NBER WORKING PAPER SERIES

THE EFFECTS OF PRIMARY CARE CHRONIC-DISEASE MANAGEMENT IN RURAL CHINA

Yiwei Chen
Hui Ding
Min Yu
Jieming Zhong
Ruying Hu
Xiangyu Chen
Chunmei Wang
Kaixu Xie
Karen Eggleston

Working Paper 26100
http://www.nber.org/papers/w26100

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
July 2019, Revised October 2021

This research was made possible by the Tongxiang CDC and local government, which provided data instrumental to this research. We would like to express gratitude to the Stanford University Freeman Spogli Institute for International Studies' Policy Implementation Lab and a Shorenstein Asia Pacific Research Center faculty research award for funding this project. Yiwei Chen also gratefully acknowledges funding from the Asia-Pacific Scholars program of Stanford University Freeman Spogli Institute. Hui Ding and Yiwei Chen contributed equally to this work and should be considered joint first author of this paper. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

The Effects of Primary Care Chronic-Disease Management in Rural China
Yiwei Chen, Hui Ding, Min Yu, Jieming Zhong, Ruying Hu, Xiangyu Chen, Chunmei Wang, Kaixu Xie, and Karen Eggleston
NBER Working Paper No. 26100
July 2019, Revised October 2021
JEL No. I11,I18

ABSTRACT

Health systems globally face increasing morbidity and mortality from chronic diseases, yet many - especially in low- and middle-income countries - lack strong chronic disease management in primary health care (PHC). We provide evidence on China's efforts to promote PHC management using unique five-year panel data in a rural county, including health care utilization from medical claims and health outcomes from biomarkers. Utilizing plausibly exogenous variation in management intensity generated by administrative and geographic boundaries, we compare hypertension/diabetes patients in villages within two kilometers distance but managed by different townships. Results show that, compared to patients in townships with median management intensity, patients in high-intensity townships have 4.8% more PHC visits, 5.2% fewer specialist visits, 11.7% fewer inpatient admissions, and 3.6% lower medical spending. They also tend to have better medication adherence and better control of blood pressure. The resource savings from avoided inpatient admissions substantially outweigh the costs of the program.

Yiwei Chen
Stanford University
Economics Department
yiwei312@stanford.edu

Hui Ding
Stanford University
khding@stanford.edu

Min Yu
Zhejiang Provincial Center for
Disease Control and Prevention
Hangzhou
PRC
myu@cdc.zj.cn

Jieming Zhong
Zhejiang Provincial Center for
Disease Control and Prevention
Hangzhou
PRC
jmzhong@cdc.zj.cn

Ruying Hu
Zhejiang Provincial Center for
Disease Control and Prevention
3399 Binsheng Road
Hangzhou
PRC
ryhu@cdc.zj.cn

Xiangyu Chen
Zhejiang Provincial Center for
Disease Control and Prevention
3399 Binsheng Road
Hangzhou
PRC
xychen@cdc.zj.cn

Chunmei Wang
Tongxiang Center for Disease Control
and Prevention
64 Maodun E Rd
Tongxiang
PRC
503156112@qq.com

Kaixu Xie
Tongxiang Center for Disease Control
and Prevention
64 Maodun E Rd
Tongxiang
PRC
154006180@qq.com

Karen Eggleston
Shorenstein Asia-Pacific Research Center
Stanford University
616 Serra Street
Stanford, CA 94305
and NBER
karene@stanford.edu

1 Introduction

Primary health care (PHC) has often been held up as vital for health systems globally to cope with many health challenges, from control of new and re-emerging infectious diseases to helping individuals prevent and manage chronic conditions. Empirical research documents a strong positive association between the "strength" of PHC and various positive health system outcomes, including better health at lower per capita cost (Scott and Jan, 2011). The efficiency-enhancing promise of quality PHC may be especially important for non-communicable chronic diseases (NCDs) such as hypertension and diabetes. Globally, NCDs account for about two-thirds of deaths and an increasing burden of morbidity that reduces productivity, increases medical spending, and shortens lives (Bloom et al., 2012). Regular monitoring of blood pressure and blood sugar and adherence to lifestyle modifications and/or medication and other therapies to control such biomarkers can reduce premature mortality from common chronic conditions like hypertension, cardiovascular disease, and diabetes. Indeed, a lower rate of avoidable admissions for NCDs is often used as an indicator of high-quality PHC.[1]

Yet many health systems — especially in low- and middle-income countries (LMIC) — lack strong PHC systems, blighting lives and livelihoods, contributing to loss of productivity, premature mortality, and growing medical spending that may undercut the sustainability of universal health coverage. How can policymakers improve PHC to increase the productivity of health systems? Not every village can host a tertiary hospital. The logic of economies of scale and scope as well as specialization push toward a referral system with coordination between PHC and other levels of care. Excellent PHC requires knowledge of common conditions, competent differential diagnosis, and knowing when to refer a patient for more specialized care. Many settings analyzed by health economists involve a similar trade-off. For example, does the quality benefit of a "center of excellence" or specialist treatment outweigh the convenience of PHC? To improve maternal health outcomes in poor countries, should we promote safer home births or universal hospital-based birth? Even in high income countries – by definition facing less stringent resource constraints than LMIC – controversy continues to swirl about the optimal model for emergency services, characterized as "stay and play" or "scoop and run" (Brun et al., 2014).

This quality-convenience trade-off is all the more of a conundrum in LMIC, where the quality and the productivity of PHC may be especially problematic. Patients eschewing perceived low quality may lead

[1] For the health care quality indicators developed for comparing the quality of PHC across countries, see for example the methods and descriptions at the OECD website https://www.oecd.org/els/health-systems/hcqi-primary-care.htm.

to village doctors sitting idle, and even those who are in demand may perform below their potential (the "know-do gap" (Mohanan et al., 2015). Interventions that increase health worker effort can clearly improve patient outcomes (Okeke and Abubakar, 2020). Yet there is little rigorous evidence from LMICs about the effectiveness of programs seeking to inspire such effort to better leverage the latent capacity of PHC for controlling chronic disease. In this study, we provide empirical evidence on the effectiveness of a program promoting PHC management of hypertension and diabetes for rural Chinese.

China provides an important case study, as a large and rapidly developing middle-income country once famous for its "barefoot doctors" but now with a hospital-based service delivery system for its aging population. PHC use in China has decreased relative to hospital-based care since 2009 national health reforms (Chen, 2009) despite policy designed to reduce crowding at hospitals, including government investments in strengthening PHC infrastructure and training, as well as more recent efforts toward implementing a "family doctor system".[2] The relative decrease in utilization of grassroots providers is a natural consequence of the unprecedented increases in living standards and improvements in financial risk protection, which increase patients' effective demand for quality care and spur self-referral to providers with higher perceived quality like hospitals. If perceived quality improvements of grassroots providers do not exceed those of hospital-based care, continued crowding at hospitals would be a predicted – albeit unintended – effect of China's reforms. Underutilized resources at the village and township levels embody a stubborn inefficiency that few policies have proven effective in remedying. According to the 2019 national health statistical yearbook, the 2018 average occupancy rate at township health centers (THCs) was 59.6%, compared to 89.8% at county-level hospitals and 100.6% at provincial-level hospitals. These low occupancy rates in PHC have changed little over the past decade, despite the expansion of capacity and double-digit growth rates in overall healthcare utilization.

Patients in China do have reason to be skeptical of the quality of PHC.[3] Most of the quality resources in medical care, both human capital and modern diagnostic and therapeutic technologies, have been concentrated in China's hospitals, which traditionally operate large outpatient departments. Both patients and providers frequently perceive hospital-based care to be of higher quality (Wu et al., 2017), including for the control of chronic disease (Yang et al., 2008; Lu et al., 2017). Indeed, the poor performance of China's PHC

[2]Nationally as utilization increased, the percentage of healthcare visits at the PHC level declined from over 63% in 2005 to less than 55% by 2017 (Statistical yearbooks, various years; `http://www.nhfpc.gov.cn/guihuaxxs/s10743/201806/44e3cdfe11fa4c7f928c879d435b6a18.shtml`). Also see Wu and Lam (2016).

[3]Li et al. (2020) summarize evidence on the quality of China's PHC and discuss several hypotheses for the causes of poor quality.

in the control of hypertension and diabetes features in a recent review of the national PHC system (Li et al., 2017).

The performance differences between PHC and hospital-based care are especially stark in rural areas. Rural residents are more likely to live with the condition without any treatment until complications develop because under-diagnosis of hypertension and diabetes is higher in rural areas (Lei et al., 2014; Zhao et al., 2016; Lu et al., 2017); and rural residents are less likely to be under control even when diagnosed (Xu et al., 2013; Lu et al., 2017). Successful PHC management of NCDs in rural areas relies heavily on grassroots physicians, who have limited medical education and training. In 2012, 84% of China's rural doctors did not have a college degree, compared to 60% in urban areas (Meng et al., 2015). The questionable quality of PHC may be one reason why China's decades-long rhetorical embrace of PHC has shied away from any imposition of gatekeeping or mandatory referrals, but rather allowed patients freely to self-refer to higher-level providers according to ability and willingness to pay. China's universal health coverage system through local monopoly social health insurance does generally provide higher reimbursement rates and low copayment requirements for PHC, but the continued crowding at hospital outpatient departments reveals patients' ongoing skepticism that PHC is an adequate substitute for hospital-based physician expertise.[4]

In light of this important and challenging context of PHC services in China, we provide new empirical evidence about the impact of the NCD-management component of the National Basic Public Health Service Program for rural Chinese. This program, launched as part of national health reforms in 2009, financially rewards PHC ("grassroots") physicians for managing local residents with chronic diseases. To study its effectiveness, we assemble a unique dataset linking administrative and health data between 2011 and 2015 at the individual level for rural Chinese diagnosed with hypertension or diabetes in a mostly rural county of Zhejiang province in southeast China.

We utilize variation in management intensity generated by administrative boundaries to study the program's effects on healthcare utilization, spending, and health outcomes. By comparing residents in neighboring villages that straddle township boundaries and thus are subject to different PHC management, we evaluate the effect of PHC management on health care utilization, medication use, and health status. From a sample of about 70,000 individuals with hypertension or diabetes across the county, we first measure each township's management intensity using the average number of years the residents have been enrolled in PHC management, controlling for residents' sociodemographic characteristics and physician capacity in

[4]For more on China's system of social health insurance and health service delivery, see Burns and Liu (2017).

each township. This management intensity index reflects the cumulative efforts of PHC physicians to screen their communities and recruit and retain patients within the PHC management programs for hypertension and diabetes. Then we derive our estimates of the effect of PHC intensity by focusing on the subsample of residents of villages that are within two kilometers of each other, but have PHC services managed by different townships. That is, we compare all pairs of villages in the county that are within two kilometers of each other but on opposite sides of a township border. This "border sample" of slightly over 12,000 rural Chinese residing in 14 pairs of border-straddling villages is balanced across observable population characteristics such as age, gender, and educational attainment, and their residents enjoy identical insurance coverage and hospital access. Each township's experience with PHC management over the 5-year study period is a case study for rural China. Focusing on neighboring border-straddling villages allows us to use only variation in PHC management within pairs of neighboring villages to identify the effect. As emphasized by Dube et al. (2010), such an approach generalizes the case study approach, essentially pooling the local comparisons and examining patient-level differences in utilization, spending, and health outcomes from 2011 to 2015, since residents of a neighboring village offer a better-matched control group than residents of all villages within each township.

Utilizing this plausibly exogenous variation, we find that patients residing in a village within a township with more intensive PHC management, compared to neighbors with less intensive management, had a relative increase in PHC visits, fewer specialist visits, fewer hospital admissions and lower spending. Comparing patients in the township with the highest management intensity with those in a median-intensity township, we see 4.8% more PHC visits, 5.2% fewer specialist visits, 11.7% lower likelihood of having an inpatient admission, and overall 3.6% lower medical spending. They also tend to have better medication adherence and better health outcomes as measured by control of blood pressure, especially among those with relatively severe disease (Stage II hypertension). Results are robust to examining differences in health outcomes over time, and to using a "leave-one-out" measure of township management intensity to mitigate any concern that unobserved differences in health demand of adjacent villages may explain the differences in management.

Overall our results suggest that PHC chronic-disease management in rural China can leverage existing resources to increase PHC utilization, decrease the growth rate of medical spending, and reduce avoidable hospitalizations, while not adversely impacting — and perhaps even improving — intermediate- and long-run health outcomes. A back-of-the-envelope estimate of the welfare implications of this program suggests that the resource savings from avoided inpatient admissions substantially outweigh the public subsidy costs of the program, even if we ignore the value of any associated improvements in quality of life and survival.

The remainder of the paper is organized as follows. Section 2 provides background, including a discussion of the related literature and the institutional setting. Section 3 presents our unique datasets and summary statistics of our sample. Section 4 describes our empirical strategy. Sections 5 reports the empirical results. Section 6 discusses the implications and concludes.

2 Background: Related Literature and Institutional Setting

2.1 Related Literature and Our Contribution

Our study contributes to several strands of related literature. First, we contribute to health economics research showing the latent potential of healthcare workers in many LMICs to be more productive (Christensen et al., 2020; Okeke and Abubakar, 2020). Many LMIC health systems suffer from high rates of absenteeism among healthcare workers (Chaudhury et al., 2006). There is also evidence of a "know-do" gap, or a disparity between what healthcare providers know that they should do and what they actually do, in some healthcare systems (Das and Hammer, 2005; Mohanan et al., 2015), and of low quality of PHC management of chronic conditions, including in China (Li et al., 2020). Our study contributes evidence about programs' potential to overcome these challenges and thereby to enhance productivity of PHC. Studies have explored the impact of financial incentives on healthcare workers' performance (Miller and Babiarz, 2013; Huillery and Seban, 2014; Gertler and Vermeersch, 2012) or other organizational and accountability mechanisms, including peer monitoring, unconditional non-monetary gifts, and community-based monitoring (Björkman and Svensson, 2009; Brock et al., 2018; Christensen et al., 2020; Okeke and Abubakar, 2020). Much of the research on the economics of PHC focuses on high-income countries in Europe or North America, although the need for rigorous study designs "is particularly acute in low- and middle-income settings where policies are being introduced that will shape the fundamentals of the future health system on the basis of little empirical evidence" (Scott and Jan, 2011). Therefore evidence from China can make a valuable contribution.

A related strand of literature focuses on financial and reputational incentives for motivating public workers and administrators to improve human capital in LMICs (Duflo et al., 2012; Luo et al., 2020). We find that budget funds earmarked for NCD management paid to townships are passed along to front-line workers in ways that lead to improvement in outcomes for individuals managed at the clinics overseen by that township.

This result is consistent with that of Luo et al. (2020), who also demonstrate how incentives in rural China can spur innovation — specifically, effort and inputs along relevant margins not dictated directly by the incentives. They focus on school administrators and incentives to improve the health of school-age children (primarily to reduce iron-deficiency anemia through supplementation at school and/or persuading parents to change nutrition provided at home). They demonstrate that incentives to Chinese rural administrators can enhance health outcomes. School administrators and health workers face similar civil servant evaluation systems administered by local governments, so both their study and ours confirm that aligning local officials' incentives and accountability mechanisms with health goals can contribute to improved outcomes.

We also contribute to economic assessment of chronic disease management, especially in middle-income countries like China. Our empirical evidence supports the effectiveness of chronic disease management programs as part of broader regional initiatives to address population health. The program we analyze focused on improving population health starting with community screening for NCDs, finding incident cases earlier in the course of disease (i.e., reducing under-diagnosis) and enhancing primary and secondary prevention with management of incident cases. The literature clearly shows the importance of such efforts. High blood pressure was the leading preventable risk factor for premature mortality in China already in 2005 (He et al., 2009). According to a study of 1.7 million Chinese aged 35-75, compared to urban residents, rural residents had slightly higher prevalence of hypertension (46.1%) and significantly lower awareness (43.8%), treatment (28.2%), and control (6.1%) in 2014-2017 (Lu et al., 2017). China is also home to about one-quarter of the world's population with diabetes, with prevalence increasing with age and urbanization, and half or more undiagnosed (Zhao et al., 2016). Among those aware of their conditions, control of blood pressure and blood sugar is still far from optimal (e.g. less than 40% of patients treated for diabetes had adequate glycemic control (Xu et al., 2013)), and spending on care for hypertension and diabetes represents a substantial burden for rural households despite basic health coverage (Liu et al., 2016). All of these factors underscore the importance of our finding that PHC can be effective for chronic disease management in rural China.

2.2 Tongxiang and PHC Management

The project examines the PHC chronic disease management program in Tongxiang, a mostly rural county[5] in Zhejiang province, eastern China. In 2015, the population totaled 687,304 registered residents, of whom 80.99% were rural (agricultural) *hukou* residents and 19.01% were urban (non-agricultural) *hukou* residents. Tongxiang is one of the richest rural counties in China. As of 2015, the annual per capita income for rural *hukou* residents was 27,357 RMB[6] ($4,392 USD based on the 2015 exchange rate[7]). This income level, though high for rural China, represents a fraction of that of high-income countries and thus could be considered representative of emerging market populations beginning to gain access to the living standards of high-middle-income countries.

Like most of rural China, Tongxiang county has provided universal health coverage for almost two decades through social health insurance programs, namely the New Cooperative Medical Scheme (NCMS) for rural residents and Urban Residents Basic Medical Insurance (URBMI) for urban residents not engaged in formal sector employment. Both programs are voluntary, financed by heavily subsidized premiums, and have been merged since before our study period into a single "resident insurance" risk pool (*jumin yibao*). For more details about the insurance programs and their benefit structure, see Appendix A1.

The Tongxiang PHC chronic disease management program is one constituent part of the essential public health services program launched under the broad guidelines of the 2009 national health reforms.[8] Started in 2009, the program expanded gradually over the years in Tongxiang county. Local public hospitals and the Center for Disease Control and Prevention collaborated in developing the program to screen and manage individuals with hypertension and type 2 diabetes. The first stages involved community-wide door-to-door canvassing and screening to identify existing chronic-disease patients and reduce under-diagnosis. Over time, newly diagnosed patients are referred to the program after being diagnosed during a health check-up or hospital visit. Enrollment is voluntary and patients can continue to access care through hospital outpatient departments without PHC management if they so desire. Those who choose to enroll in the program will

[5] The administrative structure in China is such that a given county or municipality has jurisdiction over the surrounding rural areas and residents with agricultural *hukou*.

[6] Source: http://xxgk.tx.gov.cn/xxgk/jcms_files/jcms1/web24/site/art/2016/3/25/art_3620_79724.html. Accessed on November 29 2018.

[7] The exchange rate between US dollars and Chinese RMB in this paper is set to 100 USD = 622.84 RMB, which was the exchange rate in 2015. Source: National Bureau of Statistics of China, http://data.stats.gov.cn/easyquery.htm?cn=C01&zb=A060J&sj=2017. Accessed on November 29, 2018.

[8] See the article by China's Minister of Health (Chen, 2009) and the associated policy announcement at http://www.gov.cn/ztzl/ygzt/content_1661065.htm.

be assigned to a specific PHC physician, called their responsible physician, usually a physician employed at the local public grassroots clinic such as a village clinic or township health center. That doctor is required to meet with each assigned hypertension or diabetes patient quarterly at minimum, record vital statistics, monitor blood pressure and blood glucose, and provide advice regarding medications and lifestyle, at no extra cost to enrolled patients.

Financially, the program is funded by the government essential public health service (*jiben gonggong weisheng fuwu*) budget: 45 RMB (approximately 7.2 USD) per resident, which covers chronic-disease management and other public health programs. The budget is assigned to the local township health centers (THCs). In Tongxiang, those THCs also serve as the employers of all physicians in the surrounding grassroots clinics and are in charge of providing the management service to all enrolled patients in nearby villages.

THCs and their physicians receive incentives to encourage residents to enroll in the management program. For physicians, the chronic-disease management program together with other public health programs constitute job requirements linked to their salaries. For THCs, the administrators of the program, Tongxiang county Center for Disease Control and Prevention (CDC) evaluates each township's performance and rewards each THC both financially and reputationally. Since 2013, the Tongxiang CDC started to evaluate THCs through reviewing and auditing performance annually; each THC's overall performance score represents the cumulative total of their scores for each basic health service, with chronic-disease management accounting for 10% of the total score. The ranking results assigned to each THC are disseminated within the public healthcare system as a reputational incentive. In addition, since 2014, the THCs ranked among the top 3 overall and the bottom 3 overall have a 30% payment difference in their per-capita payments. At least in part because of these incentives and the effort they spur PHC physicians to invest in recruiting and retaining patients, enrollment in the program is high. Around 90% of diagnosed patients are enrolled in the PHC management programs for hypertension and diabetes by 2015 (we will discuss this further in the next section).

3 Data and Summary Statistics

The backbone of the project is the unique administrative data collected by the Tongxiang CDC and Zhejiang CDC. The compiled database links basic demographic information, health insurance claims, PHC service logs, and health check-up records. To our knowledge, these four sets of data are rarely linked and analyzed in combination in China healthcare research.

3.1 Administrative data for the county and each township

The county is divided into twelve townships, home to 209 villages. The basic population database collects demographic information for all residents including age, gender, educational attainment, *hukou* status and residential location (name of village or town of residence). PHC doctors, nurses, and other staff maintain and update these records through community canvassing, screening, and outreach programs. The data cover the entire population.

To focus on chronic disease management for rural adults, we restrict the analytic sample to agricultural *hukou* residents who were age 40 or older in 2015.[9] We do not consider urban *hukou* residents because they have different lifecourse exposures and opportunities, and are a minority within Tongxiang.[10] We also exclude individuals who died during the study period.[11] These sample restrictions leave us 344,784 residents, among whom 85,180 people have been enrolled in the management program.[12]

Our dataset also includes basic administrative information at the township level, including the number of PHC physicians in 2010; urban/rural designation (*jiedao* or *zhen*); and Tongxiang CDC's assessment of each township's performance in providing population health services, including 2013-2016 scores for each township's PHC management programs for hypertension and diabetes.

3.2 Health insurance claims data, 2011-2015

The health insurance claims dataset contains all medical claims during 2011-2015 for individuals who enrolled in the chronic-disease management programs for hypertension or diabetes. Each claim records the date of service, clinic or hospital, any procedures or prescriptions, and payment for each visit or admission. The claims are supplemented by PHC service logs for 2015 which detail every PHC service provided for

[9] Residents with agricultural *hukou* constitute 81.1% of the Tongxiang population, and 61.3% of these rural residents are above age 40.

[10] Urban residents (those with non-agricultural *hukou*) have access to different housing, schooling, local public goods, social protection policies and jobs throughout their lifetimes.

[11] We restrict to people alive by the end of 2015 to observe their chronic disease management duration and medical utilization. Our main results are robust to excluding patients older than age 75 from the sample (as shown in Appendix Table A9), suggesting that any bias from selective survival is of second-order importance. Overall, mortality rates are relatively low (township population weighted average = 1.19%) and similar across townships (s.d. = 0.12%). Our estimates are conservative to the extent that they omit benefits from lower mortality associated with better control of blood pressure or blood sugar.

[12] More specifically, 77,543 residents are enrolled in a hypertension management program; 16,582 residents are enrolled in a diabetes management program; and 8,945 residents are enrolled in both programs.

patients in either the hypertension or diabetes program. These encounters can take various forms including checkups by phone, home visits, and visits to local clinics. The last dataset linked at the individual level consists of health check-up data for individuals who participate in voluntary health check-ups during 2013-2015.[13]

The following key variables are used in the analysis. Demographic variables such as age, gender, and education level come directly from the basic population database. Our individual-level panel data on health care utilization and spending, including number of visits, expenditure, and number of days on medication, are calculated using the medical insurance claims dataset for 2011-2015. Inpatient utilization is measured by an indicator for whether a patient was admitted as an inpatient in that year. The number of visits to specialists and PHC physicians are counted based on service date and provider ID. A specialist visit is defined as a visit to a county-level or higher-level hospital (e.g., municipal or provincial hospital), while a PHC visit is defined as a visit to a THC or a village clinic. Multiple visits on one date are analyzed as a single visit. Total expenditure is decomposed into spending on inpatient admissions (if any), outpatient visits, and prescription drugs. All expenditures are adjusted to 2015 RMB[14].

3.3 Prescription drug claims, 2015

In addition to overall healthcare utilization and spending, we analyze medications from prescription drug claims. Using 2015 health insurance claims data, we measure the number of anti-hypertensive and anti-diabetic medications prescribed and the number of days covered by these medications. Since unfortunately the 2011 claims data use a different version of the drug code and lack the necessary package information to code dosage properly, we can only construct these medication prescription and adherence measures for the year 2015.

Specifically, we extract from the claims data information on 36 anti-hypertensive drugs (Irbesartan, Telmisartan, Hydrochlorothiazide, etc.) and 23 anti-diabetic drugs (Gliclazide, Metformin, Repaglinide, etc.). We code the number of days covered for each medication claim based on the number of pills, milligrams per pill, and milligrams per day. For drugs prescribed on the same date, we take the maximum number of days covered, capped at the number of days until the next prescription, assuming each patient

[13]The collection and analysis of this unique dataset was approved by Institutional Review Boards at both the Zhejiang Provincial CDC and Stanford University, and the data was de-identified prior to analysis.

[14]Conversion of all expenditures to real 2015 RMB used rural residents' Consumer Price Index for health care services. Source: National Bureau of Statistics of China, http://data.stats.gov.cn/easyquery.htm?cn=C01. Accessed on November 29, 2018.

will adhere to the medication properly within the intervening period. We then aggregate up the days covered by each prescription to obtain the total number of days covered by anti-hypertensive and anti-diabetic medications, separately (see Appendix A2 for more details about medication coding).

In 2015, 77% of managed hypertension patients had at least one claim for an anti-hypertensive drug, and 84% of managed diabetes patients had at least one claim for an anti-diabetic drug. Among patients on any medication (i.e., conditional on having at least one corresponding claim), patients managed for hypertension on average had 214 days covered by anti-hypertensive drugs. Covered days were slightly higher among the patients managed for diabetes, with an average of 249 days covered by anti-diabetic drugs (See Table 1 for more summary statistics and Appendix Figure A1 for the distribution of covered days).

3.4 Health check-up data, 2013-2015

Finally, to study metrics of health outcomes for rural Chinese opting into PHC management, we linked our dataset to biomarker measures recorded from health check-up data at the end of the study period, including systolic blood pressure (SBP), diastolic blood pressure (DBP), and fasting blood sugar (FBS). Such linked biomarker data is available for 79.5% of individuals in PHC management programs; we examine the correlates of participation in the voluntary checkups below.

Management of blood pressure is crucial both for hypertensives and individuals with diabetes. Among the latter, control of blood sugar is also critical. Using the two blood pressure measurements, we construct two indicators for hypertension under control: the first, based on the threshold for diagnosis of Stage 1 hypertension, equals one if SBP$<$140 and DBP$<$90, and zero otherwise. The second indicator for hypertension control, based on the threshold for Stage 2 hypertension, equals one if SBP$<$160 and DBP$<$100, and zero otherwise (Lu et al., 2017)). Using the lab values for FBS from the check-up data, we construct an indicator for diabetes under control which equals one if FBS is below 7, and zero otherwise.

For the 69.2% of individuals in PHC management who have records of more than one health check-up between 2013 and 2015, we take the average across the readings and then construct dummy indicators for hypertension/diabetes under control based on that average reading. We also analyze the sub-sample of individuals with panel data on biomarkers from checkups in both 2013 and 2015, a further subsample representing 29.7% of the patients in the health check-up sample. For this biomarker panel sub-sample of individuals, we examine the relationship between more intensive PHC management and change in health outcomes between 2013 and 2015, as measured by control of blood pressure and FBS. Appendix Table A2 presents descriptive statistics for the subsamples of patients with no health checkup, cross-sectional checkup results, and panel data from health checkups.

3.5 Summary Statistics

Table 1 shows the summary statistics of our full sample and border village sample (mentioned in the introduction and discussed in detail in Section 4). There were 73,762 rural patients over the age of 40 enrolled in the hypertension management program, and 15,912 enrolled in the diabetes management program. Among the full sample of those enrolled in the hypertension (diabetes) managed program, average age is 65.1 (61.8), 48% (42%) are male, median education level is primary school, average length of management is 3.6 (3.1) years, and average number of years since diagnosis is 6.5 (5.1) years. Thus, for the majority, our 5-year study period includes observations soon after diagnosis and before enrollment in PHC management.

In terms of medical utilization, among the full sample of patients choosing PHC management for hypertension (diabetes), 13% (18%) had inpatient claims in 2015, spending on average 3,186 (5,330) RMB in total for healthcare in that year. Their number of specialist visits, defined as visits to municipal- or provincial-level hospitals, averaged 2.1 (4.8) in 2015, while the average number of PHC visits, defined as visits to THCs or village clinics, was 11.4 (14.7). As expected, the patients managed for diabetes have greater utilization and spending. For both groups, utilization measures have almost all doubled from 2011 to 2015, reflecting trends in better access to health care as well as aging of the cohort. Furthermore, as noted earlier, 77% of hypertension managed patients had at least one claim for an anti-hypertensive drug in 2015, while 84% of diabetes managed patients had a claim for at least one anti-diabetic medication in 2015. Conditional on having at least one corresponding claim, people on medication on average had 210 (249) days covered by anti-hypertensive (anti-diabetic) drugs. Finally, 46% of hypertension managed patients have their blood pressure under control during their health check-up at the end of the study period, while 43% of diabetes managed patients had their blood sugar under control at such check-ups. Some individuals have both diabetes and hypertension, as they are relatively common co-morbidities. In our sample, about 89% of patients managed for hypertension have normal blood sugar readings, while only 52% of diabetes patients have normal blood pressure readings at their health check-ups at the end of the study period.

One concern regarding the study sample is that, by looking at only patients enrolled in management programs, we are ignoring patients who are diagnosed but not enrolled. Fortunately, we are able to address this selection issue for the diabetes sample by linking the health check-up records to diabetic surveillance

Table 1: Summary Statistics

	Hypertension					Diabetes				
	Full Sample (N=73,762)		Border sample (N=11,865)		T-test	Full Sample (N=15,912)		Border sample (N=2,740)		T-test
	Mean	SD	Mean	SD	P-value	Mean	SD	Mean	SD	P-value
Demographics										
Age (years)	65.1	10.8	65.2	10.8	0.771	61.8	10.2	61.8	10.2	0.675
Male	0.48	0.50	0.49	0.50	0.006	0.42	0.49	0.41	0.49	0.146
Education level (median)	2		2		0.077	2		2		0.076
# Years in management	3.6	1.3	3.6	1.2	0.002	3.1	1.4	3.0	1.3	0.571
# Years since diagnosis	6.5	3.8	6.4	3.9	0.011	5.1	4.1	5.2	3.4	0.106
Utilization, 2015										
Any inpatient admission	0.13	0.34	0.13	0.34	0.140	0.18	0.39	0.16	0.37	0.001
# Specialist visits	2.1	4.1	1.9	3.9	0.000	4.8	6.5	4.3	6.1	0.000
# PHC visits	11.4	9.0	11.8	9.1	0.000	14.7	12.0	15.8	12.6	0.000
Expenditure (RMB)	3186	9718	3129	9569	0.482	5330	13400	4352	8296	0.000
Inpatient expenditure	1875	9047	1819	8869	0.447	2814	12488	1954	7377	0.000
Outpatient expenditure	377	781	386	915	0.255	583	1080	578	1117	0.785
Drug expenditure	933	1848	925	1914	0.600	1933	2625	1820	2072	0.003
Any anti-hypertensive drug	0.77	0.42	0.77	0.42	0.687	0.58	0.49	0.59	0.49	0.408
# Days covered	214	112	213	111	0.331	208	118	209	112	0.649
Any anti-diabetic drug	0.12	0.33	0.13	0.33	0.030	0.84	0.37	0.83	0.38	0.279
# Days covered	225	116	227	116	0.375	249	101	254	99	0.009
Utilization, 2011										
Any inpatient admission	0.07	0.26	0.07	0.26	0.194	0.10	0.30	0.09	0.29	0.074
# Specialist visits	1.5	3.4	1.1	2.7	0.000	3.4	5.5	2.7	4.9	0.000
# PHC visits	6.5	6.5	7.2	6.8	0.000	7.4	7.6	8.3	7.8	0.000
Expenditure (RMB)	1619	5580	1578	5552	0.383	2666	7233	2623	8474	0.764
Inpatient expenditure	820	5049	797	5119	0.598	1219	6464	1248	7728	0.821
Outpatient expenditure	172	455	168	601	0.349	253	512	227	350	0.000
Drug expenditure	627	1380	614	1304	0.219	1194	1865	1149	2061	0.196
Checkup										
FBS under control	0.89	0.32	0.89	0.31	0.350	0.43	0.50	0.44	0.50	0.885
(# Observations)	(55971)		(9066)			(11254)		(1960)		
BP under control (stage 1)	0.46	0.50	0.51	0.50	0.000	0.52	0.50	0.55	0.50	0.000
(# Observations)	(58418)		(9349)			(11897)		(2064)		
BP under control (stage 2)	0.87	0.34	0.90	0.31	0.000	0.88	0.33	0.90	0.29	0.000
(# Observations)	(58418)		(9349)			(11897)		(2064)		

Notes: This table presents summary statistics for patients enrolled in the PHC hypertension and diabetes management programs in our full sample and border village sample. Demographic characteristics include age, gender, education level (1=illiterate, 2=primary school, 3=secondary school, 4=college and above), number of years enrolled in the management program, and number of years since diagnosis. Healthcare utilization is measured in both 2011 and 2015, including whether the patient had any inpatient admission, the number of visits to specialists and PHC providers, total healthcare expenditure, and expenditures for inpatient services, outpatient services, and drugs (all in 2015 RMB). Utilization measures in 2015 also include the usage of anti-hypertensive and anti-diabetic drugs and the number of days covered by those medications, conditional on having at least one claim for the corresponding type of drugs. Checkup measures include the share of people with fasting blood sugar under control ($FBS < 7$) and the share of people with blood pressure (BP) under control according to their systolic BP (SBP) and diastolic BP (DBP) biomarkers (stage 1: $SBP < 140$ and $DBP < 90$; stage 2: $SBP < 160$ and $DBP < 100$). The number of people with at least one FBS or BP reading during a health checkup in 2013-2015 is shown in parentheses.

records to identify about 10% (n = 1,863) of diagnosed patients who are not enrolled in PHC management. These patients are on average younger (58.1 years old) and have a shorter duration of diagnosis (3.8 years). They are also healthier based on check-up records, with 55% having normal blood sugar levels and 64% having normal blood pressure. The diagnosed-but-not-managed sample are also more likely to be urban (either by *hukou* status or residential address), consistent with the younger working-age population with easier access to specialist care being least likely to select into PHC management for their chronic conditions. Taken together, the minority of individuals who were not enrolled in PHC management are a younger, urban, and more recently diagnosed subset of individuals, suggesting that the PHC management program experiences adverse selection (not favorable selection or cherry picking).

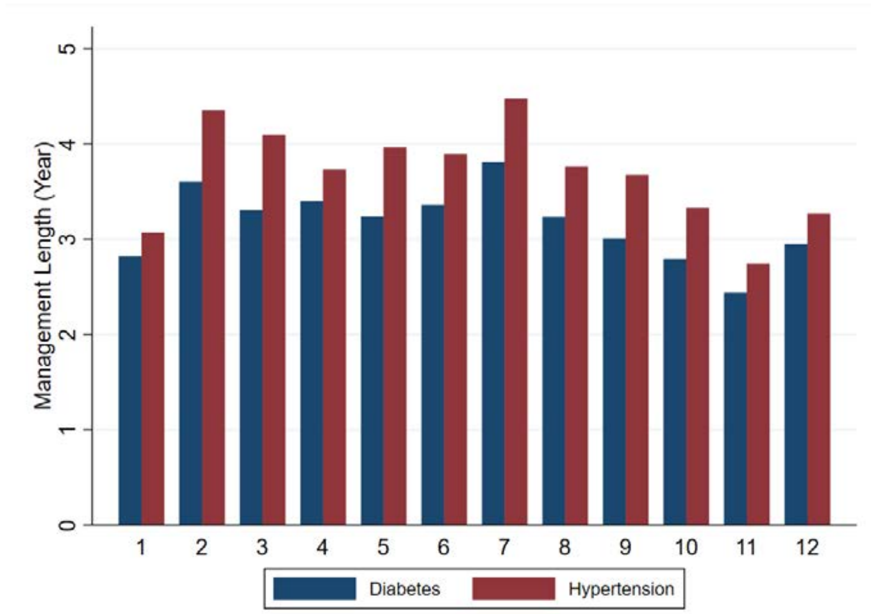
4 Empirical Strategy

4.1 Study Design

The research question we are interested in is whether rural PHC services are effective in managing chronic-disease patients. To look at this, we leverage differences in management intensity across different townships, the level at which the PHC program is administered. The first and primary dimension of management we leverage is duration of enrollment. As shown in Figure 1, the average number of years individuals are enrolled in PHC management (until 2015) varies between 2.5 years and 4.5 years across the dozen townships. In every township, the average duration of management for hypertension is longer than that for diabetes, consistent with the fact that the whole PHC management program launched first with a focus on hypertension; the diabetes program following later. Average management durations for hypertension and diabetes are strongly positively correlated, reflecting some shared features in PHC programs in a given township, with variation across townships, such as the ability to motivate PHC physicians and leverage their latent capacity for more productive use. Appendix Figure A2 shows the distribution of management-starting year by township. As discussed in the description of the PHC program, most of the enrollees were enrolled during the initial community-wide canvassing and screening in 2010-2012, depending on their township of residence.

Although management duration varies across townships, a direct analysis looking at the effect of management duration on patient outcomes might be biased if there are other confounding factors, such as population demographic composition and proximity to hospitals, that co-determine townships' management length and patients' healthcare utilization. To tackle this potential endogeneity problem, we utilize a boundary discontinuity approach. As developed by Dube et al. (2010) and others, this approach generalizes the case study method by aggregating cases of matched or neighboring jurisdictions subject to different policies. This identification strategy has also been used to study the effects of public health insurance (Sen and DeLeire, 2018), pharmaceutical direct-to-consumer advertising (Shapiro, 2018), and other policies that differ across administrative borders. In our study, we compare neighboring villages that straddle a township border, i.e. village pairs for which centroids are located within 2km of each other yet belong to different

Figure 1: Management Duration by Township



Note: This figure shows average hypertension and diabetes management duration by township. Management duration is calculated as the number of years since enrollment in the program until 2015. Townships are identified based on each patient's residential address. We keep the names of townships anonymous in this figure by assigning a numeric ID from 1-12.

townships. We exclude pairs in which one village belongs to a traditionally rural township and the other belongs to an urban township.[15] We also restrict to villages with population above 200.[16] This yields 27 villages in 14 boundary-straddling pairs across the 12 townships[17], as shown in Figure 2.

Table 1 shows that residents in these border villages are generally similar to the full sample in terms of demographics, health care utilization, and health check-up measurements.[18] In the border sample, blood pressure is slightly better under control than for the full sample, among patients managed for hypertension (51% vs 46%) as well as those managed for diabetes (55% vs 52%).[19]

[15]The average distance between villages within the same township is 4.5km.

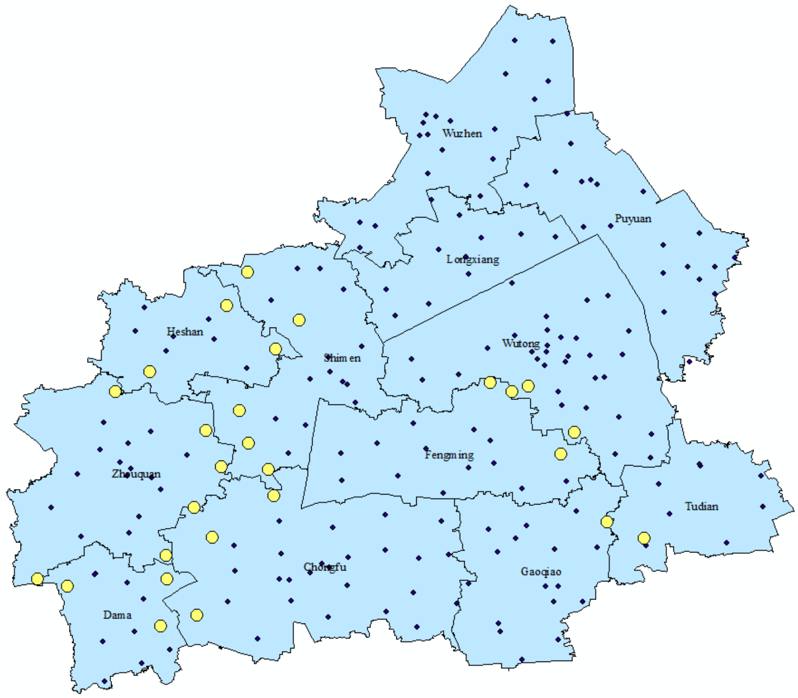
[16]The average village population is 3,298, or 1,641 if we include only the rural population above age 40 (consistent with our sample restriction). By restricting to villages with total population above 200, we drop 8 out of 215 villages, or one village pair that satisfies all other restrictions.

[17]One village in Wutong is matched to two different villages in Fengming.

[18]The border sample is slightly more male among hypertensives, but less male among diabetes patients, compared to the overall sample. The border sample also exhibits slightly longer duration of hypertension management, although only 11 days longer relative to the overall hypertensive sample, or less than 1% of mean duration of management in the overall sample. Consistent with being in PHC management longer, individuals in the border sample also exhibit higher PHC utilization and fewer specialist visits, both in 2011 and 2015; and those in diabetes management are 2 percentage points less likely to have a hospital admission in 2015, with (unsurprisingly) lower inpatient and total expenditures. The border sample exhibited a greater percentage increase in specialist visits than in the full sample (especially among hypertensives), suggesting that our finding of a slower increase in specialists visits when under more intense management, based on the border sample, may actually be conservative.

[19]We only have check-up biomarker data for 2013-15; if the border sample also were better under control at baseline, our estimates would be conservative to the extent that achieving better control while reducing the rate of spending increase may be

Figure 2: Border Sample - Villages Across Township Boundaries



Notes: This figure illustrates the 14 boundary-straddling pairs of villages used in our research design. Each black line shows a township border and each yellow dot is the google map location of each neighboring village. We identify 14 pairs of neighboring villages by restricting to villages that are located within 2km of each other yet belong to different townships. We exclude pairs in which one village belongs to a traditionally rural township and the other belongs to an urban township, as well as villages with population less than 200.

Since PHC management services are administered at the township level, villages on different sides of a township border are subject to different implementation of PHC management policies. At the same time, residents in the paired villages should be similar in individual characteristics as well as in many environmental and geographic access characteristics that may shape health care use but are not directly measured. For example, insurance coverage features — such as copayment requirements at different levels of provider — are uniform across all 12 townships. Therefore, the residents of neighboring villages enjoy exactly the same health insurance policies and may self-refer to any hospital in Tongxiang at will. Furthermore, as shown in Appendix Table A3, many spatial factors that shape healthcare demand, such as population composition and distance to healthcare facilities, are comparable between neighboring villages that are located in townships with higher versus lower management intensity. (The intensity measure is defined with more detail in Section 4.2.) Factors related to climate and pollution should also be similar across villages within

harder for those already better under control at baseline.

2km. Therefore, comparing individuals residing within pairs of boundary-straddling villages allows us to control for underlying spatial differences in healthcare demand factors and thus better disentangle those other determinants of utilization from the effects of PHC management.

Another potential threat related to this design would arise if there were endogenous moving or selection of residence in response to the PHC program. We argue that this is not much of a concern for the following reasons. First, those with agricultural hukou have allocated land holdings in a given village within a given township both for living and farming, and rarely move except for marriage or migrating to urban areas for schooling or work. The vast majority of older individuals such as those in our sample remain resident in the same county and village over a lifetime. According to nationally representative data in 2015, 85% of Chinese rural residents above age 40 still reside in the same county in which they were born. Among those who live outside their birth county or village, only 3.8% moved to the current place during the past 5 years.[20] Moreover, residents are unlikely to know much about the management intensity in neighboring townships since the rankings we discuss are not disseminated publicly but only among the PHC workforce.

4.2 Construction of the management intensity index

The strength or intensity of PHC management encompasses multiple dimensions of provider effort, from community screening, to persuading residents with abnormal blood pressure and/or blood glucose measures to agree to PHC management, to retaining individuals in PHC management despite an inpatient admission or episode of specialist treatment (after which they may be advised by specialists – with their own fee-for-service incentives – to continue care through hospital outpatient department visits). To measure the intensity of PHC management, we proxy these multiple dimensions with the average number of years patients in that disease management program have been enrolled in PHC management, accounting for differences in the demographics of enrolled individuals. We create this proxy measure in two steps. First, we estimate the following regressions for all managed patients in all townships:

$$t_i = X_i\beta + \eta_i \tag{1}$$

where t_i is the duration of management (measured in years) for patient i by 2015. Since the average management lengths of the two programs are highly correlated, we estimate a pooled regression with patients from

[20] China Health and Retirement Longitudinal Study (CHARLS). See Zhao et al. (2020) for more details about this data.

both programs included and construct one measure of general management intensity in our main analysis. X_i includes gender, age, age squared, years of schooling, years since diagnosis (duration Dx), duration Dx squared, indicators for whether enrolled in PHC management for hypertension, and for diabetes, as well as number of physicians in the township (in 2010).

Table 2 column 1 reports the results of model (1) for the pooled sample of patients enrolled in hypertension and diabetes management programs. Women and more-educated patients tend to be enrolled in the program for a longer period of time, whereas age is not significantly related to management duration after controlling for years since diagnosis. Patients under hypertension management have longer duration, reiterating the fact that the PHC management program started with a focus on hypertension. Not surprisingly, patients with earlier diagnosis have longer enrollment under the program. A coefficient less than one and negative for the second polynomial term reflects the fact that many previously diagnosed patients are first enrolled together during the initial community screening when the programs were established. Perhaps most counter-intuitively, the number of physicians within the township is negatively correlated with management duration. Individuals residing in townships with more physicians have a shorter duration of management, controlling for other factors. In Appendix Figure A3, we further show that the number of physicians (per 1,000 residents) did not significantly differ in the pre-period and did not change differently across townships with longer vs. shorter management duration. These patterns suggest that the endowment of physician capacity might not be the constraining factor in providing more PHC services. In columns 2 and 3, we report separate regressions for those enrolled only in the hypertension or diabetes management program. The results are consistent with those from the pooled regression, although diabetes management duration is increasingly larger for those diagnosed earlier and is not significantly correlated with physician capacity.

The residual η_i from regression (1) is an individual's management duration, controlling for that patient's demographics and local physician capacity. This residual captures differences in management duration that are not attributable to a patient's chronic disease condition, age, or other observed factors related to demand. It also controls for differences in physician capacity across townships, thus isolating the effect of PHC program effort conditional on the existing provider workforce. This residual may also capture the characteristics of local leadership, PHC team dynamics, and word-of-mouth recommendations from residents about their satisfaction with the program. As we discuss below, we also find that perhaps peer reputational incentives from the internal rankings of PHC management of each township promote management intensity, although the results are only suggestive. All these factors affecting the management intensity of the program

Table 2: Factors Related to PHC Management Duration (Factors controlled for in the PHC Management Intensity Score and $PHCrank_k$)

	(1) Pooled Sample	(2) HP Sample	(3) DB Sample
	Years since managed		
Age	0.00249	0.00147	0.00959
	(0.00314)	(0.00334)	(0.00940)
Age^2	-4.17e-06	2.61e-06	-5.14e-05
	(2.37e-05)	(2.51e-05)	(7.38e-05)
Male	-0.0315***	-0.0331***	-0.0467**
	(0.00682)	(0.00714)	(0.0195)
# Years of school	0.0133***	0.0135***	0.0101***
	(0.00113)	(0.00119)	(0.00308)
# Years since Diagnosis	0.682***	0.683***	0.267***
	(0.00265)	(0.00277)	(0.00263)
# Years since Diagnosis^2	-0.0295***	-0.0294***	0.000896***
	(0.000150)	(0.000156)	(1.60e-05)
# Physicians	-0.122***	-0.149***	0.0182
	(0.00896)	(0.00943)	(0.0248)
HP Management	0.224***		
	(0.0153)		
DB Management	0.00491		
	(0.0108)		
Constant	0.615***	0.895***	1.232***
	(0.103)	(0.111)	(0.299)
Observations	74,627	67,813	14,543
R-squared	0.546	0.537	0.420
Mean of Dep Var	3.532	3.617	3.055

Notes: This table presents regression results of Equation (1) in predicting individual management duration. Column 1 uses the pooled sample including all patients enrolled in the hypertension and/or diabetes management programs. Columns 2 and 3 report results from the separate samples of hypertension (HP) and diabetes (DB) management program enrollees, respectively. Predictors X_i include the patient's age, gender, educational attainment (years of schooling), management program, years since diagnosis, and physician capacity in the township of residence (in 2010). In the pooled sample, length of management is calculate as the larger of either hypertension or diabetes management program length; the length of diagnosis is constructed similarly. Residuals are used to calculate the characteristics-controlled average management duration of each township. Statistical significance at the 10%, 5% and 1% levels are denoted by *, **, and ***.

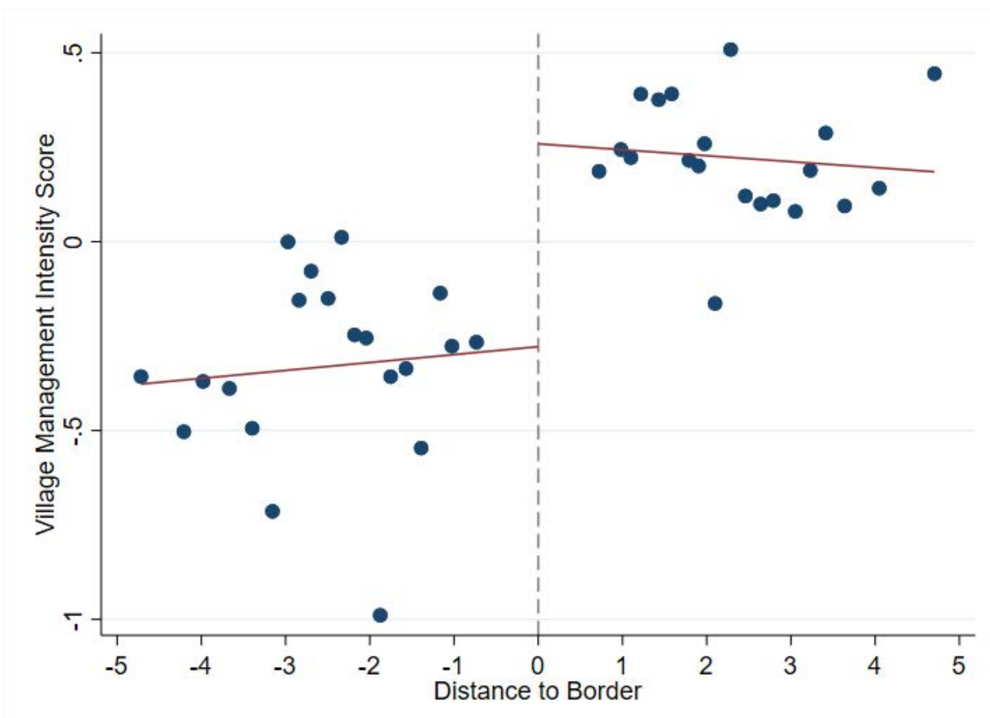
vary across township boundaries within each border village pair. Next, we take the average of individuals' residuals over each village j and township k as the management intensity score l for that village or township:

$$l_j = \frac{1}{n_j}\sum_{i\in j}\hat{\eta}_i; \quad l_k = \frac{1}{n_k}\sum_{i\in k}\hat{\eta}_i. \tag{2}$$

In the main analysis, we use the residual term estimated from the pooled sample of patients in the hypertension and diabetes management programs (as in Table 2 column 1) to construct the management intensity score. As robustness checks, we also use residuals from the separate regressions for each program (columns 2 and 3) to construct management intensity scores specific to each condition. As shown in Appendix Figure A5, these separate program intensity scores are highly correlated with each other.

The validity of the study design requires discontinuity in the management intensity score l_k across township borders. We perform the following steps to show the "jump" across township boundaries. Figure 3 plots the average village-level management intensity score (l_j) of enrolled hypertension and diabetic patients against villages' distance to the township boundary, for all adjacent township pairs. On average, across the boundaries, adjacent villages have about a 0.5 difference in intensity score, or 0.5 year difference in management duration controlling for demographics and physician capacity. Appendix Figure A4 shows a similar jump at the township boundary for hypertension and diabetic management intensity scores when constructed separately.

Figure 3: Discontinuity in Village PHC Management Intensity Score across Township Boundaries



Note: This figure is a binscatter plot of village-level management intensity scores (l_j) against villages' distance to the township border. The management intensity score is calculated as the residuals η_i from regression (1) averaged for each village. Distances to the township boundary are calculated based on the latitude and longitude of the centroid of each village. For each village, we find the village closest to it across each township boundary, calculate the distance between the two, and use half of that distance as the distance to the border. Note that this distance might be smaller than the actual distance to the boundary, especially for those far away from the border, but it does reflect the relative order of distance among villages within the same township. Villages in townships with higher township-level management intensity scores are always put on the right-hand side of the graph. The solid red line is the kernel fit of binscattered dots.

After calculating the management intensity score for each township (l_k), we transform it into a rank variable ($PHCrank_k$), ranging from 1 (for the township with the lowest l_k) to 12 (for the township with the highest l_k). We show in Appendix Figure A6 panel (a) that the continuous intensity score is dispersed fairly evenly across low and high ranks. Moreover, townships with higher management intensity rank also tend to put more effort into the program as reflected in their receiving higher evaluation scores in the CDC's assessment, as shown in Appendix Figure A6 panel (b). This indicates that the intensity score captures not only patient management duration but also other dimensions of the management program, such as the intensity of provider effort invested in follow-up via phone or home visits. Therefore, by using the rank index $PHCrank_k$ in the main analysis, we estimate the effect of overall management intensity of the program, instead of solely the margin of longer years of management. In the robustness checks, we also use the continuous version of the intensity score, i.e., the (leave-one-out) z-score of the township management intensity score (excluding the village in which each patient resides).

4.3 Patient-level first difference regression

The primary regression includes all rural residents over the age of 40 living in the matched bordering villages and enrolled in the management program by 2015. We estimate this individual-level first difference regression for their health service utilization, i.e. whether they have any inpatient care, number of specialist and PHC visits, total medical expenditure and expenditures by type:

$$\Delta Y_{ijk} = \beta_0 + \beta_1 PHCrank_k + W_i'\beta_2 + Z_k'\beta_3 + Border_{p(j)} + \varepsilon_{ijk}, \tag{3}$$

where ΔY_{ijk} represents the change in utilization between 2011 and 2015 for individual i, $(Y_{ijk}^{2015} - Y_{ijk}^{2011})$, resident of village j in township k, enrolled in a PHC management program. The key independent variable is the PHC management intensity rank ($PHCrank_k$) of township k, ranging from 1 (worst) to 12 (best) as defined in Section 4.2. The coefficient β_1 reflects changes in outcome when the management intensity rank is higher by one. We illustrate the economic significance of this PHC rank effect for some outcomes by multiplying coefficient β_1 by 6 to capture the effect magnitude that would arise if patients in the median-ranked (6th) township instead received management in the highest-ranked (12th) township. W_i captures individual characteristics as in regression (1), i.e. gender, age, age squared, years of schooling, years since diagnosis (duration Dx), duration Dx squared, indicators for whether enrolled in PHC management for hypertension and/or for diabetes. Z_k includes township-level control variables such as physician capacity in the township in 2010. Also included in the regression, $Border_{p(j)}$ contains a set of fixed effects for each boundary pair $p(j)$ that village j belongs to.

Apart from this reduced form regression for the effect of township PHC management intensity, we also undertake a 2SLS analysis, where the township PHC management intensity rank is used as an instrumental variable for an individual's management duration. Under this framework, Appendix Table A4 presents the first-stage results verifying that higher township management intensity leads to longer management duration of the local residents. Then we estimate the second-stage,

$$\Delta Y_{ijk} = \beta_0 + \beta_1 PHC\widehat{Duration}_i + W_i'\beta_2 + Z_k'\beta_3 + Border_{p(j)} + \varepsilon_{ijk}, \tag{4}$$

where the estimated coefficient β_1 for an individual's instrumented management duration ($PHC\widehat{Duration}_i$) reflects the changes in outcome when that patient is enrolled in the program for one additional year.

The regression framework captures within a boundary pair, how PHC management intensity impacts patient i's change in healthcare utilization, relative to similar patients living with hypertension and/or diabetes in the neighbor village. Recall that our metric for strength or intensity of PHC management is a constant at the township level, reflecting the cumulative efforts of PHC physicians to screen their communities and recruit and retain patients within the PHC management programs. The identification assumption is that, within a pair of neighboring villages straddling the boundary between two townships that differ in PHC management intensity, no other unobserved factors affect the change in healthcare utilization of a resident in a high management intensity village relative to an observably similar resident of a low intensity village, other than the differences in PHC management programs.

4.4 Cross-sectional data analysis: Medication use and health check-up biomarkers

As described in the previous section, for health care utilization measurements, as well as health check-up biomarkers for a subset of patients who had records in both 2013 and 2015, the panel data enables us to use the first difference of these outcome variables to account for initial differences across individual patients. For medication use and health check-up biomarkers for the majority of patients, however, we only have cross-sectional data. To examine the relationship between PHC management intensity and these outcomes, we estimate the following cross-sectional regression:

$$Y_{ijk}^{2015} = \beta_0 + \beta_1 PHCrank_k + W_i' \beta_2 + Z_k' \beta_3 + U_i^{2011'} \beta_4 + Border_{p(j)} + \varepsilon_{ijk}, \quad (5)$$

which features the same main explanatory variable $PHCrank_k$ and sets of control variables as in equation (3), except that we also add controls U_i^{2011} for baseline utilization, i.e. total medical spending and an indicator for a hospital admission in 2011. Although not ideal, these baseline measurements serve as controls for initial differences in health status across patients and help us to have a better understanding of the effect of PHC management intensity on medical use and health outcomes in the cross-sectional setting. To interpret the estimation as causal, we need to assume that no other unobserved factors affect medication use and health biomarkers differentially between neighboring villages other than the PHC management intensity and the individual's severity as captured by baseline medical utilization.

5 Results

5.1 Impact of PHC management on healthcare utilization and spending

To evaluate the effect of PHC management, we start by examining the healthcare utilization and spending of people enrolled. In Table 3, we report the individual-level first difference results estimated using equation (3). The dependent variables are the patient-level differences in healthcare utilization and expenditures between 2011 and 2015. The explanatory variable of interest is the PHC management intensity rank, ranging from 1 (worst) to 12 (best), for the patient's township of residence. We estimate this regression on three samples: the pooled sample of both hypertensive and diabetic managed patients in Panel A; the sample of hypertensive managed patients in Panel B; and the sample of diabetic managed patients in Panel C.

For the pooled sample, a township that ranks one position higher (i.e., longer management duration, given patient characteristics) has a smaller increase in the rate of inpatient admissions and a smaller relative increase in total spending, mostly attributable to less spending for inpatient care. Moreover, greater PHC management intensity leads to relatively fewer specialist visits and more PHC visits. These effects hold after controlling for patient demographics, diagnosis length, and physician capacity as in Table 2, as well as fixed effects for border village pairs. These results are primarily driven by those for hypertension patients, for whom residing in a township with higher management intensity leads to more PHC visits, fewer specialist

Table 3: Impact of PHC Management on Healthcare Utilization and Spending - Reduced Form Results

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Inpatient	Expenditure	Expenditure (Inpatient)	Expenditure (Outpatient)	Expenditure (Drug)	# Specialist Visits	# PHC Visits
Panel A: Pooled							
PHC Rank	-0.00258**	-13.62*	-13.12**	0.747	1.542	-0.0169**	0.0914**
	(0.00101)	(6.675)	(5.235)	(1.471)	(2.011)	(0.00559)	(0.0349)
Observations	12,641	12,641	12,641	12,641	12,641	12,641	12,641
Mean of Dep.Var (2015)	0.132	2295	958.1	338.5	856.2	1.968	11.48
Panel B: Hypertension							
PHC Rank	-0.00244*	-12.80*	-14.82**	1.545	2.557	-0.0183**	0.116***
	(0.00110)	(5.874)	(5.582)	(1.378)	(2.176)	(0.00573)	(0.0309)
Observations	11,432	11,432	11,432	11,432	11,432	11,432	11,432
Mean of Dep.Var (2015)	0.130	2194	947.4	326.7	799.4	1.767	11.42
Panel C: Diabetes							
PHC Rank	-0.00214	-5.144	-1.905	-5.126	-3.647	-0.0420**	-0.0100
	(0.00199)	(15.08)	(10.78)	(3.067)	(4.248)	(0.0167)	(0.0398)
Observations	2,620	2,620	2,620	2,620	2,620	2,620	2,620
Mean of Dep.Var (2015)	0.159	3397	1161	454.6	1456	3.597	14.26

Notes: This table shows the estimation results of equation (3). The sample includes all enrolled chronic-disease patients aged 40+ with agricultural hukou in the 14 pairs of adjacent villages located in different townships (the border sample). The observation is at the patient level. Panel A) shows results for the pooled sample of individuals in either or both programs; panel B) uses enrolled hypertension patients only and Panel C) uses diabetic patients only. Dependent variables are healthcare utilization and spending measures in 2015 less that of 2011. All expenditures are winsorized at the 95th percentile within our claims sample. Specialist visits include all visits at provincial or municipality level hospitals, while PHC visits includes visits at township and village clinics. All regressions include boundary pair fixed effects, individual and township-level characteristics, i.e. gender, age, age squared, years of schooling, years since diagnosis (duration Dx), duration Dx squared, indicators for whether enrolled in PHC management for hypertension and/or for diabetes (for the pooled sample), and physician capacity in the township in 2010. Standard errors are clustered at the township level. Statistical significance at the 10%, 5% and 1% levels are denoted by *, **, and ***.

visits, a relative decrease in inpatient utilization and spending, and a decrease in total health expenditures. Comparing a township ranked 6th to the 12th, the difference in inpatient admissions among hypertension patients is about 1.5 percentage points ($0.00258*6 = 0.015$), which is about 11.7% of the average inpatient admission rate among hypertension patients (Table 3 Panel B Column 1). Such a difference between the median and highest ranking township is also associated with 3.6% lower spending ($13.62*6/2295 = 0.036$), 5.2% fewer specialist visits ($0.0169*6/1.968 = 0.052$), and 4.8% more PHC visits ($0.0914*6/11.48 = 0.048$). For the smaller sample of diabetic patients, more intensive management also reduces the relative increase in number of specialist visits, while other measures of utilization or spending show changes in the same direction yet are statistically insignificant.

To interpret the effect of PHC management in terms of one additional year enrolled in the program, we also estimate equation (4) in 2SLS format using the management intensity rank as an instrumental variable for individual management duration. The first-stage regression results, reported in Appendix Table A4, confirm a strong relationship between township management intensity and length of program enrollment (PHC duration) among local residents. Using instrumented PHC duration, the second-stage regressions results reported in Table 4 show that one more year in PHC management decreases the inpatient admission rate by 23.9% (3.16 percentage points relative to a mean of 13.2) and inpatient spending by 16.7% (160.4 RMB relative to a mean of 958.5 RMB). In terms of utilization, one more year of PHC management leads to 10.7% fewer specialist visits and 10.2% more PHC visits – significantly different from the national trend of more specialist visits and fewer PHC visits.

Table 4: Impact of PHC Management on Healthcare Utilization and Spending - 2SLS Results

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Inpatient	Expenditure	Expenditure (Inpatient)	Expenditure (Outpatient)	Expenditure (Drug)	# Specialist Visit	# PHC Visit
Panel A: Pool							
PHC Duration	-0.0316**	-166.3*	-160.4**	10.02	19.54	-0.212***	1.174**
	(0.0130)	(88.80)	(70.56)	(16.86)	(23.71)	(0.0658)	(0.494)
Observations	12,580	12,580	12,580	12,580	12,580	12,580	12,580
Mean of Dep.Var (2015)	0.132	2294	958.5	338	855.6	1.968	11.47
Panel B: Hypertension							
PHC Duration	-0.0306**	-159.7**	-187.4**	20.62	33.14	-0.235***	1.497***
	(0.0140)	(80.54)	(75.35)	(14.90)	(27.30)	(0.0676)	(0.503)
Observations	11,382	11,382	11,382	11,382	11,382	11,382	11,382
Mean of Dep.Var (2015)	0.130	2193	946.8	326.4	799.5	1.769	11.41
Panel C: Diabetes							
PHC Duration	-0.0179	-39.89	-6.165	-48.11*	-34.17	-0.368***	-0.111
	(0.0170)	(128.3)	(93.90)	(25.81)	(33.42)	(0.133)	(0.344)
Observations	2,609	2,609	2,609	2,609	2,609	2,609	2,609
Mean of Dep.Var (2015)	0.160	3399	1166	453.8	1454	3.591	14.28

Notes: This table shows the estimation results of the 2SLS version for equation (4) using PHC management intensity rank as instrumental variable for individual management duration. The sample includes all enrolled chronic-disease patients aged 40+ with agricultural hukou in the 14 pairs of adjacent villages located in different townships (the border sample). The observation is at the patient level. Panel A) shows results for the pooled sample of individuals in either or both programs; panel B) uses enrolled hypertension patients only and Panel C) uses diabetic patients only. Dependent variables are healthcare utilization and spending measures in 2015 less that of 2011. All expenditures are winsorized at the 95th percentile within our claims sample. Specialist visits include all visits at provincial or municipality level hospitals, while PHC visits includes visits at township and village clinics. All regressions include boundary pair fixed effects, individual and township-level characteristics, i.e. gender, age, age squared, years of schooling, years since diagnosis (duration Dx), duration Dx squared, indicators for whether enrolled in PHC management for hypertension and/or for diabetes (for the pooled sample), and physician capacity in the township in 2010. Standard errors are clustered at the township level. Statistical significance at the 10%, 5% and 1% levels are denoted by *, **, and ***.

Thus, overall, we observe four statistically significant results: a township with higher management intensity is associated with more PHC visits, fewer specialist visits, lower inpatient admission rates, and lower inpatient spending, relative to a township with lower management intensity. These results suggest that PHC chronic-disease management in rural Tongxiang shifts the relative increase in utilization toward PHC settings, decreasing inpatient spending by reducing avoidable hospital admissions. As noted previously, a lower rate of hospitalizations for ambulatory-care sensitive conditions such as hypertension and diabetes itself serves as an indicator of high-quality PHC in many studies and health system comparative metrics, such as for OECD countries.[21] We next explore this pattern by studying prescribing, adherence, and health outcomes as measured by blood pressure and blood sugar.

[21] See https://www.oecd.org/els/health-systems/hcqi-primary-care.htm.

5.2 Prescribing and adherence to medication

To understand potential mechanisms behind these findings, we gathered and coded data on medication adherence. Although overall expenditures on medications did not statistically differ by management intensity, patterns of prescribing and adherence to specific medications might. For individuals with high blood pressure and diabetes, regular and consistent adherence to anti-hypertensive and anti-diabetic drugs can be crucial for preventing complications, yet individuals without salient symptoms often exhibit poor adherence. Improving medication adherence is therefore one primary mechanism through which PHC management can enhance health outcomes and reduce avoidable admissions for acute sequelae.

To test this hypothesis, we report results of our regression equation (5) with anti-hypertensive and anti-diabetic drug usage in 2015 as the dependent variable in Table 5. More intensive PHC management is not associated with greater prescribing of, or adherence to, anti-hypertensives. Interestingly, more intensive management appears to lead to fewer patients being prescribed and taking anti-diabetic drugs, but better adherence for those who are prescribed and taking anti-diabetic medications, as measured by number of days covered. For example, column 8 indicates that one more year of PHC management is associated with 12.86 more covered days of anti-diabetic medications in 2015, which was 5.4% of the average. These empirical results suggest that one mechanism for fewer inpatient admissions might be that the village doctors increased medication adherence for a few basic medications (without increasing overall drug spending) among patients with whom they regularly interacted.

5.3 PHC management and health outcomes

After looking at the effect of PHC management on healthcare utilization, it is also if not more important to test how management is related to patients' health outcomes. Analysis of biomarkers in this section shows that the reduced inpatient and specialist utilization associated with more intensive PHC management does not adversely impact health outcomes, but rather appears to be at least in part the result of improved health.

As discussed in the Data and Empirical Strategy sections, we only have biomarker information from health checkup records in 2013-2015. In the analyses of health outcomes in Tables 6 and 7 columns 2-4, we use the average biomarker reading taken during any health checkup during this period as the outcome variable for equation (5), controlling for total medical spending and an indicator for hospital admission in 2011 — two measures often used in risk adjustment — to account for differences in baseline severity. We

Table 5: PHC Management and Medication Use

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Reduced Form				2SLS			
	HP drug		DB drug		HP drug		DB drug	
	Any	# Days	Any	# Days	Any	# Days	Any	# Days
Panel A: Pool								
PHC Rank	0.00177	-0.759	-0.00221***	1.211*				
	(0.00241)	(0.774)	(0.000242)	(0.646)				
PHC Duration					0.0247	-9.940	-0.0287***	12.86**
					(0.0316)	(9.396)	(0.00461)	(6.310)
Observations	12,641	9,201	12,641	2,449	12,580	9,158	12,580	2,439
Mean of Dep.Var (2015)	0.728	207.1	0.194	236.6	0.728	207.1	0.194	236.5
Panel B: Hypertension								
PHC Rank	0.00223	-0.691	-0.00172***	1.938**				
	(0.00281)	(0.787)	(0.000394)	(0.711)				
PHC Duration					0.0302	-9.062	-0.0222***	19.23***
					(0.0376)	(9.478)	(0.00526)	(6.980)
Observations	11,432	8,831	11,432	1,464	11,382	8,789	11,382	1,464
Mean of Dep.Var (2015)	0.772	209.2	0.128	227.3	0.772	209.2	0.129	227.3
Panel C: Diabetes								
PHC Rank	-0.00298	-1.457	-0.00801**	1.286				
	(0.00192)	(0.950)	(0.00279)	(0.750)				
PHC Duration					-0.0248	-12.14*	-0.0751***	11.69*
					(0.0168)	(7.232)	(0.0222)	(6.630)
Observations	2,620	1,562	2,620	2,167	2,609	1,561	2,609	2,157
Mean of Dep.Var (2015)	0.596	205.9	0.827	253.7	0.598	206	0.827	253.7
Border FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
2011 Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: This table shows the estimation results of the reduced form regression following equation (5) and the 2SLS version using PHC management intensity index as instrumental variable for individual management duration. The sample includes all enrolled chronic-disease patients aged 40+ with agricultural hukou in the 14 pairs of adjacent villages located in different townships (the border sample). The observation is at the patient level. Panel A) shows results for the pooled sample of individuals in either or both programs; panel B) uses enrolled hypertension patients only and Panel C) uses diabetic patients only. Dependent variables are whether the individual had any medication claim, and if so, the number of days covered by anti-hypertensive or anti-diabetic drugs (see Appendix A2 for the detailed construction process). All regressions include boundary pair fixed effects, individual and township-level characteristics as in Table 3, as well as baseline health controls, i.e. total medical spending and whether had hospital admission in 2011. Standard errors are clustered at the township level. Statistical significance at the 10%, 5% and 1% levels are denoted by *, **, and ***.

find that, among those with at least one health checkup record, patients managed more intensively in PHC are substantially more likely to have their blood pressure under control during 2013-15, and are less likely to suffer from Stage II hypertension, compared to similar patients managed less intensively (or for fewer years). Controlling for age, duration of diagnosis, and other individual-level factors (including baseline utilization), those under PHC management for one more year are 17.5 percentage points more likely to have blood pressure under control, compared to a mean of 52.3% of managed patients having blood pressure controlled (Table 7 Panel A Column 2), a 33.5% improvement. Another year of management also leads to significantly better control of Stage 2 hypertension, with a 5.4 percentage point higher rate of control relative to a mean of 90.2%. In other words, whereas 9.8% of patients suffer from Stage II hypertension on average, only 4.4% do when under more intensive PHC management, a 55% improvement among these higher-severity patients. These results do not stem from selective testing among healthier enrollees; in fact, enrollment in a more intensive PHC management program is associated with a statistically significantly higher probability of having any checkup (Column 1 of Tables 6 and 7). Interestingly, patients managed more intensively for hypertension are slightly more likely to have FBS out of control (column 4), compared to the average PHC patient, perhaps because hypertension management programs focus on blood pressure at the expense of monitoring blood glucose (a multitasking problem), or perhaps because they screen more patients for blood glucose and thus are more prone to pick up those with elevated FBS or those who were not truly fasting at the time of the blood draw.

Table 6: PHC Management and Health Outcomes - Reduced Form Results

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
		Average Health Checkup			Difference in Health Checkup, 2013-15		
	Any Health Checkup	BP1 under control	BP2 under control	FBS under control	BP1 under control	BP2 under control	FBS under control
Panel A: Pool							
PHC Rank	0.00238**	0.0142**	0.00448***	-0.00209**	0.0278***	0.0203***	-0.00234
	(0.000984)	(0.00450)	(0.00116)	(0.000844)	(0.00198)	(0.00116)	(0.00170)
Observations	12,641	9,819	9,819	9,492	2,824	2,824	2,877
Mean of Dep.Var	0.790	0.523	0.902	0.850	-0.0875	-0.0637	-0.0372
Panel B: Hypertension							
PHC Rank	0.00205*	0.0142**	0.00462***	-0.00321***	0.0282***	0.0209***	-0.00179
	(0.00100)	(0.00460)	(0.00124)	(0.000837)	(0.00260)	(0.00107)	(0.00166)
Observations	11,432	9,014	9,014	8,743	2,626	2,626	2,674
Mean of Dep.Var	0.802	0.508	0.896	0.889	-0.0864	-0.0678	-0.0333
Panel C: Diabetes							
PHC Rank	0.00635***	0.0150**	0.00283**	0.000841	0.0165**	0.00602	-0.00474
	(0.00169)	(0.00477)	(0.00117)	(0.00379)	(0.00542)	(0.00468)	(0.00316)
Observations	2,620	1,981	1,981	1,887	581	581	604
Mean of Dep.Var	0.772	0.552	0.907	0.436	-0.0878	-0.0499	-0.0844
2011 Controls	Yes	Yes	Yes	Yes			

Notes: This table shows the estimation results for equation (5) in Column 1-4 and equation (3) in Column 5-7. Outcome variables are whether had any health checkup (Column 1), average health checkup results in 2013-2015 (Column 2-4), and the difference in blood pressure and FBS between 2013 to 2015 for those who had multiple testing results (Column 5-7). The sample includes all enrolled chronic-disease patients aged 40+ with agricultural hukou in the 14 pairs of adjacent villages located in different townships (the border sample). The observation is at the patient level. Panel A) shows results for the pooled sample of individuals in either or both programs; panel B) uses enrolled hypertension patients only and Panel C) uses diabetic patients only. Dependent variables are whether blood pressure is under control (systolic blood pressure (SBP) < 140 mm Hg and diastolic blood pressure (DBP) < 90 mm Hg), whether blood pressure is below Stage II (SBP < 160 and DBP < 100), and whether fasting blood sugar is under control (FBS < 7). All regressions include boundary pair fixed effects, individual and township-level characteristics as in Table 3. Column 1-4 also include baseline health controls, i.e. total medical spending and whether had hospital admission in 2011. Standard errors are clustered at the township level. Statistical significance at the 10%, 5% and 1% levels are denoted by *, **, and ***.

Although the regression controls of measures of baseline severity, these cross-sectional associations between average biomarkers and the intensity of PHC management may not be as persuasive as longitudinal evidence. To better examine the effect of management intensity on health outcomes, we also analyze the sub-

Table 7: PHC Management and Health Outcomes - 2SLS Results

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
		Average Health Checkup			Difference in Health Checkup, 2013-15		
	Any Health Checkup	BP1 under control	BP2 under control	FBS under control	BP1 under control	BP2 under control	FBS under control
Panel A: Pool							
PHC Duration	0.0297**	0.175***	0.0544***	-0.0258***	0.285***	0.208***	-0.0226
	(0.0126)	(0.0628)	(0.0178)	(0.00760)	(0.0355)	(0.0102)	(0.0142)
Observations	12,580	9,775	9,775	9,453	2,813	2,813	2,867
Mean of Dep.Var	0.790	0.523	0.902	0.850	-0.0878	-0.0636	-0.0370
Panel B: Hypertension							
PHC Duration	0.0260**	0.177***	0.0567***	-0.0398***	0.291***	0.217***	-0.0181
	(0.0125)	(0.0660)	(0.0199)	(0.00619)	(0.0408)	(0.0107)	(0.0147)
Observations	11,382	8,977	8,977	8,709	2,616	2,616	2,665
Mean of Dep.Va	0.802	0.508	0.897	0.889	-0.0868	-0.0677	-0.0334
Panel C: Diabetes							
PHC Duration	0.0571***	0.135***	0.0239**	0.00759	0.119***	0.0434	-0.0314*
	(0.0136)	(0.0418)	(0.0105)	(0.0328)	(0.0356)	(0.0339)	(0.0173)
Observations	2,609	1,974	1,974	1,882	580	580	603
Mean of Dep.Var	0.772	0.551	0.906	0.436	-0.0879	-0.0500	-0.0829
Border FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
2011 Controls	Yes	Yes	Yes	Yes			

Notes: This table shows the estimation results for the 2SLS version of equation (5) in Columns 1-4, and equation (3) in Columns 5-7, using PHC management intensity index as an instrumental variable for individual management duration. Outcome variables are whether the patient had any health checkup (Column 1), average health checkup results in 2013-2015 (Columns 2-4), and the difference in blood pressure and FBS between 2013 to 2015 for those who had multiple testing results (Columns 5-7). The sample includes all enrolled chronic-disease patients aged 40+ with agricultural hukou in the 14 pairs of adjacent villages located in different townships (the border sample). The observation is at the patient level. Panel A) shows results for the pooled sample of individuals in either or both programs; panel B) uses enrolled hypertension patients only and Panel C) uses diabetic patients only. Dependent variables are whether blood pressure is under control (systolic blood pressure (SBP) $<$ 140 mm Hg and diastolic blood pressure (DBP) $<$ 90 mm Hg), whether blood pressure is below Stage II (SBP $<$ 160 and DBP $<$ 100), and whether fasting blood sugar is under control (FBS $<$ 7). All regressions include boundary pair fixed effects, individual and township-level characteristics as in Table 3. Columns 1-4 also include baseline health controls, i.e. total medical spending and whether the individual had a hospital admission in 2011. Standard errors are clustered at the township level. Statistical significance at the 10%, 5% and 1% levels are denoted by *, **, and ***.

sample of individuals with panel data on biomarkers from checkups between 2013 to 2015, who represent a 29.7% of the patients with biomarkers in the health check-up sample. These results, reported in columns 5-7 of Tables 6 and 7, are in the same direction as those found in the cross-sectional regressions, consistent with PHC management conferring benefits in terms on health status. For example, Table 6 Column 5 implies that compared to a patient in the median-intensity PHC program, patients in the most-intensive PHC management program are about 16.7% more likely to have their blood pressure get under control from 2013 to 2015 ($0.0278 * 6 = 16.68\%$), relative to a mean 8.8% increase in the probability of control. Given the approximate half-year difference in management between the median and most-intensive program, this implies about one-third better likelihood of having blood pressure under control per year of PHC management. This order of magnitude is confirmed in the 2SLS estimation shown in Table 7 Column 5, showing that a patient with one more year of PHC management is 28.5% more likely to have blood pressure remain under control, compared to the 8.8% decline in having blood pressure under control for the average patient. Notably, more intense or longer PHC management leads to better health outcomes for both hypertensive and diabetes patients, as measured by their probability of having measured blood pressure under control over three years. FBS control, by contrast, shows little indication of improvement, and most differences are not statistically significant.

This subsample is not a random sample of enrolled patients and thus not necessarily representative of the border sample across townships with different levels of PHC management intensity. Nevertheless, the fact that blood pressure control improves when controlling for observable patient differences is suggestive that PHC management improves net value, achieving better health outcomes for the same or lower resource use – or that the reductions in spending do not arise at the expense of quality, but rather as the result of improved quality.

Our regression analyses also suggest that better health is one mechanism explaining the relative reduction in hospital admissions for those under more intensive PHC management. Appendix Table A5 shows that better control is strongly associated with fewer inpatient admissions in 2015. Among patients enrolled in either hypertension or diabetes management, those with blood pressure below the stage II threshold were 1.09 percentage points (7.57%) less likely to experience a hospitalization in 2015, while those with FBS under control were 3.45 percentage points (23.6%) less likely to be hospitalized in 2015.

5.4 Robustness

The estimation results are generally robust to a series of robustness checks using different forms of the PHC management intensity index: (1) separate management intensity ranks for the hypertension and diabetes programs (these intensity ranks are estimated from columns 2-3 in Table 2, with results reported in Table A6); (2) z-score of township management intensity (results reported in Table A7); (3) leave-one-out z-score township management intensity (results reported in Table A8). Demonstrating robustness to the first set of alternative intensity indices addresses the potential concern that hypertension and diabetes management programs may have different intensity scores and effects. In fact, the two intensity scores are highly correlated with each other, as shown in Appendix Figure A5. The two intensity rank variables are used to show robustness of results when separately estimating impacts for hypertensive and diabetic patients, respectively. The latter two sets of intensity indexes are both continuous metrics (rather than ranks). Based on the township-level intensity score calculated in equation (2), we first derive the z-score with mean 0 and standard deviation 1. To construct the leave-one-out intensity score, we calculate the average residual among townships, excluding the village where focal patients reside, and calculate the standardized z-score. This helps mitigate any concern that unobserved differences in health demand within adjacent villages may explain the differences in management.

In order to compare coefficients estimated using different intensity indices, we report 2SLS results where each index is used as instrumental variable for individual management duration. Our primary empirical results continue to hold: more intensive management leads to fewer specialist visits, more PHC visits, and more patients having their blood pressure under control, relative to similar patients residing in townships with less intensive management. The coefficients are also of similar magnitude compared to the main results in Tables 4 - 7.

Another series of robustness checks focuses on the distance threshold for selecting neighboring villages. In the main analysis, we identify villages pairs whose centroids are located within 2km yet belong to different townships. As shown in Appendix Figure A7 panel (a), the number of village pairs changes when we use different distance thresholds, ranging from 6 pairs within 1.5km to 72 pairs within 2.9km. Nevertheless, as shown in Appendix Figure A7 panel (b) to (g), our main results remain consistent across these distance thresholds. Villagers in high management intensity townships have lower probability of having an inpatient admission, more PHC visits, lower overall medical spending, and a higher blood pressure control rate, compared to residents in the neighboring village across the township border (with lower-intensity PHC management). Results for specialist visits are less stable when more village pairs are included, but coefficients remain significantly negative when the distance threshold is less than 2.3km.

5.5 Welfare implications

While we do not have comprehensive information to be able to assess the overall welfare impact of these programs, it may nevertheless be instructive to examine a back-of-the-envelope welfare calculation to understand the impacts of the chronic-disease management program along the dimensions that we do observe. The central government allocated fiscal funds of 45 RMB per capita in 2015 to support essential population health services, and 10% of the administrative evaluation score for those services is assigned to chronic-disease management. Therefore, we assume that 10% of the 45 RMB per capita goes to chronic-disease management. Thus, if benefits exceed 4.5 RMB per capita, the chronic-disease management program would be delivering a positive return. According to the results reported in Table 3, savings of 13.6 RMB on expenditures per rank implies a "median effect" of saving 81.6 (=6*13.6) RMB per hypertensive if those in the median township instead received care of the same intensity as patients in the highest-ranked township. This implies medical expenditure savings of more than 20 RMB per capita (81.6/4=22.2), given that more than one in four residents have hypertension, substantially exceeding the cost of 4.5 RMB per capita.

Alternatively, consider that with about 687,000 residents in Tongxiang county, the total cost of the program for one year is slightly over 3 million RMB (Cost 2015 = 10% * 45 * 687k = 3.09 million RMB). With medical expenditure savings of 81.6 RMB on average for each patient managed for hypertension, the benefit of managing 72,000 hypertension patients would be 5.9 million RMB. This estimation of saved medical expenditures arguably represents a lower bound of the benefits of the program, as it does not include any cost savings from managing diabetes, nor does it account for the potential value of improved quality of life and survival for individuals suffering from either or both chronic diseases we study.

6 Discussion and Conclusion

In this paper, we provide empirical evidence on the effect of PHC chronic disease management on healthcare use and outcomes for rural Chinese. Isolating the impact of PHC management requires disentangling it from other factors shaping healthcare use and health outcomes for chronic disease patients in rural China. Following the boundary design following Dube et al. (2010) and others, we compare healthcare utilization, spending, and health outcomes for about 12,000 rural Chinese living with hypertension or diabetes who reside in neighboring villages that straddle township boundaries and thus are subject to different PHC management. By looking at changes in healthcare use and spending over time for these individuals, and comparing differences in their end-of-period biomarker outcomes, we seek to disentangle the effect of PHC management from other factors. The empirical results suggest that better PHC management of chronic disease in rural China can reduce spending while contributing to better health.

Specifically, we find that enrolled patients residing in a township with higher management intensity have more PHC visits, fewer specialist visits, a relative decrease in inpatient utilization and spending, and a decrease in total health expenditures. These impacts are economically meaningful: one more year in PHC management decreases inpatient admissions by 24% (3.2 percentage points relative to a mean of 13.2), inpatient spending by 17%, and visits to specialists by 11%, while boosting visits to PHC by a similar amount and yielding 5% better adherence to diabetes medications (i.e., 12.9 more covered days of prescribed anti-diabetic medications).

Examining health outcomes as measured by biomarker values at health checkups, we find that patients managed more intensively in PHC (or managed in PHC for more years) are substantially more likely to have their blood pressure under control, and less likely to suffer from Stage II hypertension, compared to similar patients managed less intensively (or for fewer years). Thus, PHC appears to improve average health outcomes in part by bringing under better control those at greatest risk of severe complications (i.e., those in the upper tail of the blood pressure distribution). Moreover, in the sub-sample of individuals with 2013-15 panel data on biomarkers, one more year of PHC management is associated with 33.5% greater likelihood of blood pressure remaining under control over three years.

This improvement in hypertension control seems especially promising, since high blood pressure is the leading preventable risk factor for premature mortality in China (He et al., 2009). In a recent review, Lu et al. (2017) document "remarkably low" blood pressure control of less than 20% across all sub-groups, and only 6.1% for rural individuals. Therefore, the 46% level of blood pressure control in Tongxiang among hypertensives enrolled in PHC management (see Table 1) is well above the national average; and the 17.5 percentage point improvement in control achieved through one more year of management (a 33.5% improvement) provides 'proof of concept' regarding efforts to strengthen PHC to manage the chronic disease burden for China's aging population.

Less encouragingly, in our sample the control of blood glucose for individuals living with type 2 diabetes does not appear to differ significantly by intensity of PHC management (although the panel of those with diabetes biomarkers is too small to have much power). In fact, hypertensive patients managed more intensively are slightly less likely to have their average FBS under control (Table 6 column 4), compared to similar patients in less-intensive hypertension management programs. We hypothesize that this discrepancy arises through two mechanisms that are not necessarily mutually exclusive. First, the better control of hypertension than FBS may reflect wider screening for pre-diabetes and diabetes among hypertensives in the more-intensive programs. With wider testing of FBS among hypertensives, some individuals may be found to have undiagnosed diabetes and thus recommended for enrollment in diabetes management earlier in the course of the disease. More comprehensive blood glucose monitoring among hypertensives may also lead to testing of individuals who were not truly fasting at the time of the blood draw, artificially inflating the FBS measurement. Second, the better control of hypertension than FBS may reflect multitasking: highly-ranked hypertension management programs clearly must be providing their PHC teams with incentives to focus on hypertension control, and thus may be less likely to monitor blood glucose among their hypertensive patients.[22]

While probing these mechanisms is left to future research, it is important to point out the central role of provider incentives for leveraging latent capacity in PHC. Indeed, aligning provider incentives with the goals of improved PHC management was a key component of program effectiveness in the Tongxiang case. Although providing some resources and training, the studied program mostly relied on supply-side incentives and augmented workload for the existing PHC workforce, while allowing patients to choose between enhanced PHC with no additional out-of-pocket payment, on the one hand, or hospital-based care previously available with co-insurance, on the other. The supply-side incentives took the form of reputational and financial incentives for teams of PHC providers, stemming from annual evaluation scores for each township's program, shared among all the townships. Suggestive evidence supports our conjecture that these evaluation incentives were key in prompting more intensive PHC management: townships given low scores for NCD management in year 1 enroll more patients in PHC management in year 2 (Appendix Figure A7), presumably because their PHC physicians exerted more effort to recruit and retain patients to improve their evaluation scores.[23] Moreover, funding for this PHC management program stems from per capita allocations, rather than the fee-for-service payments under social health insurance, so that the social health insurance savings from reduced utilization does not reduce these providers' revenues or incentives.[24]

Our measure of PHC management intensity captures the patient-years of enrollment in the NCD-control component of the essential population health services package. This intensity measure presumably captures the efforts of village doctors and THC staff to convince community residents of the desirability and effectiveness of regular community-level management of their conditions, relative to regular management through hospital outpatient department visits (or lack of management until complications develop). Of course our management intensity metric could also capture patient-to-patient spillover effects within the village—letting others know about the program and persuading each other of its benefit. Any such spillovers reinforce the positive effects from village management, indirectly amplifying physician effort through positive interaction among neighbors. We find statistically significant effects of differing management intensity

[22] On multitasking incentives generally and in healthcare payment, see Holmstrom and Milgrom (1991) and Eggleston (2005).

[23] Current year enrollment numbers and evaluation scores exhibit positive correlation since enrollment represents one of the evaluation factors. Year 1 enrollment and year 2's evaluation score exhibit a flatter relationship, suggesting that mean-reversion is perhaps not the driving force; see Appendix Figure A7.

[24] This fragmentation of financing has been decried as inefficient, and it may contribute to obscuring the PHC program benefits that offset social health insurance expenditures on specialists and hospitalizations; but it does provide a blend of payment incentives in a system dominated until recently by fee-for-service.

in neighboring villages despite any such spillovers across township boundaries.

Although physicians in China have long had financial incentives to prescribe medications because clinics derived revenue from drug dispensing (Currie et al., 2011, 2014), China has removed that 'drug mark-up' revenue from government-owned PHC providers since national health reforms in 2009. Moreover, any such incentive was not differential across the townships in our study, so cannot explain the pattern of adherence improvements that we observe—fewer prescriptions but better adherence to the anti-diabetic medications prescribed, with no significant change in overall drug spending.

Our study contributes evidence on the potential for strengthened PHC to provide accessible, affordable, and decent quality management of common chronic diseases, thereby reducing use of specialist and inpatient services in China. Localities elsewhere in Zhejiang and other urban areas (such as Shanghai and Xiamen) have strengthened PHC over decades with some success, but there has been little evidence of effective interventions in rural areas to date. Our findings imply that appropriately aligning PHC providers' incentives and knowledge with the goal of convincing patients to make appropriate use of PHC can contribute to better health and reduced medical spending, improving welfare. Indeed, a back-of-the-envelope estimate suggests that the resource savings from the Tongxiang PHC management program, in terms of avoided inpatient admissions, substantially outweighed the fiscal costs of the program.

China's health policymakers continue to experiment with incentive reforms to improve the efficiency and sustainability of its universal health coverage, such as through provider payment reforms (Jian et al., 2015; Powell-Jackson et al., 2015) and strengthening PHC, including introduction and nationwide implementation of a voluntary "family doctor" system. The success of these efforts will rest partially on the appropriate crafting of incentives for PHC providers. In Tongxiang, the program initially rewarded effort in enrolling new patients; but gradually as the program enrolled the stock of existing NCD patients in the community—all but those most devoted to hospital-based management—the incentive structure needed to shift to focus on quality of care for those patients rather than the number of new patients. Like most of China, policymakers eschewed mandatory referrals or strict gatekeeping. Our study suggests that it may be promising in China to retain patient choice of care setting while deepening supply-side incentives to strengthen PHC enough to attract patients and increase the net value of care.

Beyond the China context, our study contributes to the health economics literature on PHC in general, as well as to economic assessment of chronic disease management in diverse settings. Regarding the specific challenges facing LMICs, our empirical evidence confirms the latent potential of healthcare workers to be more productive (Okeke and Abubakar, 2020), as well as how financial and reputational incentives can motivate front-line workers and administrators to realize that latent potential (Duflo et al., 2012; Luo et al., 2020).

In sum, this study provides empirical evidence of the causal effects of PHC for improving health outcomes—as measured by control of blood pressure as well as reducing avoidable hospitalizations—while reducing total medical spending for patients with hypertension and diabetes. This evidence from a relatively high per capita income part of rural China suggests that further efforts to strengthen PHC hold considerable promise for improving the control of chronic disease in rural areas as China develops, and the quality, efficiency, and convenience of basic healthcare services in other LMICs as well. Further research with panel data from other settings would be valuable for assessing the ongoing efforts to improve the quality and accountability of PHC in China and other LMICs around the globe.

References

M. Björkman and J. Svensson. Power to the people: evidence from a randomized field experiment on community-based monitoring in Uganda. *The Quarterly Journal of Economics*, 124(2):735–769, 2009.

D. E. Bloom, E. Cafiero, E. Jané-Llopis, S. Abrahams-Gessel, L. R. Bloom, S. Fathima, A. B. Feigl, T. Gaziano, A. Hamandi, M. Mowafi, et al. The global economic burden of noncommunicable diseases. Technical report, Program on the Global Demography of Aging, 2012.

J. M. Brock, A. Lange, and K. L. Leonard. Giving and promising gifts: experimental evidence on reciprocity from the field. *Journal of Health Economics*, 58:188–201, 2018.

P.-M. Brun, J. Bessereau, H. Chenaitia, A.-L. Pradel, C. Deniel, G. Garbaye, R. Melaine, O. Bylicki, and C. Lablanche. Stay and play, or scoop and run? That is the question! *The American Journal of Emergency Medicine*, 32(2):166–170, 2014.

N. Chaudhury, J. Hammer, M. Kremer, K. Muralidharan, and F. H. Rogers. Missing in action: teacher and health worker absence in developing countries. *Journal of Economic perspectives*, 20(1):91–116, 2006.

Z. Chen. Launch of the health-care reform plan in China. *The Lancet*, 373(9672):1322–1324, 2009.

D. Christensen, O. Dube, J. Haushofer, B. Siddiqi, and M. Voors. *Building Resilient Health Systems: Experimental Evidence from Sierra Leone and the 2014 Ebola Outbreak*. The World Bank, 2020.

J. Currie, W. Lin, and W. Zhang. Patient knowledge and antibiotic abuse: Evidence from an audit study in China. *Journal of health economics*, 30(5):933–949, 2011.

J. Currie, W. Lin, and J. Meng. Addressing antibiotic abuse in China: an experimental audit study. *Journal of Development Economics*, 110:39–51, 2014.

J. Das and J. Hammer. Which doctor? combining vignettes and item response to measure clinical competence. *Journal of Development Economics*, 78(2):348–383, 2005.

A. Dube, T. W. Lester, and M. Reich. Minimum wage effects across state borders: Estimates using contiguous counties. *The review of economics and statistics*, 92(4):945–964, 2010.

E. Duflo, R. Hanna, and S. P. Ryan. Incentives work: Getting teachers to come to school. *American Economic Review*, 102(4):1241–78, 2012.

K. Eggleston. Multitasking and mixed systems for provider payment. *Journal of Health Economics*, 24(1): 211–223, 2005.

P. Gertler and C. Vermeersch. *Using performance incentives to improve health outcomes*. The World Bank, 2012.

J. He, D. Gu, J. Chen, X. Wu, T. N. Kelly, J.-f. Huang, J.-c. Chen, C.-S. Chen, L. A. Bazzano, K. Reynolds, et al. Premature deaths attributable to blood pressure in China: a prospective cohort study. *The Lancet*, 374(9703):1765–1772, 2009.

B. Holmstrom and P. Milgrom. Multitask principal-agent analyses: Incentive contracts, asset ownership, and job design. *JL Econ. & Org.*, 7:24, 1991.

E. Huillery and J. Seban. Performance-based financing, motivation and final output in the health sector: Experimental evidence from the Democratic Republic of Congo. 2014.

W. Jian, M. Lu, K. Y. Chan, A. N. Poon, W. Han, M. Hu, and W. Yip. Payment reform pilot in Beijing hospitals reduced expenditures and out-of-pocket payments per admission. *Health Affairs*, 34(10):1745–1752, 2015.

X. Lei, X. Sun, J. Strauss, Y. Zhao, G. Yang, P. Hu, Y. Hu, and X. Yin. Health outcomes and socio-economic status among the mid-aged and elderly in china: Evidence from the CHARLS national baseline data. *The Journal of the Economics of Ageing*, 3:29–43, 2014.

X. Li, J. Lu, S. Hu, K. Cheng, J. De Maeseneer, Q. Meng, E. Mossialos, D. R. Xu, W. Yip, H. Zhang, et al. The primary health-care system in china. *The Lancet*, 390(10112):2584–2594, 2017.

X. Li, H. M. Krumholz, W. Yip, K. K. Cheng, J. De Maeseneer, Q. Meng, E. Mossialos, C. Li, J. Lu, M. Su, et al. Quality of primary health care in china: challenges and recommendations. *The Lancet*, 395(10239): 1802–1812, 2020.

X. Liu, X. Sun, Y. Zhao, and Q. Meng. Financial protection of rural health insurance for patients with hypertension and diabetes: repeated cross-sectional surveys in rural china. *BMC health services research*, 16(1):481, 2016.

J. Lu, Y. Lu, X. Wang, X. Li, G. C. Linderman, C. Wu, X. Cheng, L. Mu, H. Zhang, J. Liu, et al. Prevalence, awareness, treatment, and control of hypertension in China: Data from 1· 7 million adults in a population-based screening study (China peace million persons project). *The Lancet*, 390:2549–2558, 2017.

R. Luo, G. Miller, S. Rozelle, S. Sylvia, and M. Vera-Hernández. Can bureaucrats really be paid like CEOs? Substitution between incentives and resources among school administrators in China. *Journal of the European Economic Association*, 18(1):165–201, 2020.

Q. Meng, H. Yang, W. Chen, Q. Sun, and X. Liu. People's Republic of China health system review. vol. 5 no.7 *World Health Organization, Manila*, 2015.

R. Miller and K. S. Babiarz. Pay-for-performance incentives in low-and middle-income country health programs. Technical report, National Bureau of Economic Research, 2013.

M. Mohanan, M. Vera-Hernández, V. Das, S. Giardili, J. D. Goldhaber-Fiebert, T. L. Rabin, S. S. Raj, J. I. Schwartz, and A. Seth. The know-do gap in quality of health care for childhood diarrhea and pneumonia in rural india. *JAMA pediatrics*, 169(4):349–357, 2015.

E. N. Okeke and I. S. Abubakar. Healthcare at the beginning of life and child survival: Evidence from a cash transfer experiment in Nigeria. *Journal of Development Economics*, 143:102426, 2020.

T. Powell-Jackson, W. C.-M. Yip, and W. Han. Realigning demand and supply side incentives to improve primary health care seeking in rural China. *Health Economics*, 24(6):755–772, 2015.

A. Scott and S. Jan. Primary care. In S. Glied and P. C. Smith, editors, *The Oxford handbook of health economics*, chapter 20, pages 463–485. Oxford University Press, Oxford, 2011.

A. P. Sen and T. DeLeire. How does expansion of public health insurance affect risk pools and premiums in the market for private health insurance? evidence from medicaid and the affordable care act marketplaces. *Health Economics*, 27(12):1877–1903, 2018.

B. T. Shapiro. Positive spillovers and free riding in advertising of prescription pharmaceuticals: The case of antidepressants. *Journal of Political Economy*, 126(1):381–437, 2018.

D. Wu and T. P. Lam. Underuse of primary care in china: the scale, causes, and solutions. *The Journal of the American Board of Family Medicine*, 29(2):240–247, 2016.

D. Wu, T. P. Lam, K. F. Lam, X. D. Zhou, and K. S. Sun. Challenges to healthcare reform in China: profit-oriented medical practices, patients' choice of care and guanxi culture in Zhejiang province. *Health Policy and Planning*, 32(9):1241–1247, 2017.

Y. Xu, L. Wang, J. He, Y. Bi, M. Li, T. Wang, L. Wang, Y. Jiang, M. Dai, J. Lu, et al. Prevalence and control of diabetes in Chinese adults. *JAMA*, 310(9):948–959, 2013.

G. Yang, L. Kong, W. Zhao, X. Wan, Y. Zhai, L. C. Chen, and J. P. Koplan. Emergence of chronic non-communicable diseases in China. *The Lancet*, 372(9650):1697–1705, 2008.

Y. Zhao, E. M. Crimmins, P. Hu, Y. Shen, J. P. Smith, J. Strauss, Y. Wang, and Y. Zhang. Prevalence, diagnosis, and management of diabetes mellitus among older chinese: results from the China health and retirement longitudinal study. *International journal of public health*, 61(3):347–356, 2016.

Y. Zhao, J. Strauss, X. Chen, Y. Wang, J. Gong, Q. Meng, G. Wang, and H. Wang. China health and retirement longitudinal study wave 4 user's guide. *National School of Development, Peking University*, 2020.

Appendix

A1 Tongxiang Social Health Insurance

Tongxiang *hukou* residents as well as full-time students are eligible to enroll in the Tongxiang resident insurance plan, which provides access to medical care at all local village and township-level clinics, 12 Tongxiang county-level hospitals, and selected city-level hospitals in Zhejiang province and Shanghai. In 2015, the plan provided access to PHC with a co-insurance rate of 50% at local community health centers (village clinics and township health centers [THC]), in contrast to 90% at county- or city-level hospital outpatient departments. The coverage of inpatient admissions is relatively generous. For example, in 2015 the co-insurance rate patients paid for an inpatient admission at a THC was 15% and at a hospital was 35%. Pharmaceutical expenditures are covered as well in the program. To enroll in the program, a resident in 2015 needed to pay 260 RMB (41.7 USD) for the annual premium and the local government would supplement the premium with 540 RMB (86.7 USD) per resident. In 2012, more than 95% of eligible residents enrolled in this program.

Eligibility to enroll in the Tongxiang employee insurance plan is restricted to employees who contribute towards social security benefits (i.e., formal sector employees) and eligible retired employees who have contributed to social security for enough time. The program is more generous than the resident insurance plan. In 2015, for employees that are not retired, the plan provided access to PHC with co-insurance rates of 30% at local community health centers (village clinics and THCs) and 50% at county- or city-level hospital outpatient departments. The co-insurance rate patients paid for an inpatient admission at a THC was 10% and at a hospital was 20%. The co-insurance rates were even lower for a retired employee.

The rural residents in Tongxiang are traditionally farmers who do not have formal employment and mostly enroll in the resident insurance program. The overall enrollment rate for private health insurance plans among Chinese residents is low and in general concentrated among the richer population. This paper focuses on the rural residents only, which constitute the majority of the Tongxiang population and are more homogeneous in demographics.

A2 Prescription Drug Coding

Prescription drug information is contained in medical claims data which includes drug ID, filling date, package information and physician prescribed usage. Based on these numbers, we calculate medication indicator and number of days covered by medication for each patient. Detailed procedures are listed below.

First, we identify 36 anti-hypertensive drugs and 23 anti-diabetic drugs based on drug ID in the claims, listed below in Appendix Table A1 together with number of claims in our data:

Table A1: List of anti-hypertensive and anti-diabetic drugs

Anti-Hypertensive		Anti-Diabetic	
Generic Name	# Claims (2015)	Generic Name	# Claims (2015)
Telmisartan	149019	Metformin	99481
Amlodipine	137077	Repaglinide	67610
Irbesartan	113476	Gliclazide	58594
Valsartan	80767	Acarbose	40111
Levamlodipine	69209	Insulin Aspart	17177
Nifedipine	58852	Pioglitazone	16349
Candesartan	42117	Glipizide	15326
Spironolactone	40776	Glimepiride	14759
Triamterene	40507	Insulin Lispro	12255
Felodipine	38174	Pre-Mixed Insulin	11295
Furosemide	37903	Insulin	5585
Metoprolol	34070	Insulin Glargine	5378
Enalapril	23766	Voglibose	4756
ZhenJuJiangYa Tablet	23008	Insulin Detemir	1642
Indapamide	22421	Gliquidone	1310
Hydrochlorothiazide	11610	Rosiglitazone	985
Reserpine	7406	NPH Insulin	880
Losartan	4421	Long-Acting Insulin	538
Clonidine	4126	Recombinant Insulin	157
Fosinopril	3213	Glibenclamide	132
Captopril	3089	Nateglinide	44
Benazepril	2950	Recombinant Insulin Glargine	39
Diltiazem	2361	Insulin from Animal	<11
Bisoprolol	1509		
Lacidipine	1258		
Perindopril	492		
Propranolol	392		
Nitrendipine	362		
Nicardipine	153		
Ramipril	136		
Verapamil	129		
Lercanidipine	56		
Lisinopril	27		
Benidipine	27		
Amiloride	23		
Betaxolol	<11		

Notes: This table the number of claims in 2015 for 36 anti-hypertensive drugs and 23 anti-diabetic drugs based on drug ID in our dataset. Prescription drug information is extract from medical claims data based on drug ID and aggregated at generic name level.

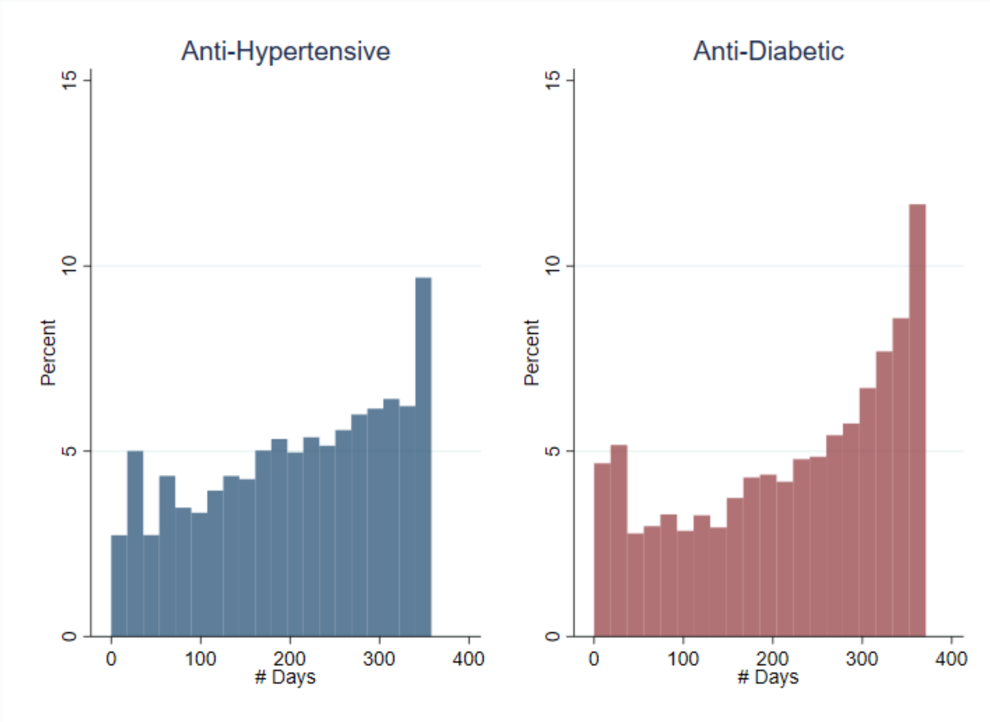
Then, we manually extract package information and physician prescribed usage from the claims data for each medication to code the number of pills, mg per pill, and mg per day from text data associated with that medication. For example, the combination of "25mg*100" and "1 pill a time, 3 times per day" will be coded as 100 pills per pack, 25mg per pill, and 75mg per day. In the original data, 15.07% of claims lack package information, and 83.55% of claims lack physician prescribed usage. To fill in missing values, we applied the three imputation methods: 1) Fill in number of pills and mg per pill using the mode value among claims with the same generic name and price. This allows us to differentiate drugs with different packages across claims. This leaves only 3.26% missing values in package information; 2) Fill in mg per day using the mode value among claims by the same patients and generic name. Assuming patients have stable usage for the same drug within a single year, this allows imputation with patient-specific information. This leaves 47.56% missing values in physician prescribed usage; and 3) Fill in mg per day using the mode value among claims with the same generic name. Here we assume claims without specific listing physician prescribed usage should have the most common usage. However, it does ignore any patient- specific usage and might bias our calculation of medication days. In the end, there are only 0.76% missing values for mg per day (these claims are for drugs that do not have one single claim with physician prescribed usage recorded).

Finally, we calculate number of days covered for each drug claim as # Days = Q * number of pills * mg per pill / mg per day. We trimmed the number at the 95th percentile of each drug to remove extreme values presumably caused by errors in the claim. For multiple drugs prescribed in the same day, we take the maximum length of days (anti-hypertensive and anti-diabetic drugs separately). We then calculate the gap in number of days until the next prescription and take the minimum value of the two, assuming patients will take their drugs properly during this intervening period and refill the prescription when they have almost consumed all previously prescribed drugs. By capping the number of days of a given prescription with the days until next prescription, we fill in the number of days for claims that still have missing values after imputation, and further mitigate problems with using imputed values of mg per day which ignores patient-specific usage. This approach also accounts for switching medication if they stop taking the previously prescribed ones to switch to a better-matched medication before consuming all of the first one.

The number of days with anti-hypertensive and anti-diabetic drugs is aggregated over the year of 2015. We do not calculate the medication coverage before 2015 since claims data in the early years use a different version of drug code and do not contain necessary information on packaging and use. The aggregate numbers are again censored at the 95th percentile to account for extreme values. Figure A1 summarizes the distribution of days covered by anti-hypertensive and anti-diabetic drugs among the patients in our data for

2015.

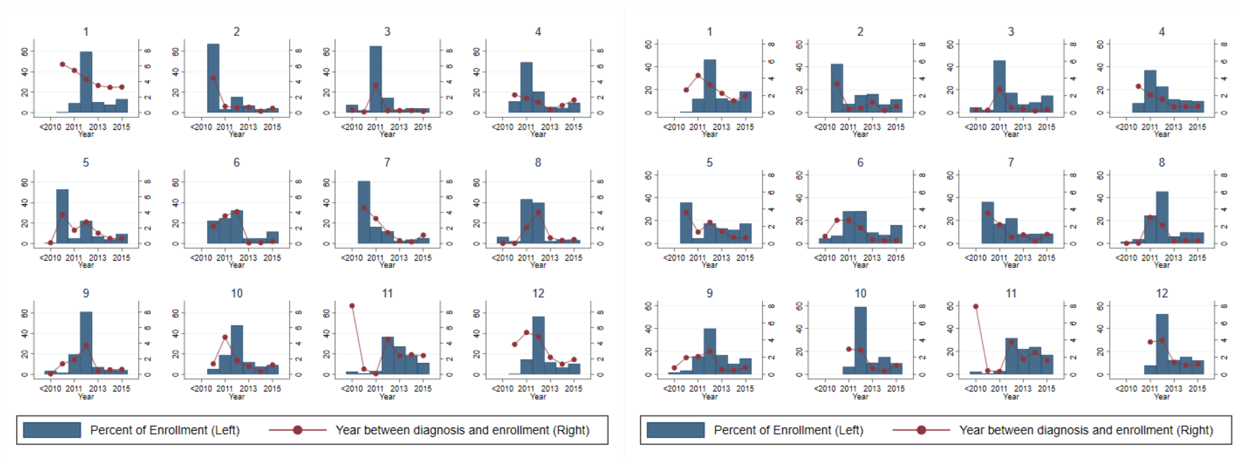
Figure A1: Distribution of Days Covered by Medication



Note: The figure shows the distribution of the number of days covered by anti-hypertensive and anti-diabetic medication among patients with at least one prescription of these respectively. The information is collected from medical claims data in 2015 and calculated following the procedure in Appendix A2

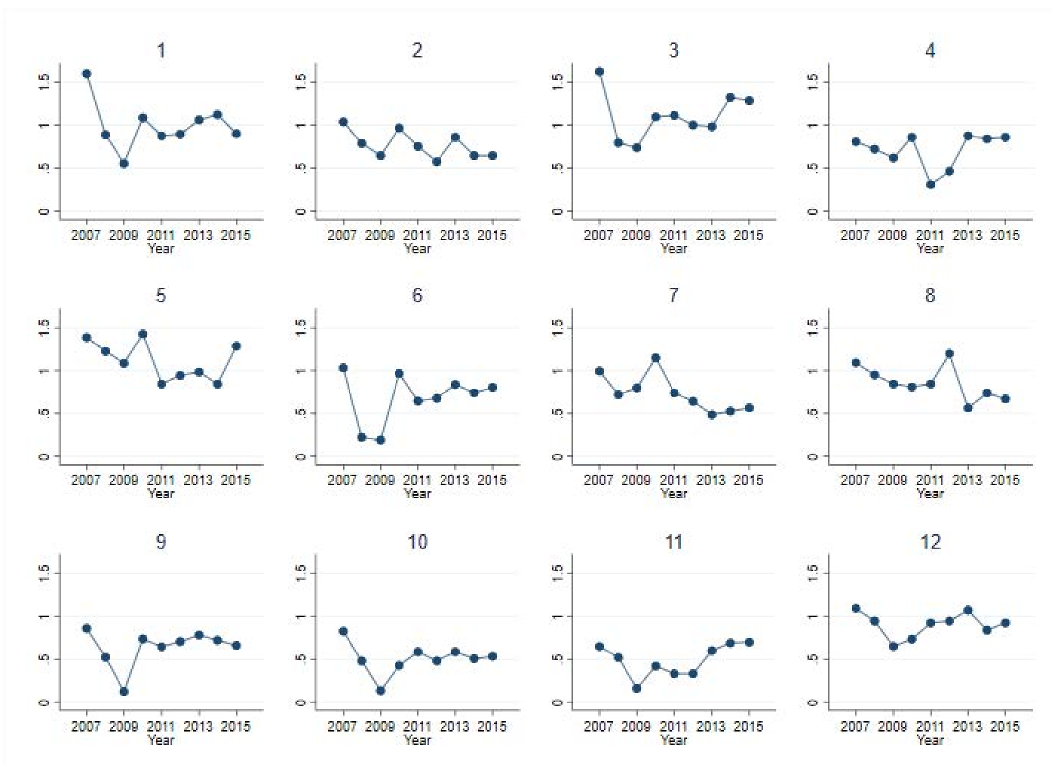
Figure A2: Management Starting Time by Township

(a) Hypertension Management (b) Diabetes Management



Note: The figure shows average starting time of hypertension and diabetes management program by township.

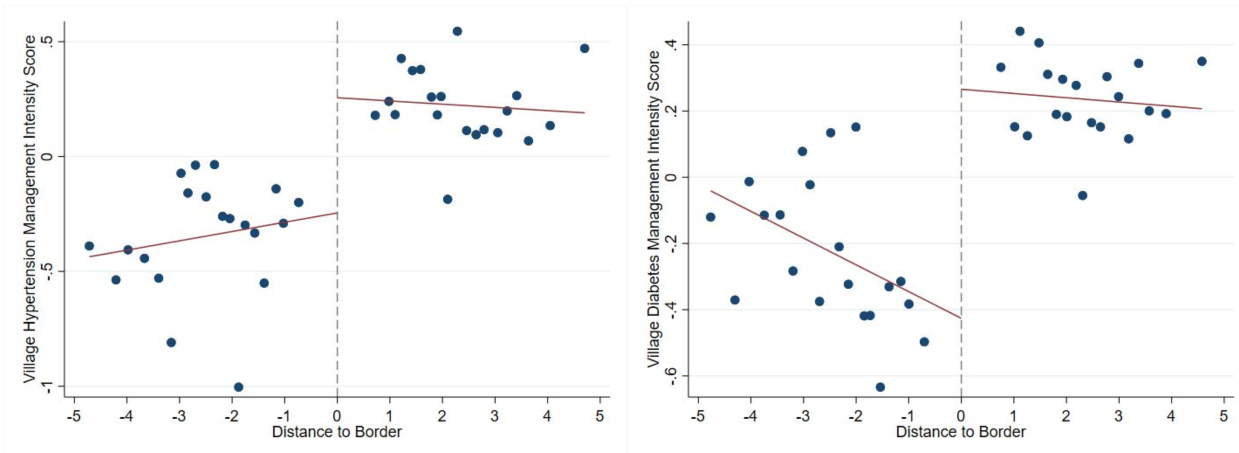
Figure A3: Number of Physicians by Township over Years



Note: The figure shows number of physicians per 1,000 residents in each township during 2011-2015. The townships are labelled 1-12 same as Figure 1 and Appendix Figure A2. Number of physicians are counted as the PHC doctors who maintain, update the basic population database and are assigned to residents as ”responsible doctors”.

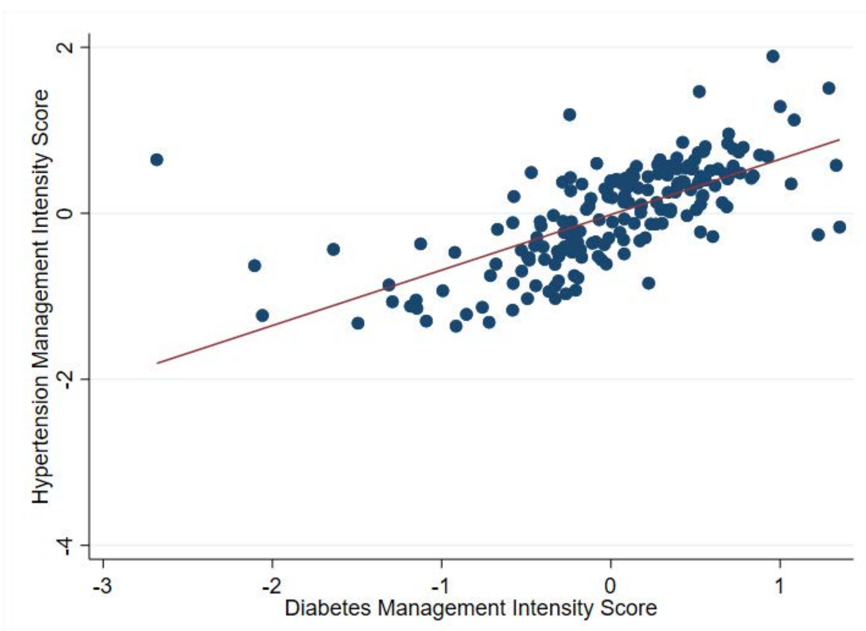
Figure A4: Discontinuity in Village Management Intensity Score across Township Boundary - Hypertention and Diabetes Management Program Separately

(a) Hypertension Management (b) Diabetes Management



Note: The figures are binscatter plots of village-level hypertension and diabetes management intensity score against villages' distance to township border. Panel A shows hypertension management intensity score (I_j^{HP}) while Panel B shows diabetes management intensity score (I_j^{DB}). Management intensity score is calculated as the average residual η_i from Equation 1 among each village for hypertension and diabetes management programs separately. The solid red line is the kernel fit of binscattered dots.

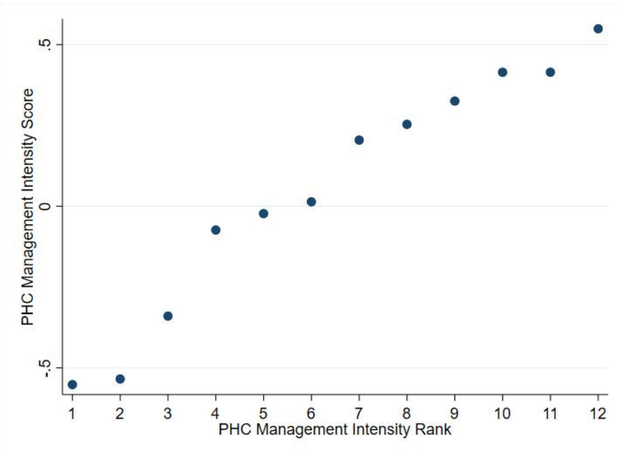
Figure A5: Correlation between Village Hypertension and Diabetic Management Intensity Score



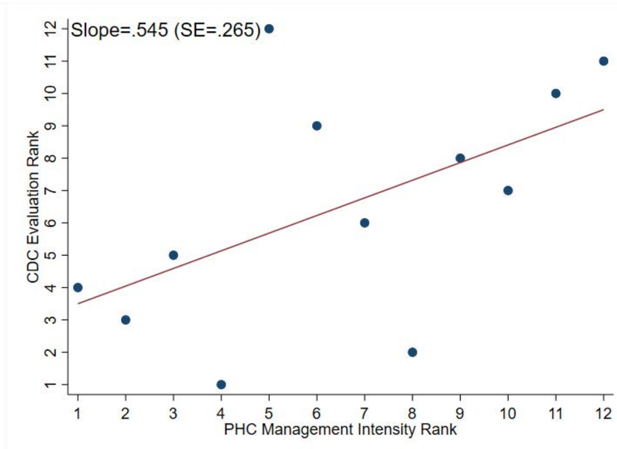
Note: The figure is scatter plots of villages' hypertension management intensity score (I_j^{HP}) and diabetes management intensity score (I_j^{DB}). Management intensity scores are calculated as Equation 2 using residual η_i from Equation 1 estimated in hypertension and diabetes management program separately.

Figure A6: Township Management Intensity Rank, Score, and CDC Evaluation

(a) Management Intensity Score



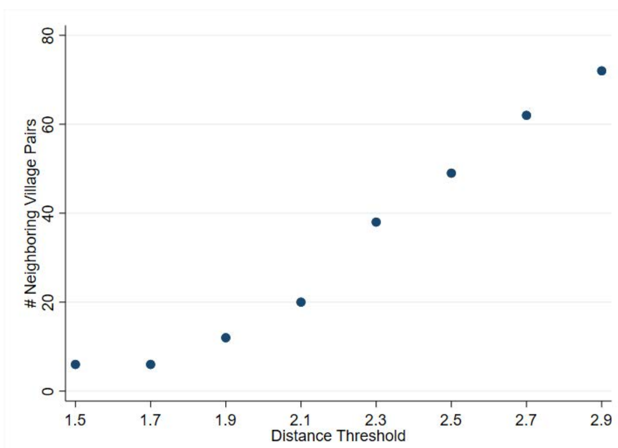
(b) CDC Evaluation Rank

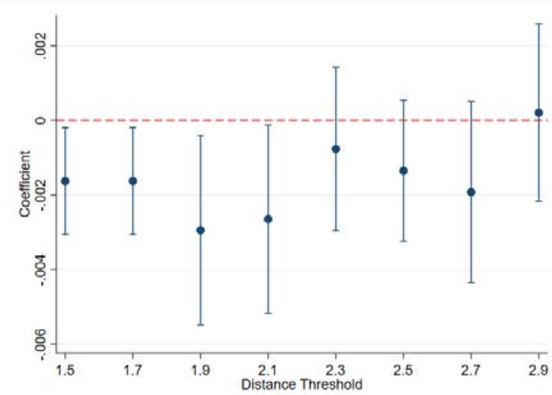


Note: These figures are scatter plots for the correlation between township's management intensity rank ($PHCrank_k$), intensity score (l_k), and CDC evaluation rank. Twelve townships are ordered by their management intensity rank along the x-axis. Panel (a) shows the actual management intensity score (calculated as equation (2) using residual from regression (1). Panel (b) shows CDC evaluation rank (based on average score in 2013-2016) for each township.

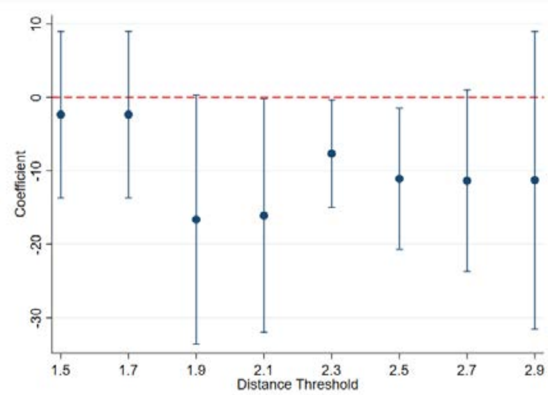
Figure A7: Robustness to Different Threshold Distances for Defining Neighboring Village Pairs

(a) Number of village pairs by distance threshold

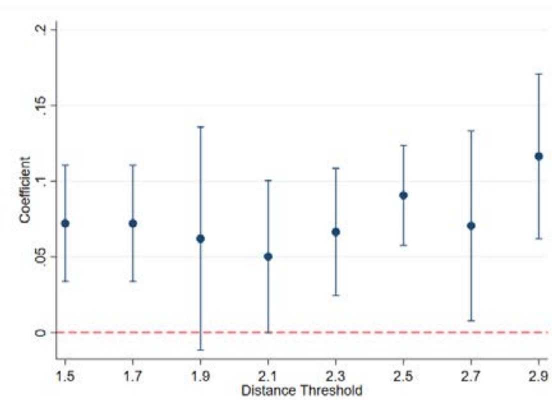




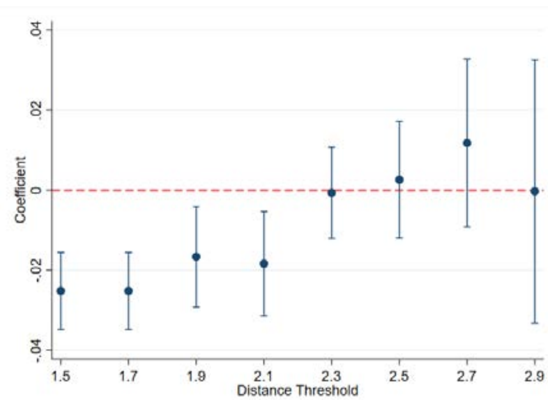
(b) Inpatient



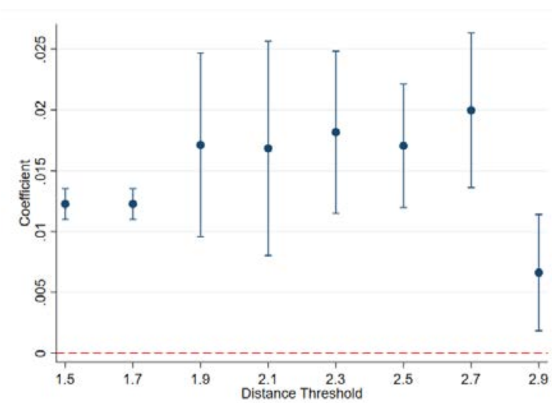
(c) Medical spending



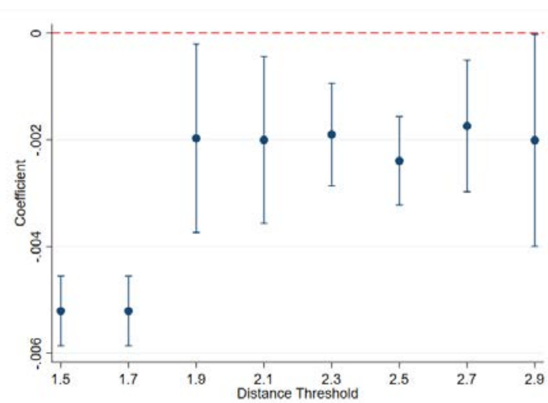
(d) PHC visit



(e) Specialist visit



(f) BP1 under control

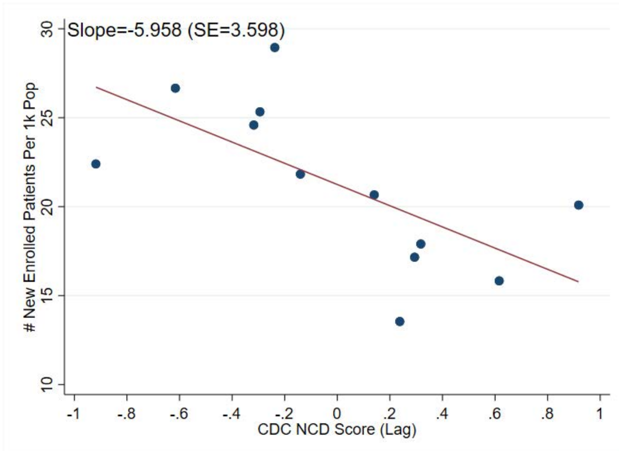


(g) FBS under control

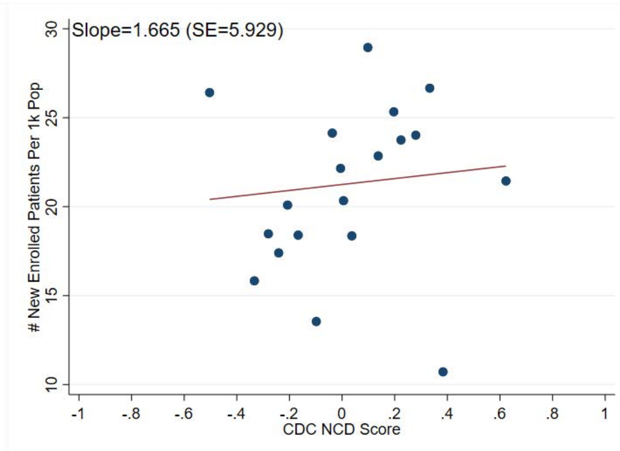
Note: These figures show the robustness check using a series of different distances as the distance threshold for identifying neighboring villages. Panel (a) plots the number of village pairs identified under distance thresholds between 1.5km and 2.9 km. Note that 1.5km and 1.7km thresholds produce the same set of six village pairs. Panels (b)-(g) plot the coefficients for township management intensity rank estimated by equation (3) for care use and equation (5) for blood pressure and FBS indicators. Vertical lines represent the 95 percent confidence intervals for those coefficients, based upon standard errors clustered at the township level.

Figure A7: Correlation between Township Enrollment Effort and Evaluation Score

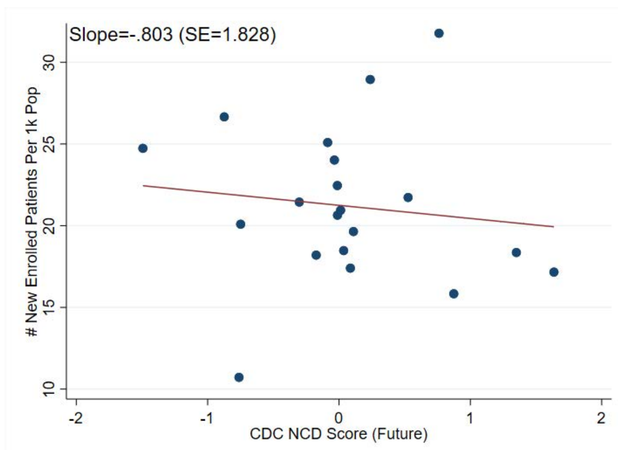
(a) Last Year's Evaluation



(b) This Year's Evaluation



(c) Next Year's Evaluation



Note: These figures are binscatter plots for the correlation between township's evaluation score and the number of new enrolled patients. The level of observation is township by year. Panel A correlates evaluation score in year 1 with enrollment effort in year 2; Panel B correlates evaluation score and enrollment effort of the same year; Panel C correlates evaluation score in year 2 with enrollment effort in year 1. All binscatter plots controlled for township and year fixed effects. The coefficient and standard error (clustered at township level) are reported on the top-left corner of each plot.

Table A2: Summary Statistics

	No Health Checkup (N=16,676)		Health Checkup (Cross-section) (N=45,308)		Health Checkup (Panel) (N=19,180)	
	Mean	SD	Mean	SD	Mean	SD
Demographic						
Age	56.9	10.6	66.1	10.8	67.4	8.0
Male	0.72	0.45	0.43	0.49	0.39	0.49
Education (median)	2		2			
# Year in management	3.4	1.4	3.6	1.3	3.7	1.3
# Year since diagnosis	5.8	3.5	6.5	3.8	6.8	3.9
Utilization						
2015						
Inpatient	0.11	0.31	0.15	0.36	0.13	0.34
Expenditure	3311	12093	3529	10444	2950	7290
Expenditure (Inpatient)	1829	11172	2136	9769	1555	6731
Expenditure (Outpatient)	371	837	396	826	421	766
Expenditure (Drug)	1111	2736	997	1839	975	1338
# Specialist visit	2.5	4.8	2.4	4.5	2.3	4.2
# PHC visit	8.2	8.6	11.8	9.1	13.3	9.3
HP drug	0.65	0.48	0.74	0.44	0.77	0.42
# Days covered by HP drug	212	118	212	113	211	112
DB drug	0.20	0.40	0.18	0.39	0.18	0.39
# Days covered by DB drug	225	112	235	112	237	110
2011						
Inpatient	0.06	0.24	0.08	0.27	0.08	0.27
Expenditure	1658	6916	1695	5361	1708	5405
Expenditure (Inpatient)	847	6034	850	4869	819	5070
Expenditure (Outpatient)	162	605	177	441	197	341
Expenditure (Drug)	650	2060	667	1259	692	1016
# Specialist visit	1.5	3.6	1.7	3.7	1.8	3.5
# PHC visit	4.7	5.9	6.7	6.5	7.3	6.7
Checkup						
FBS under control			0.84	0.36	0.86	0.34
BP under control			0.50	0.50	0.44	0.50
BP under control (stage 2)			0.86	0.34	0.89	0.32

Notes: This table shows summary statistics for patients enrolled in the PHC hypertension and diabetes management programs by whether they had health checkup record. Column 1-2 include patients with no health checkup record. Column 3-4 include patients with at least one health checkup record during 2013-2015 but not both 2013 and 2015. Column 5-6 include patients with health checkup record in both 2013 and 2015.

Table A3: Balance Check between Neighboring Village Pairs

	Village in low intensity township	Village in high intensity township
Township management intensity (l_k)	-0.24	0.30
Village management intensity (l_j)	-0.17	0.22
Population	3,613	3,295
Age	56.6	57.7
Male	0.50	0.49
Education (Median)	2	2
Cost Sharing (Village Clinic / THC)	0.52	0.51
Cost Sharing (County or Higher Level Hospital)	0.76	0.75
Distance to Nearest THC	2.83	2.76
Distance to County or Higher Level Hospital	5.48	5.96
# Physicians	1.22	1.18

Notes: This table shows summary statistics for neighboring villages located in townships with low versus high management-intensity. Management intensity at the township and village levels is defined following Equations (1) and (2). The cost sharing ratio is calculated as the patient's annual out-of-pocket spending divided by total annual spending, and then averaged across patients within the same village. Distances to different levels of hospitals are calculated based on the latitude and longitude of the hospital relative to the centroid of each village. Locations of village clinics are not recorded, but most of the villages have their own clinics.

Table A4: PHC Management Duration - First Stage Results

	(1)	(2)	(3)	(4)	(5)
	Pooled Sample	HP Sample		DB Sample	
	PHC Management Duration				
PHC Rank (HPDB Management)	0.104***	0.106***		0.0891***	
	(0.00724)	(0.00785)		(0.00860)	
PHC Rank (HP Management)			0.106***		
			(0.00785)		
PHC Rank (DB Management)					0.0947***
					(0.00491)
Age	-0.000222	-0.00136	-0.00136	0.0105	0.00928
	(0.00626)	(0.00696)	(0.00696)	(0.00761)	(0.00754)
Age^2	8.22e-06	1.58e-05	1.58e-05	-7.08e-05	-6.30e-05
	(4.56e-05)	(4.91e-05)	(4.91e-05)	(5.77e-05)	(5.75e-05)
Male	-0.0204*	-0.0208	-0.0208	-0.0341*	-0.0384*
	(0.0107)	(0.0132)	(0.0132)	(0.0181)	(0.0181)
# Years of school	0.00268	0.00249	0.00249	0.00152	0.000519
	(0.00366)	(0.00414)	(0.00414)	(0.00486)	(0.00435)
# Years since Diagnosis	0.662***	0.659***	0.659***	0.267***	0.266***
	(0.0419)	(0.0429)	(0.0429)	(0.0238)	(0.0238)
# Years since $Diagnosis^2$	-0.0289***	-0.0286***	-0.0286***	0.000901***	0.000900***
	(0.00189)	(0.00192)	(0.00192)	(6.80e-05)	(6.85e-05)
# Physicians	-0.193***	-0.216***	-0.216***	-0.0720	0.0683
	(0.0518)	(0.0534)	(0.0534)	(0.0792)	(0.0561)
HP Management	0.222***				
	(0.0387)				
DB Management	0.0263				
	(0.0329)				
Constant	0.327	0.609*	0.609*	0.870***	0.730**
	(0.271)	(0.290)	(0.290)	(0.224)	(0.312)
Observations	74,627	67,813	67,813	14,543	14,543
R-squared	0.625	0.621	0.621	0.468	0.471
Mean of Dep Var	3.532	3.617	3.617	3.055	3.055
F-stat	204.6	181.4	181.4	107.2	371.6

Notes: This table shows the first-stage result of the 2SLS analysis with township management intensity rank as the instrument for individual management duration. Column 1 uses the pooled sample including all patients enrolled in the hypertension (HP) and/or diabetes (DB) management programs. Column 2-3 and 4-5 use separate samples for hypertension and diabetes management programs respectively. PHC Rank (HPDB Management) is the township management intensity rank based on the average residual from estimating Equation 1 using the pooled sample, while PHC Rank (HP Management) and PHC Rank (DB Management) are the management intensity ranks estimated using the hypertension and diabetes program samples separately. All regressions include patients' age, gender, educational attainment (years of schooling), management program, years since diagnosis, and physician capacity in the township of residence (in 2010) as controls. In the pooled sample, length of management is calculated as the larger of either the hypertension or diabetes management program; length of diagnosis similarly is defined by the earliest diagnosis, whether of hypertension or diabetes. Standard errors are clustered at the township level. Statistical significance at the 10%, 5% and 1% levels are denoted by *, **, and ***.

Table A5: Correlation between Biomarker Measurements and Inpatient Care

	(1)	(2)	(3)
	Any Inpatient (2015)		
BP1 under control	-0.00338		
	(0.00394)		
BP2 under control		-0.0109**	
		(0.00413)	
FBS under control			-0.0345***
			(0.00461)
Observation	58,901	58,901	56,305
Mean of Dep Var	0.144	0.144	0.146

Notes: This table presents the correlation between whether having inpatient in 2015 and biomarker measurements from health checkup record, i.e. whether blood pressure and FBS are under control. The sample includes all patients enrolled in hypertension or diabetes PHC management program with at least one health checkup record during 2013-2015. All regressions include township fixed effects, individual and township-level characteristics, i.e. gender, age, age squared, years of schooling, years since diagnosis (duration Dx), duration Dx squared, indicators for whether enrolled in PHC management for hypertension and/or for diabetes (for the pooled sample), and physician capacity in the township in 2010. Statistical significance at the 10%, 5% and 1% levels are denoted by *, **, and ***.

Table A6: Impact of PHC Management - 2SLS, HP/DB separately

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
								HP drug		DB drug					
	Inpatient	Expenditure	Expenditure (Inpatient)	Expenditure (Outpatient)	Expenditure (Drug)	# Specialist Visit	# PHC Visit	Any	# Days	Any	# Days	Any Health Checkup	BP1 under control	BP2 under control	FBS under control
Panel A: Hypertension															
PHC Duration	-0.0323**	-187.8**	-199.8**	18.07	22.49	-0.260***	1.447***	0.0282	-9.767	-0.0380***	20.28	0.0270**	0.182***	0.0594***	-0.0397***
	(0.0143)	(87.46)	(78.09)	(16.45)	(24.18)	(0.0699)	(0.499)	(0.0372)	(9.768)	(0.00868)	(13.14)	(0.0127)	(0.0655)	(0.0205)	(0.00808)
Observations	11,432	11,432	11,432	11,432	11,432	11,432	11,432	11,432	8,831	11,432	1,464	11,432	9,014	9,014	8,743
Mean of Dep.Var (2015)	0.130	2194	947.4	326.7	799.4	1.767	11.42	0.772	209.2	0.128	227.3	0.802	0.508	0.896	0.889
Panel B: Diabetes															
PHC Duration	-0.0379	-88.87	-57.63	-88.79**	-22.23	-0.618***	-0.371	0.0310	-14.90	-0.124***	18.78*	0.108***	0.199***	0.0318*	0.0422
	(0.0274)	(218.9)	(148.2)	(41.10)	(58.55)	(0.230)	(0.606)	(0.0294)	(14.68)	(0.0357)	(10.44)	(0.0291)	(0.0665)	(0.0177)	(0.0477)
Observations	2,620	2,620	2,620	2,620	2,620	2,620	2,620	2,620	1,562	2,620	2,167	2,620	1,981	1,981	1,887
Mean of Dep.Var (2015)	0.159	3397	1161	454.6	1456	3.597	14.26	0.596	205.9	0.827	253.7	0.772	0.552	0.907	0.436
Border FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
2011 Controls								Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: This table replicates 2SLS estimates in Table 4, Table 5 and 7 using management intensity rank estimated separately for hypertension and diabetes management program. The sample includes all enrolled chronic-disease patients aged 40+ with agricultural hukou in the 14 pairs of adjacent villages located in different townships (the border sample). The observation is at the patient level. Panel A) shows results for the pooled sample of individuals in either or both programs; panel B) uses enrolled hypertension patients only and Panel C) uses diabetic patients only. Three sets of dependent variables are presenting together. Column 1-7 are healthcare utilization and spending changes in 2015 less that of 2011. All expenditures are winsorized at the 95th percentile within our claims sample. Specialist visits include all visits at provincial or municipality level hospitals, while PHC visits includes visits at township and village clinics. Column 8-11 are whether had any and the number of days covered by anti-hypertensive or anti-diabetic drugs in 2015 (see Appendix A2 for detailed construction process). Column 12-15 are whether having health checkup, whether blood pressure is under control (systolic blood pressure (SBP) < 140 mm Hg and diastolic blood pressure (DBP) < 90 mm Hg), whether blood pressure is below Stage II (SBP < 160 and DBP < 100), and whether fasting blood sugar under control (FBS < 7). These measurements are based on average reading in all health checkup during 2013-2015. All regressions include boundary pair fixed effects, individual and township-level characteristics, i.e. gender, age, age squared, years of schooling, years since diagnosis (duration Dx), duration Dx squared, indicators for whether enrolled in PHC management for hypertension and/or for diabetes (for the pooled sample), and physician capacity in the township in 2010. Column 8-15 further include baseline health controls, i.e. total medical spending and whether had hospital admission in 2011. All regressions include boundary pair fixed effects, individual and township-level characteristics, i.e. gender, age, age squared, years of schooling, years since diagnosis (duration Dx), duration Dx squared, indicators for whether enrolled in PHC management for hypertension and/or for diabetes (for the pooled sample), and physician capacity in the township in 2010. Standard errors are clustered at the township level. Statistical significance at the 10%, 5% and 1% levels are denoted by *, **, and ***.

Table A7: Impact of PHC Management - 2SLS, Z-score

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
								HP drug		DB drug					
	Inpatient	Expenditure	Expenditure (Inpatient)	Expenditure (Outpatient)	Expenditure (Drug)	# Specialist Visit	# PHC Visit	Any	# Days	Any	# Days	Any Health Checkup	BP1 under control	BP2 under control	FBS under control
Panel A: Pool															
PHC Duration	-0.0222	-89.98	-112.7	22.95	31.53	-0.166*	1.326**	0.0181	-5.312	-0.0298***	16.57**	0.0229	0.158***	0.0546***	-0.0263***
	(0.0136)	(92.75)	(71.95)	(16.69)	(22.48)	(0.0851)	(0.604)	(0.0373)	(10.17)	(0.00476)	(7.921)	(0.0164)	(0.0528)	(0.0186)	(0.00714)
Observations	12,580	12,580	12,580	12,580	12,580	12,580	12,580	12,580	9,158	12,580	2,439	12,580	9,775	9,775	9,453
Mean of Dep.Var (2015)	0.132	2294	958.5	338	855.6	1.968	11.47	0.728	207.1	0.194	236.5	0.790	0.523	0.902	0.850
Panel B: Hypertension															
PHC Duration	-0.0213	-93.69	-140.2*	32.01**	41.18	-0.194**	1.630***	0.0251	-4.252	-0.0246***	23.27**	0.0216	0.159***	0.0563***	-0.0388***
	(0.0141)	(85.30)	(74.67)	(15.17)	(26.16)	(0.0860)	(0.615)	(0.0436)	(10.36)	(0.00581)	(9.333)	(0.0157)	(0.0550)	(0.0204)	(0.00634)
Observations	11,382	11,382	11,382	11,382	11,382	11,382	11,382	11,382	8,789	11,382	1,464	11,382	8,977	8,977	8,709
Mean of Dep.Var (2015)	0.130	2193	946.8	326.4	799.5	1.769	11.41	0.772	209.2	0.129	227.3	0.802	0.508	0.897	0.889
Panel C: Diabetes															
PHC Duration	-0.00658	83.03	65.21	-35.86	-12.42	-0.260	-0.0194	-0.0313	-9.127	-0.0697***	15.00*	0.0490***	0.127***	0.0228**	0.00188
	(0.0206)	(150.6)	(115.0)	(23.59)	(31.99)	(0.160)	(0.371)	(0.0205)	(7.588)	(0.0201)	(7.834)	(0.0145)	(0.0359)	(0.0114)	(0.0300)
Observations	2,609	2,609	2,609	2,609	2,609	2,609	2,609	2,609	1,561	2,609	2,157	2,609	1,974	1,974	1,882
Mean of Dep.Var (2015)	0.160	3399	1166	453.8	1454	3.591	14.28	0.598	206	0.827	253.7	0.772	0.551	0.906	0.436
Border FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
2011 Controls								Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: This table replicates IV estimates in Table 4, Table 5 and 7 using the z-score of township management intensity rank. The sample includes all enrolled chronic-disease patients aged 40+ with agricultural hukou in the 14 pairs of adjacent villages located in different townships (the border sample). The observation is at the patient level. Panel A) shows results for the pooled sample of individuals in either or both programs; panel B) uses enrolled hypertension patients only and Panel C) uses diabetic patients only. Three sets of dependent variables are presenting together. Column 1-7 are healthcare utilization and spending changes in 2015 less that of 2011. All expenditures are winsorized at the 95th percentile within our claims sample. Specialist visits include all visits at provincial or municipality level hospitals, while PHC visits includes visits at township and village clinics. Column 8-11 are whether had any and the number of days covered by anti-hypertensive or anti-diabetic drugs in 2015 (see Appendix A2 for detailed construction process). Column 12-15 are whether having health checkup, whether blood pressure is under control (systolic blood pressure (SBP) < 140 mm Hg and diastolic blood pressure (DBP) < 90 mm Hg), whether blood pressure is below Stage II (SBP < 160 and DBP < 100), and whether fasting blood sugar under control (FBS < 7). These measurements are based on average reading in all health checkup during 2013-2015. All regressions include boundary pair fixed effects, individual and township-level characteristics, i.e. gender, age, age squared, years of schooling, years since diagnosis (duration Dx), duration Dx squared, indicators for whether enrolled in PHC management for hypertension and/or for diabetes (for the pooled sample), and physician capacity in the township in 2010. Column 8-15 further include baseline health controls, i.e. total medical spending and whether had hospital admission in 2011. All regressions include boundary pair fixed effects, individual and township-level characteristics, i.e. gender, age, age squared, years of schooling, years since diagnosis (duration Dx), duration Dx squared, indicators for whether enrolled in PHC management for hypertension and/or for diabetes (for the pooled sample), and physician capacity in the township in 2010. Standard errors are clustered at the township level. Statistical significance at the 10%, 5% and 1% levels are denoted by *, **, and ***.

Table A8: Impact of PHC Management - 2SLS, Leave-one-out Z-score

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
								HP drug		DB drug					
	Inpatient	Expenditure	Expenditure (Inpatient)	Expenditure (Outpatient)	Expenditure (Drug)	# Specialist Visit	# PHC Visit	Any	# Days	Any	# Days	Any Health Checkup	BP1 under control	BP2 under control	FBS under control
Panel A: Pool															
PHC Duration	-0.0249	-108.9	-129.1	23.54	33.07	-0.166*	1.344**	0.0128	-6.225	-0.0307***	17.57**	0.0271	0.178***	0.0583***	-0.0269***
	((0.0152)	(110.6)	(85.33)	(18.45)	(23.99)	(0.0966)	(0.660)	(0.0390)	(10.66)	(0.00532)	(8.559)	(0.0180)	(0.0592)	(0.0206)	(0.00748)
Observations	12,580	12,580	12,580	12,580	12,580	12,580	12,580	12,580	9,158	12,580	2,439	12,580	9,775	9,775	9,453
Mean of Dep.Var (2015)	0.132	2294	958.5	338	855.6	1.968	11.47	0.728	207.1	0.194	236.5	0.790	0.523	0.902	0.850
Panel B: Hypertension															
PHC Duration	-0.0243	-113.0	-160.2*	33.75**	43.62	-0.200**	1.692**	0.0185	-5.604	-0.0251***	24.63**	0.0261	0.181***	0.0606***	-0.0404***
	(0.0157)	(104.2)	(88.92)	(16.68)	(28.12)	(0.0982)	(0.693)	(0.0451)	(10.78)	(0.00641)	(10.34)	(0.0169)	(0.0626)	(0.0229)	(0.00652)
Observations	11,382	11,382	11,382	11,382	11,382	11,382	11,382	11,382	8,789	11,382	1,464	11,382	8,977	8,977	8,709
Mean of Dep.Var (2015)	0.130	2193	946.8	326.4	799.5	1.769	11.41	0.772	209.2	0.129	227.3	0.802	0.508	0.897	0.889
Panel C: Diabetes															
PHC Duration	-0.00749	67.90	55.48	-41.34*	-9.289	-0.263	-0.246	-0.0334	-7.781	-0.0720***	15.98*	0.0507***	0.130***	0.0217*	0.00391
	(0.0207)	(152.4)	(113.9)	(24.28)	(32.14)	(0.171)	(0.426)	(0.0219)	(8.133)	(0.0203)	(8.307)	(0.0150)	(0.0360)	(0.0113)	(0.0306)
Observations	2,609	2,609	2,609	2,609	2,609	2,609	2,609	2,609	1,561	2,609	2,157	2,609	1,974	1,974	1,882
Mean of Dep.Var (2015)	0.160	3399	1166	453.8	1454	3.591	14.28	0.598	206	0.827	253.7	0.772	0.551	0.906	0.436
Border FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
2011 Controls								Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: This table replicates 2SLS estimates in Table 4, Table 5 and 7 using leave-one-out z-score management intensity. The sample includes all enrolled chronic-disease patients aged 40+ with agricultural hukou in the 14 pairs of adjacent villages located in different townships (the border sample). The observation is at the patient level. Panel A) shows results for the pooled sample of individuals in either or both programs; panel B) uses enrolled hypertension patients only and Panel C) uses diabetic patients only. Three sets of dependent variables are presenting together. Column 1-7 are healthcare utilization and spending changes in 2015 less that of 2011. All expenditures are winsorized at the 95th percentile within our claims sample. Specialist visits include all visits at provincial or municipality level hospitals, while PHC visits includes visits at township and village clinics. Column 8-11 are whether had any and the number of days covered by anti-hypertensive or anti-diabetic drugs in 2015 (see Appendix A2 for detailed construction process). Column 12-15 are whether having health checkup, whether blood pressure is under control (systolic blood pressure (SBP) $<$ 140 mm Hg and diastolic blood pressure (DBP) $<$ 90 mm Hg), whether blood pressure is below Stage II (SBP $<$ 160 and DBP $<$ 100), and whether fasting blood sugar under control (FBS $<$ 7). These measurements are based on average reading in all health checkup during 2013-2015. All regressions include boundary pair fixed effects, individual and township-level characteristics, i.e. gender, age, age squared, years of schooling, years since diagnosis (duration Dx), duration Dx squared, indicators for whether enrolled in PHC management for hypertension and/or for diabetes (for the pooled sample), and physician capacity in the township in 2010. Column 8-15 further include baseline health controls, i.e. total medical spending and whether had hospital admission in 2011. All regressions include boundary pair fixed effects, individual and township-level characteristics, i.e. gender, age, age squared, years of schooling, years since diagnosis (duration Dx), duration Dx squared, indicators for whether enrolled in PHC management for hypertension and/or for diabetes (for the pooled sample), and physician capacity in the township in 2010. Standard errors are clustered at the township level. Statistical significance at the 10%, 5% and 1% levels are denoted by *, **, and ***.

Table A9: Impact of PHC Management - Reduced Form, Age below 75

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
	Inpatient	Expenditure	Expenditure (Inpatient)	Expenditure (Outpatient)	Expenditure (Drug)	# Specialist Visit	# PHC Visit	HP drug		DB drug		Any Health Checkup	BP1 under control	BP2 under control	FBS under control
								Any	# Days	Any	# Days				
Panel A: Pool															
PHC Rank	-0.00196*	-10.84**	-9.241**	0.490	-0.210	-0.0145*	0.0816**	0.00234	-0.875	-0.00207***	1.660**	0.00215*	0.0143**	0.00385**	-0.00250**
	(0.000903)	(4.499)	(3.421)	(1.543)	(1.935)	(0.00677)	(0.0352)	(0.00216)	(0.826)	(0.000326)	(0.639)	(0.00103)	(0.00495)	(0.00117)	(0.000768)
Observations	10,161	10,161	10,161	10,161	10,161	10,161	10,161	10,161	7,327	10,161	2,150	10,161	7,598	7,598	7,272
Mean of Dep. Var (2015)	0.119	2264	868.4	356.6	887.5	2.093	11.38	0.721	206.7	0.212	239.5	0.761	0.498	0.896	0.839
Panel B: Hypertension															
PHC Rank	-0.00182*	-10.25**	-10.98**	1.250	0.795	-0.0145*	0.104**	0.00260	-0.764	-0.00186***	2.584***	0.00159	0.0142**	0.00381**	-0.00362***
	(0.000972)	(3.280)	(3.815)	(1.404)	(1.970)	(0.00652)	(0.0323)	(0.00261)	(0.828)	(0.000349)	(0.750)	(0.00113)	(0.00516)	(0.00123)	(0.000696)
Observations	9,036	9,036	9,036	9,036	9,036	9,036	9,036	9,036	6,996	9,036	1,226	9,036	6,864	6,864	6,592
Mean of Dep. Var (2015)	0.116	2143	852.3	344	821.9	1.860	11.32	0.774	209	0.136	231.1	0.772	0.478	0.889	0.885
Panel C: Diabetes															
PHC Rank	-0.00359	-22.20	-13.05	-6.802*	-8.141*	-0.0665***	-0.00898	-0.00263	-1.709	-0.00676**	1.422*	0.00600***	0.0167***	0.00318**	-0.00168
	(0.00231)	(16.49)	(12.87)	(3.266)	(4.198)	(0.0196)	(0.0410)	(0.00292)	(1.054)	(0.00288)	(0.711)	(0.00148)	(0.00468)	(0.00124)	(0.00384)
Observations	2,318	2,318	2,318	2,318	2,318	2,318	2,318	2,318	1,338	2,318	1,926	2,318	1,705	1,705	1,613
Mean of Dep. Var (2015)	0.150	3361	1081	465.3	1473	3.686	14.05	0.577	204.3	0.831	254.5	0.752	0.549	0.904	0.427
Border FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
2011 Controls								Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: This table replicates reduced form estimates in Table 3, Table 5 and 6 using a restricted sample of managed patients below age 75. Panel A) shows results for the pooled sample of individuals in either or both programs; panel B) uses enrolled hypertension patients only and Panel C) uses diabetic patients only. Three sets of dependent variables are presenting together. Column 1-7 are healthcare utilization and spending changes in 2015 less that of 2011. All expenditures are winsorized at the 95th percentile within our claims sample. Specialist visits include all visits at provincial or municipality level hospitals, while PHC visits includes visits at township and village clinics. Column 8-11 are whether had any and the number of days covered by anti-hypertensive or anti-diabetic drugs in 2015 (see Appendix A2 for detailed construction process). Column 12-15 are whether having health checkup, whether blood pressure is under control (systolic blood pressure (SBP) $<$ 140 mm Hg and diastolic blood pressure (DBP) $<$ 90 mm Hg), whether blood pressure is below Stage II (SBP $<$ 160 and DBP $<$ 100), and whether fasting blood sugar under control (FBS $<$ 7). These measurements are based on average reading in all health checkup during 2013-2015. All regressions include boundary pair fixed effects, individual and township-level characteristics, i.e. gender, age, age squared, years of schooling, years since diagnosis (duration Dx), duration Dx squared, indicators for whether enrolled in PHC management for hypertension and/or for diabetes (for the pooled sample), and physician capacity in the township in 2010. Column 8-15 further include baseline health controls, i.e. total medical spending and whether had hospital admission in 2011. All regressions include boundary pair fixed effects, individual and township-level characteristics, i.e. gender, age, age squared, years of schooling, years since diagnosis (duration Dx), duration Dx squared, indicators for whether enrolled in PHC management for hypertension and/or for diabetes (for the pooled sample), and physician capacity in the township in 2010. Standard errors are clustered at the township level. Statistical significance at the 10%, 5% and 1% levels are denoted by *, **, and ***.