Doctor competence and the demand for healthcare: evidence from rural China

May 14, 2016

Abstract

The agency problem between patients and doctors has long been emphasised in the health economics literature, but the empirical evidence on whether patients can evaluate and respond to better quality care remains mixed and inconclusive. Using household data linked to an assessment of village doctors' clinical competence in rural China, we show that there is no correlation between doctor competence and patients' healthcare utilisation, with confidence intervals reasonably tight around zero. Household perceptions of quality are an important determinant of care seeking behaviour yet patients appear unable to recognise more competent doctors -there is no relationship between doctor competence and perceptions of quality.

Key Words: *Doctor Competence, Quality of Care, Perceptions, Item Response Theory*

JEL Classification: *I12, I15, I18, C29.*

1 Introduction

Economists have long emphasized the problem of asymmetrical information in the provision of healthcare with patients finding it difficult to evaluate the quality of services (Arrow, 1963;Arrow, 1986; Chalkley and Malcomson, 1998; Dranove and White, 1987; Gaynor, 1994; Glied and Remler, 2002; Mooney and Ryan, 1993). Patients do not have specialist knowledge of their condition and there is an inherent unpredictability in how individuals respond to medical care (Arrow, 1963). Indeed patients seek care precisely because they lack the medical training and expertise of their doctor. In many countries it falls to regulatory bodies to enforce quality of care standards, partly because patients are not thought to be able to recognize and respond to better quality.

At the same time patients are not considered completely ignorant. Policies seeking to expand patient choice, for example, are seen as a way of sharpening provider incentives to deliver quality care and rest on the idea of the patient as an intelligent consumer. Such policies encouraging competition, however, remain controversial and the evidence is mixed (Cooper et al., 2011; Bloom et al., 2010; Propper et al., 2004; Kessler and McClellan, 2000; Pollock et al., 2011). Many theoretical analyses model patient demand as a function of quality and assume the market can impose some discipline on the behavior of physicians (Chalkley and Tilley, 2005; Chalkley and Khalil, 2005; Ma and T. McGuire, 1997). There are multiple channels through which patients can infer quality, the most obvious being through repeated exposure to health services (Corno, 2014). However, as emphasised by the literature on empirical industrial organisation, (see seminal work by Perloff and Salop, 1985, Berry, Levinsohn, and Pakes, 1995 as well as the survey article by Einav and Levin, 2010) the accuracy of consumers' inferences, and those of patients in particular, are likely to be variable and ridded by unobserved heterogeneity. As a result, how the demand for healthcare is affected by quality remains an empirical question.

In this paper, we examine the correlates of demand for primary healthcare in rural China, using novel data on health providers linked to household information on health care seeking. Our analysis focuses on the relationship between doctor competence and utilization of healthcare at the primary level. Village doctors are meant to be the first point of contact under the rural health system in China, providing primary care services to the rural population under the management of township health centres (Barbiarz et al., 2012). Strengthening primary care

is a key focus of the latest health care reforms (Chen, 2009; Yip et al., 2012) and ultimately the intention is that village doctors will act as gatekeepers to higher tiers of healthcare in the public system (Yip and Hsiao, 2009; Yip et al., 2012).

Much of the existing literature on determinants of health care utilization focuses on the more observable dimensions of health care, such as prices, location and the availability of staff, drugs, equipment and infrastructure. Our contribution relates to a second strand of the literature which examines whether patients are sensitive to technical aspects of quality that are harder for patients to observe and typically difficult to measure. Such measures of technical quality usually relate to diagnosis and treatment taking place during the interaction between the patient and clinician, referred to process of care in the quality of care literature (Donabedian, 1988; Donabedian, 2005). Process measures are generally informed by well-conducted clinical trials or observational studies that have demonstrated their effectiveness on health outcomes (Donabedian, 1988).

The paper is organized as follows. We first provide some context by describing the study setting and ongoing health reforms at the time of the research. We then outline our approach to measuring doctor competence as well as other important features of the data, before describing our statistical methods. Finally, we present the results, interpret the findings, and discuss possible implications for policy.

2 Study Setting

Our study is located in Ningxia, a low income province in the Northwest part of China. It has a population of 6.25 million. The ethnic composition is 63 percent Han and 36 percent Hui. With a per capita annual income of 15,551 RMB in urban areas and 6,627 RMB in rural areas, Ningxia was the third poorest province in China in 2009 (Ningxia Statistical Yearbook, 2010). Life expectancy at the time of the study was 74 years. Expansion of health insurance through the New Cooperative Medical Scheme (NCMS) has been rapid. Introduced in 2003, the NCMS is now estimated to cover more than 95 percent of the rural population.

Healthcare provision at the village level has a long history in China. Since the 1950s Great Leap Forward, and until Deng Xiaopings economic reforms in the 1980s, agricultural production in rural China was organised in communes, each encompassing several production

brigades. During this period land was owned by the state and collectivism was prioritised over individual labour. Within this system, rural health financing was organised around communal medical funds which contribute to the Rural Cooperative Medical System. Yet, despite the existence of this formal system, by the 1960s delivery of health care in rural areas was still scarce (McConnell, 1993, Zhang and Unschuld, 2008, White, 1988). This led to the introduction of a national policy on barefoot doctors by which teams of doctors from urban hospitals were mobilised to rural areas to care and train local paramedics.

The village healthcare providers or barefoot doctors of the 1960s were selected and paid by the Rural Cooperative Medical System. These individuals received between 3 and 6 months of training, with an emphasis on prevention of epidemics and the treatment of common diseases. These doctors had limited capabilities, earning about 50 percent less than regular physicians and, as members of a commune they had to earn work points to pay for their food and other goods at the end of the year. With the introduction of Deng Xiaopings economic reforms, the commune system collapsed and with it the cooperative medical insurance, and so barefoot doctors lost their institutional and financial support. Barefoot doctors were renamed "village doctors"and they were required to pass a formal examination before they could practice, and allowed to charge for services and drug dispensation. More recently, in recognition of the problem of over-prescription and expenditure growth, the Government of Ningxia introduced in 2006 a zero drug profit policy for village clinics and township health centres. This had a substantial impact on these providers' income, especially village clinics, that used to earn over 95 percent of their revenue from drug sales.

While the public sector dominates healthcare provision in our study province and indeed China, patients have choice whether and at what level to seek care. In the province where this study was located, policymakers expressed particular concern over the escalating costs associated with increasing numbers of patients going directly to hospitals to seek care. There is no gatekeeper system preventing patients from bypassing primary care providers and going directly to higher level hospitals.

3 Data

We use primary data collected from several sources in five counties of Ningxia province. The first source is a household survey conducted in 2011. The sampling frame consisted of the universe of all towns, villages and households in the study area. In each county we sampled every town and then stratified villages according to their economic situation (i.e. rich, middle, poor). We selected 40 percent of villages in each stratum using a random number generator. Within each sampled village, we randomly selected between 20 and 33 households. In total, the universe of the sample was 75 towns, comprising 917 villages and 357,400 households. A total of 6,612 households (28,890 individuals) were interviewed in 260 villages.

We exploit data from several modules in the household survey on: i) basic household and individual characteristics; ii) illness, injury and outpatient visits; and iii) perceptions of quality of care. Table 1 provides descriptive statistics on the household data. Educational attainment is relatively low with 27 percent of individuals never attending school. Average consumption in the sample is 5,407 RMB per capita, substantially below the national average. Healthcare utilization is measured as a binary variable indicating whether the individual sought care in the past two weeks if ill. More than 10 percent of individuals were ill in the past two weeks. Of these, 15 percent sought no treatment and 34 percent self-treated. Individuals are most likely to seek care from village doctors (18 percent), followed by county hospitals (13 percent) and township health centres (13 percent). The private sector caters to only 5 percent of sick individuals.

The head of each household was asked to evaluate the quality of care provided by the village doctor serving their community, irrespective of whether care had been sought. Perceptions of quality were assessed on score of 1 (poor) to 5 (excellent) across a number of dimensions including waiting times, facility environment, politeness of doctor, description of illness by doctor, drug availability, sophistication of the equipment, and ability of the doctor. To generate an aggregate score we convert the score for each dimension to a z-score with mean zero and standard deviation one and take the unweighted average across all z-scores. Table 1 presents the data, showing that village doctors score worst on sophistication of the equipment and best on waiting times.

We link data on individuals interviewed in the household survey to data on their village

doctor collected at the same time. This matching is performed at the village level. Of the 260 clusters sampled in the household survey, village doctors in 247 villages were interviewed. Two tools were used. First, a questionnaire was administered to collect information on the characteristics of the doctor, the services provided and sources of revenue. Second, respondents were asked to complete a 35 item knowledge test that sought to assess the clinical competence of the village doctor. Designed specifically for the study, questions were developed following the national clinic guidelines on primary health care by two qualified Chinese doctors. Each question concerned patient cases that village doctors are expected to be able to manage. The test involved multiple choice questions and a short case study involving a number of questions based on the description of the case. The questionnaires are all available at www.publichealth.ox.ac.uk/ningxiaproject.

Table 2 provides summary statistics on the sample of village doctors. The majority of doctors are male with an average age of 44 years, having worked as a doctor for 22 years. Two-thirds of the village doctors were barefoot doctors prior to the reforms of village healthcare. Over one-third have less than middle school education. Two-thirds have specialised in Chinese medicine, Western medicine or both.

4 Methods.

4.1 Measuring competence.

The responses of doctors to the 35-question, multiple choice questionnaire were used to produce a synthetic index of competence. We regard competence or technical knowledge of the doctor as a pre-requisite for implementing best practice. In the absence of a fixed standard (such as a national benchmark or a history of testing) we cannot assess the difficulty of the test or its ability to discern competence levels. For that reason, our index was computed using Item Response Theory (IRT), a technique widely used in the fields of psychometry and education (Lord, 1962, Lord, 1969). Under certain parametric assumptions, IRT produces a competence index that takes into consideration the difficulty and discrimination ability of each question in the test. IRT assumes the existence of a latent competence parameter for each individual, θ_i, which is imperfectly revealed by respondents' answers to the questions in a particular test.

This competence parameter determines the likelihood of answering each item correctly, which is summarised in the Item Characteristic Curve,

$$\mathbb{P}(X_{ig} = 1|\theta_i) = \mathbb{P}(X_{ig} = 1; a_g, b_g, c_g|\theta_i) = c_g + \frac{(1 - c_g)}{1 + \exp(-a_g(\theta_i - b_g))} \tag{4.1}$$

where X is an indicator taking value 1 if item g has been answered correctly. The scalar c_g is a *guessing parameter* incorporating the fact that on multiple choice exams, even the least capable individual $(\theta \to -\infty)$ can guess correctly. The scalar b_g measures the *difficulty* of an item. It is a location parameter marking the point of the latent variable where the probability of correctly answering the question is halfway between the floor given by c_g and the maximum. The lower b_g the more easily somebody with ability θ will answer correctly. Finally, a_g is the *discrimination parameter*, which is proportional to the slope of $Prob(\theta)$ at $\theta = b_j$ and measures the ability of the item to distinguish between values of the latent variable that are close ot each other in the neighbourhood of its difficulty. Under the above parameterisation, the likelihood function for the sample equals

$$\ell(a, b, c|\theta_i) = \prod_{i=1}^{N} \prod_{g=1}^{G} \mathbb{P}(X_{ig} = 1|\theta_i)^{x_{ig}} (1 - \mathbb{P}(X_{ig} = 1|\theta_i))^{1-x_{ig}} \tag{4.2}$$

Unlike standardised scores, the synthetic score provided by IRT takes into account the traits of the test. In particular, note that as the discrimination parameter tends to 0, $a_i \to 0$, the distribution in (4.1) converges to a constant, the gradient of the likelihood function with respect to item i converges to 0. Estimation of the models is discussed in Das and Hammer (2004). The literature has developed three main approaches. Conditional maximum likelihood, Bayesian estimation and multiple imputation. The latter method, also referred to as *Plausible Values* by Das and Hammer (2004), was introduced in Mislevy et al. (1992), and treats θ_i as a missing observations. If θ_i is missing at random, imputation can produce unbiased estimates of θ_i, a_i, b_i and c_i as well as good estimates of the variance of the competence scores, while both Bayesian and Likelihood criteria produce biased estimates of the variance and, therefore, cannot be informative about the distribution of knowledge. In particular, Bayesian methods provide a lower bound for the variance of θ_i while maximum likelihood produces an upper bound for the variance of θ_i. Therefore, our synthetic index of competence is constructed using the Plausible

Values approach[1].

For robustness, we complete the analysis including a second index of competence, namely the average of standardized scores. This indicator weights each question (or item) equally regardless of its difficulty or ability to discriminate competence.

The results of the multiple choice test are given in figures 1 to A2 and table A1 in the Appendix. To assess the overall reliability of the test, we computed Cronbach's Alpha, which took a value of 0.74, indicating a good (albeit low stakes) test (see Cronbach, 1951). The number of questions answered correctly per individual was generally low. Figure 1 shows that the best score was 29 correct answers, while the median of the distribution was located at 16. Table A1 lists the 35 questions asked as well as the proportion of individuals who correctly responded each of the questions. This table reveals that no question was answered correctly by all doctors, although several questions were poorly addressed. This can be due to a general lack of competence or a too difficult test. To cast some light on this matter, we plotted figure 2 and figures A1 and A2 in the Appendix. Figure 2 explores the relationship between the estimated difficulty and discrimination parameter from the three parameter logistic model. We plot the predicted three parameter logistic model for each of the times listed in table A1. The figure reveals that the estimated difficulty of the items (β in equation 4.1) is generally large and positive (typically exceeds 1) which suggests that, indeed, most questions were difficult for the population under study. Nonetheless, the figure reveals that there is a good spread of difficulty among items. This is important as it determines the ability of the test to discriminate competence among doctors. Although difficult questions are, on average, better at discriminating levels of competence, the U-shape of the local polynomial regression reveals that some of the easiest questions have very high discrimination power. This is the case of items 14 (suffocation of injured individuals), 12 (treatment for infantile diarrhoea), 1 (treatment of senile hypertension) and 19 (the WHO's recommendation of a persons daily intake of salt). The

---

[1]The additional information provided by IRT comes at the expense of three additional assumptions. Firstly, competence is assumed to be a unidimensional object and, therefore, any test is assumed to capture this parameter only and not other factors such as, for example, motivation. Secondly, individuals' traits are uninformative about the likelihood of answering an item correctly, once competence has been established. In particular, environmental factors (such as access to medical technology) are assumed innocuous. Finally, for a fixed level of competence, the likelihood of answering any item correctly is independent of the likelihood of answering any other item correctly. Despite the restrictive nature of the above assumptions, all doctors in our sample face very similar working conditions and reimbursement rates, so that the work in relatively homogeneous professional environments and our questionnaires did not have any tree structure so that answers to each item can safely be assumed to be independent, given ability. Therefore, the assumptions underpinning IRT do not seem conflicting with our scenario

figure also identifies which questions have above average discrimination power, given difficulty. Among these we find items 28 (effect of bezoar stones), 24 (sciatica treatment for lumbar disc injury) 21 (principles of treatment of high blood pressure) 5 (characterisation of bile duct stone). Figures A1 and A2 in the Appendix plot estimated Item Characteristic Curves and the empirical proportions of correct answers given estimated competence. The figures reveal that the questions with above average discrimination power are those with a rich distribution of correct answers, given competence, where the logistic model can provide a good fit. These are the questions that will receive more weight when constructing the synthetic competence index. On the contrary, too easy items (such as items 2 and 3) or too difficult items (such as items 8, 28 or 29) are treated by the logistic model as uninformative. Overall, the individuals in the sample found the test difficult but the test succeeded in producing enough variation in results as to be informative and discriminate different level of competence among doctors.

4.2 Estimation strategy

There are three stages to the analysis. First, we estimate the effect of competence, denoted θ , on demand for village clinic services (Q). Then we explore if individuals can infer the quality of the local doctor by estimating the effect of competence on household perceptions of quality (S). Finally, we examine whether perceptions and demand are correlated.

Our primary interest is in the relationship between utilization of healthcare at the village level and doctor competence. To estimate the sign and strength of the relationship between demand and competence we estimated a probit model with conditional mean

$$E(Q_{i,j,k}|\mathbf{x}_{i,j,k}) = \Phi(\alpha\theta_k + \beta'\mathbf{x}_{i,j,k})$$

where $\mathbf{x}$ is a vector of individual's characteristics (including disease), doctor's characteristics and county dummies. The model was estimated via maximum likelihood with standard errors clustered at village level. We consider a model in which use of healthcare is conditional on being ill. A similar specification was used to estimate the correlation between demand and the head of households' valuations of the local clinic. In particular,

$$E(Q_{i',j,k}|\mathbf{x}_{i',j,k}) = \Phi(\alpha S_{j,k} + \beta'\mathbf{x}_{i',j,k})$$

where i' now refers to the head of the household only.

Finally, the relationship between clinic valuation and doctor's competence was estimated using a standard least squares model with clustered standard errors (at village level) and conditional mean,

$$E(S_{i',j,k}|\mathbf{x}_{i',j,k}) = \alpha\theta_k + \beta'\mathbf{x}_{i',j,k}$$

5 Results

5.1 Demand and doctor competence

We first examine the relationship between demand for healthcare and the competence of village doctors[2]. The main results are shown in Table 3 as marginal effect estimates from a probit model with standard errors clustered by village. The dependent variable is any visit to the village doctor conditional on being ill in the past two weeks.

Column 1 of Table 3 shows no evidence of a correlation, positive or negative, between utilization of village doctors and doctor competence. Columns 2 and 3, respectively, introduce controls for individual characteristics and categories of disease. In both instances, the lack of any correlation remains and the marginal effects move closer to zero. In column 4, we include controls for characteristics of the doctor and see that the result remains robust. Finally, we include county fixed effects in column 5 that control for unobserved heterogeneity at that level. The marginal effect on doctor competence remains small with no evidence of statistical significance.

We have sufficiently tight confidence intervals to rule out reasonably small effect sizes. The results from column 5 suggests that a one standard deviation increase in doctor competence has an insignificant effect of 1.7 percentage points on the probability of visiting a village doctor, relative to a mean of 18 percent. The confidence intervals imply that we can rule out a positive effect size of 4.7 percentage points. Taken together, the findings suggest that healthcare seeking behaviour is unresponsive to changes in doctor competence at the village level. We scrutinise

---

[2]The determinants of doctors competence are explored in Table A4 in the Appendix. The only statistically significant variable is the indicator for specialisation in nursing, which appears with a negative coefficient, suggesting that those individuals trained as nurses obtained lower competence scores on average. In general, the low adjusted R^2 suggests that much of the variation in doctors' competence was left largely unexplained by the data. We arrived to the same conclusion when using the standardised competence score.

our core result with further robustness tests in Section 5.3.

5.2 Perceptions of quality

If health care seeking is unresponsive to more competent doctors, this may be because house-holds find it difficult to assess the ability of doctors. We therefore study the relationship between household perceptions of quality of care and doctor competence. Table 4 reports OLS estimates in which the dependent variable is our overall measure of perceptions of quality.

Column 1 simply regresses perceptions of quality on our measure of doctor competence without any controls. The coefficient is small and not statistically significant at conventional levels. When we include controls for individual characteristics and disease groups, in columns 2 and 3 respectively, the coefficient approaches zero. The inclusion of doctors' characteristics in column 4 and county fixed effects in column 5 makes little difference to the coefficient and statistical significance. The estimate in column 5 indicates that a one standard deviation increase in doctor competence has an insignificant effect of 0.02 standard deviations on perceptions of quality[3].

Table 5 considers whether perceptions are an important determinant of health care seeking. We see in column 1 that the coefficient on perceptions of quality is strongly positive and significant, indicating that perceptions of quality matter for health care seeking. A one standard deviation increase in perceptions of quality is associated with 6 percentage point increase in use of village doctors. The inclusion of controls for individual characteristics (column 2), disease groups (column 3), and doctor characteristics (column 4) does not change the result substantively, if anything increasing the magnitude of the coefficient. Finally, in column 5, the result is robust to the inclusion of county fixed effects.

Together these results paint a consistent picture. Our measure of doctor competence is not correlated with utilization of village doctors and this can be reconciled with the fact that households appear unable to recognise more competent doctors. At the same time perceptions of quality are an important determinant of healthcare seeking behaviour.

---

[3]We also ran separate regressions for each of the seven dimensions of perceptions of quality that make up our overall score. In no instance is the coefficient on doctor competence significant. In particular, we find no relationship between household perceptions of the doctor ability and the doctor competence score.

5.3 Robustness

In this section we perform further robustness checks on our main findings to address a number of concerns. The results of these tests are shown in Table A2 in the Appendix . First, motivated by the findings of a study in the Philippines (Peabody et al., 2011), we examine whether there are non-linearities in the relationship between healthcare utilization and doctor competence by splitting competence into terciles. Columns 1 to 3 suggests that there is no evidence of a non-linear relationship. Second, we define the dependent variable as utilization across the entire sample, irrespective of whether the individual was ill in the past two weeks. As shown in column 2, the marginal effect is close to zero and far from statistical significance. Third, we examine the robustness of the results to simpler weighting schemes to construct the measure of doctor competence. In column 3 we use the raw score and in column 4 the standardised score. Both the marginal effects of interest are close to zero. Fourth, in column 5, we include township instead of county fixed effects and find no qualitative change in the result.

Another concern is that doctors may sort into villages based on the characteristics of the population in that locality. If these same factors are correlated with healthcare utilization, OLS estimates of the relationship between demand for healthcare and doctor competence will be biased. Given that village doctors have been in their position for on average 22 years, there is clearly little turnover and therefore limited scope for sorting. We also show in Table A3 of the Appendix that doctor competence is not associated with the characteristics of households in the catchment area of the village doctor.

The focus of this article is the correlation between doctors competence and demand. Of secondary importance was the relationship between perceptions of quality and demand. A concern with the latter is that demand and peoples perceptions of quality might be simultaneously determined, in which case the positive association may be biased. In the absence of valid instrumental variables in our data, we explored this question from a more structural point of view[4]. To this end, we used follow-up household data in 2012 which included similar indicators of demand and perceptions of quality (but no data on village doctors). This enabled us to fit two

---

[4]In our analysis we considered using the village clinic average valuation as an instrument for an individual's own valuation of quality. For each individual in the sample, this instrument was calculated as the village clinic average valuation excluding the respondent's own valuation. Under random sampling, this instrument is statistically independent of a respondent's demand. Unfortunately, the instrument was statistically insignificant in the first stage equation ($p = 0.2023$) of the IV-probit procedure, which is suggestive of a very weak instrument.

parametric models (results available on request) . We firstly estimated a random effects panel probit using 2011 and 2012 data. This allowed to explore the question of whether valuation of health services is still correlated with demand once time invariant unobserved heterogeneity is taken into account. Secondly, we estimated the effect of perceptions of quality in 2011 on demand in 2012 using a probit model. Under the assumption that health shocks are exogenous and unforeseen, the coefficient of perceptions of quality in this model would be informative about the causal effect of valuation on demand. The estimated marginal effect of valuation on health in all models was around 0.02 and significant (p-value =0.025), thus supporting our finding that demand and perceptions of quality are positively correlated.

6   Discussion

We studied the relationship between doctor competence and the utilization of health care at the primary care level using household data on care seeking behaviour linked to an assessment of health providers' clinical competence in rural areas of Ningxia province.

Our main finding is that we find no evidence of a correlation between doctor competence and utilization of healthcare, either positive or negative, with confidence intervals reasonably tight around zero. In other words, patients in this particular setting do not appear to seek more care from doctors with higher competence. While we lack exogenous sources of variation in doctor competence to make causal claims, we show that this result is robust to a large number of additional controls. In further support of the main result, we find that household perceptions of quality are an important determinant of care seeking behaviour, yet patients appear unable to recognize more competent doctors (there is no relationship between doctor competence and perceptions of quality).

A plausible explanation for this finding is that patients appeared unable to recognize more competent doctors. There was no relationship between doctor competence and perceptions of quality. Thus, while patient perceptions played an important role in explaining the demand of health provider, these perceptions did not reflect the competence of village doctors. Most likely they reflect more observable aspects of care that are less important to health outcomes. Note, in particular that, because 98% of our sample are covered by the rural health insurance program (the New Cooperative Medical Scheme, NCMS) and the benefit package is identical

within each county, variation in insurance coverage cannot explain variations in demand as our analyses have controlled for county fixed effects..

Several studies have previously looked at the relationship between the technical quality of care and patient satisfaction, with mixed findings. Leonard (2008) provides convincing evidence of a positive relationship in Tanzania. Results in Hanson et al. (2004), which are consistent with our findings, suggest that patients demand of provider is sensitive to quality in terms of household perceptions of quality, and that interpersonal quality is more important than either technical quality or system-related factors in Cyprus. Evidence from the Philippines also suggests that quality is a predictor of patient satisfaction, but the relationship appears non-linear (Peabody et al., 2011). At low levels of the quality score, improvements in quality exert a negative influence on patient satisfaction that becomes positive at higher levels of quality. By contrast, other studies find no evidence of a link in a variety of different settings (Chang et al., 2006; Das and Sohnesen, 2006; Erikson, 1986).

Our study goes beyond patient satisfaction to study whether individuals respond to more competent doctors in their health care seeking behaviour. The fact they do not suggests that there are information asymmetries for health providers to exploit if the incentives are there to do so. Several recent studies in China highlight well the agency problem inherent in the relationship between doctor and patient. Lu (2014) experimentally manipulated the incentives faced by doctors to show that when doctors could retain drug profits, they prescribed more drugs, more expensive drugs and more unnecessary drugs. Meanwhile, Currie et al. (2011) varied the knowledge of simulated patients showing that doctors prescribed fewer antibiotics when faced with more knowledgeable patients. Pay-for-performance offers a potential solution to the agency problem although such schemes can be complex to design and implement. Evidence from a study of the introduction of pay-for-peformance in Ningxia province suggests that incentives can be improved to reduce the excessively high antibiotic prescription rates (Yip et al., 2014).

The findings of the study imply that improvements in the technical quality of village doctors alone would be ineffective in shifting the pattern of healthcare seeking away from higher tiers of care towards the primary care level, other things being equal. This means that, firstly, patients are unlikely to receive treatment in the most adequate facilities and, secondly, market and competitive forces devised to incentivise providers towards better technical quality may

not have much effect on patients' demand. Of course, the government could take an active role by improving and publicising information on the quality of providers but this would require a functioning management information system. Even then, it would only be effective if the patients have adequate knowledge to differentiate good and bad clinical quality of care. How then could the current situation of inefficient use of higher level facilities for case that could be dealt with at the primary care level be improved? Gatekeeping is an obvious candidate and there are already pilots of such a policy in China. However, the concern is that if quality of care provided by villages doctors do not satisfy patients expectation, such a system will not succeed or sustain.

In thinking about the implications of the findings we note a number of limitations. First, we are cautious in claiming the results are free of omitted variable bias. It remains possible that the finding of zero correlation between utilization and doctor competence is driven by unobserved factors although there is no evidence that this the case from our robustness checks[5]. Second, our data and analysis are cross-sectional in nature. However, patients' assessment of the village doctors' technical quality can change overtime with repeated encounters and thus affect their utilization behaviour, which our cross-sectional data cannot capture. The empirical literature on this is quite scarce because longitudinal data with the same individuals having repeated visits over the observed period is required (Corno, 2014). Third, we measured competence by clinical knowledge not the actual quality of care delivered to patients. Our findings may be spurious if the actual quality of care provided by doctors is unrelated to this measure of competence. This could be because more competent doctors fail to exert more effort or lack the resources to translate better knowledge into clinically effective care. While it would have been preferable to have process measures of quality from, say, standardised patients (Sylvia et al., 2014; Das et al., 2012), it is likely that knowledge is a strong predictor of practice. Fourth, the external validity of the results can always be challenged. Ningxia is the third least developed region in China and it is possible people living elsewhere where education levels are higher are more able to evaluate the quality of care they receive.

The evidence we have presented is suggestive, and is sufficiently interesting to spur future

---

[5]The ability of our results to unveil a causal effect depends on whether utilisation levels would remain the same regardless of doctors' competence. This would not be the case if, for example, doctors' competence positively affect public health in an area or if the characteristics of the area attract a more or less skilful type of doctor. Thus the overall selection bias can be negative or positive, and we have no way of determining its sign.

research with stronger methods. For example, one of the authors of this paper is currently conducting a Randomized Control Trial of a quality improvement programme in which it will be possible to see whether perceptions of quality and demand are influenced by exogenous changes in clinical quality. Given the current lack of evidence on the question at the heart of the paper, we believe these findings to be an important contribution to the debate on whether patients and the public at large are able to evaluate quality beyond the more observable aspects of care.

References

Arrow, K. (1963). Uncertainty and the welfare economics of medical care. *American Economic Review 53*, 941–973.

Arrow, K. (1986). Rationality of self and others in an economic system. *Journal of Business 59*, S385S399.

Berry, S., J. Levinsohn, and A. Pakes (1995). Automobil prices in market equilibrium. *Econometrica 63*, 841–889.

Bloom, N., C. Propper, S. Seiler, and J. Van Reenen (2010). The impact of competition on management quality: evidence from public hospitals. Technical report, CEP Working Paper, London School of Economics.

Chalkley, M. and Khalil (2005). Third party purchasing of health services: Patient choice and agency. *Journal of Health Economics 24*, 11321153.

Chalkley, M. and J. Malcomson (1998). Contracting for health services when patient demand does not reflect quality. *Journal of Health Economics 17*, 1–19.

Chalkley, M. and C. Tilley (2005). The existence and nature of physician agency: evidence of stinting from the british national health service. *Journal of Economics and Management Strategy 14*, 64766.

Chang, T., R. Hays, P. Shekelle, C. MacLean, D. Solomon, D. Reuben, D. Roth, C. Kamberg, J. Adams, R. Young, and N. Wenger (2006). Patients global ratings of their health care are not associated with the technical quality of their care. *Annals of Internal Medicine 144*, 665672.

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

Cooper, Z., S. Gibbons, S. Jones, and A. McGuire (2011). Does hospital competition save lives? evidence from the english nhs patient choice reforms. *Economic Journal 121*, F288–F260.

Corno, L. (2014). Learning (or not) in health-seeking behaviour: Evidence from rural Tanzania. *Economic Development and Cultural Change 63*, 27–72.

Cronbach, L. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika 16*, 297–334.

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

Das, J. and J. Hammer (2004). Which doctor? Combining vignettes and Item Response Theory to measure doctor quality. Technical report, World Bank Working Paper 3301.

Das, J., A. Holla, V. Das, M. Mohanan, D. Tabak, and C. B. (2012). In urban and rural india, a standardized patient study showed low levels of provider training and huge quality gaps. *Health Affairs 31*, 2774–2784.

Das, J. and T. Sohnesen (2006). Patient satisfaction, doctor effort, and interview location: Evidence from Paraguay. Technical report, Technical report, World Bank, Policy Research Working Paper WPS 4086.

Donabedian, A. (1988). The quality of care: how can it be assessed? *Journal of the American Medical Association 260*(12), 1743–1748.

Donabedian, A. (2005). Evaluating the quality of medical care. 1966. *Milbank Quarterly 83*(4), 691–729.

Dranove, D. and W. White (1987). Agency and the organization of health care delivery. *Inquiry 24*, 405–415.

Einav, L. and J. Levin (2010). Empirical industrial organization: A progress report. *Journal of Economic Perspectives 24*, 145–162.

Erikson, L. (1986). Patient satisfaction: an indicator of nursing care quality. *Nursing Management 18*, 3135.

Gaynor, M. (1994). Issues in the industrial organization of the market for physician services. *Journal of Economics and Management Strategy 3*, 211255.

Glied, S. and D. Remler (2002). What every public finance economist needs to know about health economics: Recent advances and unresolved questions. *National Tax Journal 55*, 771788.

Hanson, K., W. Yip, and W. C. Hsiao (2004). The impact of quality of the demand for outpatient services in Cyprus. *Health Economics 13*, 1167–1180.

Kessler, D. and M. McClellan (2000). Is hospital competition socially wasteful? *Quarterly Journal of Economics 115*, 577615.

Leonard, K. (2008). Is patient satisfaction sensitive to changes in quality of care? an exploitation of the hawthorne effect. *Journal of Health Economics 27*, 444–459.

Lord, F. (1962). Estimating norms by item-sampling. *Educational Psychological Measurement 22*, 259–267.

Lord, F. (1969). Estimating true score distributions in psychological testing (and empirical Bayes problem). *Psychometrika 34*, 259–299.

Lu, F. (2014). Insurance coverage and agency problems in doctor prescriptions: Evidence from a field experiment in china. *Journal of Development Economics 106*, 156–167.

Ma, A. and T. T. McGuire (1997). Optimal health insurance and provider payment. *American Economic Review 87*, 685704.

McConnell, J. (1993). Barefoot no more. *The Lancet 341*, 1275.

Mislevy, R., A. Beaton, B. Kaplan, and K. Sheehan (1992). Estimating population characteristics from sparse matrix samples of item responses. *Journal of Educational Measurement 29*, 133–161.

Mooney, G. and M. Ryan (1993). Agency in health care: Getting beyond first principles. *Journal of Health Economics 12*, 125135.

Peabody, J., R. Shimkhada, S. Quimbo, J. Florentino, M. Becate, C. McCulloch, and O. Solon (2011). Financial incentives and measurement improved physicians quality of care in the phillippines. *Health Affairs 30*, 773781.

Perloff, J. and S. Salop (1985). Equilibrium wiht product differentiation. *Review of Economic Studies 52*, 107–120.

Pollock, A., A. MacFarlane, G. Kirkwood, F. Majeed, I. Greener, C. Morelli, S. Boyle, H. Mellett, S. Godden, D. Price, and P. Brhlikova (2011). No evidence that patient choice in the nhs saves lives. *Lancet 378*, 20572060.

Propper, C., S. Burgess, and K. Green (2004). Does competition between hospitals improve the quality of care? hospital death rates and thenhs internal market. *Journal of Public Economics 88*, 12471272.

Sylvia, S., Y. Shi, H. Xue, X. Tian, H. Wang, Q. Liu, A. Medina, and S. Rozelle (2014). Survey using incognito standardized patients shows poor quality care in china's rural clinics. *Health Policy and Planning*.

White, S. (1988). From barefoot doctor to village doctor in tiger springs village: A case study of rural health care transformations in socialist china. *Human Organization 57*, 19.

Yip, W. and W. Hsiao (2009). China's health care reform: a tentative assessment. *China Economic Review*.

Yip, W., W. Hsiao, and W. Chen (2012). Early appraisal of China's huge and complex health-care reforms. *The Lancet*.

Yip, W., T. Powell-Jackson, W. Chen, M. Hu, E. Fe, M. Hu, W. Jian, M. Lu, W. Han, and W. Hsiao (2014). Capitation combined with pay-for-performance iimprove antibiotic prescribing practices in rural China. *Health Affairs*.

Zhang, D. and P. Unschuld (2008). Chinas barefoot doctor: Past, present and future. *The Lancet 372*, 18651867.

Table 1: Household Survey: Summary statistics

Variable	Mean	(Std. Dev.)	N
Male	0.519	(0.5)	28886
Age	31.543	(19.957)	28886
Never attended school	0.269	(0.443)	22219
Elementary school	0.328	(0.469)	22219
Middle school	0.292	(0.455)	22219
High school or more	0.112	(0.315)	22219
Hui	0.52	(0.5)	28886
Log-consumption	9.860	(0.671)	28882
Health care utilisation (Individual).			
Ill in the past 14 days	0.135	(0.342)	28886
No treatment	0.151	(0.358)	3898
Self-treatment	0.336	(0.472)	3898
Attended village clinic	0.181	(0.385)	3898
Attended private clinic	0.051	(0.221)	3898
Attended township health centre	0.125	(0.330)	3898
Attended county hospital	0.131	(0.338)	3898
Perceptions of Quality (Head of Household).			
Waiting time	4.188	(0.875)	6428
Environment	3.694	(0.906)	6427
Physican politeness	3.866	(0.908)	6428
Physican description of illness	3.67	(0.965)	6427
Drug availability	2.93	(1.068)	6417
Equipment sophistication	2.705	(1.11)	6414
Physician ability	3.42	(0.956)	6484
Standardised Score	0.007	(0.980)	6404

Table 2: Summary statistics: Village Doctors

Variable	Mean	Std. Dev.	N
Male	0.813	0.391	240
Age	44.227	12.3	238
Barefoot doctor	0.663	0.474	240
Education: < Middle School	0.389	0.489	239
Education: High School	0.126	0.332	239
Education: Vocational School	0.448	0.498	239
Education: > College	0.038	0.191	239
Occupation: Village doctor	0.951	0.217	239
Occupation: Midwife	0.03	0.167	239
Specialisation: Chinese medicine	0.057	0.233	239
Specialisation: Western medicine	0.254	0.437	239
Specialisation: Western and Chinese medicine	0.352	0.479	239
Specialisation: Nursing	0.012	0.110	239
Specialisation: Public Health	0.169	0.375	239
Experience	21.554	14.317	240

Table 3: Conditional Demand (Marginal Effects)

	Basic	Individual & Household Characteristics	Disease Groups	Doctor's Characteristics	Regional Indicators
Competence	0.0164	0.0120	0.0146	0.0178	0.0172
	0.0169	0.0162	0.0160	0.0157	0.0152
Individual Characteristics.					
Male		0.0114	-0.0001	-0.0007	0.0011
		0.0100	0.0099	0.0100	0.0101
Age		-0.0039***	-0.0011	-0.0014	-0.0013
		0.0012	0.0011	0.0011	0.0011
Age^2		0.0000*	0.0000	0.0000	0.0000
		0.0000	0.0000	0.0000	0.0000
Hui		-0.0513*	-0.0450	-0.0396	-0.0565*
		0.0257	0.0248	0.0244	0.0282
Head of household/household characteristics.					
Elementary School		0.0096	0.0124	0.0057	0.0075
		0.0215	0.0212	0.0215	0.0213
Middle School		-0.0107	-0.0095	-0.0134	-0.0101
		0.0226	0.0224	0.0230	0.0231
High School		-0.0378	-0.0406	-0.0483	-0.0466
		0.0308	0.0294	0.0273	0.0274
Migrant		0.0044	0.0070	0.0088	0.0056
		0.0171	0.0168	0.0170	0.0172
Log-consumption		-0.0086	-0.0169	-0.0201	-0.0203
		0.0131	0.0131	0.0131	0.0132
Doctors' characteristics.					
Experience				0.0019*	0.0019*
				0.0009	0.0009
Barefoot doctor				0.0238	0.0259
				0.0276	0.0276
High School				0.1031*	0.1115*
				0.0462	0.0455
Vocational School				0.0486	0.0538
				0.0275	0.0285
College				-0.0250	-0.0174
				0.0457	0.0434
Disease Dummy	N	N	Y	Y	Y
County Fixed Effects	N	N	N	N	Y
N	3811	3811	3774	3720	3720

Table 4: Perception and competence.

	Basic	Individual & Household Characteristics	Disease Groups	Doctor's Characteristics	Regional Indicators
Competence	0.0372	0.0171	0.0162	0.0138	0.0167
	0.0419	0.0416	0.0417	0.0407	0.0387
Individual traits.					
Male		0.0754	0.0893	0.0665	0.0617
		0.0944	0.0936	0.1003	0.0993
Age		-0.0090	-0.0099	-0.0093	-0.0146
		0.0080	0.0080	0.0080	0.0078
Age^2		0.0001	0.0001	0.0001	0.0001
		0.0001	0.0001	0.0001	0.0001
Hui		-0.1676**	-0.1696**	-0.1496*	-0.1636*
		0.0638	0.0638	0.0647	0.0685
Head of household/household characteristics.					
Elementary School		0.0340	0.0335	0.0345	0.0601
		0.0396	0.0396	0.0392	0.0390
Middle School		-0.0374	-0.0371	-0.0394	-0.0015
		0.0502	0.0500	0.0505	0.0520
High School		0.1231	0.1217	0.1220	0.0986
		0.0703	0.0705	0.0693	0.0686
Migrant		-0.0533	-0.0518	-0.0488	-0.0561
		0.0376	0.0375	0.0371	0.0359
Log-consumption		-0.0205	-0.0170	-0.0320	0.0245
		0.0273	0.0272	0.0274	0.0241
Doctor's characteristics.					
Experience				0.0008	0.0025
				0.0029	0.0028
Barefoot				0.1072	0.1127
				0.0913	0.0843
High School				0.0865	0.1043
				0.1006	0.0930
Vocational School				0.1476	0.1672*
				0.0805	0.0836
College				-0.0975	-0.0642
				0.2681	0.2524
Disease dummy	N	N	Y	Y	Y
County Fixed Effects	N	N	N	N	Y
N	6353	6351	6351	6237	6237

Table 5: Conditional Demand and Perception.

	Basic	Individual Characteristics	Disease Groups	Doctor's Characteristics	Regional Indicators
Valuation of clinic	0.0674***	0.0655***	0.0648***	0.0660***	0.0657***
	0.0111	0.0105	0.0098	0.0098	0.0095
Individual traits.					
Male		0.0113	0.0000	-0.0012	-0.0001
		0.0100	0.0098	0.0098	0.0099
Age		-0.0039***	-0.0010	-0.0012	-0.0012
		0.0012	0.0012	0.0012	0.0011
Age^2		0.0000*	0.0000	0.0000	0.0000
		0.0000	0.0000	0.0000	0.0000
Hui		-0.0457	-0.0402	-0.0360	-0.0498
		0.0239	0.0228	0.0226	0.0264
Head of household/household characteristics.					
Elementary School		0.0054	0.0087	-0.0012	-0.0004
		0.0203	0.0199	0.0204	0.0203
Middle School		-0.0061	-0.0046	-0.0094	-0.0066
		0.0218	0.0214	0.0221	0.0221
High School		-0.0460	-0.0477	-0.0563*	-0.0536*
		0.0291	0.0276	0.0258	0.0261
Migrant		0.0018	0.0039	0.0092	0.0070
		0.0174	0.0171	0.0172	0.0175
Log-consumption		-0.0103	-0.0188	-0.0178	-0.0203
		0.0126	0.0125	0.0124	0.0123
Doctor's characteristics.					
Experience				0.0019*	0.0018*
				0.0008	0.0008
Barefoot doctor				0.0193	0.0219
				0.0260	0.0255
High School				0.1019*	0.1090*
				0.0441	0.0438
Vocational School				0.0461	0.0504
				0.0250	0.0257
College				0.0042	0.0102
				0.0494	0.0473
Disease dummy	N	N	Y	Y	Y
County Fixed Effects	N	N	N	N	Y
N	3843	3843	3806	3685	3685
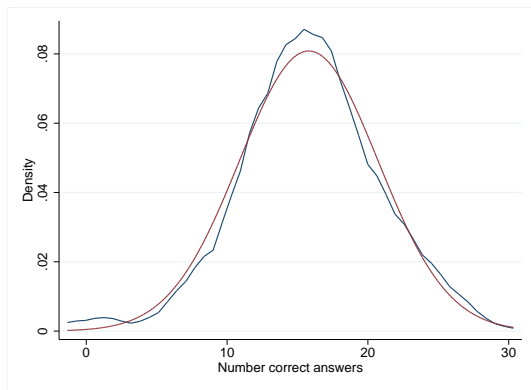
(a) # Right Answers
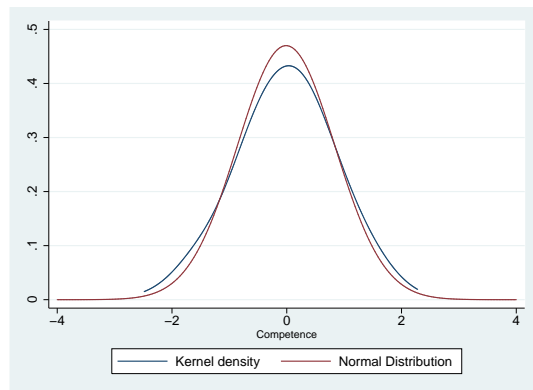
(b) Standardised Competence

Figure 1: Density estimators: Number of questions answered correctly (left panel) and the Estimate of standardised competence (right panel).
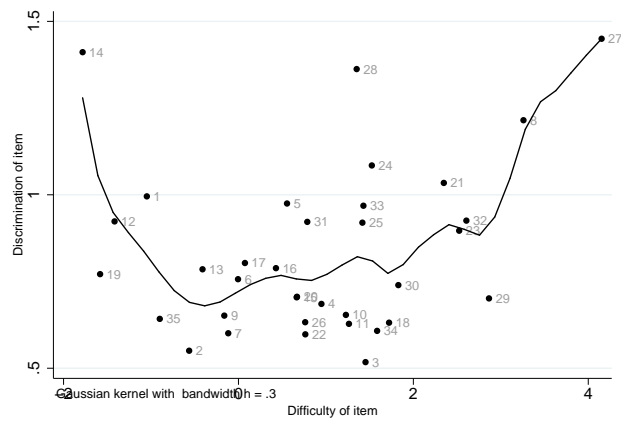
Figure 2: Local linear regression of the discrimination parameter to the difficulty parameter. Gaussian kernel and default bandwidth

A   Appendix: Additional Materials.

Table A1: Village doctor Multiple Choice Test (% Doctors who gave the right answer(s))

Question	
Q1: The key principle of treating senile hypertension is	0.812
Q2: The abnormal range of BP for adults is	0.922
Q3: The most important pathogen of acute upper respiratory infection is	0.690
Q4: Acute upper respiratory infection should be treated by	0.335
Q5: Common bile duct stone with acute cholangitis are characterized by	0.469
Q6: Which of the following statement is incorrect about urinary tract infection	0.58
Q7: The emergency treatment of ischemic cerebrovascular disease should focus on	0.649
Q8: When presenting with diarrhoea, carefully examine if the patient has	0.045
Q9: The most common clinical manifestations of chronic gastritis are	0.735
Q10: When treating chronic active gastritis, special attention should be given to the usage of	0.322
Q11: When diabetic patients present with strong sense of hunger, palpitations, hand tremors and sweating, this may be caused by	0.371
Q12: The correct treatment for infantile diarrhoea is	0.845
Q13: The most important principle to be considered when treating duodenal ulcers is	0.698
Q14: When suffocation occurs to an injured individual, immediate first aid measures should be	0.922
Q15: When organophosphorus pesticide poisoning occurs, the best drug to reduce or eliminate muscarinic symptoms is	0.694
Q16: Which of the following presentation signals the existence of herniation in children with seizures	0.514
Q17: Drug of first choice for anaphylactic shock due to penicillin allergy is	0.735
Q18: The core principle of diet therapy for diabetes is	0.212
Q19: The WHO recommends that a persons daily intake of salt should not exceed	0.886
Q20: The treatment plan to clear helicobacter pylori is	0.441
Q21: The principles of the treatment for high blood pressure are	0.078
Q22: Suspicious symptoms of TB include	0.412
Q23: Patients with TB-like symptoms should be treated in the following method	0.245
Q24: Which of the following statements are true about the sciatica treatment for lumbar disc herniation	0.127
Q25: When administering first-aid and transporting trauma victims, which of the following statements are correct	0.196
Q26: When performing a neonatal visit, the examinations you should perform on the baby include	0.331
Q27: Which of the following statements are incorrect	0.024
Q28: Which of the following is the effect of bezoar	0.147
Q29: Major risk factors for coronary heart disease include	0.065
Q30: Which of the following represents irregular or delayed motor development in children	0.257
Q31: Which of the following represents are the ways of transmission of Hepatitis B	0.306
Q32: Regarding acute attacks and persistent symptoms of chronic bronchitis, the correct treatment include :	0.073
Q33: Case study 1 [a]	0.437
Q34: Case study 2 [b]	0.388
Q35.A: Case study 3[c] what is your diagnosis?	0.837
Q35.B: Case study 3. What facility should she be referred to?	0.799
Q35.C: Case study 3. For what investigations and treatment?	0.298

[a] Male, 60 years old, after meal he suddenly had headache vomiting, and unconsciousness. Examination found his left pupil was enlarged and he had lost his light reflex in that eye (but the right eye was normal).The primary diagnosis should be?

[b] Male, 38 years old, left paranasal boil swollen for 5 days. Use of topical medication and antibiotics did not improve the situations. After 2 days, pus began to appear, pain has not relieved and the body temperature is 38 degrees. The risk of the focus to the patient is that

[c] A patient (female, 25 years old) came for consultation her chief complaint/medical history is: for the last month she has been coughing without any specific cause. She was coughing with small amount of white sputum. There was no bloody sputum. She felt feverish in the afternoon but self-tests body temperature did not exceed 38.0 degrees. She also had fatigue, night sweats, poor appetite and weight loss. Use of oral anti-inflammatory drugs was not obviously effective. Previously healthy with no history of drug allergy.Physical examination: body temperature 37.5 degrees, pulse 79bpm, respiration rate 22 /min, BP 120/70mmHg. She has chronic facies, wasting and no enlargement of superficial lymph nodes. Right upper lung breathing sounds coarse without blisters; heart rate 79 bpm with regular rhythm; soft abdomen, liver/spleen not felt when palpated. According to the patients description, medical history, and physical exam...

Table A2: Robustness Check: Demand Equation.

	Tertiles	Unconditional Demand	Raw Scores	Standardised Scores	Township Indicators
Competence		0.0006			0.0254
		0.0004			0.0144
Competence (raw scores)			0.0043		
			0.0028		
Competence (Standardised scores)				0.0207	
				0.0130	
2nd Tertile	-0.0046				
	0.0274				
3nd Tertile	0.0455				
	0.0307				
Individual traits.					
Male	0.0008	-0.0002	-0.0012	-0.0015	0.0008
	0.0101	0.0003	0.0100	0.0100	0.0101
Age	-0.0013	-0.0000	-0.0012	-0.0012	-0.0015
	0.0011	0.0000	0.0012	0.0012	0.0011
Age^2	0.0000	0.0000	0.0000	0.0000	0.0000
	0.0000	0.0000	0.0000	0.0000	0.0000
Hui	-0.0589*	-0.0010	-0.0611*	-0.0609*	-0.0282
	0.0278	.	0.0283	0.0282	0.0230
Head of household/household characteristics.					
Elementary School	0.0042	0.0003	0.0005	0.0003	0.0135
	0.0210	0.0006	0.0210	0.0211	0.0206
Middle School	-0.0112	-0.0003	-0.0122	-0.0124	-0.0087
	0.0229	0.0006	0.0232	0.0232	0.0224
High School	-0.0503	-0.0011	-0.0491	-0.0502	-0.0493
	0.0268	0.0006	0.0273	0.0272	0.0263
Migrant	0.0069	0.0003	0.0069	0.0064	0.0044
	0.0170	0.0005	0.0173	0.0173	0.0165
Log-consumption	-0.0196	-0.0007	-0.0206	-0.0205	-0.0201
	0.0133	0.0004	0.0135	0.0135	0.0123
Doctor's characteristics.					
Experience	0.0020*	0.0000	0.0020*	0.0021*	0.0022*
	0.0009	0.0000	0.0009	0.0009	0.0009
Barefoot doctor	0.0199	0.0006	0.0265	0.0272	0.0309
	0.0275	0.0007	0.0280	0.0280	0.0277
High School	0.1132*	0.0036*	0.1156*	0.1142*	0.1095*
	0.0457	0.0018	0.0460	0.0458	0.0459
Vocational School	0.0559*	0.0015	0.0531	0.0523	0.0626*
	0.0282	0.0008	0.0286	0.0286	0.0271
College	-0.0201	-0.0008	-0.0542	-0.0561	-0.0035
	0.0433	0.0010	0.0457	0.0454	0.0546
Disease Indicators	Y	Y	Y	Y	Y
County Dummy	Y	Y	Y	Y	N
Township Dummy	N	N	N	N	Y
N	3725	27605	3655	3645	3720

Table A3: Regression of Competence and household characteristics.

Male	0.0651
	0.4792
Age	-0.0140
	0.0694
Age^2	0.0001
	0.1350
Elementary School	0.0642
	0.0966
Middle School	0.0064
	0.8991
High School	0.1080
	0.1170
Migrant	-0.0566
	0.1112
Hui	-0.1700*
	0.0108
Log-consumption	0.0336
	0.1533
Parasitic disease	-0.0494
	0.9344
Malignant Neoplasm	0.6780**
	0.0079
Benign Neoplasm	0.7581
	0.2273
Nutritional/Endocrine/Metabolic	0.1534
	0.5374
Disease of blood/pancreas	0.5959***
	0.0000
Mental disease	-0.4399
	0.1375
Nervous system	0.0683
	0.7348
Eye	0.4683
	0.0737
Ear	0.8365***
	0.0000
Circulatory system	0.0564
	0.6346
Respiratory	0.0012
	0.9812
Digestive	0.1499
	0.0705
Genitourinary	-0.1497
	0.5416
Skin	-0.2431
	0.3036
Musculoskeletal	0.1309
	0.0977
Congenital	-0.1953*
	0.0104
Injury/poisoning/burn	-0.2693
	0.1253
Other	0.4595*
	0.0357
County Dummies	Y
N	6489

Table A4: Least squares estimates and adjusted R^2 of a regression of doctors' estimated competence. Education and experience both have a positive effect on doctors' competence, however this effect is not robust to specification and is, in general, not statistically significant. The only statistically significant variable is the indicator for specialisation in nursing, which appears with a negative coefficient, suggesting that those individuals trained as nurses obtained lower competence scores on average. In general, the low adjusted R^2 suggests that much of the variation in doctors' competence was left largely unexplained by the data. We arrived to the same conclusion when using the standardised competence score.

	Education	Experience	Specialty	Major	County
High School	0.0274	-0.0337	-0.0087	0.0069	-0.0207
	0.8837	0.8604	0.9637	0.9724	0.9204
Vocational School	0.2641	0.1772	0.1966	0.1543	0.0631
	0.0746	0.2375	0.1859	0.3339	0.6894
College	1.0961***	0.6950*	0.7279*	0.5136	0.5199
	0.0000	0.0157	0.0119	0.1016	0.1093
Experience: 5-9 Years		0.3104	0.3153	-0.0349	-0.0797
		0.4918	0.4690	0.9274	0.8258
Experience: 10-19 Years		0.7498	0.7444	0.4581	0.4475
		0.0902	0.0845	0.2155	0.1983
Experience: 20-29 Years		0.1986	0.2204	-0.1073	-0.0991
		0.6552	0.6002	0.7712	0.7774
Experience: 30-39 Years		0.2461	0.2644	-0.0618	-0.0646
		0.5831	0.5329	0.8687	0.8564
Experience: 40 +		0.2229	0.2349	-0.1322	-0.2207
		0.6299	0.5974	0.7456	0.5739
Specialty: Village Doctor			0.1631	0.1683	0.3379
			0.6607	0.6783	0.4152
Specialty: Midwife			0.4769	0.4891	0.7078
			0.4682	0.4508	0.2842
Major: Western Medicine				-0.0351	-0.1211
				0.8977	0.6873
Major: Chinese + Western Medicine				0.2836	0.1217
				0.2984	0.6845
Major: Nursing				-1.5820**	-1.5649**
				0.0049	0.0039
Major: Public Health				0.1847	0.1419
				0.5507	0.6606
Major: Other				0.0569	-0.0173
				0.8465	0.9576
Barefoot doctor					-0.0649
					0.6177
Male					0.1220
					0.4488
Yanchi					0.3601
					0.0605
Tongxin					-0.1905
					0.3266
Pengyang					-0.1440
					0.4859
Xiji					-0.1150
					0.5371
Intercept	-0.1655	-0.4592	-0.6511	-0.4313	-0.4654
	0.1594	0.3041	0.1649	0.4079	0.4118
N	239	239	239	239	239
R^2	0.0468	0.0975	0.1008	0.1491	0.1787
Adjusted R^2	0.0346	0.0661	0.0614	0.0919	0.0993

(a) Item 1   **(b)** Item 2   **(c)** Item 3

(d) Item 4   **(e)** Item 5   **(f)** Item 6

(g) Item 7   **(h)** Item 8   **(i)** Item 9

(j) Item 10   **(k)** Item 11   **(l)** Item 12

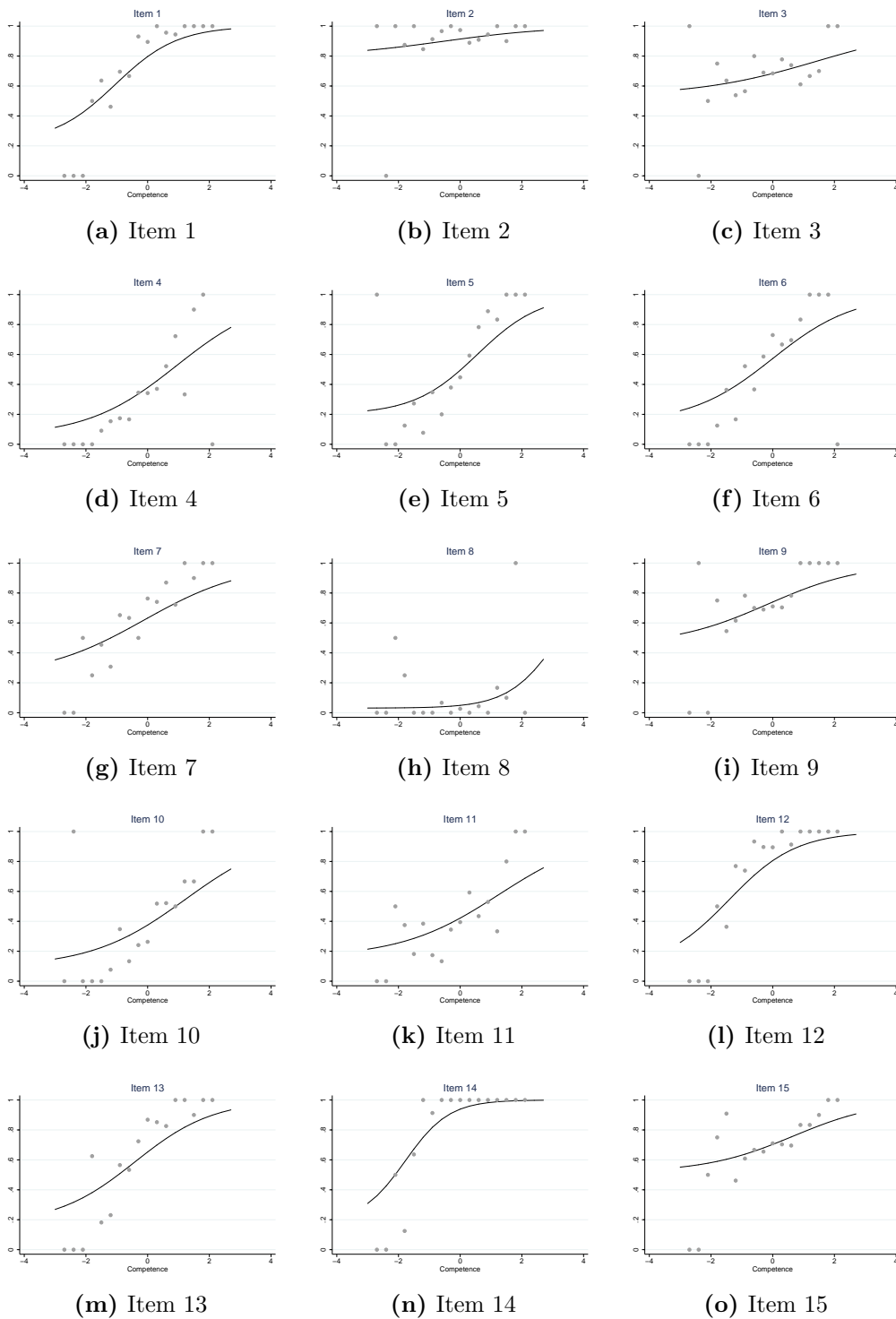**(m)** Item 13   **(n)** Item 14   **(o)** Item 15

Figure A1: Item characteristic curves and proportion of correct answers by estimated competence (Part I). Actual values are computed as the proportion of correct answers in intervals of length 0.3 in the competence domain.

(a) Item 16 **(b)** Item 17 **(c)** Item 18

(d) Item 19 **(e)** Item 20 **(f)** Item 21

(g) Item 22 **(h)** Item 23 **(i)** Item 24

(j) Item 25 **(k)** Item 26 **(l)** Item 27

(m) Item 28 **(n)** Item 29 **(o)** Item 30

Figure A2: Item characteristic curves and proportion of correct answers by estimated competence (Part II). Actual values are computed as the proportion of correct answers in interva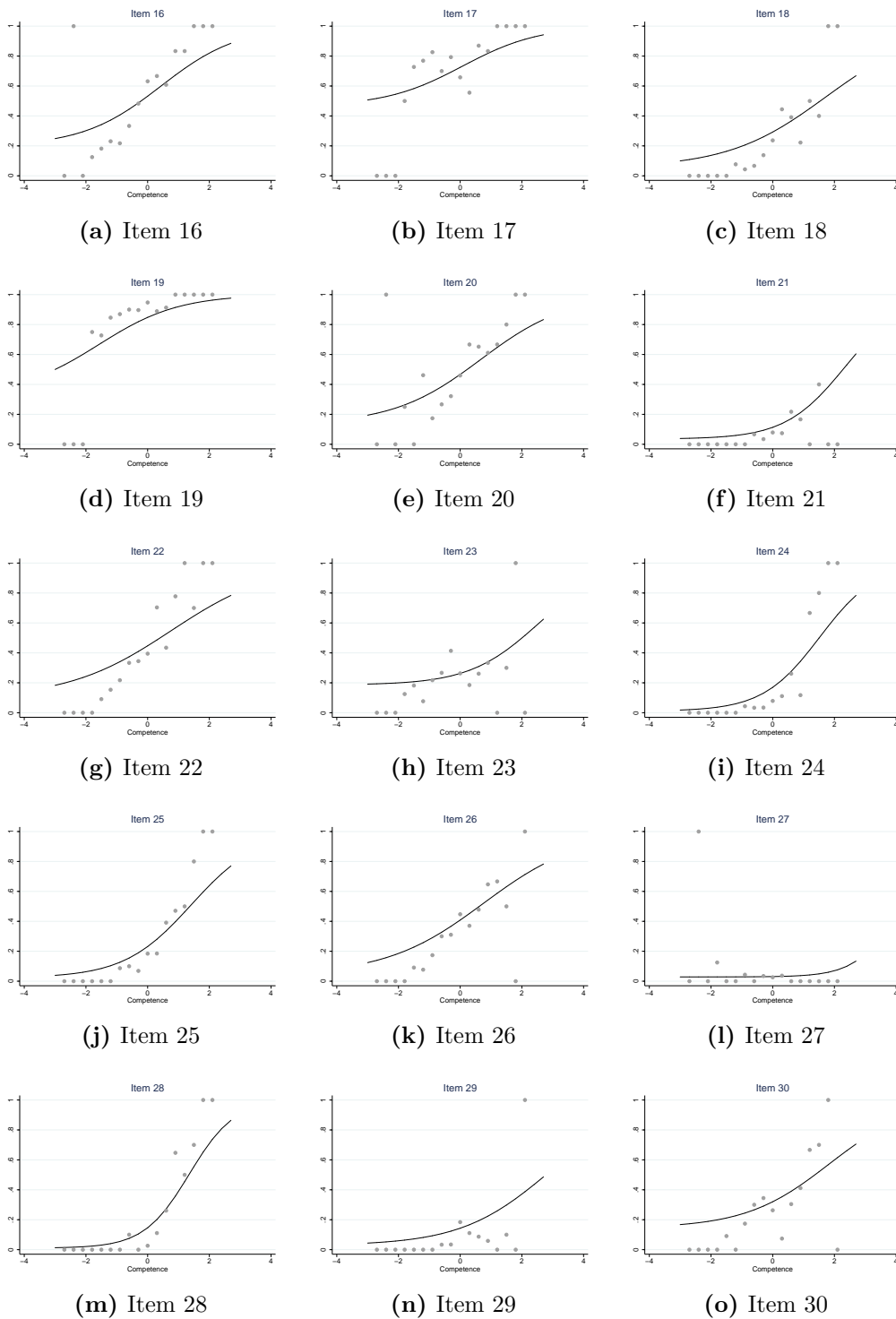ls of length 0.3 in the competence domain.