatric facilities remain scarce (Patel et al., 2018)..

Second, for conditions requiring in-person treatment—such as severe psychotic disorders or complex medication management, policymakers should consider locating specialized facilities outside patients' immediate social networks rather than within neighborhood settings. This strategy reduces "social distance" costs by minimizing the risk of encountering neighbors, colleagues, or acquaintances during treatment visits (Jorm and Oh, 2009). For example, provincial-level psychiatric hospitals located in district centers, rather than community clinics embedded in residential neighborhoods, allow patients to seek care with greater anonymity. The trade-off between physical accessibility and privacy protection reverses for high-risk individuals: the marginal utility gain from anonymity exceeds the marginal utility loss from longer travel distances.

Third, given the high price sensitivity of severely ill individuals, direct financial subsidies may be more effective than accessibility improvements for this population. Specific interventions could include eliminating out-of-pocket costs for psychiatric diagnosis and first-line treatments such as antidepressants and anxiolytics, providing means-tested subsidies for psychotherapy sessions (which remain poorly covered by China's public insurance system), and establishing catastrophic expenditure protections for patients requiring long-term psychiatric care (Patel et al., 2018). Our DCE estimates suggest that for high-risk individuals, the marginal effect of price reductions on adoption probability substantially exceeds the marginal effect of commuting time reductions. This tension echoes broader debates in health policy about equity versus efficiency in service delivery, but our evidence suggests that mental health policy requires explicit stratification: different delivery modalities for different severity levels, rather than a uniform accessibility target applied to all populations.

In the specific context of China, the observed heterogeneity may also reflect the dual structure of the psychological counseling market. This market is sharply divided between professional psychiatric institutions holding Medical Institution Practicing Licenses and unregulated counseling agencies operating without medical credentials. Following the abolition of the national unified qualification certification for psychological counselors in 2017, entry barriers for non-medical agencies shifted from formal credentials to market reputation and self-regulation. This transition has generated significant quality heterogeneity, making it difficult for consumers to identify competent providers. Our finding that the Plan primarily benefited individuals with better baseline mental health may partly reflect this structural feature: the rapid supply expansion consisted largely of general counseling agencies rather than specialized psychiatric facilities, which require longer time horizons for professionalization and workforce training. This context underscores that quality regulation must accompany quantity expansion. Specific measures could include mandatory licensing and continuing education requirements for all mental health providers, public disclosure of provider credentials and patient satisfaction ratings while protecting patient privacy, stricter enforcement of advertising regulations to prevent misleading claims by unlicensed agencies, and differential reimbursement rates favoring licensed psychiatric professionals over unregulated counselors.

7.2 Limitations

Several limitations should be noted when interpreting our findings. First, we acknowledge a potential identification challenge regarding the temporal overlap between the Plan's implementation and the COVID-19 pandemic in 2020. This coincidence raises the concern that our results might be confounded by the pandemic's uneven impact. However, we believe several features of our analysis mitigate these concerns. Our baseline model includes year fixed effects that absorb common shocks related to the pandemic. Furthermore, we find no significant difference in COVID-19 mortality rates between treated and control cities, suggesting the treatment was not systematically correlated with pandemic severity. Most importantly, the estimated policy effects remain statistically significant and similar in magnitude when we restrict the sample to cities with either low or high COVID-19 mortality. This provides strong evidence that our findings are driven by the policy intervention rather than pandemic-related confounders

Second, we do not provide administrative data on patient visits, prescriptions, or diagnostic records due to data access restrictions. Instead, we triangulate three sources of evidence: confirmed supply expansion (psychiatric bed capacity growth), suggestive demand signals (online review platform engagement), and experimental stated preferences (DCE). While this convergence strengthens our inferences, future research with access to medical claims data or electronic health records could provide more definitive evidence of utilization changes and their clinical impacts.

In addition, our mechanism analysis relies on hypothetical DCE scenarios rather than observed behavior. Although we formally demonstrate that the Plan increased the number of mental health providers (subsubsection 5.2.2), we cannot directly observe changes in out-of-pocket prices, actual commuting times, or experienced stigma levels. Combining revealed preference data with stated preference estimates would provide a more robust test of the proposed mechanisms.

Third, we do not present a cost-benefit analysis of the Plan. Implementing the intervention required substantial public expenditure, including infrastructure investments, workforce training, and subsidy programs. While improving population mental health generates potential productivity gains and welfare benefits, the extent to which these benefits outweigh fiscal costs remains uncertain. A comprehensive cost-benefit analysis would be valuable for informing resource allocation decisions in future policy design.

Finally, our findings are situated in China's unique institutional context: rapid urbanization, centralized policy implementation, and a dual-structure mental health market. The extent to which our severity-contingent policy prescriptions apply to other settings—particularly high-income countries with mature mental health systems and different stigma profiles—requires empirical investigation.

8 Conclusion

This study evaluates the impact of the National Social Psychological Service System Construction Pilot Work Plan in China, a nationwide intervention designed to simultaneously expand mental healthcare capacity and reduce public stigma. We document significant improvements in population mental health following the Plan's implementation, but these gains are concentrated among individuals with better baseline mental health conditions. Using a discrete choice experiment, we identify the mechanism underlying this heterogeneity: individuals with worse mental health exhibit stronger price sensitivity but weaker responsiveness to accessibility improvements, reflecting proximity-induced privacy concerns that make community-based services less attractive to high-risk populations.

Our findings challenge the prevailing one-size-fits-all approach to mental health policy. For the general population, expanding accessible community-based services remains an effective strategy, with each 10-minute reduction in commuting time increasing adoption probability by approximately 7%. For high-risk individuals, however, policymakers should prioritize anonymity-preserving delivery mechanisms—telehealth platforms, strategically located specialized facilities, and direct financial subsidies—rather than physical proximity. Mandating universal community-based care may inadvertently exclude those most in need by increasing involuntary disclosure risks. Effective mental health policy requires explicit stratification across severity levels, balancing equity and efficiency through differentiated delivery modalities tailored to the distinct preferences and constraints of heterogeneous populations.

References

Abadie, A., Chingos, M. M., and West, M. R. (2018). Endogenous stratification in randomized experiments. *Review of Economics and Statistics*, 100(4):567–580.

Andersen, M. (2015). Heterogeneity and the effect of mental health parity mandates on the labor market. *Journal of health economics*, 43:74–84.

Angelucci, M. and Bennett, D. (2024). Depression, poverty, and economic shocks: Evidence from india. In *AEA Papers and Proceedings*, volume 114, pages 412–417. American Economic Association 2014 Broadway, Suite 305, Nashville, TN 37203.

Baranov, V., Bhalotra, S., Biroli, P., and Maselko, J. (2020). Maternal depression, women's empowerment, and parental investment: Evidence from a randomized controlled trial. *American economic review*, 110(3):824–859.

Barker, N., Bryan, G., Karlan, D., Ofori-Atta, A., and Udry, C. (2022). Cognitive behavioral therapy among ghana's rural poor is effective regardless of baseline mental distress. *American Economic Review: Insights*, 4(4):527–545.

Bharadwaj, P., Pai, M. M., and Suziedelyte, A. (2017). Mental health stigma. *Economics Letters*, 159:57–60.

Bi, K., Chen, P., and Chen, S. (2023). Validating the 8-item center for epidemiological studies depression scale-chinese (cesd-chinese): Data from the china family panel studies (cfps).

Bower, P., Kontopantelis, E., Sutton, A., Kendrick, T., Richards, D. A., Gilbody, S., Knowles, S., Cuijpers, P., Andersson, G., Christensen, H., et al. (2013). Influence of initial severity of depression on effectiveness of low intensity interventions: meta-analysis of individual patient data. *BMJ*, 346:f540.

Brewer, M., Dang, T., and Tominey, E. (2024). Universal credit: Welfare reform and mental health. *Journal of Health Economics*, page 102940.

Cassidy, M., Currie, J., Glied, S., and Howland, R. E. (2025). Universal access to counsel, housing court filings, and child mental health: Evidence from new york city. In *AEA Papers and Proceedings*, volume 115, pages 96–102. American Economic Association 2014 Broadway, Suite 305, Nashville, TN 37203.

Chen, S., Oliva, P., and Zhang, P. (2024). Air pollution and mental health: evidence from china. In *AEA Papers and Proceedings*, volume 114, pages 423–428. American Economic Association 2014 Broadway, Suite 305, Nashville, TN 37203.

Costantini, S. (2024). How do mental health treatment delays impact long term mortality? *American Economic Review*.

de Bekker-Grob, E. W., Ryan, M., and Gerard, K. (2012). Discrete choice experiments in health economics: a review of the literature. *Health economics*, 21(2):145–172.

Dias, M. and Fontes, L. F. (2024). The effects of a large-scale mental health reform: evidence from brazil. *American Economic Journal: Economic Policy*, 16(3):257–289.

Ding, H. (2023). Geographic variation in mental health treatment utilization: evidence from migration. *Available at SSRN 4487660*.

Feyman, Y., Figueroa, S. M., Yuan, Y., Price, M. E., Kabdiyeva, A., Nebeker, J. R., Ward, M. C., Shafer, P. R., Pizer, S. D., and Strombotne, K. L. (2023). Effect of mental health staffing inputs on suicide-related events. *Health services research*, 58(2):375–382.

Finkelstein, A., Taubman, S., Wright, B., Bernstein, M., Gruber, J., Newhouse, J. P., Allen, H., Baicker, K., and Oregon Health Study Group, t. (2012). The oregon health insurance experiment: evidence from the first year. *The Quarterly journal of economics*, 127(3):1057–1106.

Furukawa, T. A., Kessler, R. C., Slade, T., and Andrews, G. (2003). The performance of the k6 and k10 screening scales for psychological distress in the australian national survey of mental health and well-being. *Psychological medicine*, 33(2):357–362.

Gruber, J., Lin, M., and Yi, J. (2023). The largest insurance program in history: Saving one million lives per year in china. *Journal of Public Economics*, 226:104999.

Hainmueller, J. (2012). Entropy balancing for causal effects: A multivariate reweighting method to produce balanced samples in observational studies. *Political analysis*, 20(1):25–46.

Harrell, B., Fumarco, L., Button, P., Schwegman, D. J., and Denwood, K. (2023). The impact of covid-19 on access to mental health care services. In *AEA Papers and Proceedings*, volume 113, pages 420–422. American Economic Association 2014 Broadway, Suite 305, Nashville, TN 37203.

Huang, Y., Wang, Y., Wang, H., Liu, Z., Yu, X., Yan, J., Yu, Y., Kou, C., Xu, X., Lu, J., et al. (2019). Prevalence of mental disorders in china: a cross-sectional epidemiological study. *The lancet psychiatry*, 6(3):211–224.

Jackson, L., Al-Janabi, H., Roberts, T., and Ross, J. (2021). Exploring young people's preferences for sti screening in the uk: a qualitative study and discrete choice experiment. *Social Science & Medicine*, 279:113945.

Jorm, A. F. and Oh, E. (2009). Desire for social distance from people with mental disorders. *Australian & New Zealand Journal of Psychiatry*, 43(3):183–200.

Kessler, R. C., Barker, P. R., Colpe, L. J., Epstein, J. F., Gfroerer, J. C., Hiripi, E., Howes, M. J., Normand, S.-L. T., Manderscheid, R. W., Walters, E. E., et al. (2003). Screening for serious mental illness in the general population. *Archives of general psychiatry*, 60(2):184–189.

Luo, C.-y., Liu, Z., and Song, H. (2023). The impact of economic growth slowdown on public mental health-evidence from households in china. *China Economic Quaterly*, 2:604–621.

Patel, V., Saxena, S., Lund, C., Thornicroft, G., Baingana, F., Bolton, P., Chisholm, D., Collins, P. Y., Cooper, J. L., Eaton, J., et al. (2018). The lancet commission on global mental health and sustainable development. *The lancet*, 392(10157):1553–1598.

Radloff, L. S. (1977). The ces-d scale: A self-report depression scale for research in the general population. *Applied psychological measurement*, 1(3):385–401.

Rüsch, N., Brohan, E., Gabbidon, J., Thornicroft, G., and Clement, S. (2014). Stigma and disclosing one's mental illness to family and friends. *Social psychiatry and psychiatric epidemiology*, 49(7):1157–1160.

Schwandt, H. (2018). Wealth shocks and health outcomes: Evidence from stock market fluctuations. *American Economic Journal: Applied Economics*, 10(4):349–377.

Shafer, P. R., Yuan, Y., Feyman, Y., Price, M. E., Kabdiyeva, A., Figueroa, S. M., Shen, Y.-J., Nebeker, J. R., Ward, M. C., Strombotne, K. L., et al. (2024). Effect of mental health staffing inputs on initiation of care among recently separated veterans. *Health Services Research*, 59:e14333.

Singla, D. R., Kohrt, B. A., Murray, L. K., Anand, A., Chorpita, B. F., and Patel, V. (2017). Psychological treatments for the world: lessons from low-and middle-income countries. *Annual review of clinical psychology*, 13(1):149–181.

Solomon, K. T. and Dasgupta, K. (2022). State mental health insurance parity laws and college educational outcomes. *Journal of health economics*, 86:102675.

Thornicroft, G. (2008). Stigma and discrimination limit access to mental health care. *Epidemiology and Psychiatric Sciences*, 17(1):14–19.

Vlassopoulos, M., Siddique, A., Rahman, T., Pakrashi, D., Islam, A., and Ahmed, F. (2024). Improving women's mental health during a pandemic. *American Economic Journal: Applied Economics*, 16(2):422–455.

Wang, S. and Yang, D. Y. (2025). Policy experimentation in china: The political economy of policy learning. *Journal of Political Economy*, 133(7):000–000.

World Health Organization (2025). Mental health.

Xie, Y. and Hu, J. (2014). An introduction to the china family panel studies (cfps). *Chinese sociological review*, 47(1):3–29.

A Appendix

Table 1: Downstream treatment effects

	(1)	(2)	(3)	(4)
	Self-rated health status	Health status relative to last year	Smoked in the last month	Drank 3 times a week last month
Treat × Post	0.032	0.011	-0.010**	-0.007
	(0.021)	(0.009)	(0.004)	(0.006)
Observations	122,711	127,021	127,528	127,522
R-squared	0.489	0.483	0.851	0.642
Individual FE	YES	YES	YES	YES
Survey Wave FE	YES	YES	YES	YES
Environmental Control	YES	YES	YES	YES
Control mean	3.145	1.769	0.314	0.155
Std Dev	1.298	0.620	0.464	0.362

Notes: Standard errors are clustered at city level and are in parentheses. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

Table 2: Summary statistics: Survey respondents

Variable	Obs	Mean	Std. Dev.	Min	Max
Seconds spent filling the survey	310	287.613	157.054	73	1004
Gender (male = 1)	310	0.368	0.483	0	1
Age (years old)	310	31.932	7.035	18	62
Married	310	0.735	0.442	0	1
Unemploy	310	0.003	0.057	0	1
Third industry	310	0.016	0.126	0	1
Industry 2	310	0.387	0.488	0	1
Industry 3	310	0.516	0.501	0	1
Student	310	0.074	0.263	0	1
Edu yr	310	16.226	1.188	12	22
Agricultural hukou	310	0.474	0.500	0	1
Annual household income (10,000 yuan)	310	21.993	21.661	1	300
Working hours per day (hr)	310	8.426	1.312	0	15
Family size	310	3.723	1.129	1	8
Hours taken from recovering from illness	309	4.343	7.493	0	55
Received mental healthcare consultation	310	0.297	0.458	0	1
Diagnosed with mental illness	92	0.337	0.475	0	1
Mental counseling in psychiatry	31	0.613	0.495	0	1
Mental counseling in community centers	31	0.258	0.445	0	1
Mental counseling in professional settings	31	0.677	0.475	0	1
Mental counseling in other places	31	0	0	0	0
Cost of mental counseling (yuan)	31	506.71	1404.329	20	8000
Rate: quality of experienced mental health services	31	3.903	0.79	2	5
CESD-8 total score	310	10.106	2.554	4	19
Severe mental disorder	310	0.919	0.273	0	1
Standardized values of score	310	-0.008	0.996	-2.391	3.462
% People who discriminate against mental illness	310	34.313	25.697	0	99
Helpfulness of mental consultation	310	4.090	0.807	1	5

Note: This table reports summary statistics for survey respondents.

Table 3: Difference between piloting and regular cities at baseline (2018)

Variable	(1) Mean Control	(2) Mean Treat	(3) Treat - Control	(4) Std. Diff.
Total number of registered mental consultation organizations	78.305	310.636	232.331**	0.342
	(345.080)	(585.111)	(90.359)	
GDP per capita (yuan)	56,723.922	82,740.953	26,017.031***	0.494
	(31,906.062)	(41,905.527)	(6,603.822)	
Unemployment rate (Registered unemployed/total population)	0.006	0.007	0.001**	0.235
	(0.004)	(0.004)	(0.001)	
Internet penetration rate (#households with access to internet/total population)	0.289	0.350	0.061***	0.340
	(0.122)	(0.132)	(0.021)	
Observations	236	44	285	

Note: Std Diff = $\frac{\bar{X}_{\text{treatment}} - \bar{X}_{\text{control}}}{\text{SE}_{\text{cluster}}}$. Standard errors are clustered at city level.

Table 4: Mixed logit estimates: Baseline preferences

Attribute and Levels	**Coef.**	**SE**	**p**	**SD**	**SE**	**p**
Price (ref: ¥500)						
¥300	0.346	0.130	0.008	0.971	0.159	0.000
¥100	1.478	0.124	0.000	0.755	0.180	0.000
¥0	1.514	0.144	0.000	1.732	0.146	0.000
Access (ref: 100 min)						
30 minutes	0.132	0.101	0.191	0.961	0.116	0.000
0 minutes	0.681	0.104	0.000	0.602	0.171	0.000
Stigma (ref: 90%)						
40%	0.852	0.120	0.000	1.357	0.135	0.000
10%	1.882	0.174	0.000	2.404	0.173	0.000
N			9,240			
Log likelihood			−631.69			
Pseudo R^2			0.111			

Notes: The table reports mixed logit estimates. Columns (2)–(4) show mean coefficients; columns (5)–(7) show standard deviations of random parameters. All attribute levels are effects-coded relative to the reference level (highest cost/burden).

Table 5: Conditional logit results: add three-way interaction

	(1)
Price (500 yuan as reference level)	
Price_300	0.205*
	(0.113)
Price_100	0.531***
	(0.156)
Price_0	0.875***
	(0.150)
Price_300 × Bad	0.183
	(0.163)
Price_100 × Bad	0.575***
	(0.195)
Price_0 × Bad	0.305*
	(0.175)
Access (90 min as reference level)	
Access_30	0.221***
	(0.081)
Access_0	0.331
	(0.219)
Access_30 × Bad	-0.045
	(0.107)
Access_0 × Bad	0.217
	(0.299)
Access_0 × Stigma_40	0.250
	(0.247)
Access_0 × Stigma_10	0.534*
	(0.306)
Access_0 × Stigma_40 × Bad	-0.729**
	(0.355)
Access_0 × Stigma_10 × Bad	-0.817*
	(0.431)
Stigma (90% as reference level)	
Stigma_40	0.745***
	(0.133)
Stigma_10	1.225***
	(0.181)
Stigma_40 × Bad	-0.048
	(0.167)
Stigma_10 × Bad	-0.003
	(0.199)
Observations	9,240

Notes: Standard errors are clustered at respondent level and are in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table 6: Difference in COVID-related outcomes in 2020

Variable	(1) Mean Control	(2) Mean Treat	(3) Treat - Control	(4) Std. Diff.
COVID19 mortality rate (‱)	0.009	0.004	-0.001	-0.120
	(0.045)	(0.014)	(0.002)	
Cumulative confirmed cases	131.313	152.556	7.856	0.031
	(621.803)	(301.669)	(19.060)	
Cumulative cured cases	124.929	146.600	6.553	0.034
	(563.990)	(285.816)	(18.579)	
Cumulative death counts	2.972	1.689	-0.040	-0.078
	(15.355)	(5.728)	(0.345)	
Observations	281	45	326	

Notes: Standard errors are clustered at city level and are in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table 7: Difference between survey and CFPS sample (high risk/severe group)

	(1) DCE Sample (High Risk)	(2) CFPS Sample (High Risk)	(3) Difference
Age (years old)	31.989	51.412	19.423***
	(7.173)	(14.099)	(0.000)
Years of education	16.219	6.184	-10.035***
	(1.171)	(5.009)	(0.000)
Gender (male = 1)	0.367	0.371	0.004
	(0.483)	(0.483)	(0.890)
Married	0.731	0.815	0.084***
	(0.444)	(0.388)	(0.000)
Observations	283	42,401	42,684

Note: Standard deviations in parentheses for columns (1) and (2). p-values in parentheses for column (3). *** p¡0.01.

Table 8: Robustness check: Heterogeneous treatment effects by LOO group (with EB weights)

	Low Risk		Middle Risk		High Risk	
	(1) Score	(2) High risk of caseness	(3) Score	(4) High risk of caseness	(5) Score	(6) High risk of caseness
DID	-0.097*	-0.097*	-0.015	-0.015	-0.020	-0.020
	(0.058)	(0.058)	(0.016)	(0.016)	(0.018)	(0.018)
Observations	28,849	28,849	32,765	32,765	35,152	35,152
R-squared	0.898	0.898	0.805	0.805	0.738	0.738
Individual FE	YES	YES	YES	YES	YES	YES
Survey Wave FE	YES	YES	YES	YES	YES	YES
Environmental Control	YES	YES	YES	YES	YES	YES

Notes: This table reports the results of the heterogeneous treatment effects re-estimated using Entropy Balancing weights. Standard errors are clustered at the city level and reported in parentheses. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

Table 9: Robustness check

	Add baseline × year		LTCI		CESD score only		EB	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Dep. Var.	Score	Caseness	Score	Caseness	Score	Caseness	Score	Caseness
DID	-0.050**	-0.030***	-0.065**	-0.036***	-0.073**	-0.035**	-0.066	-0.028*
	(0.025)	(0.011)	(0.029)	(0.013)	(0.029)	(0.014)	(0.066)	(0.016)
Observations	126,200	126,200	127,201	127,201	79,172	79,172	121,921	121,921
R-squared	0.549	0.476	0.548	0.473	0.645	0.551	0.804	0.757
Individual FE	YES	YES	YES	YES	YES	YES	YES	YES
Survey Wave FE	YES	YES	YES	YES	YES	YES	YES	YES
Environmental Control	YES	YES	YES	YES	YES	YES	YES	YES

Notes: Standard errors are clustered at city level and are in parentheses. Variable abbreviations: Score = Standardized mental health score; Caseness = High risk of caseness. LTCI = Long-term Care Insurance. EB = Entropy Balance Weights. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table 10: Robustness Check: Heterogeneity by COVID-19 Lockdown Stringency

	Standardized score	Caseness
	(1)	(2)
Treat × Post × Lockdown	−0.007	0.012
	(0.038)	(0.018)
Treat × Post	−0.062*	−0.051***
	(0.037)	(0.016)
Observations	127,201	127,201
Individual FE	YES	YES
Survey Wave FE	YES	YES
Environmental Control	YES	YES

Notes: This table presents the results from estimating the following triple-difference specification:

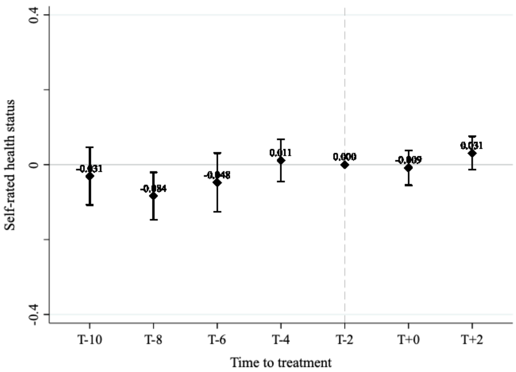
$$y_{ict} = \alpha + \beta(\text{Treat}_c \times \text{Post}_t) + \beta_1(\text{Treat}_c \times \text{Post}_t \times \text{Lockdown}_c) + \gamma i + \eta_t + \epsilon_{ict}$$

In this model, the first difference arises from comparing pilot cities with non-pilot cities, capturing the differential impact of the treatment. The second difference contrasts the mental health outcomes in cities with varying levels of COVID-19 stringency, highlighting the impact of lockdown conditions. The third difference considers the timing of policy implementation, distinguishing between the pre- and post-implementation periods. We include individual and survey year fixed effects in our model. These fixed effects account for all unobserved variables that vary at the individual level (which inherently includes city-level effects, as individuals in our sample do not relocate across cities) and those that vary at the survey year level. Standard errors are clustered at city level and are in parentheses. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

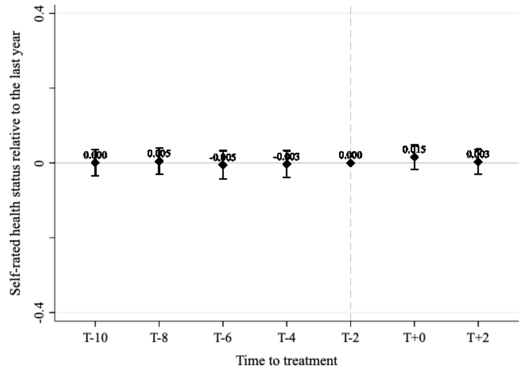
Table 11: Mediation: Regression with the number of counseling organizations as control

	(1)	(2)
	Score	Caseness
Treat × Post	-0.042	-0.014
	(0.030)	(0.014)
Observations	61,622	61,622
Individual FE	YES	YES
Survey Wave FE	YES	YES
Environmental Control	YES	YES
# Counseling organizations	YES	YES

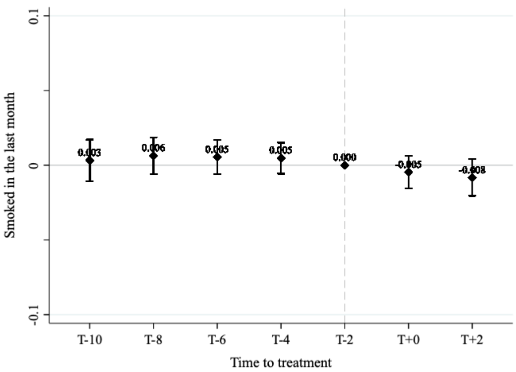
Notes: Standard errors are clustered at city level and are in parentheses. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.



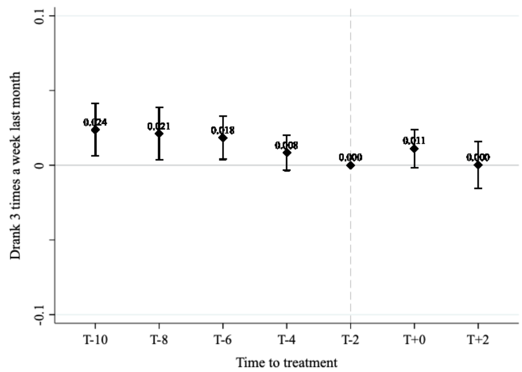
(a) Self-rated health status



(b) Self-rated health status relative to the last year



(c) Smoked last month



(d) Drank more than 3 times last month

Notes: The figure presents the estimated coefficients (β_k) from an event study model derived from Equation 2: $y_{ict} = \alpha + \gamma_i + \lambda_t + \mathbf{Envi'_{ct}}\eta + \sum_{n=-10}^{2} \beta_n Plan_{ct}^n + \epsilon_{ict}$. The coefficients show the effect of the Plan on the mental health outcome in event time k, relative to the reference year (the reference period, $k = -2$). The panel title represents the outcome of interest. The vertical bars represent the 95% confidence intervals, calculated using standard errors clustered at the city level. Despite the DID estimates in Column (3) of Table 1 show a statistically significant negative average effect of the treatment on "Smoked in the last month" at the 5% level ($\beta = -0.010$), the corresponding event study plot (Figure 1) indicates that none of the individual post-treatment coefficients ($k \geq 0$) are statistically significant at the 10% level. We argue that this finding does not constitute a contradiction, but rather reflects the dynamic characteristics of the effect and the differences in the statistical power of the two specifications. The treatment generated a sustained but dispersed small effect. This effect achieves statistical significance when aggregated into the overall average DID estimate, but it lacks the necessary statistical precision when tested on a survey year basis. Importantly, the lack of significance for all pre-treatment coefficients in Panel (c) validates the parallel trends assumption.

Figure 1: Downstream effects

64. Based on the following scenario, would you choose to receive psychological counseling?

	Option A	Option B
Cost (RMB/hour)	100	300
Travel Time (minutes, round trip)	30	0
Perceived Social Stigma Rate: Receiving mental health counseling will make it known to those around you. Approximately X% of the population are likely to hold stigmatizing attitudes toward individuals with mental health conditions.	10%	90%

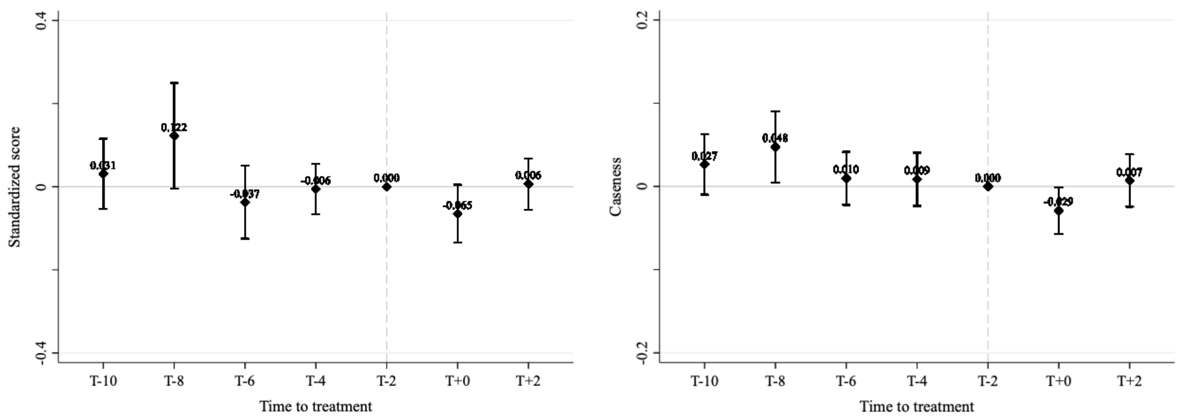
[Single Choice] *

○ Option A

○ Option B

○ Neither

Figure 2: An example of choice set

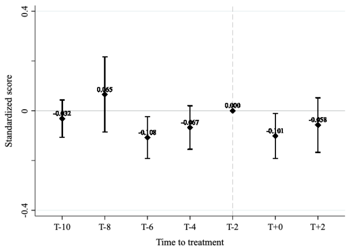


(a) Standardized score

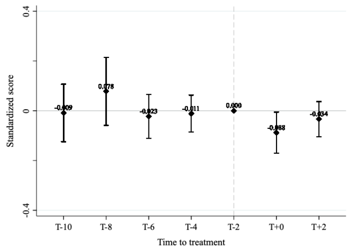
(b) Caseness

Notes: The figure presents the estimated coefficients (β_k) from an event study model derived from $y_{ict} = \alpha + \gamma_i + \lambda_t + \mathbf{Envi_{ct}}\eta + \sum_{n=-10}^{2} \beta_n Plan_{ct}^n + \mathbf{X_{cp0}} \times \lambda_\mathbf{t}\psi + \epsilon_{ict}$. The coefficients show the effect of the Plan on the mental health outcome in event time k, relative to the reference year (the reference period, $k = -2$). The panel title represents the outcome of interest. The vertical bars represent the 95% confidence intervals, calculated using standard errors clustered at the city level.

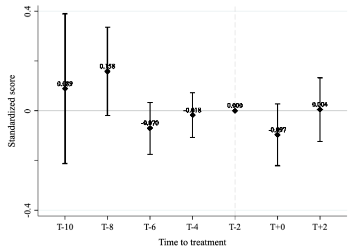
Figure 3: Dynamic effects of the Plan on mental health: add time trend of city baseline features



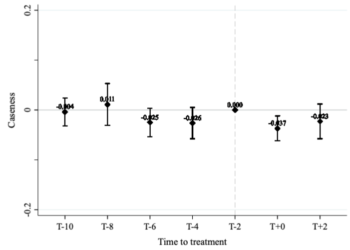
(a) Low risk of caseness



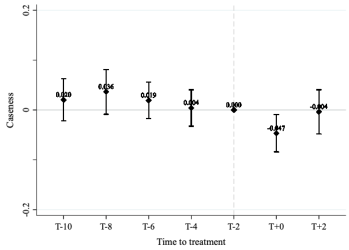
(b) Middle risk of caseness



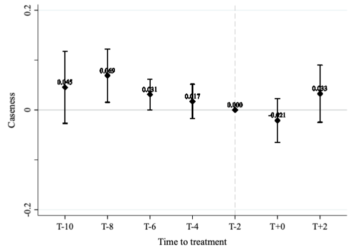
(c) High risk of caseness



(d) Low risk of caseness



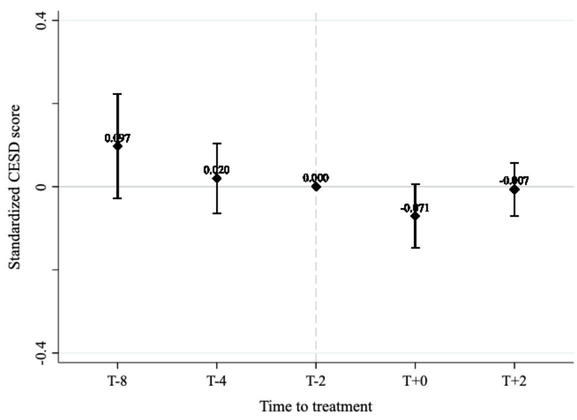
(e) Middle risk of caseness



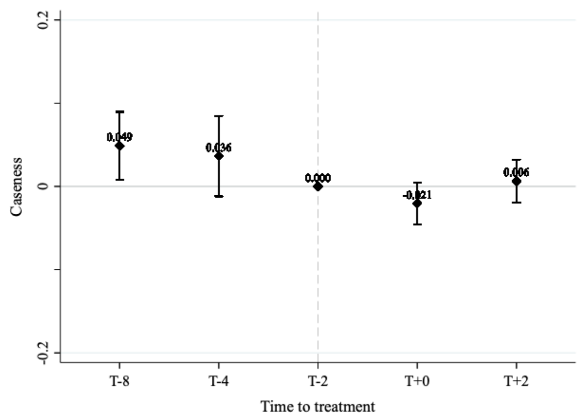
(f) High risk of caseness

Notes: The figure presents the estimated coefficients for LOO subgroups (β_k) from an event study model derived from $y_{ict} = \alpha + \gamma_i + \lambda_t + \mathbf{Envi_{ct}}\eta + \sum_{n=-10}^{2} \beta_n Plan_{ct}^n + \mathbf{X_{cp0}} \times \boldsymbol{\lambda_t}\psi + \epsilon_{ict}$. The coefficients show the effect of the Plan on the mental health outcome in event time k, relative to the reference year (the reference period, $k = -2$). The title on the left of each plot represents the outcome of interest. The panel title represents the corresponding subgroup. The vertical bars represent the 95% confidence intervals, calculated using standard errors clustered at the city level.

Figure 4: Heterogeneous treatment effects: add time trend of city baseline features



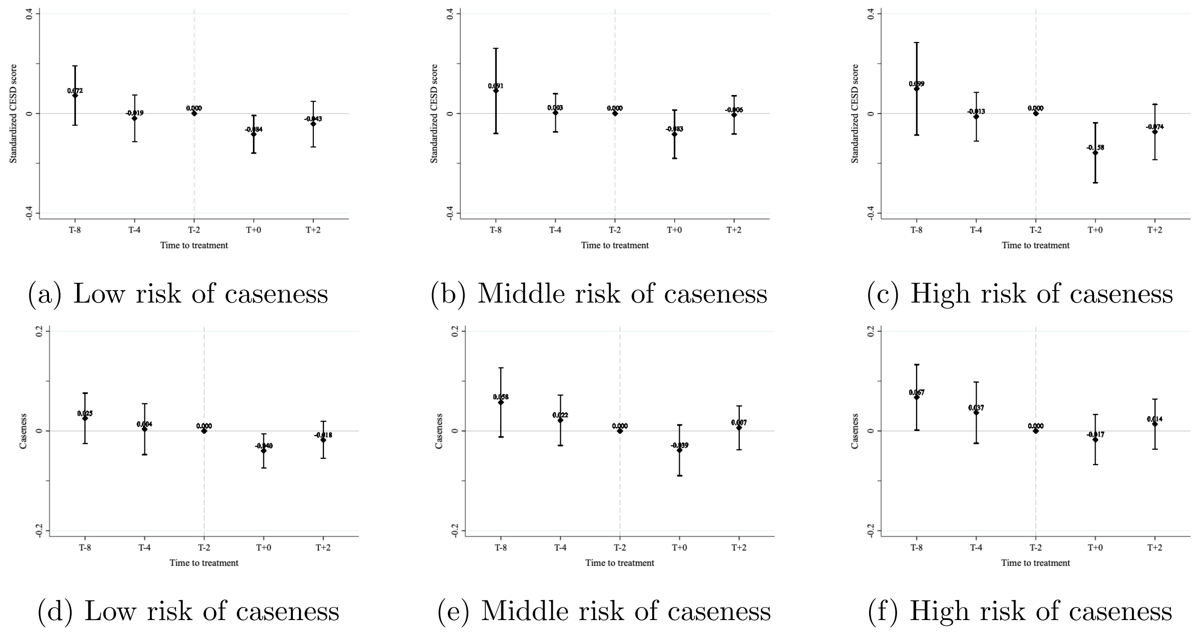
(a) Standardized score



(b) High risk of severe disorder

Notes: The figure presents the estimated coefficients (β_k) from an event study model derived from $y_{ict} = \alpha + \gamma_i + \lambda_t + \mathbf{Envi_{ct}}\eta + \sum_{n=-10}^{2} \beta_n Plan_{ct}^n + \epsilon_{ict}$. The coefficients show the effect of the Plan on the standardized CESD score in event time k, relative to the reference year (the reference period, $k = -2$). The panel title represents the outcome of interest. The vertical bars represent the 95% confidence intervals, calculated using standard errors clustered at the city level.

Figure 5: Dynamic effects of the Plan on mental health measured by CESD: Full sample



(a) Low risk of caseness (b) Middle risk of caseness (c) High risk of caseness

(d) Low risk of caseness (e) Middle risk of caseness (f) High risk of caseness

Notes: The figure presents the estimated coefficients (β_k) from an event study model derived from $y_{ict} = \alpha + \gamma_i + \lambda_t + \mathbf{Envi_{ct}}\eta + \sum_{n=-10}^{2} \beta_n Plan_{ct}^n + \epsilon_{ict}$. The coefficients show the effect of the Plan on the standardized CESD score in event time k, relative to the reference year (the reference period, $k = -2$). The panel title represents the outcome of interest. The vertical bars represent the 95% confidence intervals, calculated using standard errors clustered at the city level.

Figure 6: Heterogeneous treatment effects: mental health outcomes based on CESD score