Contracting for Health:
Evidence from Cambodia

Erik Bloom
Asian Development Bank

Indu Bhushan
Asian Development Bank

David Clingingsmith
Harvard University

Rathavuth Hong
ORC Macro

Elizabeth King
World Bank

Michael Kremer
Harvard University

Benjamin Loevinsohn
World Bank

J. Brad Schwartz
University of North Carolina—Chapel Hill

Abstract: In 1999, Cambodia contracted out management of government health services to NGOs in five districts that had been randomly made eligible for contracting. The contracts specified targets for maternal and child health service improvement. Targeted outcomes improved by about 0.5 standard deviations relative to comparison districts. Changes in non-targeted outcomes were small. The program increased the availability of 24-hour service, reduced provider absence, and increased supervisory visits. There is some evidence it improved health. The program involved increased public health funding, but led to roughly offsetting reductions in private expenditure as residents in treated districts switched from unlicensed drug sellers and traditional healers to government clinics.

We are grateful for comments from Jeff Kling, Esther Duflo, Abhijit Banerjee, Angus Deaton, Sendhil Mullainathan, Ben Olken, Karthik Muralidharan, Emily Oster, and seminar participants at the BREAD Conference (May 2006), the IMF Institutional Reform Conference (July 2005), the LEaF Conference (May 2005), Columbia University, New York University, the Project on Justice, Welfare, and Economics at Harvard, and the Harvard Development Lunch. We also thank Loraine Hawkins, Krang Sun Lorn, Sao Chhorn, Peng Kok, Char Meng Chor, Sheryl Keller, Fred Griffiths, Rob Overtoom, Bart Jacobs, Gerry Pais, Reggie Xavier, Jim Tulloch, Steven Schomberger, and Sabine Heinrich for their insight into the contracting project and health care in Cambodia. Clingingsmith acknowledges support from the Project on Justice, Welfare, and Economics at Harvard University.

1. Introduction

Health care systems in many developing countries artfully combine the worst aspects of government and private provision. Incentives in government clinics are notoriously weak: Chaudhury et al. (2006) found an average absence rate among staff of 35% in surprise visits to health facilities in six developing countries. Private practitioners' incentives are strong, but are often not well aligned either with the interests of their patients (due to information asymmetries between providers and patients), or with larger public health concerns (due to externalities related to infectious diseases). Cross-country evidence shows little to no relation between public health spending and child mortality (Musgrove 1996; Filmer and Pritchett 1999). Banerjee, Deaton, and Duflo (2004) found that private medical practitioners in rural Rajasthan gave patients injections 68% of the time and IV drips 12% of the time, but tests only 3% of the time. An estimated 30%-50% of prescriptions written in India are unnecessary or contraindicated (Phadke, 1998; Das and Sanchez 2000). Is this private sector? If not, may want better example.

Starting in 1999, Cambodia tried an alternative approach in which the government tendered management of government health services for contract in certain districts to private bidders, and increased public health expenditure to pay for these bids. Contractors were required to provide all preventive, promotional, and simple curative health care services mandated for a district by the Ministry of Health, known as the Minimum Package of Activities (MPA). They were responsible for services at district hospitals, subdistrict health centers, and more remote health posts. Performance was measured against eight service delivery indicators. Inadequate performance could lead to sanctions and would reduce the likelihood that the contract would be renewed.

The district-level contracting approach is potentially attractive because it offers the opportunity to strengthen incentives for government workers while reducing potentially harmful incentives associated with private fee-for-service provision, such as the incentive to over prescribe antibiotics or to provide glucose drips, which do not improve health but make patients feel better in the short run. In rural areas of developing countries with limited mobility, contracting at the district level can allow substantial sharing of risks from health shocks without inducing the adverse selection associated with individual-level insurance. Contracting at the district level, rather than the national level, allows benchmark competition between providers. At the same time, offering contracts based on eight targeted outcomes runs the risk of inducing multi-tasking problems, in which contractors divert effort from measured to unmeasured outcomes.

Estimation of program impact is hampered by two difficulties. First, not all districts initially randomly assigned to be treated were in fact treated. The government randomly selected 8 districts from a set of 12 in which to introduce the program. However, bids that met technical and cost requirements were received in only five of the eight districts, and hence the program was only implemented in these districts. We therefore estimate the causal effects of the program by using the initial random assignment of treatment as an instrument for actual treatment status.

Second, a very small number of units were randomized. We compute average effects across families of outcomes to help alleviate the limited statistical power of our twelve-unit randomization. We report standard errors both with clustering and with randomization inference.

Despite the limited power associated with the small sample, estimated effects are large enough that many are statistically significant. The contracting program caused large increases in the service outcomes targeted by it, on average about one baseline standard deviation. To cite two examples, the receipt of vitamin A by children under 5 was increased by 42 percentage points and receipt of antenatal care by pregnant women was increased by 36 percentage points. The project improved the management of government health centers, particularly in the availability of 24-hour service, the actual presence of staff scheduled to be there, supervisory visits, and the presence of supplies and equipment. The program did not have large effects on health services indicators not explicitly mentioned in the contract. There is some limited evidence the program improved self-reported health. The program led individuals to shift curative care visits to public facilities, and reduce visits to untrained service providers such as drug sellers and traditional healers. Decreased out-of-pocket spending on curative care offset increased public spending, so the program did not increase, and probably decreased, overall health spending.

We draw on several previous studies of the contracting project (Keller and Schwartz 2001; Bhushan, Keller, and Schwartz 2002, Schwartz and Bhushan 2004). Using a 1997 baseline survey and 2001 midterm survey, these studies found that contracted outcomes improved in all treated and comparison districts, and that the improvements for treated districts were much greater than the comparison districts. However, the 2001 midterm survey did not collect data in the three districts initially assigned to treatment where the program was not implemented. This meant that previous work could not take advantage of the initial randomization of treatment eligibility to estimate causal impacts. Estimates will potentially be subject to bias if districts that received technically responsive bids differed from those that did not in unobserved variables that influence outcomes. For example, if potential contractors were more likely to bid on districts in which it appeared to be easiest to reach the contract targets, the effects of the program could be overestimated. To compute causal estimates, we draw on a new survey that covered all of the districts initially randomized into treatment. This paper also differs from earlier work in accounting for the cluster-randomized nature of the design, and in presenting new evidence on health center management, non-contracted outcomes, expenditures, and perception of the quality of care.

Our paper is organized as follows. Section 2 provides background on the health care system in Cambodia and the contracting project. Section 3 presents a model of health care provision. Section 4 discusses the empirical methods we employ in our analysis. Section 5 presents estimates of the project impact on health center management, targeted outcomes, non-targeted outcomes, careseeking behavior, and perception of care quality. Second 6 discusses the effects of the program on public, private, and overall health care expenditures. Section 7 concludes.

2. Health Care in Cambodia and the Contracting Project

This section provides background on health care in Cambodia and the contracting project, including randomization of treatment, the bidding process and its results, the contract terms, monitoring provisions, and budgets. We give an overview of some of the management practices used by the contractors once the program began. The section concludes with a review of the data collected as part of the project and what baseline data says about the quality of the randomization.

a. Health Care in Cambodia

Nearly 20% of Cambodia's population perished during the genocidal Khmer Rouge regime, which lasted from 1975 to 1979. When the Vietnamese drove the Khmer Rouge from power in 1979, only 50 doctors were left in the country. New medical personnel were trained during the Vietnamese-backed regime that ruled the country from 1979 until 1993, though the quality of training was poor. Private practice was banned, but public facilities were sparse and poorly equipped. Little was invested in rural health infrastructure, and many communes lacked a building to house a health clinic. Fighting continued in some areas of the country until 1998, just before the introduction of the program. Governance, corruption, and politicization of the civil service have been seen as serious issues in Cambodia. Political allegiance plays a role both in the selection of individuals to join the public service and in promotion. Political work is expected of public employees around election time. Promotion depends in part on such political work and political connections, and so managers are not necessarily chosen on the basis of merit or managerial ability alone.

In 1993, following the departure of the Vietnamese, UN-sponsored elections, and the adoption of a market economy, Cambodia began to build up its first universal public health infrastructure. International NGOs established a large presence in the country, delivering high-quality health-care services in limited geographical areas. Private practice, now legal, boomed. Nearly all trained providers in the country worked for the government, but many employees of the public system ran private medical practices on the side. Absenteeism in public facilities was high and diversion of patients to private practice and under-the-table payments were common. Private practice was very attractive because public sector salaries for health staff averaged only about 85% of Cambodia's US$283 per-capita GDP (Conway 2000). Importation of pharmaceuticals into the country was freely allowed after 1993 and their sale remains *de facto* unregulated. One can easily find a wide range of drugs from a wide variety of countries in most Cambodian markets, almost invariably sold by a vendor with no formal pharmaceutical or medical training. While the country made tremendous improvements to public health between 1993 and 1997, the contracting project's 1997 baseline survey found population coverage of preventative health care measures were still low even by developing country standards. For example, only a third of children under two were fully immunized. Only 4% of patients in need of curative care used the public system. About 33% of sick individuals who sought curative care went only to a drug seller.

Traditional understanding of health and disease remain strong. A 2000 survey asked mothers who reported the death of a child under five about the primary cause of death (RACHA 2000). Supernatural causes were cited by 39% of mothers. Patient beliefs no doubt affect both care choice and the practices of medical providers. Van de Put (1991) reports on a medical anthropology field study of Cambodian medical practices. His evidence suggests that for ordinary rural Cambodians, quality medical care involves being dispensed drugs. Diagnostic skill and medical expertise are not viewed as critical in deciding which drugs, if any, are appropriate to a given complaint, and there is little perception that drugs can be potentially harmful. Injections and intravenous drips are perceived as being more powerful and hence better than pills. Van de Put (1991) also argues that patients feel it is important to have input on which drugs they are given, and will view a practitioner who provides the medications they request favorably. Practitioners commonly give medically unnecessary vitamin injections and glucose drips that may make patients feel better in the short run.

b. The Contracting of Health Services Pilot Project

In 1996, Cambodia launched the Basic Health Services Project (BHSP), which focused primarily on the construction of rural health centers and referral hospitals and on the improvement of district-level health service management. The contracting project was an element of the effort to improve management. During the period in which the contracting project was in operation, the number of functioning rural health centers in all of Cambodia increased from 60 to more than 900.

The contracting project ran from 1999 to 2003 and covered a total population of about 1.26 million people, or about 11% of Cambodia's population (Schwartz and Bhushan 2001, Cambodia 1998 Census). The approach was then expanded to additional districts, though not as a randomized experiment.

There were two variants of the approach, contracting-in and contracting-out. They differed in the degree of control to be given the contractors. Contracting-in districts were expected to work within the existing government system for procurement of drugs, equipment, and supplies. Their operating expenses were financed through the government budget in the same manner as ordinary districts. They were required to use existing Ministry of Health personnel; they could request transfers of personnel but not hire or fire. Contracting-out district management had pretty much full authority for and responsibility over their districts. They were allowed to hire and fire staff, could bring in health workers from other parts of the country, and were responsible for their own procurement of drugs, supplies, and equipment. Existing Ministry of Health staff in the contracting-out districts could join the contractor's organization and take leaves of absence from the civil service. If the contractor decided to fire these staff, they would be transferred to a government post in a different district. In the end, only a few staff members in contracting out districts were fired. The project designers' initial intention was that salaries in the contracting-in districts would be based on the civil service pay structure, plus additional amounts decided by the contractors that would be raised from user fees. Contracting-out contractors, in contrast, could implement the pay structure of their choosing.

In treated districts, the management of government health care services was put out to competitive bid for qualified organizations, such as NGOs and private firms. For each district the organization with the highest combined score on technical quality of their proposal and price was awarded a contract to manage the district's government health care service. In the end, only international NGOs, firms, and universities submitted bids. All the winners were international NGOs, which is not surprising as there were almost no local NGOs working in the health sector at the time. The comparison districts continued to be managed by the local employees of the Ministry of Health.

The twelve districts in the project came from three provinces in south-central Cambodia: three districts came from the province of Takeo, four districts came from the province of Prey Veng, and five districts came from the province of Kampong Cham. The twelve rural districts participating in the contracting experiment were chosen because they did not contain the provincial capital and were not significant recipients of other development assistance.

The twelve districts were randomly assigned to three groups: Four were eligible to receive contracting-in bids, four were eligible to receive contracting-out bids, and the remaining four served as a comparison group. Randomization was quasi-stratified by province. A project team visited each provincial health department and, in the presence of district managers, had its director randomly draw one district to become part of the contracting-in group, one district to become part of the contracting-out group, and one district to become part of the comparison group. The remaining three districts were randomly assigned later in Phnom Penh to one of the three eligibility groups.

The Ministry of Health put out a request for bids in early 1998. Ten bidders submitted a total of 16 proposals for the eight districts (bidders were able to apply for more than one contract). The 10 bidders represented 14 different organizations, as some bids came from partnerships among different organizations. Of the 14 organizations involved in the bidding process, 8 were NGOs working in Cambodia, 4 were consulting firms, and 2 were university-affiliated groups. A two-envelope system was used to evaluate the bids. First, a committee of Ministry of Health officials and outsiders assigned technical scores to the proposals based on criteria that were explicit in the request for proposals. Bids had to receive a minimum technical score to be considered "technically responsive." The price envelopes of the technically responsive bidders were opened in public. Then the technical scores were combined with the bid price (the lower the price the higher the score) using an explicit formula and the bidder with the highest combined score was awarded the contract. The technical criteria included the prior experience of the contractor in similar projects, the quality of the key staff proposed to run the project, and the quality of the contractor's management plan. The contracts were awarded in late 1998.

Only five of the eight districts randomized into the treatment group were successfully contracted. In two districts, the bids received were not technically responsive. In the third, two technically responsive bids were received, but both were judged to be too expensive. These remaining three districts remained under government management during the contract period. Contractors took over management of the five successfully contracted districts by April 1999.

The contracts signed by the Ministry of Health and the successful bidders targeted improvements in the delivery of the minimum package of activities and large increases in eight health-service indicators primarily related to maternal and child health: childhood immunization, administration of vitamin A to children, antenatal care for pregnant women, child delivery by a trained professional, delivery in a health facility, the knowledge and use of birth control, and use of public facilities when seeking curative care. Note that since most of these services are preventative and several create positive externalities, they arguably would be under-supplied by private providers working under fee-for service contracts.

Baseline levels for the targeted outcomes are shown in Table 1 along with program goals. The goals are between 160% and 450% of the baseline levels and were to be achieved within the four-year term of the contract. The substantial gains expected reflect the very poor initial state of the system as well as optimism about the potential of the contracting approach.

The contract also made a provision for a Ministry of Health monitoring group to survey the contracted districts to determine progress toward the targeted outcome goals, and allowed the Ministry of Health to withhold payments to the contractors if progress was not satisfactory. Costs of monitoring were included in the project budget but not in the contracts. Monitoring teams visited each district quarterly, inspected the district hospital, conducted village surveys to measure targeted outcomes, and visited a sample of patients listed in health center registers to see if they were actually treated and assess their experience. In at least two instances the Ministry of Health judged that a contractor was not making satisfactory progress toward the contract goals, and sent a letter to the contractor outlining their concerns. One problem was resolved without further action. The other was not, and led to the Ministry of Health sanctioning the contractor by suspending payment for one quarter, until the problem was remedied.

The contracting-out districts received their funds directly from the ADB after the Ministry of Health made a payment request. The contracting-in districts received the management fee portion of their contract budget in the same manner. Operating funds and supplies were provided to the contracting-in and comparison districts through normal government channels. In addition, contracting in and comparison districts were eligible to receive an operating supplement of $0.25 per capita per year paid directly from the ADB after submitting an acceptable plan.[1] The comparison districts were also given health care management consulting services and management training as part of the Basic Health Services Project. Those districts randomized into treatment that were not successfully contracted received neither the supplement, management consulting, nor management training (see Table 2).

c. Human Resource Practices under the Project

The most detailed account of NGO management under the contracting project is Soeters and Griffiths's (2003) article about Pereang district of Prey Veng province, operated under the contracting-in model by HealthNet International (HNI). The contractor viewed staff motivation as the key challenge it needed to overcome, and implemented a performance-based incentive system. After trying to implement direct subcontracts with all staff members, the contractor decided to subcontract with the managers of the health centers and hospital under its control, who would in turn establish contracts with the staff members they supervised. Staff members received a guaranteed supplement of 55% of their government salaries plus a 30% performance bonus and a 15% punctuality bonus. HNI designed a user-fee system with the goal of formalizing part of the substantial out-of-pocket patient expenditure, bringing it into a system where it could be monitored. Service prices were set with patient community consent at about 60% of the prevailing market price, which the NGO determined was adequate to make private practice

---

[1] Most of the comparison districts proposed to utilize the budget supplement to conduct immunization outreach activities. Audit irregularities were often encountered in three of the four districts. The amounts actually used by comparison districts was often less than the full $0.25 per capita.

unattractive for most. User fees thus collected paid about half of staff incentives, with the other half coming through the NGO. In addition to the monitoring activity of the Ministry of Health, which was shared with all contractors, HNI conducted its own surveys and spot checks.

While the contracting-in budget supplement was intended to boost the operating funds sent through government channels, in general the contracting-in contractors used it to provide incentives to their staff instead. Within the first few weeks of taking over their districts, the contracting-in contractors discovered it was impossible to motivate their staff members to work or to enforce regulations without salary supplements. Contractors in all five treated districts implemented performance-based incentives for staff. In the three contracting-in districts, this typically consisted of a fixed supplement to staff members' government salaries plus a performance-based bonus. The two contracting-out districts implemented fixed salaries considerably higher than the government previously paid, with the incentive provided by the possibility of dismissal. Two of the five treated districts attempted to ban private practice by employees, while the other three tried to restrict it by forcing staff to attend their assigned hours in the health center. Encouraged by the Ministry of Health, the contractors and the district health management teams in the comparison districts implemented user-fee systems. By 2003, nearly all facilities in treated and comparison districts had established user-fee systems that contributed to the payment of staff salaries and incentives.

All of the successful bidders hired expatriates for some management and advisory roles, with about 0.5 to 3.0 expatriates per district at any given time. They typically filled the role of district manager, overseeing between about 60 and 120 local staff. Annual salaries for expat managers ranged widely between about US$15,000 to US$60,000. Expatriate staff included Europeans, North Americans, and Asians.

d. Data and Randomization Quality

We collected data on individual health care outcomes and care-seeking behavior from a random sample of 30 villages in each of the 12 districts involved in the contracting project. About 20,000 individuals in 3,700 households are included in the samples. A baseline survey was conducted in 1997 before interventions were made, and a full follow-up was conducted in 2003. While the same villages were sampled in both survey years, within villages a new random sample of households was taken each time. The data is thus a panel at the village level and a repeated cross-section at the household level. The 2003 follow-up also included questions about respondents' perceptions of the quality of care at government facilities. A separate survey of the 143 health centers in the project area was also conducted in 2003. Administrative data on public expenditures during the project years was compiled from Ministry of Health records in 2004.

We have baseline data for 22 of the outcomes we will examine in the paper, and can use it to examine how well balanced our randomization was. Our treatment effects regressions show that baseline levels of three outcomes for each variant have coefficients that are statistically significant at 5% under clustering, and that one for each variant have coefficients that are statistically significant at 5% under randomization inference. Under random assignment, we would expect one significant coefficient at 5% for each of the two variants.

Contracting out had significant baseline coefficients on the following outcomes at the following levels: receipt of vitamin A by children under 5 (5% under clustering, 10% under randomization inference), treatment of diarrhea in children under 5 who have symptoms (5% under clustering only), and private health spending (5% under both clustering and randomization inference). Contracting in had significant coefficients on the number of village outreach visits during the past month (5% under clustering and randomization inference), and a negative coefficient significant at 10% under clustering on not giving water away from newborns under one month old.

3. A Model of Health Service Provision

The pre-existing Cambodian health system involved a combination of government clinics with very flat incentives and de facto unregulated fee-for-service private practice, in which providers had steep incentives, but these incentives were only to provide health services privately beneficial to the patient, rather than to take into account public health benefits (for example of vaccination). Moreover, since patients observe only a noisy signal of health, private providers may have had incentives to provide services that made people feel better in the short-run, like glucose drips. The program created incentives to focus on the targeted indicators, and led to restrictions on private practice.

A Holmstrom-Milgrom (1991) framework suggests that contracts linking incentives to the 8 targeted outcomes will lead to better performance on those measures, but how it affects other outcomes depends on whether effort directed at those non-targeted outcomes is a complement or substitute with the targeted outcomes. Either scenario is plausible. For example, it could be that the incentives provided to the contractor cause contractors to create incentives for health workers to reduce absence from the facilities, and that this is complementary with providing other types of care. On the other hand, facilities might shift resources away from unmeasured care to targeted outcomes.

We will formalize this idea in a simplified Holmstrom-Milgrom (1991) framework. Suppose there are two health outcomes. The agent has control over two kinds of effort that are costly to exert. Suppose only one of the outcomes is contractible. Denote the outcomes C and NC and the effort types e_1 and e_2 and let them be produced as follows

$$\begin{aligned} C &= f(e_1, e_2) + \varepsilon \\ NC &= g(e_1, e_2) + \eta \end{aligned} . \tag{1}$$

The agent cares about compensation w as well as the cost of exerting effort,

$$u(w, e_1, e_2) = w - c(e_1, e_2) . \tag{2}$$

Agents are paid a linear wage in the amount of the contracted outcome produced

$$w = \alpha + B \cdot C . \tag{3}$$

The agent's first order conditions are

$$\frac{dc}{de_1} = B\frac{df}{de_1}; \ \frac{dc}{de_2} = B\frac{df}{de_2} . \tag{4}$$

Note that the function $g(e_1, e_2)$ does not appear in the first order conditions. The agent chooses effort only according to the tradeoff between the cost of effort and the marginal increase in C output that results from effort. Increasing B will typically increase C, but may increase or decrease NC.

This simple framework does not consider the possibility that the NGO and Ministry of Health view themselves as being in a repeated game setting. In such a setting, the Ministry would plausibly learn both C and NC after the contract finishes, and could use that information in deciding whether to accept another bid from the NGO for a further contract. Substituting away from non-targeted outcomes that the Ministry of Health cares about would tend to be muted to the extent an additional contract is desirable to the NGO.

4. Empirical Approach

There are two main empirical challenges to estimate program impact. First, not all districts initially assigned for treatment were treated. Second, random assignment was at the district level, reducing power. Simple comparisons of actually treated districts with others could yield biased estimates of program impact if districts dropped out because they were known by NGOs to be difficult to work in, or because few NGOs had ever worked there before. As discussed in Subsection 4a, we use the initial randomization of treatment as an instrumental variable for actual treatment. This produces an estimate of the effect of treatment on the treated (TOT) that, in expectation, expunges the variation in actual treatment that may have come from sources other than random assignment.

As discussed in Subsection 4b, we address the fact that randomization at the district level requires us to take care in computing standard errors. We use both the techniques of clustering at the district level and of randomization inference. We also report the average effect of contracting across a family of related outcomes that comprise a domain of interest. An example of a family is the eight service-delivery outcomes explicitly targeted by the program for improvement. We discuss this technique in Subsection 4c.

a. Estimation of per-comparison causal effects

We report intention-to-treat (ITT) and treatment-on-treated (TOT) causal effects (Imbens and Angrist 1994, Angrist, Imbens, and Rubin 1996). The ITT tells us the effect on a health care outcome of being in a district randomly selected for health-care contracting. The TOT tells us the local average treatment effect (LATE) on a health care outcome of actually being subject to a contracted health care system due to random assignment. (Thus, we can't necessarily generalize about what would have happened if the program were implemented in districts where no acceptable bids were received.)

Our basic regression ITT model for the effect of the program on a particular outcome y_{ivdpt} for an individual i from village v, district d, and province p at time t is:

$$y_{ivdpt} = \beta_0 + \beta_1 I_d^{CI-R} + \beta_2 I_d^{CO-R} + \beta_3 I_t^{2003} + \beta_4 I_d^{CI-R} \times I_t^{2003} + \beta_5 I_d^{CO-R} \times I_t^{2003} + X'_{ivdpt}\theta + p_{pt} + \varepsilon_{ivdpt} . \quad (5)$$

The dummy variables I_d^{CI-R} and I_d^{CO-R} indicate a district's random assignment. X'_{ivdpt} is a vector of individual characteristics which we include in some specifications. The ITT effect of contracting-in and contracting-out are β_4 and β_5, respectively. The province-by-year fixed effects p_{pt} reflect the quasi-stratification of the randomization by province and absorb time-varying shocks at the province level.

Our regression model for the TOT is similar:

$$y_{ivdpt} = \beta_0 + \beta_1 I_d^{CI-T} + \beta_2 I_d^{CO-T} + \beta_3 I_t^{2003} + \beta_4 I_d^{CI-T} \times I_t^{2003} + \beta_5 I_d^{CO-T} \times I_t^{2003} + X'_{ivdpt}\theta + p_{pt} + \varepsilon_{ivdpt} . \quad (6)$$

In this case, I_d^{CI-T} and I_d^{CO-T} are dummy variables for a district's actual treatment status. We estimate this latter equation using I_d^{CI-R} and I_d^{CO-R} as instrumental variables for actual treatment status.

Two notes of caution are warranted in interpreting the results. First, in many tables, constants in our regressions differ from baseline values of variables for the comparison districts because of the inclusion of province-by-year effects. Second, we estimate linear effects while most of our dependent variables are bounded between zero and one. IV estimates may occasionally lead to effects that seem to imply that the level of the dependent variable has risen above one or fallen below zero.

b. Randomization at district level

Our analysis examines the effects of district-level independent variables on individual-level dependent variables. When treatment is constant within the aggregate unit, we must allow for errors to be correlated at the aggregate level (Moulton 1990, Donald and Lang 2001). In our case, individual outcomes may be correlated due to a district-level trend, such as the development of the local transport system or differential recovery from conflict.

One approach to computing standard errors is to use the cluster-correlated Huber-White covariance matrix estimator. Donald and Lang (2001) and Wooldridge (2004) have pointed out that asymptotic justification of this estimator assumes a large number of aggregate units. Simulations in Bertrand, Duflo, and Mullainathan (2002) show the cluster-correlated Huber-White estimator performs poorly when the number of clusters is small (<50), leading to over-rejection of the null hypothesis of no effect.

Randomization inference provides an alternative approach to hypothesis testing (Rosenbaum 2002). Consider a simplified setting in which we have observations Y_{ij} on an outcome for individuals *i* residing in clusters *j*. Suppose we randomly allocate each of the clusters to a treatment group and a comparison group, and that $T_j = 1$ for the treatment group and $T_j = 0$ for the comparison group. We apply a treatment to the treatment group. Then β_T in regression equation (7) represents the average effect of treatment on Y_{ij},

$$Y_{ij} = \alpha + \beta_T T_j + \varepsilon_{ij} . \tag{7}$$

Denote the set of all possible assignments from the randomization process $\{P_j\}$. We call the P_j placebo random assignments. Now consider β_P in the following regression equation:

$$Y_{ij} = \delta + \beta_P P_j + \nu_{ij} . \tag{8}$$

Since P_j is a randomly generated placebo, $E(\beta_P) = 0$. Let $F(\hat{\beta}_P)$ be the empirical c.d.f. of $\hat{\beta}_P$ for all elements of $\{P_j\}$. We can now perform a hypothesis test by checking if our measured treatment effect is in the tails of the distribution of placebo treatments. We can reject $H_0 : \hat{\beta}_T = 0$ with a confidence level of $1-\alpha$ if $\hat{\beta}_T \le F^{-1}(\frac{\alpha}{2})$ or $\hat{\beta}_T \ge F^{-1}(1-\frac{\alpha}{2})$. Since the placebo assignments P_j only vary across clusters, this method takes intracluster correlations into account.

In practice, we use the method employed to make the actual random assignment, described in Section 3, to generate the full set of potential unique assignments of the twelve districts to the comparison group and contracting-in and contracting-out treatment groups. There are 6,480 unique random assignments, which were equally likely to occur. We then compute placebo treatment effects for each of these random assignments using the placebo version of our ITT estimating equation (5). We compute a p-value by noting where the true effect lies in the distribution of placebo effects.

We note that this technique has low power relative to more parametric approaches when the true effect is large because it puts not even minimal structure on the error term. To see this, consider a hypothetical example with six clusters. The randomization process selects two clusters to be in the treatment group and the remaining clusters to be in the comparison group. There are 15 combinations in the set of placebo randomizations $\{P_j\}$, each of which has an equal probability of being selected of $\frac{1}{15} = 6.7\%$. This means that neither the largest nor the smallest $\hat{\beta}_P$ can fall within the 2.5% tails of the distribution $F(\hat{\beta}_P)$. Randomization inference would thus fail to reject the null hypothesis of no treatment effect with 95% confidence no matter how large the difference between treatment and comparison and how small the differences within groups, since it is always impossible to reject the hypothesis that errors take the form $-x$ with probability 1/3 and $x/2$ with probability 2/3, where x is the difference between groups. If one were willing to impose that errors were single-peaked, however, one would reject the null if the difference between the treatment and comparison groups were large enough relative to the differences within the groups. Randomization inference is in this sense a low power test relative to one that imposes even minimal structure on the error term. We present hypothesis tests based both on clustering and on randomization inference below.

c. Summary measures of causal effects

We have very low power in testing individual effects, but many of the outcomes we believe will be affected by contracting fall into families. To the extent that outcomes within families are not perfectly correlated, looking at families can help increase our power. We are interested in the magnitude of the average effect of treatment on entire families of outcomes and whether it can be distinguished from the null hypothesis of no average effect. In order for the average effect to be meaningful, each of the outcomes $k = 1 \ldots K$ in each family is scaled so that the treatment effects π_k are positive if they are desirable.

More specifically, we seek to test the one-sided hypothesis

$$H_0 : (\pi_1,\ldots,\pi_K)' = \mathbf{0} \text{ vs. } H_a : (\pi_1,\ldots,\pi_K)' \in O^+, \tag{9}$$

where O^+ is the positive orthant. Following O'Brien (1984), Tamhane and Logan (2003), and Kling, Katz, Leibman, and Sonbanmatsu (2004), consider the measure of average effect size τ over the family of K outcomes $\{y_{kdt} \mid k = 1 \ldots K\}$ in which each treatment effect is normalized by the standard deviation σ_k of the change in the outcome Δy_{kdt}, where

$$\tau = \frac{1}{K}\sum_{k=1}^{K}\frac{\pi_k}{\sigma_k}. \tag{10}$$

[2]

[2] Alternative ways of weighting the treatment effects include using the first principal component. If the first principal component accounted for a good deal of the variation in outcomes, we might be persuaded that a single dimension, such as provider effort, drives the outcomes. We have explored this possibility, but found that first principal component accounts for only a small part of the variation in our outcomes.

O'Brien (1984) showed that τ could be used to test a restricted version of hypothesis (9) that includes the additional assumption of a constant treatment effect across outcomes within the family: $\pi_k / \sigma_k = \lambda \; \forall k$. O'Brien's method thus tests the modified hypothesis $H_0 : \lambda = 0$ against $H_A : \lambda > 0$.

Let t_K be the $K \times 1$ vector per-comparison t-ratio for each of the treatment effects. Let R be a $K \times K$ matrix with elements $\rho_{k\ell} = Corr(y_{kdt}, y_{\ell dt})$ and j be a $K \times 1$ vector of ones. Then the then the t-ratio for τ in testing hypothesis (9) is given by equation (11)

$$t_\tau = \frac{j^T t_K}{\sqrt{j^T \hat{R} j}}. \tag{11}$$

O'Brien showed that the ratio t_τ is t-distributed with $n-2$ degrees of freedom.

Because our outcomes are defined over different groups of individuals (e.g. children 12-23 months old, women who have given birth in the past year), we aggregate outcomes to a common level, such as the village, to obtain a consistent unit of observation:

$$\begin{aligned} \overline{y}_{vdtk} &= \alpha_1 + \alpha_2 I_d^{CI-T} + \alpha_3 I_d^{CO-T} \alpha_3 I_t^{2003} + \alpha_4 I_d^{CI-T} \times I_t^{2003} + \alpha_5 I_d^{CO-T} \times I_t^{2003} + p_{pt} + \eta_{vdkt} \\ &= W\theta_k + \eta_{vdkt} \end{aligned}. \tag{12}$$

Our joint equation is estimated by instrumental variables, again using random assignment as an instrument for treatment

$$\begin{pmatrix} \overline{y}_{vdt1} \\ \vdots \\ \overline{y}_{vdtK} \end{pmatrix} = \left(I_K \otimes W \right) \theta + \eta_{vdkt}. \tag{13}$$

Coefficients in θ form the inputs into our average effect size calculations.

5. Health Care Results

We explore the causal effects of the program on five families of health care outcomes, looking at both effects at the discrete outcome and family level. Subsection 5a considers health center management, the domain under the most immediate control of the contractors. Subsections 5b, 5c, 5d, 5e and 5f consider targeted outcomes, non-targeted outcomes, final health outcomes, curative care seeking behavior, and finally consumers' perception of quality.

a. Health center management

We would initially like to understand something about what practices contractors engaged in as a result of the program. Health centers in contracting project districts were surveyed in 2003 to collect information on how they ran their facilities, what services were available, and how well supplied they were. The survey visits were unannounced to help ensure an accurate account of the condition of the health centers. A baseline survey was not conducted because very few permanent facilities were operating in 1997. We therefore analyze the survey using the simple differences between the treated groups and the comparison groups.

We construct 18 measures of health center activities (Table 3). Both variants had positive point estimates for there being a permanent, functioning health center building. The 23.6 percentage point contracting-in effect was significant at 5% under clustering. The comparison mean was 74%. (It is possible the contracting effects on facility construction were greater earlier in the program.) Because opportunity and travel costs to visit the health center can be high, people may be unlikely to go if they are unsure if staff will be available. Round-the-clock service could be an important factor in patients' shift toward public facilities. Contracting-in made it much more likely that round the clock services would be reported to be available at the health center. The contracting-in effect is a very large 83 percentage points and is statistically significant. The contracting-out effect is 47 percentage points but not significant. Both variants had large positive point estimates on health centers being open and treating patients during an unannounced visit. The 48 percentage point contracting-in effect was significant at 5% under clustering. Contracting-in and contracting-out increased the probability that all scheduled staff would be present by 50 and 79 percentage points, respectively. These effects were significant at 5% and 1% under clustering. Unfortunately we do not have more detailed information on the fraction of staff present. Neither the receipt of support from additional NGOs or the accuracy of health center registers was statistically significantly affected by the contracting variants.

Contracting-in had a point estimate of 25 percentage points on the health centers offering delivery services, against a comparison mean of 52%. The contracting-out point estimate was 40 percentage points. Neither was statistically significant.

User fees were common in 2003, with 71% of comparison health centers charging them. The contracting variants did not have statistically significant positive effects on them being in place.

Having user fees clearly posted can help prevent staff members from overcharging patients. Contracting-in and contracting-out respectively had 24 and 28 percentage point effects on user fees being clearly posted, against a comparison mean of 77%. The contracting-in effect was statistically significant at 5% under clustering.

Contracting out increased the level of supervision of health centers. This variant saw an increased number of supervisor visits to the health center during the previous three months by a dramatic 5.7 visits against a comparison mean of 2.5 visits. This effect was significant at 1% under clustering and 10% under randomization inference. Contracting-in had a small and statistically insignificant point estimate. The contracting variants did not have an effect on the supervisor's reported activities during the visits to the extent we can measure them. Contracting-out NGOs had larger budgets under their control, which may have made it easier for them to apply more resources to site visits. Similarly, contracting-in caused a decrease of 2.7 in the number of outreach visits in the past month against a comparison mean of 14.3. While this figure was not significant, management efficiency as measured by number of actual visits less scheduled visits was a statistically significant and 0.2 higher for contracting-in against a comparison mean of -0.1. So contracting-in resulted in significantly fewer scheduled visits yet a higher ratio of actual to scheduled visits.

We constructed indices that measured how many of 22 required pieces of equipment were present and functioning in the health center and how many of 41 required supplies were present. Point estimates of the contracting effect on the equipment index were positive for both program variants. Contracting-in showed an increase in the index of 3.5 against a comparison mean of 15.0 that was significant at 1% under clustering and 10% under randomization inference. Contracting out had an effect of 3.0 that was significant at 10% under clustering. Contracting-in and contracting-out had statistically significant effects of 5.5 and 8.9 on the index presence of required supplies, with the comparison districts averaging 25. The effects were significant under clustering at 1% and 5%, respectively.

We view 11 of the 18 health center management outcomes as positively related to management quality in a clear way, and construct an average effect size measure based on them

(Table 3).[3] Both contracting-in and contracting-out had statistically significant effects under clustering and randomization inference. The contracting-in average effect was about 0.6 comparison group standard deviations, while the contracting-out effect was 1.1 comparison group standard deviations. The two effects are different from one another ($p<0.01$), so we conclude the while both variants substantially improved management, contracting out did so even more than contracting in.

b. Targeted outcomes

This subsection first presents histograms on targeted outcomes, and then per comparison and average effect estimates. To show robustness, we also report an alternative estimator that controls for household wealth.

Before plunging into statistical analysis of program impact, it is worth examining histograms of the percentage point changes in district level average outcomes (Figures 1a and 1b). Note first the tremendous overall improvement in Cambodia over the period. It is also clear from histogram that the changes in some variables, such as use of public facilities, are much larger for the treated districts than either the comparison districts or the districts in which treatment was not taken up. The variance within the treated and not-treated groups is clearly less than the variation across groups. The picture for antenatal care receipt by women who had a child in the past year is similar. Changes in comparison districts and not-taken-up districts are

[3] The 11 outcomes are whether the permanent health center building is constructed and open, the availability of 24-hour service at the health center, on an unannounced visit whether the health center was open and seeing patients and whether all scheduled staff were present, availability of child delivery service, for health centers with user fees whether the fees are clearly posted, the number of supervisor visits in the past three months, the number of outreach trips undertaken in the past month, an index of equipment that should be installed and functional, an index of drugs and other supplies available, and whether the health center offers the full set of childhood immunizations.

roughly comparable. Note that the large differences between actually treated and not-treated districts will show up as large standard errors in our regression.

Per-comparison ITT regressions are shown in Table 4. Recall that they are all estimated in levels. All outcomes show large and statistically significant increases between the baseline and follow-up, in five of eight cases more than doubling. The background for the treatment effects is thus a strong secular increase in service provision, driven perhaps by general recovery in Cambodia, the large expansion of national health infrastructure under the Basic Health Services Project, competition between NGOs and the Ministry of Health, and improvements in the management of vertical programs like immunization. Over the study period, the number of functional rural health centers in Cambodia increased from 60 to more than 900.

The bolded figures show estimated program impact. The point estimates of ITT effects are nearly all positive, though the standard errors are fairly large, reflecting the limits put on the sample size by the twelve-district randomization. They are statistically significant for the use of public facilities when sick under both contracting-in and contracting-out (1% under clustering; 10% and 5% under randomization inference, respectively), for adequate antenatal care for those recently pregnant under contracting-in, and for vitamin A receipt by a child under contracting-out. P-values under randomization inference are substantially smaller than those under clustering, but do not change the overall pattern of statistical significance.

TOT effects are large and statistically significant in four of eight outcomes (see Table 5). Contracting-in and contracting-out led to 18 and 29 percentage point increases in the choice of a public sector facility when needing a curative care consultation (significant at 1% under clustering and 10% and 5%, respectively, under randomization inference), very substantially above the 4% baseline level for the comparison group and its 8 percentage point increase by

2003. Contracting-in caused a statistically significant 36 percentage point increase in the receipt of antenatal care by pregnant women, compared to a baseline level of 11% in the comparison group and over a 22 percentage point comparison increase by 2003. Contracting-out caused a statistically significant 42 percentage point increase in vitamin A receipt by children over a comparison increase of 23 percentage points and against a baseline comparison ratio of 43%. Contracting-in and contracting out led to 18 and 30 percentage point increases in the use of a health care facility for delivery against baseline level of 4% (significant at 1% under clustering in both cases and 10% and 5%, respectively, under randomization inference).

Both variants had substantial positive effects on child immunization, though only the contracting –in effect was significant (10% under clustering). Note that immunization increased very strongly in comparison districts between the surveys, going from 34% in 1997 to 81% in 2003. In this and a number of similar cases in the paper, there is limited scope for increasing immunization over and above the comparison group change, even though the final few percent may contain the most difficult cases to administer.

Contracting-in and contracting-out improved targeted outcomes an average of about 0.51 and 0.54 standard deviations more than the comparison group (Table 5, final column). These are very large average effects. The hypothesis of no average positive effect is rejected at 1% under clustering for both variants and at 1% and 10% under randomization inference for contracting in and contracting out, respectively. Contracting clearly caused large gains in the coverage ratios targeted by the program. These gains are even more impressive when we consider that comparison districts were encouraged by the Ministry of Health to compete with the contracted districts in these areas. We have explored whether the average effects of contracting differ across variants or between the upper and lower halves of the wealth distribution within variants, and fail to reject the null hypothesis of no difference in both cases.

It is worth considering results from an alternative specification that seeks to control for time-varying district-level economic shocks, for example, from differential economic recovery following 25 years of conflict or shocks to agricultural output and prices. Table 6 shows specifications that include controls for household assets in the per-comparison results. This will not be our main specification due to concerns about the endogeneity of household assets, though we believe that they are unlikely to have been significantly affected by the program. To the extent that they were affected, we would expect the bias induced by including them in our treatment regressions to be toward zero: As discussed below, the program reduced private health expenditure, presumably freeing up some income to be spent in part on the assets in our index, assuming they are normal goods. People with more assets are likely to be better able to access publicly provided services. With asset controls, the expected value of the estimated treatment effects will be the true effect less the product of two positive quantities—the effect of the program on assets and the effect of assets on service receipt. Our argument about bias is more fully developed in Appendix A.

Average TOT effects of the program on wealth are small and statistically insignificant, suggesting the program itself did not have a measurable effect on wealth (Table 6, Panel A). Controlling for assets improves the fit of the per-comparison TOT regressions slightly (Table 6, Panel B). For example, contracting-in now has TOT effect of 12.4 percentage points on delivery in a health facility that is statistically significant at 10% under both inference methods. Overall, the size of the per-comparison effects is similar to regressions that do not control for assets. The average TOT effects of contracting-in and contracting-out on targeted outcomes controlling for assets are 0.63 and 0.67 standard deviations, respectively, and are statistically significant at 5% under both methods. It is reassuring to know that effects are similar under this specification.

c. Non-Targeted outcomes

In section 4 we described the possibility that contractors could take advantage of the incompleteness of their contracts to divert resources away from outcomes on which they were not being explicitly evaluated. We now consider a set of six outcomes that were not explicitly part of the contracts between the Ministry of Health and the contracted NGOs but are nonetheless likely to be important. They are treatment of diarrhea in children, the number of antenatal services (excluding a blood pressure check, which was targeted), whether individuals report that an outreach team has visited the village in the previous four weeks, whether a mother breastfeed a newborn within six hours of birth, whether a mother gave a newborn water in the first month of life, and knowledge of AIDS risk factors.[4] Data was collected about them along with the contracted outcomes in the baseline and follow-up surveys. While it is unclear how they fit into the Ministry of Health's objective function, they all can have significant impact on the well being of individuals.

Overall, the program had little effect on non-targeted outcomes (Table 7). I think we might want to rephrase this, here and elsewhere as saying effects are not statistically significant. Contracting-in had a statistically significant positive TOT effect of 18 percentage points on an outreach visit to the village in the past month (significant at 5% under clustering and 1% under randomization inference). Contracting in improved knowledge of HIV risk factors by a

[4] The additional antenatal services considered are checking the abdomen and feet, blood and urine test, anemia eye check, dispensing iron tablets, and advice about food during pregnancy and the danger signs of pregnancy and child birth. Breastfeeding within the first hours of birth allows the transmission of colostrum to the baby, which important for its immune system. Water should not be given to babies under six months old because of the danger of disease transmission.

statistically significant 21 percentage points. Contracting out had a positive effect on the treatment of diarrhea, which we define as the administration of oral rehydration solution, salted soup, or an IV. The point estimate was 14 percentage points and is significant at 10% under clustering. Both contracting-in and contracting-out had positive point estimates for mothers being more likely to give newborns water during the first month of life, an undesirable practice since water can carry disease, though only the contracting-out coefficient of 9.3 percentage points was statistically significant (10% under randomization inference). This practice was universal in 1997, and declined to 95% in comparison districts in 2003.

Overall, contracting-in had a positive average estimated effect on non-targeted outcomes of about 0.3 baseline comparison group standard deviations, though the effect was not significant and much smaller than its 1.0 standard deviation average effect on targeted outcomes. Contracting-out had a negative average point estimate of 0.3 standard deviations, compared to 1.1 standard deviations on the targeted outcomes, though this estimate is also not statistically significant. Treatment effects were not statistically significantly different between the top and bottom of the wealth distribution for either variant. These results alleviate the concern that the incomplete nature of explicit contracts would lead the variants to have large negative effects on non-targeted outcomes.

d. Final health outcomes

Improvement in final health outcomes such as mortality and morbidity are the ultimate goals of an improved public health system. Logically, intermediate outcomes such as increased vaccination should be associated with reduced mortality, but it is difficult to pick up mortality effects without prohibitively large samples. For example, given the variation we observe and our cluster design, we would need a sample at least eight times larger to have a 50% chance of observing a 20% change in infant mortality due to contracting. It may easier to detect effects on more common outcomes, such as illness or diarrhea.

We have data on three final health outcomes. We found that contracting out reduced the chance of an individual reporting they were sick in the past month and this was significant at 5% under randomization inference on (see Table 8). Contracting out also reduced the incidence of diarrhea in children under five at 10% significance under randomization inference. We found no significant effect on whether a child born in the past year remained alive, though our sample is too small to detect typical changes in mortality.

To examine the average effect of contracting on final health outcomes, we recode diarrhea incidence and reported illness so that a good outcome is one and a bad outcome zero. Contracting out had a positive effect on final health of 0.62 standard deviations that was significant at the 5% level under clustering. Contracting in had a very small and insignificant negative point estimate.

e. Care-seeking behavior

We have already seen that one goal of the contracting experiment was to increase the use of public health facilities for curative care, and that both the contracting-in and contracting-out variants did so. We noted earlier that in 1997 patients had high out of pocket expenses. Patients frequently consulted unqualified drug sellers and traditional healers and saw government workers in their own private practices. In this section we will investigate how the program affected individual care-seeking behavior in more detail.

We have data on the type of provider sought for all family members who were ill and consulted a care provider during the previous month. There are four categories of providers: traditional healers, drug sellers, trained providers in private practice, and trained providers in public practice. Trained public providers are members of the government health service operating in their official capacity at public facilities. Trained private providers are qualified medical personnel operating on their own account, some of whom may be members of the government health service working outside or instead of their official duties. Trained providers include medical doctors, medical assistants, nurses, and midwives. Drug sellers include all vendors that sell drugs. Many also dispense advice, and almost none are pharmacists. Most are simply traders.

We first examine care-seeking behavior for all household members during the past month. We do not condition on whether they were sick, since this is potentially endogenous. At baseline, about 17% of individuals visited a provider in the past month. Of those who did, about 48% saw unqualified private providers (drug sellers and traditional healers), 44% saw private providers, and 8% saw public providers,. Household members were about 3.6 and 5.4 percentage points more likely to consult a public provider under contracting-in and contracting-out, respectively, compared to a comparison group baseline of 0.6% (Table 9). Both effects are statistically significant, though the contracting-in effect only reaches the 10% level under randomization inference. Both variants had negative point estimates on visits to unqualified providers; the contracting-out effect of negative 9.8 percentage points was statistically significant at the 5% level under randomization inference. Contracting-out discouraged visits to a private provider by 9.0 points against a baseline comparison level of 9.5% (statistically significant at 10%).

We perform a similar analysis of the care choices of those household members who were sick and visited a provider and found similar results (Table 9). The conditional average effects analysis treats visits to biomedically trained providers as positive and visits to drug sellers and traditional healers as negative. Contracting-in and contracting-out increased visits to a trained provider conditional on visiting a provider by 0.67 and 0.70 standard deviations, respectively. (The latter effect is statistically significant at only 10% under randomization inference.) The program did not explicitly target moving patients away from drug sellers or traditional healers, but reducing the use of these untrained providers is likely to improve public health, whereas shifting from private to public providers may or may not. We fail to reject the null hypothesis that the contracting-in and contracting-out average effects are equal. Point estimates of the impact of both variants are larger in the bottom half of the wealth distribution, though the differences are not statistically significant.

Interestingly, although "revealed preference" suggests the program made public clinics more attractive, we will see in the following subsection that patients had a poorer perception of the quality of care in those districts.[5]

f. Consumer Perception of Quality

The 2003 follow-up survey asked a set of questions about the experience of household members who had sought care from public health facilities in the past year. The survey asked respondents: "Based on the experience of your household members, please give me your honest

[5] It is at least theoretically possible that some of the switches from public to private health care are due to reductions in the supply of private care from restrictions imposed by the NGOs, so this may not reflect increases in the attractiveness of the public clinics.

opinion about the quality of the services at the health center/outreach?" Respondents were asked to give their opinion of whether in the most recent visit the staff was honest, caring, and polite, whether the staff was competent, whether the facility was well supplied, and whether the cost was low.

The TOT point estimates for contracting-in and contracting-out on quality perception are mostly negative for health centers (Table 10). Contracting-out had statistically significant negative effects of 20 percentage points on views of staff attitudes (significant at 10% under clustering and 5% under randomization inference), 18 percentage points on competence (significant at 10% under randomization inference and 5% under clustering), and 12 percentage points on how the facility was supplied (significant at 10% under clustering). Comparison means were 63%, 50%, and 50% respectively. Average effects point estimates on quality perception for both health centers and outreach for both contracting-in and contracting-out were negative. The 0.25 standard deviation negative effect contracting out had on perception of health center quality was the only one with statistical significance (5% under clustering). Average effects on perception of health center quality were not statistically significantly different across the top and bottom of the wealth distribution, though the effect size coefficients were only negative for the richer half. Point estimates were also negative for outreach (not shown), though none were statistically significant.

Given the results we have seen on residents' revealed preference for public providers and the improvement of facility management under contracting, these findings are striking. Several explanations are possible. First, this may be a statistical artifact. Contracting changed the composition of people visiting health centers, in part drawing in people who would otherwise be visiting drug sellers. The set of patients was thus not held constant. One possible explanation for lower perception of care quality is thus simple sorting: the group of individuals visiting health centers now has a different idea on average about what constitutes quality care. There may be more substantive reasons as well: 1) There are fewer health centers than drug sellers, so people have to travel a greater distance and may have to wait longer to receive care at a health center. 2) Discussions with contracting managers revealed they believed they had a different view of appropriate care than typical Cambodian providers, who may be more willing to provide treatments such as vitamin injections and glucose drips, corroborating the anthropological evidence discussed earlier. The contracting treatment emphasized a more biomedical approach, which is not necessarily what patients want. The managers may have imposed a different standard of care in the health centers more in line with their own views. For example, some contractors forbade the practice of giving vitamin injections when they were not clinically warranted, even though this practice reportedly made patients feel they had been given a more powerful treatment. Contracting-out providers had more influence over the behavior of their staff in this regard, since views about such matters could be a criterion for hiring.

6. Health Expenditures

The contracting project affected public and out-of-pocket health care spending. In this section, we will explore the effects of the program on spending in detail. We will first discuss the decrease caused by the program in out-of-pocket spending by individuals on curative care procedures. We will then show the project led to increases in public spending. Lastly we will put together our information on out-of-pocket and public spending together to show the project probably had a neutral to negative effect on overall health care spending.

a. Out-of-pocket spending

The baseline and follow-up surveys asked respondents about the out-of-pocket curative health care expenditures made by each individual in the household during the previous month. We will use this data to analyze the effect of contracting on annualized per capita out-of-pocket spending. There are some complications with this exercise. First, self-reported medical expenditures are thought to be very sensitive to framing effects and can vary considerably across surveys. Examining changes in spending recorded using the same survey instrument should mitigate this bias to some extent, though we should expect the data to be noisy. Second, the individual spending data is non-normally distributed with large outliers. In 1997, 87% of individuals reported no expenditure during the past month, while the 99.5th percentile of the distribution was $927 (2003 USD). Recall that Cambodia's annual per-capita GDP is only about US$280. The mean of such a distribution will be very sensitive to changes in the upper tail. Inspection of these upper-tail observations show that some appear to be plausibly large given the symptoms reported and providers consulted—for example, a 78-year old man who reported chest pains and was treated at a hospital in Vietnam. However, others appear inconsistent with those characteristics, such as a young woman complaining of a sore throat who spent $80 at a drug-seller.

Mean spending levels were very high, perhaps implausibly so for a country as poor as Cambodia. At baseline mean out-of-pocket medical spending was $26.98 (USD 2003), or nearly 10% of per-capita GDP. The high mean is largely due to a small number of reports of extremely high spending. District-level spending in 1997 and 2003 is positively but weakly correlated when one looks at the full data, but highly correlated when one trims the tails. We will report results both for the whole data and when we trim potentially implausible observations.

Regardless of whether one uses trimmed or untrimmed data, contracting out had a negative effect on out-of-pocket health spending, while contracting in had no statistically significant effect (Table 11). The contracting out TOT effect is a large and statistically significant negative $55.86 (2003 USD) when we look at the full, untrimmed data in column 1, though note that there is also a large and statistically significant difference in the initial level of out of pocket spending of $47.26.[6] Column 2 shows annualized private spending excluding the upper and lower 0.5% tails of the expenditure distribution, while column 3 shows spending with suspiciously large observations dropped.[7] Both of the trimmed specifications show smaller though still significant negative effects from contracting out (-$21.12 and -$15.61) and no significant effect for contracting in. (The contracting out effect when suspiciously large observations are dropped is only significant at 10% under clustering.)

We further explore this issue by computing the ITT effects of contracting on quantiles of the private expenditure distribution (Table 14). The pattern here is consistent with the mean TOT regressions: Contracting in does not have a significant effect while contracting out has a large and significant negative effect. Interestingly, higher baseline spending in the contracting out treated districts is present in all the upper quantiles we examine, and the treatment effect grows roughly proportionally.

Our overall conclusion is that contracting in had no significant effect on private spending while contracting out had a strong negative effect. This result is consistent with the reduced use of private providers. Given that contracting out was more successful at attracting patients into

---

[6] Throughout this paper we use a specification that allows for initial differences. In the case of this variable there seems to have been such differences, and they are persistent in the trimmed data. We therefore do not think it is appropriate to use only the data from the 2003 survey.
[7] We defined as suspiciously large those observations with more than US$100 per capita spending per month. There were 114 such observations—about 0.2% of all observations and about 1.6% of nonzero observations.

government clinics, its plausible the strong negative effect for this variant may have been due to lower patient payments in public facilities.

b. Public health spending

The overall change in public health spending associated with the program can be broken down into various components. The data on public expenditures cover 1999 through 2003, and are based on the actual receipt of funds at the district level. Table 13 presents a breakdown of public spending by major funding sources for comparison, contracting in, and contracting out districts. We look at the average spending for 2000-2003 and express spending in 2003 USD per capita to facilitate comparison with private spending. The project only ran in part of 1999, and we will use 1999 spending as a rough baseline later. Direct payments to contractors are reflected in row 1. Some other project expenditures, such as ministry-level management and monitoring, are reported in row 3. Total spending was $2.56 per capita in contracting in districts, 61% higher than the $1.59 per capita spent in comparison districts. Contracting out districts spent $2.94 per capita, 85% higher than comparison.

The differences in public spending between contracting in, contracting out, and comparison districts are mainly due to the contracting program payments. Part of the difference also comes from lower budget supplement spending in comparison districts. Recall that projects using the budget supplements in comparison districts were subject to auditing irregularities, and a consequence was that comparison districts did not receive the entire $0.25 per capita per year for which they were eligible.

Contracting out increased public health spending by a statistically significant and very substantial $2.93 per capita in 2003, against a comparison mean of $1.59 (Column 4, Table 11).

Contracting in had a smaller effect of $1.44, though this was not significant under randomization inference. Neither variant had a statistically significant effect on changes in public spending when we use 1999 as a rough baseline and the 2000-2003 average as follow-up (column 5), though both coefficients were positive and the contracting out coefficient was large.

Note also that average total *government* spending per capita for all of Cambodia was $3.14 in this period.[8] This is more than double the government funds actually received in our comparison districts, suggesting at least one of the following must be true: 1) there is substantial spending at the central Ministry level, 2) there is substantial regional inequality, and 3) there is substantial leakage. It is possible that the contracting program reduced leakage, so the figures may overstate the gap in public funding from the central government. Because the contractors were accountable for service delivery targets and were not as entangled in local politics, they may have been more successful at getting their budget allocations from the government. XX So? XX

c. Total health spending

Combining the administrative data on public spending with the survey data on out-of-pocket spending provides little evidence that the program increased total resources flowing to health. Program effects on total spending are shown in columns 6 to 11 of Table 11. Column 6 shows program effects on total spending for 2003. Neither program effect is statistically distinguishable from zero, and while positive the size of the contracting in coefficient is fairly small compared with mean total spending in the comparison districts. Columns 7 and 8 trim the distribution of private spending as discussed above. The coefficients on contracting in and

[8] Based on Cambodia (2004) and World Bank (2006).

contracting out naturally are closer to zero, and still small and insignificant. Columns 9 through 11 suggest that contracting in had no effect on the change in public spending over the program period, while contracting out had a negative and significant effect. These last columns should be viewed especially cautiously since they match public spending data from 1999 (the first partial year of the project) with private spending data from 1997.

7. Conclusion

The Contracting of Health Services Pilot Project contracted out management of public facilities to NGOs and increased public health spending on those facilities. The project led to increases in targeted service outcomes of about one-half standard deviation on average. The contracting-in and contracting-out approaches produced similar results, though the greater managerial autonomy afforded contracting-out managers appears to have enabled them to make greater strides in improving health center management. There is no evidence that contractors shifted resources away from non-targeted outcomes, though non-targeted service outcomes did not show any larger improvement than the comparison group. There is some limited evidence of improvements in individual health. Although the program increased use of public providers, contracting led to lower perceived quality of care among users. It is possible this resulted from providers adhering more rigidly to the biomedical model, or from the expectations about care held by patients who previously visited a traditional healer or a drug seller.

The project reduced private health expenditure, so total health expenditure likely decreased or stayed constant, suggesting that the approach was a cost-effective way to improve health service delivery in the Cambodian districts where it was implemented.

As we have data on only one (or two) of the set possible policy alternatives to the current system, it is impossible to say definitively how results would differ with a purely institutional or purely financial intervention. However, we think that the combination of institutional change and some additional public spending that we examine is of considerable policy interest because it is feasible. Given the low public salaries of government health workers, it may not have been feasible to implement a purely institutional change without increased spending. Requiring providers to be present more often without paying them more might have violated individual participation constraints, and almost certainly would have been politically infeasible.

It is difficult to assess to what extent, could health services be improved simply by spending more money in the existing public sector: One reason to such an intervention would not produce good results is that financial controls within the government of Cambodia are weak. Using national level cross-sectional data Filmer and Pritchett (1999) found that the impact of public spending on health as measured by child and infant mortality is statistically insignificant. Close to all differences in outcomes were explained by income and poverty, while independent variation in expenditures explained less than $1/7^{th}$ of 1 percent

Overall, the contracting project was very effective in improving service delivery in the project area. Loevinsohn and Harding (2005) review the global experience with health care contracting so far. The approach has been implemented on a large scale (covering 50,000 to 30 million individuals) in nine countries, and they argue it has proven workable in a wide range of environments. However, only the Cambodia project implemented contracting using a randomized design. We believe this makes it a particularly valuable example to learn from. It is difficult to assess external validity, particularly since our estimates of treatment effects apply only to districts where bids are received. Based on the promising results from Cambodia, additional trials seem warranted, both in other (hopefully) post-conflict environments, such as Afghanistan and the Democratic Republic of the Congo, and in more stable countries such as India, which nonetheless have serious problems in health care delivery.

Bibliography

Angrist, Joshua D.; Guido W. Imbens; Donald B. Rubin. 1996. Identification of Causal Effects Using Instrumental Variables. *Journal of the American Statistical Association*, Vol. 91, No. 434.

Abhijit Banerjee, Angus Deaton, and Esther Duflo. 2004. "Health Care Delivery in Rural Rajasthan." *Poverty Action Lab Paper No. 7*

Bertrand, Marianne; Duflo, Esther and Mullainathan, Sendhil. 2002. "How Much Should We Trust Differences-in-Differences Estimates?" *NBER Working Paper* No. 8841.

Bhushan, Indu, Sheryl Keller, and Brad Schwartz. 2002. "Achieving the twin objectives of efficiency and equity: contracting health services in Cambodia." *ERD Policy Brief Series* Number 6, Asian Development Bank.

Cambodia 2004. The Medium-Term Expenditure Framework for Cambodia 2005-2007. Ministry of Economy and Finance. http://www.mef.gov.kh/SpeechDr.Naron/2004/mtef1.htm

Conway, Tim. 2000. Current Issues in Sector-Wide Approaches for Health Development. Cambodia Case Study. World Health Organization.

Das, Jishnu and Jeffery Hammer. 2004. Strained Mercy: The Quality of Medical Care in Delhi World Bank Policy Research Working Paper 3228.

Filmer, Deon, and Lant Pritchett. "The impact of public spending on health: does money matter?" *Social Science and Medicine*, 49 (1999), p 1309-1323

Holmstrom, Bengt and Paul Milgrom. 1991. "Multi-Task Principal-Agent Analysis: Incentive Contracts, Asset Ownership, and Job Design," *Journal of Law, Economics and Organization*, vol. 7, no. 0, pp. 24-52.

Imbens, Guido, and Joshua Angrist. 1994. Identification and Estimation of Local Average Treatment Effects. *Econometrica* 62(2), pp. 467-475.

Jones, Gareth, Richard W Steketee, Robert E Black, Zulfiqar A Bhutta, Saul S Morris and The Bellagio Child Survival Study Group. 2003. "How many child deaths can we prevent this year?" *The Lancet*, (362) 9377, pp. 65-71.

Keller, Sheryl, and Brad Schwartz. 2001. *Final Evaluation Report: Contracting for Health Services Pilot Project.* Unpublished Asian Development Bank Report on Loan No. 1447-CAM.

Loevinsohn Benjamin. 2000 "Contracting for the delivery of primary health care in Cambodia: design and initial experience of a large pilot-test." *World Bank Institute Flagship Program On-line Journal*. http://info.worldbank.org/etools/docs/library/48616/oj_cambodia.pdf.

Loevinsohn, Benjamin and April Harding. 2005. Buying Results? Contracting for Health Service Delivery in Developing Countries. *The Lancet.* (366) 676-81

Moulton, B. 1990. "An Illustration of a Pitfall in Estimating the Effects of Aggregate Variables on Micro Units", *Review of Economics and Statistics,* 72(2), pp. 334-338.

Musgrove, P., 1996. "Public and Private Roles in Health. Theory and Financing Patterns." World Bank Discussion Paper No. 339. Washington, DC.

Schwartz, Brad and Indu Bhushan. 2004. *Reducing Inequity In The Provision Of Primary Health Care Services: Contracting In Cambodia.* Paper presented at World Bank Conference. http://wbln0018.worldbank.org/HDNet/hddocs.nsf/vtlw/BE2F3B5B9E4D164385256E00006863DC?OpenDocument.

Van de Put, Willem. 1991. Empty Hospitals, Thriving Businesses: utilization of health services and health seeking behavior in two Cambodian districts. Medecins Sans Frontiers research report.
RACHA. 2000. The Pathway to Child Health (Siem Reap, Pursat, Stung Treng, and Kampot).
Rosenbaum, Paul R. 2002. *Observational studies.* New York: Springer.
World Bank. 2006. *World Development Indicators Online.*

Appendix A: Bias from including household assets in the treatment effects regressions

What will be the effect of controlling household assets in our treatment effects regressions if asset holdings are themselves affected by the program? We are interested in the relationship between the true program effects β_4 and β_5, given by the model

$$y_{ijpt} = \beta_0 + \beta_1 I_j^{CI-R} + \beta_2 I_j^{CO-R} + \beta_3 I_t^{2003} + \beta_4 I_j^{CI-R} \times I_t^{2003} + \beta_5 I_j^{CO-R} \times I_t^{2003} + p_{pt} + \varepsilon_{ijpt}, \qquad \text{(A1)}$$

and γ_4 and γ_5 from the reduced form regression that includes the asset measure a_{ijpt}:

$$\begin{aligned} y_{ijpt} &= \gamma_0 + \gamma_1 I_j^{CI-R} + \gamma_2 I_j^{CO-R} + \gamma_3 I_t^{2003} \\ &\quad + \gamma_4 I_j'^{CI-R} \times I_t^{2003} + \gamma_5 I_j^{CO-R} \times I_t^{2003} + \gamma_6 a_{ijpt} + p_{pt} + \zeta_{ijpt}. \end{aligned} \qquad \text{(A2)}$$

Let's suppose that the program affects asset holdings because the program reduces out-of-pocket spending on medical care, freeing up income for other purposes. If the assets are normal goods, people will buy more of them. Therefore we expect that in the regression of the asset measure on the treatment dummies

$$a_{ijpt} = \chi_0 + \chi_1 I_j^{CI-R} + \chi_2 I_j^{CO-R} + \chi_3 I_t^{2003} + \chi_4 I_j^{CI-R} \times I_t^{2003} + \chi_5 I_j^{CO-R} \times I_t^{2003} + p_{pt} + \upsilon_{ijpt}, \qquad \text{(A3)}$$

that both the coefficients χ_4 and χ_5 will be positive. It also seems reasonable to assume that individuals who have more assets will have superior access to services. For example, if a new road is built in a village, it should improve both their income and their accessibility for medical services. We therefore expect coefficient γ_6 from equation (A2) to be positive. Combining equations (A2) and (A3), we find that the relationship between our estimated effects and the true program effects will be

$$\begin{aligned} \text{CI: } & \gamma_4 = \beta_4 - \gamma_6 \chi_4 \\ \text{CO: } & \gamma_5 = \beta_5 - \gamma_6 \chi_5 \, . \end{aligned} \tag{A4}$$

We thus expect our program effect estimates when controlling for assets to be underestimates of the true effect.

Table 1: Contracted Health Outcomes Definitions and Coverage Goals with Baseline Levels by Random Assignment Status

Indicator	Definition	Program Goal %	Baseline (%) Comparison	Contracting In	Contracting Out
Fully immunized child	Full immunization for children 12 - 23 months.	70	34	28	31
Vitamin A	High-dose Vitamin A received twice in the past 12 months by children aged 6 - 59 months	70	42	46	41
Antenatal care	≥2 antenatal care visits with blood pressure measurement at least once for women who gave birth in the prior year.	50	9	11	13
Delivery by trained professional	Birth attendant was qualified nurse, midwife, doctor, or medical assistant for women with a delivery in past year.	50	24	27	32
Delivery in a health facility	Birth was in a private or public health facility for women with a delivery in the past year.	10	3	6	5
Use modern contraception method	Women with a live child age 6-23 months old currently using a modern method of	30	13	12	18
Knowledge of modern contraception	Women who gave birth in the prior 24 months know four or more modern contraception methods and where to obtain them.	70	22	27	20
Use of public health care facilities	Use of district public health care facilities (district hospital or primary health care center) for illness in the prior 4 weeks.	Increase	4	4	3

Notes: Baseline statistics are averages by randomization status, not actual treament status.

Table 2: Characteristics of Project Districts

Treatment Status	Staffing	Procurement	Budget Supplement	Technical Assistance & Management Training	Number of districts
Contracting Out	Hired at market rates. MOH staff could take leave of absence	NGO responsible	No	No	2
Contracting In	MOH staff on government salary, usually given performance based bonus	NGO responsible but through MOH system	Yes	No	3
Comparison	MOH staff on government salary, often given supplement from user charges	Through MOH system	Yes	Yes	4
Not Successfully Contracted	MOH staff on government salary, often given supplement from user charges	Through MOH system	No	No	3

Table 3: TOT Effects for Health Facility Management

	Facilities and Staffing				Supervision				Delivery
	Permanent health center building open	24 hour service at health center	Unann. visit: health center open w/ patients	Unann. visit: All sched staff present	Num. supervisor visits in 3m	Last visit: discuss MOH progs	Last visit: discussed problems	Registers match HIS reports	Delivery services offered?
CI--Treated	0.236**	0.826***	0.477**	0.496**	0.028	0.102	0.090	0.308	0.246
Clustered S.E.	(0.08)	(0.11)	(0.22)	(0.17)	(0.49)	(0.09)	(0.08)	(0.19)	(0.16)
Ran. inf. p-value	0.22	0.03	0.17	0.13	0.96	0.25	0.23	0.17	0.11
CO--Treated	0.170	0.467	0.711	0.787***	5.654***	0.197	-0.123	0.127	0.403
Clustered S.E.	(0.22)	(0.27)	(0.44)	(0.24)	(1.34)	(0.19)	(0.18)	(0.36)	(0.36)
Ran. inf. p-value	0.60	0.46	0.25	0.11	0.08	0.21	0.52	0.72	0.15
Observations	143	121	143	143	143	112	116	143	143
R-squared	0.23	0.57	0.52	0.43	0.51	0.12	0.13	0.25	0.02
Comparison Mean	0.74	0.21	0.45	0.24	2.52	0.77	0.81	0.67	0.52

	User Fees				Outreach		Equipment and Supplies			
	Health center has user fee system	User fees clearly posted	User fee income (2003 US$)	Health center support from other NGOs?	Number outreach last month	Outreach: actual less scheduled	Health center equipment index	Health center supplies index	All child vaccs available at health	Average Effect Size
CI--Treated	0.164	0.238**	93.925	-0.061	-2.690	0.193**	3.530***	5.531***	-0.155*	0.599***
Clustered S.E.	(0.15)	(0.09)	(82.83)	(0.31)	(2.06)	(0.07)	(0.66)	(1.37)	(0.08)	(0.14)
Ran. inf. p-value	0.59	0.12	0.35	0.87	0.32	0.02	0.10	0.10	0.48	0.04
CO--Treated	0.301	0.284*	92.345	0.245	3.414	0.139	2.990*	8.863**	0.146	1.128***
Clustered S.E.	(0.23)	(0.15)	(81.63)	(0.80)	(3.19)	(0.12)	(1.37)	(3.10)	(0.18)	(0.38)
Ran. inf. p-value	0.28	0.38	0.28	0.78	0.61	0.45	0.29	0.13	0.84	0.04
Observations	143	108	89	143	143	124	143	143	143	H0: CO=CI, p-value
R-squared	0.31	0.17	0.19	0.2	0.16	0.02	0.33	0.38	0.3	
Comparison Mean	0.71	0.77	92.83	0.45	14.31	-0.06	15.02	25.02	0.36	<0.01

Notes: All columns except average effect are IV regressions in levels with province fixed effects. Standard errors presented in parentheses are corrected for clustering at the district level. Stars indicate significance under clustering: * at 10%; ** at 5%; *** at 1%. P-values for treatment effects computed by randomization inference. Treatment effects are in bold. Average effect is average differential increases caused by treatment in baseline comparison-group standard deviations. Regressions include province fixed effects. Eleven outcomes are: health center open with patients present on surprise visit, all scheduled staff present on surprise visit, child delivery service available at health center, user fees clearly posted if charged, number of supervisor visits made to the health center in the past month, number of outreach trips made by health center personnel in the past month, required equipment installed and functional (index), drugs and other supplies available (index), and the availability of all childhood immunizations. Null hypothesis is zero average effect.

Table 4: ITT Effects on Changes in Contracted Outcomes

	Full Immuni-zation	Vitamin A	Antenatal Care	Trained Delivery	Delivery in Facility	Use Contracep-tion	Know Contracep-tion	Use Public Facilities
Contracting In	-0.075	-0.013	-0.006	0.012	0.016	-0.003	0.036	-0.006
	(0.06)	(0.02)	(0.03)	(0.10)	(0.02)	(0.02)	(0.06)	(0.01)
Contracting Out	-0.051	-0.066**	0.014	0.063	0.008	0.055	-0.030	-0.001
	(0.08)	(0.03)	(0.04)	(0.07)	(0.01)	(0.03)	(0.06)	(0.01)
Contracting In X 2003	**0.105**	**0.060**	**0.280*** **	**0.049***	**0.091**	**0.063**	**-0.020**	**0.140*** **
Clustered S.E.	**(0.07)**	**(0.04)**	**(0.06)**	**(0.03)**	**(0.06)**	**(0.04)**	**(0.06)**	**(0.04)**
Randomization inference p-value	**0.28**	**0.58**	**0.04**	**0.40**	**0.13**	**0.21**	**0.79**	**0.07**
Contracting Out X 2003	**0.076**	**0.203*** *	**0.138**	**-0.055**	**0.040**	**-0.015**	**0.033**	**0.166*** **
Clustered S.E.	**(0.09)**	**(0.07)**	**(0.10)**	**(0.03)**	**(0.04)**	**(0.03)**	**(0.07)**	**(0.05)**
Randomization inference p-value	**0.46**	**0.02**	**0.35**	**0.33**	**0.61**	**0.78**	**0.67**	**0.02**
Year 2003	0.333***	0.232***	0.412***	0.183***	0.142**	0.146***	0.600***	0.195***
	(0.10)	(0.05)	(0.05)	(0.02)	(0.05)	(0.03)	(0.04)	(0.04)
Observations	5,100	11,213	4,993	4,993	4,976	6,994	9,537	11,223
R-squared	0.27	0.13	0.24	0.04	0.05	0.03	0.34	0.12
Comparion mean 2003	0.81	0.61	0.35	0.34	0.10	0.23	0.80	0.13
Comparion mean 1997	0.34	0.43	0.09	0.24	0.03	0.13	0.22	0.04

Notes: All regressions include province X year fixed effects. Standard errors presented in parentheses are corrected for clustering at the district level. Stars indicate significance under clustering: * at 10%; ** at 5%; *** at 1%. P-values for treatment effects computed by randomization inference. Treatment effects are in bold.

Table 5: TOT Effects on Changes in Contracted Outcomes

	Full Immunization	Vitamin A	Antenatal Care	Trained Delivery	Delivery in Facility	Use Contraception	Know Contraception	Use Public Facilities	Average Effect Size
Contracting In--Treated	-0.099	-0.021	-0.006	0.020	0.021	0.001	0.043	-0.007	
	(0.08)	(0.03)	(0.03)	(0.12)	(0.03)	(0.04)	(0.08)	(0.01)	
Contracting Out --Treated	-0.101	-0.138**	0.030	0.134	0.014	0.116	-0.070	-0.003	
	(0.14)	(0.06)	(0.10)	(0.17)	(0.03)	(0.12)	(0.12)	(0.03)	
Contracting In--Treated X 2003	**0.139**	**0.091**	**0.364*** **	**0.057**	**0.118**	**0.077**	**-0.022**	**0.176*** **	**0.505*** **
Clustered S.E.	**(0.08)**	**(0.06)**	**(0.08)**	**(0.04)**	**(0.07)**	**(0.06)**	**(0.07)**	**(0.04)**	**(0.09)**
Randomization inference p-value	**0.28**	**0.58**	**0.04**	**0.40**	**0.13**	**0.21**	**0.79**	**0.07**	**0.03**
Contracting Out--Treated X 2003	**0.150**	**0.417*** **	**0.263**	**-0.123**	**0.074**	**-0.038**	**0.073**	**0.289*** **	**0.537*** **
Clustered S.E.	**(0.12)**	**(0.09)**	**(0.16)**	**(0.11)**	**(0.07)**	**(0.09)**	**(0.13)**	**(0.05)**	**(0.12)**
Randomization inference p-value	**0.46**	**0.02**	**0.35**	**0.33**	**0.61**	**0.78**	**0.67**	**0.02**	**0.02**
Year 2003	0.297**	0.153***	0.343***	0.203***	0.122	0.148**	0.587***	0.143***	
	(0.10)	(0.04)	(0.11)	(0.04)	(0.07)	(0.05)	(0.06)	(0.02)	
Observations	5,100	11,213	4,993	4,993	4,976	6,994	9,537	11,223	
R-squared	0.27	0.13	0.25	0.03	0.05	0.02	0.34	0.12	
Comparion mean 2003	0.81	0.61	0.35	0.34	0.10	0.23	0.80	0.13	
Comparion mean 1997	0.34	0.43	0.09	0.24	0.03	0.13	0.22	0.04	
H_0: CO=CI, p-value									0.83

Notes: IV regressions including province X year fixed effects. Average effects are differential increases caused by treatment in units of standard deviations of changes in the outcomes. Standard errors presented in parentheses are corrected for clustering at the district level. Stars indicate significance under clustering: * at 10%; ** at 5%; *** at 1%. P-values for treatment effects computed by randomization inference. Treatment effects are in bold. Average effects tests null of zero effect against positive alternative.

Table 6: Robustness Check: Wealth Controls

Panel A: Average Effect Size for 15 Wealth Measures

	Contracting In (CI)	Contracting Out (CO)	H_0: CO=CI, p-value
Average Effect	0.018	-0.052	0.41
Clustered S.E.	(0.02)	(0.06)	
Randomization inference p-value	0.23	0.46	

Panel B: TOT Estimates

	Full Immuni-zation	Vitamin A	Antenatal Care	Trained Delivery	Del. in Facility	Use Contracep-tion	Know Contracep-tion	Use Public Facilities	Average Effect Size
CI--Treated	-0.097	-0.022	-0.001	0.026	0.023	0.006	0.045	-0.004	
	(0.07)	(0.03)	(0.04)	(0.10)	(0.03)	(0.04)	(0.08)	(0.01)	
CO--Treated	-0.097	-0.133*	0.025	0.138	0.013	0.120	-0.065	-0.002	
	(0.13)	(0.07)	(0.09)	(0.15)	(0.03)	(0.11)	(0.12)	(0.03)	
CI--Treated X 2003	**0.141***	**0.091**	**0.368*** **	**0.067**	**0.124***	**0.085**	**-0.021**	**0.173*** **	**0.629*** **
Clustered S.E.	**(0.08)**	**(0.06)**	**(0.08)**	**(0.05)**	**(0.07)**	**(0.06)**	**(0.07)**	**(0.04)**	**(0.12)**
Randomization inference p-value	**0.27**	**0.54**	**0.04**	**0.34**	**0.09**	**0.17**	**0.80**	**0.08**	**0.02**
CO--Treated X 2003	**0.157**	**0.412*** **	**0.267**	**-0.110**	**0.078**	**-0.028**	**0.075**	**0.288*** **	**0.673*** **
Clustered S.E.	**(0.12)**	**(0.08)**	**(0.16)**	**(0.12)**	**(0.06)**	**(0.09)**	**(0.14)**	**(0.05)**	**(0.14)**
Randomization inference p-value	**0.44**	**0.02**	**0.35**	**0.43**	**0.60**	**0.85**	**0.65**	**0.02**	**0.05**
Year 2003	0.261**	0.145***	0.286**	0.121**	0.084	0.127**	0.568***	0.140***	
	(0.10)	(0.04)	(0.10)	(0.05)	(0.06)	(0.05)	(0.06)	(0.02)	
Observations	5,084	11,178	4,979	4,979	4,962	6,975	9,510	11,191	
R-squared	0.28	0.14	0.27	0.09	0.08	0.03	0.34	0.12	
Comparion mean 2003	0.81	0.61	0.35	0.34	0.10	0.23	0.80	0.13	
Comparion mean 1997	0.34	0.43	0.09	0.24	0.03	0.13	0.22	0.04	
H_0: CO=CI, p-value									0.71

Notes: All IV regressions in Panel B include province X year fixed effects and wealth controls. Standard errors presented in parentheses are corrected for clustering at the district level. Stars indicate significance under clustering: * at 10%; ** at 5%; *** at 1%. P-values for treatment effects computed by randomization inference. Treatment effects are in bold. Average effects tests null of zero effect against positive alternative.

Table 7: TOT Effects on Changes in Noncontracted Health Service Outcomes

	Diarrhea Treatment (0/1)	Add'l Antenatal Checks	Village Visit <4wk	Breastfeed Newborn within 6h	water to <1 Month Old	AIDS Knowledge	Average Effect Size
Contracting In--Treated	-0.003	0.370	-0.097**	0.010	-0.007	-0.016	
	(0.04)	(0.40)	(0.04)	(0.03)	(0.00)	(0.05)	
Contracting Out--Treated	-0.144**	0.556	-0.113	0.067	0.000	-0.075	
	(0.06)	(0.97)	(0.09)	(0.07)	(0.01)	(0.07)	
Contracting In--Treated X 2003	**0.018**	**1.119****	**0.180***	**0.015**	**-0.037**	**0.211****	**0.281**
Clustered S.E.	**(0.04)**	**(0.39)**	**(0.08)**	**(0.07)**	**(0.03)**	**(0.07)**	**(0.12)**
Randomization inference p-value	**0.82**	**0.27**	**0.05**	**0.89**	**0.37**	**0.02**	**0.12**
Contracting Out--Treated X 2003	**0.144***	**0.578**	**-0.029**	**-0.064**	**-0.093**	**0.196**	**0.087**
Clustered S.E.	**(0.08)**	**(1.23)**	**(0.07)**	**(0.15)**	**(0.06)**	**(0.12)**	**(0.26)**
Randomization inference p-value	**0.17**	**0.70**	**0.96**	**0.76**	**0.10**	**0.23**	**0.74**
Year 2003	0.059	2.864***	0.201***	0.438***	0.090***	0.269***	
	(0.06)	(0.59)	(0.04)	(0.08)	(0.03)	(0.07)	
Observations	2,962	4,993	9,582	4,942	4,884	8,775	
R-squared	0.03	0.25	0.05	0.13	0.01	0.10	
Comparion mean 2003	0.93	2.79	0.77	0.35	0.95	0.42	
Comparion mean 1997	0.89	0.65	0.76	0.08	1.00	0.20	
H_0: CO=CI, p-value							0.78

Notes: All regressions include province X year fixed effects. Standard errors presented in parentheses are corrected for clustering at the district level. Stars indicate significance under clustering: * at 10%; ** at 5%; *** at 1%. P-values for treatment effects computed by randomization inference. Treatment effects are in bold. Average effects tests null of zero effect against two-sided alternative.

Table 8: TOT Effects on Final Health Outcomes

	Reported ill during past month	Diarrhea Incidence (0/1)	Child <1 Alive	Average Effect
Contracting In--Treated	0.004	-0.013	0.026	
	(0.04)	(0.06)	(0.02)	
Contracting Out--Treated	0.135	0.166	0.030	
	(0.10)	(0.15)	(0.02)	
Contracting In--Treated X 2003	**0.001**	**0.010**	**-0.011**	**-0.045**
Clustered S.E.	**(0.03)**	**(0.06)**	**(0.02)**	**(0.18)**
Randomization inference p-value	**0.92**	**0.85**	**0.78**	**0.82**
Contracting Out--Treated X 2003	**-0.145**	**-0.252**	**-0.043**	**0.616****
Clustered S.E.	**(0.10)**	**(0.19)**	**(0.03)**	**(0.56)**
Randomization inference p-value	**0.05**	**0.07**	**0.36**	**0.18**
Year 2003	0.077	-0.026	0.016	
	(0.04)	(0.07)	(0.02)	
Observations	54,062	9,850	4,930	
R-squared	0.01	0.01	0.01	
Comparison mean 2003	0.185	0.26	0.97	
Comparison mean 1997	0.202	0.35	0.97	
H_0: CO=CI, p-value				

Notes: IV regressions with provinceXyear effects. Standard errors presented in parentheses are corrected for clustering at the district level. Stars indicate significance under clustering: * at 10%; ** at 5%; *** at 1%. P-values for treatment effects computed by randomization inference. Treatment effects are in bold. Average effects tests null of zero effect against two-sided alternative.

Table 9: TOT Effects on Changes in Care-Seeking Behavior

	Was any provider consulted in the past month?				Conditional on consulting a provider, what type was visited?			
	None	Unqualified Provider	Qualified Private Provider	Qualified Public Provider	Unqualified Provider	Qualified Private Provider	Qualified Public Provider	Average Effect Size (Qualified Provider)
Contracting In--Treated	-0.010	0.004	-0.020	0.000	0.104**	-0.076**	-0.006	
	(0.04)	(0.04)	(0.02)	(0.00)	(0.04)	(0.03)	(0.01)	
Contracting Out--Treated	-0.120	0.135	0.052	0.015	0.099	-0.063	0.003	
	(0.10)	(0.10)	(0.04)	(0.01)	(0.09)	(0.07)	(0.04)	
Contracting In--Treated X 2003	**0.003**	**-0.042***	**0.015**	**0.036*****	**-0.214*****	**0.029**	**0.171*****	**0.665*****
Clustered S.E.	**(0.03)**	**(0.22)**	**(0.01)**	**(0.01)**	**(0.07)**	**(0.08)**	**(0.03)**	**(0.16)**
Randomization inference p-value	**0.96**	**0.20**	**0.52**	**0.06**	**0.01**	**0.76**	**0.03**	**0.01**
Contracting Out--Treated X 2003	**0.118**	**-0.098**	**-0.090***	**0.054*****	**-0.205**	**-0.101**	**0.279*****	**0.701*****
Clustered S.E.	**(0.09)**	**(0.06)**	**(0.05)**	**(0.01)**	**(0.14)**	**(0.12)**	**(0.05)**	**(0.33)**
Randomization inference p-value	**0.10**	**0.05**	**0.07**	**0.05**	**0.18**	**0.46**	**0.01**	**0.08**
Year 2003	-0.077*	0.077	0.006	0.047***	0.078	-0.144	0.129***	
	(0.04)	(0.04)	(0.01)	(0.01)	(0.09)	(0.09)	(0.02)	
Observations	54,062	54,062	54,062	54,062	11,070	11,879	11,879	
R-squared	0.01	0.01	0.01	0.02	0.02	0.02	0.09	
Comparison mean 2003	0.826	0.081	0.080	0.031	0.44	0.41	0.16	
Comparison mean 1997	0.824	0.085	0.095	0.009	0.48	0.44	0.06	
H_0: CO=CI, p-value								0.88

Notes: IV regressions with provinceXyear effects. Standard errors presented in parentheses are corrected for clustering at the district level. Stars indicate significance under clustering: * at 10%; ** at 5%; *** at 1%. P-values for treatment effects computed by randomization inference. Treatment effects are in bold.

Table 10: TOT Effects on Consumer Perception of Quality at Health Centers

	Staff honest, polite, caring	Staff competent.	Supplies at facility/service	Cost	Average Effect Size
Contracting In--Treated	**-0.076**	**-0.052**	**-0.048**	**0.009**	**-0.128**
S.E.	**(0.07)**	**(0.06)**	**(0.05)**	**(0.05)**	**(0.14)**
Randomization inference p-value	**0.36**	**0.54**	**0.48**	**0.89**	**0.24**
Contracting Out--Treated	**-0.199***	**-0.175****	**-0.119***	**0.037**	**-0.249****
S.E.	**(0.09)**	**(0.07)**	**(0.06)**	**(0.06)**	**(0.14)**
Randomization inference p-value	**0.05**	**0.07**	**0.15**	**0.65**	**0.15**
Observations	2526	2499	2479	2524	
R-squared	0.07	0.13	0.12	0.01	
Comparison Mean	0.63	0.50	0.50	0.87	
H_0: CO=CI					0.38

Notes: The survey gives the following instruction and then asks the opinion in the four categories in this table: "Based on the experience of your household members in the visit you just told me about, please give me your honest opinion about the quality of the services at the Health Center." All regressions include province fixed effects. Standard errors presented in parentheses are corrected for clustering at the district level. Stars indicate significance under clustering: * at 10%; ** at 5%; *** at 1%. P-values for treatment effects computed by randomization inference. Treatment effects are in bold.

Table 11: TOT Effects on Health Care Spending

	Annualized Individual Curative Care Spending			Annual Public Health Spending Per Capita		Total Annualized Health Spending Per Capita					
	All Individuals	All Individuals excluding 0.5% tails	All Individuals excluding those who spent >US$100 last month	Average 2000-03	Change 99-Avg. 00-03	2003 (Public Avg. 2000-03)	2003 (Public Avg. 2000-03, Private excl.0.5% tails)	2003 (Public Avg. 2000-03, Private <$100 last month)	Change 97/9-03	Change 97/9-03 (Private excl. 0.5% tails)	Change 97/99-03 (Private <$100 last month)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
CI-Treated	2.324	0.102	-0.082								
	(06.71)	(03.06)	(02.24)								
CO-Treated	47.286**	17.003*	10.507								
	(16.83)	(8.03)	(5.93)								
CI-Treated X 2003	**-3.321**	**-0.521**	**-0.711**	**1.442****	**0.704**	**1.690**	**1.089**	**0.687**	**-2.277**	**0.124**	**0.034**
Clustered S.E.	**(4.96)**	**(2.12)**	**(1.91)**	**(0.49)**	**(0.57)**	**(4.20)**	**(2.77)**	**(2.63)**	**(11.26)**	**(5.79)**	**(5.00)**
Randomization inference p-value	**0.95**	**0.98**	**0.96**	**0.28**	**0.26**	**0.62**	**0.88**	**0.62**	**0.99**	**0.86**	**0.88**
CO-Treated X 2003	**-55.856*****	**-22.122****	**-15.608***	**2.927*****	**1.121**	**-6.366**	**-2.072**	**-2.081**	**-57.933****	**-18.247***	**-14.186**
Clustered S.E.	**(16.76)**	**(8.56)**	**(7.24)**	**(0.79)**	**(0.90)**	**(6.68)**	**(4.41)**	**(4.19)**	**(17.91)**	**(9.21)**	**(7.95)**
Randomization inference p-value	**0.01**	**0.01**	**0.02**	**0.14**	**0.24**	**0.36**	**0.80**	**0.60**	**0.01**	**0.02**	**0.02**
Year 2003	3.186	-2.259	-2.458								
	(07.76)	(03.67)	(02.89)								
Observations	54,062	53,529	53,948	12	12	12	12	12	12	12	12
R-squared	0.01	0.01	0.01	0.89	0.35	0.63	0.66	0.71	0.65	0.19	0.32
Comparison mean 2003	$12.12	$7.51	$9.84	$1.59	$0.42	$13.80	$9.20	$11.36			
Comparison mean 1997	$18.76	$10.42	$15.17								
Comparison mean change									-$7.84	-$4.79	-$4.97

Notes: IV regressions with provinceXyear effects. Standard errors presented in parentheses are corrected for clustering at the district level. Stars indicate significance under clustering: * at 10%; ** at 5%; *** at 1%. P-values for treatment effects computed by randomization inference. Treatment effects are in bold.

Table 12: Average Actual Public Per Capita Spending by Source, 2000-2003 (USD 2003)

	Comparison	Contracting In	Contracting Out
Contracting Project Payments	0.00	0.79	1.99
Contracting Budget Supplement	0.05	0.23	0.00
Other Basic Health Services Project Spending	0.08	0.12	0.09
Government	1.37	1.17	0.70
NGO/Donation	0.06	0.14	0.10
Cost Recovery	0.03	0.10	0.06
TOTAL excluding Cost Recovery	1.59	2.56	2.94
Percentage of Comparison		161%	185%

Notes: We exclude public revenue derived from user fees from the public spending data since it is captured in our out-of-pocket spending data.

Table 13: ITT Effects on Individual Curative Care Spending (USD 2003)

	Mean	90th percentile	95th percentile	99th percentile
CI-Treated	0.484	0.000	0.000	42.656
	(3.54)	(7.70)	(18.95)	(82.10)
CO-Treated	21.832***	18.546**	37.092**	310.464***
	(3.56)	(7.71)	(15.10)	(97.95)
CI-Treated X 2003	**-1.012**	**-0.604**	**-2.416**	**-57.758**
Clustered S.E.	**(5.09)**	**(6.07)**	**(12.88)**	**(70.19)**
Randomization inference p-value	**0.95**	**0.98**	**0.90**	**0.62**
CO-Treated X 2003	**-25.890*** **	**-21.567*** **	**-46.154*** **	**-355.770*** **
Clustered S.E.	**(5.12)**	**(7.32)**	**(15.37)**	**(78.70)**
Randomization inference p-value	**0.01**	**0.02**	**0.06**	**0.01**
Year 2003	-7.324	-1.59	-26.463**	-266.483***
	(5.00)	(5.08)	(12.22)	(50.32)
Observations	54,062	54,062	54,062	54,062
R-squared	0.01			
Comparison mean 2003	$12.12			
Comparison mean 1997	$18.76			
Comparison quantile 2003		$9.06	$40.76	$265.79
Comparison quantile 1997		$12.98	$74.12	$445.11

Notes: Mean and quantile regressions with provinceXyear effects. Standard errors presented in parentheses are corrected for clustering at the district level using bootstrap. Stars indicate significance under clustering: * at 10%; ** at 5%; *** at 1%. P-values for treatment effects computed by randomization inference. Treatment effects are in bold.

Figure 1A: Percentage Point Changes in Contracted Outcomes by District 1997-2003

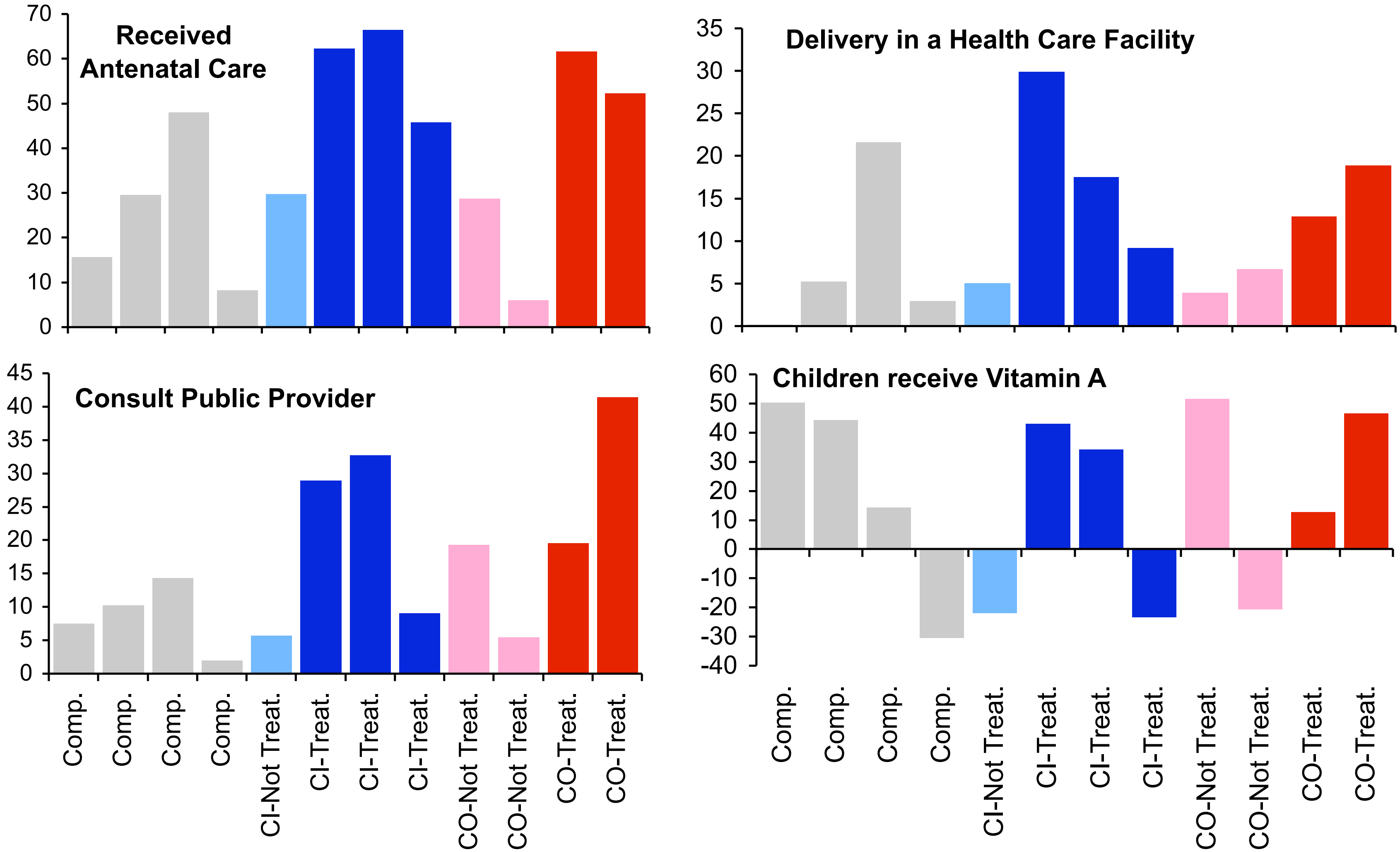

Figure 1B: Percentage Point Changes in Contracted Outcomes by District 1997-2003